- Registration of polymers with regard to structure; more consistency in registration (CA)
- More complete indexing of essential components; closer attention to claims; handling of generics (CA)
- Fuller description of scope of patent (CA abstracts)
- Searchable abstracts (CA)
- *More indexing terms, especially bound copolymers, catalyst components* (Plasdoc)
- Generic systems that distinguish between components that are alternatives, and those that must both be present.
- More attention to process details and equipment
- More use of linking logic

Figure 15. Some desired database improvements.

develops the Registry to the point where it deals with special polymer structures such as blocks and grafts and gets rid of indexing policies that omit important ingredients from the indexing. I hope, too, that they can develop a system to deal with generics and have their abstracts show more fully what the scope of a patent is. CA really needs searchable abstracts—and the searchable abstracts should be available not only on STN but on DIALOG, SDC, and other hosts.

I hope that *Plasdoc* will add more indexing terms beyond those added recently, especially terms to describe bound copolymers and catalyst components. I think that we desperately need generic systems that can distinguish components that are present together from those that are alternatives. Otherwise, file growth will further magnify the noise that is often intolerable even today.

I would certainly like to see more attention paid to process details and equipment, something that *Plasdoc* probably does best today. I want to be able to know with certainty when polymerization was run in a tubular reactor and when it was run in gas phase. And I would give a great deal for a system that could tell me with assurance that a patent really involved grafting in an extruder, rather than just the extrusion of a grafted polymer.

Finally, I would like to see an extension of linking logic, one of the best ways of cutting down on noise. There is certainly room for linking of nonfragment terms in *CLAIMS-CDB* and of the new key serial and registry terms in *Plasdoc*. Others could surely add to the list. I look forward to future enhancements of today's invaluable services.

REFERENCES AND NOTES

(1) POST-P and POST-J are no longer published, but they served as a pilot development that led to improvements in CA's polymer coverage.
(2) Kaback, S. M. "The IFI/Plenum Chemical Indexing Systems". "Computer Handling of Generic Chemical Structures"; Barnard, J. M., Ed.; Gower: Aldershot, Hampshire, England, 1984; pp 49-65.

The Paradox of Patentability Searching†

EDLYN S. SIMMONS

Merrell Dow Pharmaceuticals Inc., Cincinnati, Ohio 45215-6300

Received October 10, 1984

Patentability searching is, by its nature, a paradox. To be patentable, a new product or process must never have been described or suggested in the prior art. The searcher charged with confirming the patentability of an invention is thus presented with three dilemmas: how to design an exhaustive search for a reference that is believed not to exist, how to recognize a reference describing the invention in a different context or vocabulary, and, if no relevant references are found, how to tell when the search is complete.

Most research and development in science and engineering is conducted for the purpose of making inventions, in the sense that any new product or procedure and any improvement in an existing product or procedure are inventions. Having invested large sums of money in its development, it is natural for the successful developers of a new product or process to wish to reserve for themselves the right to profit from their invention. The accepted way to prevent others from making, using, or selling an invention is by patenting it. But not every new product or process is patentable. In the U.S. and in most other countries, only an invention that is new, useful, and nonobvious can be granted a valid patent.

Applying for a patent on an unpatentable invention represents an unnecessary expense and an unnecessary risk. The patent offices of many countries publish patent applications before determining whether or not the invention being claimed satisfies the national standards of patentability. If a patent application claiming an unpatentable invention is filed in one of these countries, the unpatented technology disclosed in the published application is freely available for use by the public in that country and in every other country where patent protection has not been obtained. The developers may well wish to maintain the details of the invention as a trade secret or to abandon the invention entirely so as to avoid sharing it with their competitors. It is possible to minimize the risk of public disclosure of an unpatentable invention by performing a patentability search before drafting the patent application.

For a hypothetical example, let us take the development of a new chemical compound by a Dr. Smith of the Panacea Cosmetic Co. Dr. Smith has had the new compound tested and knows that it prevents dandruff in rats. He has devised a new shampoo formulation that includes the new compounds and found that these compounds also reduce the incidence of dandruff, although not as well as the original compound. If possible, the Panacea Cosmetic Co. would like to patent Dr. Smith's whole genus of antidandruff agents, the method by which the compounds were synthesized, the shampoo formulation containing any of the compounds in the genus, and the method for preventing dandruff by shampooing with such a formulation, all of which would be patentable inventions under U.S. patent law if the standards of utility, novelty, and nonobviousness are met. Since it is an international company, Panacea intends to patent the invention in each of 20 countries. If it is unable to obtain such broad coverage, Panacea will settle for a patent that claims only the most active compound, and if the compound per se is not patentable, it will make do with a patent that protects only the shampoo formulation or the method of treating dandruff with it.

† Presented at the Herman Skolnik Award Symposium in honor of Montagu Hyams before the Division of Chemical Information, 188th National Meeting of the American Chemical Society, Philadelphia, PA, August 28, 1984.

0095-2338/85/1625-0379$01.50/0

If Panacea cannot protect the whole genus of related compounds, it is not willing to teach its competitors that the unpatentable compounds are useful as antidandruff agents by including that information in a patent application. Before drafting a patent application claiming the invention in its broadest form, the Panacea Patent Department has a patentability search made to determine whether any of the new compounds has ever been described before and, if it has, whether anyone has ever suggested using the compound in a dandruff shampoo.

It is hoped that the patentabilty search will confirm that Dr. Smith's newly synthesized chemical compounds have never before been described in a published reference and also that the references to the nearest related known compounds do not suggest to a person of ordinary skill in the art, which would be an experienced research chemist in our example, that the related compounds should be modified in a way that would produce Dr. Smith's new compounds. The invention embodied in the compounds would thus satisfy the statutory requirements for novelty and nonobviousness. If the compounds themselves are patentable, the synthetic method, the shampoo, and the method of treating dandruff will probably also be patentable in countries that permit claims of those types. If references to some of the compounds are found, the information in the references will be used by the Panacea Patent Department to decide whether or not a patent application should be written, how much it should disclose, what it should claim, and where it should be filed.

References published before the filing of a patent application are known as the "prior art" of the invention. In most countries, the prior art includes everything published anywhere in the world at any time before the filing of the patent application. Public use or sale of an invention prior to filing is also a bar to patentability under most circumstances. Patent law varies somewhat from country to country; the law defining prior art and other bars to patentabilty in the U.S. permits the inventor to establish that he or she made the invention or imported it into the U.S. before the publication of a prior art reference that was published less than 1 year before the filing of his or her patent applicaton,[1] but such a reference would bar the issuance of a patent in most other countries.

By its nature, a patentability search embodies a paradox. A patentability search is initiated only when it is believed that the invention being searched is patentable. In other words, a patentability search is begun only when it is expected that there are no references to find. The searcher hopes to confirm that belief by performing the most exhaustive and comprehensive search possible and failing to discover any references to the invention. Ideally, the searcher will search everywhere and find nothing.

A patentability search involves the same procedures as any other retrospective search. First, the searcher must determine the scope of the invention being searched. A patentability search should be broad enough in scope to retrieve not only references that describe the invention precisely but also those related references that would render the actual invention obvious. Second, the searcher must decide which databases ought to be searched for references to the invention. For a patentability search, compilations of data of all kinds—computerized databases, papers in technical journals, shelves of reference books, stacks of patent copies, even old Sears and Roebuck catalogs—may contain a previous disclosure of an invention and may serve as appropriate databases to search for references. To ensure that the subject matter is actually absent from the prior art, it is desirable to search every database in which it may have been described. Third, the searcher must create a strategy for searching each of the databases.

The final step, the actual search of a database, consists of two stages, (1) screening the database to identify a set of documents whose indexing suggests that they may describe the invention and (2) evaluating the full text of the candidate documents to eliminate false drops. In some cases, the two stages of the search are inseparable. Once a strategy for searching the files of the United States Patent and Trademark Office Public Search Room has been created by selecting the subclasses under which the invention may have been classified, the searcher proceeds by scanning the full text of each patent in the subclasses, without an intervening screening step. Screening a printed or computerized index often fails to identify any documents that meet the requirements of the search strategy, eliminating the need to evaluate candidate documents.

A search whose object is to prove that no relevant references exist involves a number of inconsistencies and contradictions that may not be encountered in other searches.

To begin with, the searcher cannot initiate a patentability search without putting aside the basic premise that the invention is new and nonobvious. Without a working hypothesis that an earlier inventor has published a reference to the invention, it would be difficult for the searcher to predict which databases would contain the reference and how it would be indexed. This may require that the searcher engage in a bit of "doublethink" or a bit of role-playing.

A searcher employed by the developer of the invention is not the only person who will search for prior art references if a patent application on the invention is eventually filed. Patentabilty searches are routinely performed by patent examiners. The examiner searches without any preconceptions as to the patentability of the invention to determine whether the claims in the patent application meet the standards of novelty and nonobviousness of the national patent law. If the result of the prefiling patentability search are to be used to determine the scope of the patent claims, it is crucial that it find all of the relevant prior art that will ultimately be found by the patent examiner.

After the invention is patented, it is possible that a potential competitor will wish to exploit the invention without paying royalties to the patent owner. The competitor may attempt a validity search, an even more thorough search of the prior art whose object is to prove that the patent is not valid and enforceable by showing that the invention was not truly patentable when the patent application was filed. The ideal prefiling patentabilty search would be as thorough as such a validity search and would uncover every reference available to the competitor's searcher. This would enable the writer of the patent application to focus on the novel aspects of the invention and minimize the chance that the patent would someday be declared invalid in the courts, which are the ultimate judges of patentability.

One aspect of the paradox is that a prefiling patentability search cannot possibly be fully exhaustive. The prior art includes everything published before the filing of the patent application, but not all of the prior art is accessible to the searcher. Comprehesive indexing of the scientific literature is a relatively modern development and has rarely been done retrospectively. The chemical literature of the 19th century is not retrievable through *Chemical Abstracts*, for example. The Beilstein and Gmelin handbooks[2] cover the preexisting chemical literature, but they do so selectively and ignore any publications deemed by their editors to be insignificant or redundant. Some of the prior art is so old that it has never been reported in the scientific literature. Patent examiners are aware of this and have been known to extend their searches into classical mythology or the Bible in order to cite the labors of Hercules or the nutritional use of milk and honey as prior art. Information that is not indexed for retrieval is difficult, if not impossible, to find, and it is inevitable that some details of any document will be omitted from an index.

Before a patent application has been filed, some of the prior art cannot be retrieved because it does not yet exist. There is obviously no way for a searcher to identify the prior art that will be published after the search is done and before the patent application is filed. In addition, some of the recent prior art is not fully accessible to the searcher because it was published so recently that abstracting and indexing services have not yet processed the references for retrieval. Both validity searches and searches by patent examiners are done after the patent application has been filed, when all of the modern prior art published before the effective filing date of the patent application has been published and indexed for retrieval.

Establishing the scope of the patentability search also presents a dilemma. Because the law requires the invention to be nonobvious as well as novel[3] and because no references to the invention are expected to be found, it is important to search broadly enough to retrieve the nearest related references as well as those that describe the invention itself. But deciding in advance which related concepts to search in order to find the nearest prior art is difficult because it is impossible to know what kinds of related disclosures are present in the prior art until after the search has been done. Familiarity with the field of technology and guesswork are the two tools that are most helpful in determining the search scope.

Not infrequently, the nearest prior art found in a search is an artifact of the indexing employed by the databases in which the search was done. Substructure searches using fragmentation and topological codes, for example, require that many details of a chemical structure be left undefined. As a result, the searcher may discover references to analogous prior art compounds that bear an unexpected substituent or that are position isomers of the compound being searched. The more precise the indexing terms, the less related art one will retrieve. If a search were performed with only the CAS Registry Number of the the compound of interest, even its optical isomers would be excluded from the results. One of the contradictions in patentability searching is that the most valuable information one retrieves may well be found in the false drops. False drops, references that satisfy the requirements of the search strategy but do not identically describe the invention of interest, may be the nearest related prior art or they may suggest refinements to the search strategy that will lead to even more closely related references.

Once the searcher has decided upon a scope for the search and assumed the role of the competitor's searcher, he or she must decide where to look for references to the invention. Typically the search will include collections of patents as well as the scientific literature in the field of technology to which the invention belongs. If the invention is truly novel, however, it may not fit conveniently into any recognized field of technology, and databases that are likely to cover similar technology may be difficult to identify. In choosing databases to search, it is important to consider that an earlier disclosure of the invention may have been made in a context quite different from the one contemplated by the present inventor. A prior art teaching of a claimed chemical compound for a different use or of a claimed machine in a different process will bar the issuance of a patent on the compound or the machine per se even though the new process or use may still be patentable. Dr. Smith thinks of his new compound as a cosmetic ingredient. If the compound is useful for treating skin disorders, it may have been described previously by a dermatologist as a treatment for psoriasis or by a veterinarian as a sheep dip. His search should not be limited to the cosmetic literature, but, at the very least, should also include databases that would allow the retrieval of pharmaceutical and agricultural literature.

Searching several databases is one of the most valuable techniques for overcoming the paradoxical nature of patentability searching. A successful search of one database, one that results in a "zero postings" message, is not necessarily a sign that no references exist. The reference, if it exists, may not be included in the chosen database. Even when a pertinent reference is in the database, it is likely that the innovative feature of the invention being searched is not indexed retrievably. Existing databases are indexed by means of classification systems or thesauri that were designed to describe preexisting technologies. An innovation that gives rise to a patentabilty search may be sufficiently novel that the database does not have an appropriate term or classification for it. If this is the case, a prior art reference to the invention will probably have been indexed in terms of the next broader generic classification or under a thesaurus term referring to an analogous, but different, concept. If some other feature of the reference is conveniently categorized by the indexing system, the reference will probably have been indexed according to that feature rather than the innovative concept of the present search. The lack of indexing for a prior disclosure of the novel feature of the invention is not necessarily the result of incorrect indexing. In many cases, the author of the prior art has treated the subject of the present search as peripheral to the subject of his or her publication. Unless the database uses extremely deep indexing indeed, only the ideas that were central to the reference will have been indexed. Let us imagine an earlier disclosure of one of Dr. Smith's new antidandruff compounds in a Japanese patent for a shampoo that combines the novel compound with the established antidandruff agent, zinc pyrithione, and that teaches that the combination gave a shampoo with a pleasing green color. The reference will be indexed only in a database that covers Japanese cosmetic patents. Such a reference is likely to have been indexed as a zinc pyrithione containing shampoo, both because the novel compound will not have been present in the database thesaurus and because the reference patent did not make it clear that the unfamiliar compound was not an inert ingredient. Each database has different selection policies and different indexing criteria. The more redundant the search, the more likely it becomes that an elusive reference will be retrievable in one of the databases searched.

The searcher may also have failed to retrieve an existing reference by using a search strategy that is inappropriate for the database being searched. To retrieve all relevant references and a minimum of false drop, the searcher must understand the indexing system and the indexing philosophy of each database he or she may elect to search.[4] The various available databases may be searchable by means of controlled thesaurus terms, uncontrolled index terms, free text, or classification systems. Some databases, particularly interactive online databases, may be searched by combinations of these. But all of these retrieval tools are unreliable for the identification of the novel features of inventions, because each of them is predicated upon the existence of predictable nomenclature for the subject being searched.

Nomenclature is probably the most perverse aspect of patentability searching. For a completely novel concept, no accepted terminology exists, so the inventors have to supply their own vocabulary. It is axiomatic in U.S. patent law that "the patentee is his own lexicographer".[5] The inventor of a new product or process is entitled to define it in any way that is unambiguous and descriptive. This allows new technology to be assigned an appropriate new vocabularly and allows inventors who are unfamiliar with art-recognized terminology to patent their inventions in their own words. It appears from

United States Patent [19]
Dunkelberger

[11] 4,440,905
[45] Apr. 3, 1984

[54] DUNKELSPERSERS

[75] Inventor: David L. Dunkelberger, Newtown, Pa.

[73] Assignee: Rohm and Haas Company, Philadelphia, Pa.

[21] Appl. No.: 256,202

[22] Filed: Aug. 13, 1981

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 135,945, Mar. 31, 1980, abandoned.

[51] Int. Cl.[3] C08L 51/04
[52] U.S. Cl. 525/66; 525/63; 525/64; 525/67; 525/69; 525/70; 525/80; 525/84; 525/86
[58] Field of Search 525/63, 66, 67, 69, 525/80, 84, 64, 70, 86

[56] References Cited

U.S. PATENT DOCUMENTS

3,652,483 3/1972 Tanaka et al. 525/84
3,956,424 5/1976 Muroyama et al. 525/84

Primary Examiner—J. Ziegler
Attorney, Agent, or Firm—Marc S. Adler

[57] ABSTRACT

Improved impact modifier composition comprising elastomeric impact modifier polymer and dunkelsperser in a weight ratio of about 99.5/0.5 to 96/4 which is subject to substantially reduced gel formation in rigid thermoplastic matrix polymer formulations. Also disclosed are such formulations, as well as processes for preparing the improved impact modifier compositions.

17 Claims, 2 Drawing Figures

Figure 1.

the patent shown in Figure 1 that David Dunkelberger invented a class of polymer additives that had no previous generic name and named the additives for himself. Since much of the prior art in a patentability search (and in other searches as well) is in the patent literature, this kind of creativity can cause the searcher enormous difficulties in devising a search strategy and in evaluating the references retrieved by the search. If the present inventor is unfamiliar with the vocabulary of the prior art, he may be mistaken in his belief that the invention had never before been described. On the other hand, the inventor may be thoroughly conversant with the vocabulary of the prior art, but a prior inventor may have described the same invention in nonstandard terms. Unless the term enters the standard vocabulary of the technology to which it pertains, a subsequent inventor is unlikely to recognize that a reference to a dunkelsperser is relevant to the polymer additive that he has just developed. The searcher must therefore precede the search for references to the invention with a search for alternative nomenclature that may have been used for it in the prior art and must always keep in mind the possibility that any unfamiliar term encountered during the search may describe a relevant concept. Even when the concept being searched is well-known, there are likely to be several terms that express it. Writers on the subject of dandruff, for example, differ extensively in their evaluation of the cause and significance of a flaking scalp and have described this well-known phenomenon in diverse ways, some of which are listed in Figure 2.

In addition to the general problem of naming a product or process that is believed not to have been described before is the problem that the customary language of patents is different from the everyday language of scientists and engineers. Whenever possible, patents speak in generic language, so that the inventions they protect can be claimed as broadly as possible. A patent will not describe a solvent as "methanol" if any "aliphatic alcohol" will work in the process it describes, and it will not say "aliphatic alcohol" if any "protic solvent" will do. There will probably be a specific example in the patent specification of the preferred solvent, but it is unlikely to turn up in the abstract or in the broadest claim, which are the most likely parts of the patent to be reproduced in a database. In a patentability search, the subject of the search as well as the patents retrieved as references is likely to be in generic language. It is necessary for the searcher to retrieve and recognize a description of the invention in generic terms both broader

```
DANDRUFF

Alternative Nomenclature:

        Flaking of the scalp

        Scurf

        Seborrhea sicca

        Seborrheic keratosis

        Pityriasis
```

Figure 2.

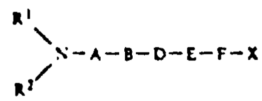
and narrower than those used to define the scope of the search.

Patents use generic descriptions not only for natural language but for structures. Generic chemical structures in patents are called Markush structures. A Markush structure represents a multiplicity of chemical compounds by means of a single skeleton that bears one or more groups of alternative substituents or skeletal components. The design of systems for indexing Markush structures that will permit a structure to be retrieved whenever the indexed structure overlaps with the structure being searched represents a major challenge for the developers of databases.[6] After retrieving a reference with a Markush structure, the searcher may have great difficulty in recognizing whether the compounds being searched are actually described by the reference structure. There are no enforceable standards for constructing Markush structures, so the odds are astronomical against finding a prior art reference to a genus of compounds that discloses an identical Markush structure. It is entirely possible that the Markush structure in the prior art will bear no resemblance at all to the one being searched, as in the case illustrated in Figure 3, taken from a recent patent.

Hierarchical indexing systems, such as the U.S. patent classification system, would seem to be ideal for the indexing

4,421,744
AMINOACYL DERIVATIVES
James J. Gormley, Holmes Chapel, United Kingdom, assignor to Imperial Chemical Industries PLC, London, England
Filed May 18, 1982, Ser. No. 379,614
Claims priority, application United Kingdom, Jun. 22, 1981, 8119163
Int. Cl.³ A61K 37/00; C07C 103/52
U.S. Cl. 424—177 9 Claims

1. A compound of the formula:



I

wherein:

R^1 stands for an alk-2-enyl, haloalk-2-enyl or alk-2-ynyl radical of not more than 5 carbon atoms, or a furylmethyl or tetrahydrofurylmethyl radical;

R^2 stands for an alk-2-enyl, haloalk-2-enyl, alk-2-ynyl or alkyl radical of not more than 5 carbon atoms, a phenylalkyl radical of not more than 10 carbon atoms, or a furylmethyl or tetrahyrofurylmethyl radical;

or R^1 and R^2 are joined to form, together with the adjacent nitrogen atom, a morpholino, piperidino, methylpiperidino or 1-aza-3,6-methanocycloheptan-1-yl radical;

>N-A stands for the residue of D-, L-, D,L- or aza-tyrosine, phenylalanine or p-aminophenylalanine;

B stands for a single valency bond or for the residue of D-, L- or D,L-, where the amino acid contains a chiral centre, or aza-, glycine, methionine, alanine, serine or sarcosine;

D stands for a single valency bond or for the residue of D-, L- or D,L-, where the amino acid contains a chiral centre, or aza-, glycine, alanine, phenylalanine, sarcosine, serine, O-benzylserine, cysteine or S-benzylcysteine;

E stands for a single valency bond or for the residue of D-,L- or D,L-, where the amino acid contains a chiral centre, or aza-, glycine, phenylalanine, N-methylphenylalanine, p-nitrophenylalanine, p-chlorophenylalanine or tryptophan;

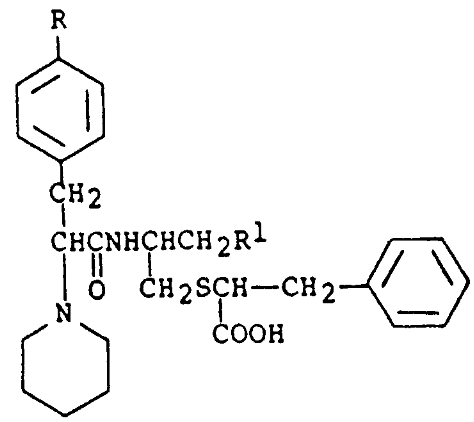
F stands for the residue of D-,L- or D,L-, where the amino acid contains a chiral centre, or aza-, glycine, leucine, methionine, alanine, phenylalanine, proline, serine, O-benzylserine or norleucine, or a dipeptide residue which is D-,L-, D,L- or aza-, leucine-arginine, leucine-glutamic acid, leucine-leucine, leucine-phenylalanine or leucine-threonine; and

X stands for a group of the formula $—CO_2R^3$ or $—CONHR^4$, wherein R^3 stands for hydrogen or an alkyl or alkenyl radical of not more than 4 carbon atoms, and R^4 stands for hydrogen, an alkyl, hydroxyalkyl, cycloalkyl or alkoxycarbonylalkyl radical of not more than 6 carbon atoms, a phenylalkyl or phenyl(hydroxy)alkyl radical of not more than 9 carbon atoms, or a phenyl, phenylcyclopropyl, 2-benzylthioethyl, 2-(2-phenylethylthio)ethyl or indanyl radical;

and wherein the linkages between the amino acid residues are peptide linkages or at least one of said linkages is a pseudo linkage selected from $—CH_2S—$, $—NHCO—$, $—CO.NH.O—$, trans-$CH{=}CH—$ and $—CH_2CH_2—$;

or a pharmaceutically-acceptable salt thereof.

Subgenus of Claim 1
US 4,421,744

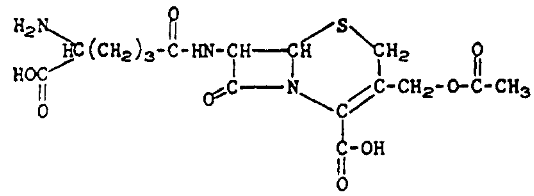


R = H, OH or NH_2

R^1 = H, OH or SH

Figure 3. Variability in Markush structure representation.

and retrieval of innovations without established terminology. To search a hierarchically indexed database, one reads through the definitions in the hierarchy until the first definition that includes the subject of the search is recognized and then scrutinizes every document under that highest priority classification to find out whether the invention is disclosed there. Hierarchical indexing assumes that future developments will fit comfortably into the categories that defined those of the past. Hierarchical systems serve well as indexes of existing inventions; most inventions can be categorized under one or another of the existing classifications, and new classifications can be added to the system to accomodate technological breakthroughs. They are less useful, however, as retrieval tools in searches for prior references to new inventions. The usefulness of such systems breaks down when the novel feature of the search is a low-priority concept in the system's hierarchy. In such cases, the innovative feature can have been combined in the invention with the features that define many of the higher ranking classifications, and all of these will have to be searched. To locate all embodiments of a Markush structure that includes a selection from a group of heterocyclic ring systems, one must usually search subclasses based upon each of the optional ring systems. The usefulness of a hierarchical indexing system also breaks down when the innovation being searched is so novel that none of the predefined categories describe it. The U.S. patent classification code describes only the invention defined by the claims of the patent to which the classification is applied, the classification system is entirely



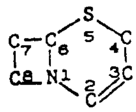
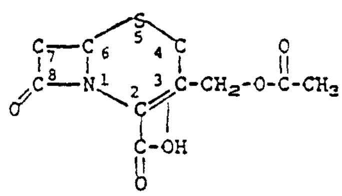
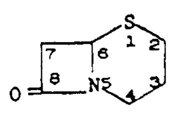
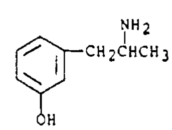
PARENT STRUCTURE	SYSTEMATIC NAME OF CEPHALOSPORIN C
 5-THIA-1-AZABICYCLO[4.2.0]OCT-2-ENE	7-(D-5-AMINO-5-CARBOXYVALERAMIDO)-3-(HYDROXYMETHYL)-8-OXO-5-THIA-1-AZABICYCLO[4.2.0]OCT-2-ENE-2-CARBOXYLIC ACID, ACETATE.
 CEPHALOSPORANIC ACID	7-(5'-AMINO-N'-ADIPAMYL)CEPHALOSPORANIC ACID OR 7-(5-AMINO-5-CARBOXYVALERAMIDO)CEPHALOSPORANIC ACID.
 CEPHAM	7-(5'-AMINOADIPAMIDO)-3-ACETOXYMETHYL-3-CEPHEM-4-CARBOXYLIC ACID.

Figure 4. Cephalosporin C nomenclature.

useless as an index to any details in a patent specification that are not reflected in the claims.

Searchers in the field of chemistry have the advantage that there are formal systems for chemical nomenclature and of chemical structure representation that can describe almost any chemical substance unambiguously. Novel chemical compounds can be indexed by means of topological codes and fragmentation codes that can identify any references to the compounds in the database without retrieving an unreasonable number of false drops. Unfortunately, there are many different nomenclature conventions and structure indexing systems. The searcher must be familiar with all of them in order to retrieve and recognize the compound in question wherever or however it is expressed. For example, as illustrated in Figure 4, there are at least three systems for naming cephalosporin derivatives, not counting the trivial names assigned to many individual cephalosporins. Two of the naming conventions were devised independently by research groups in the early days of cephalosporin research, a case of inventors acting as lexicographers. The other was established by Chemical Abstracts Service, which seeks to create order from the chaos of chemical nomenclature by naming all compounds according to a purely structural convention. It is possible to retrieve any cephalosporin reference indexed in *Chemical Abstracts* without knowing the earlier systems, but without learning both of them, it is impossible to read and understand the references. A retrospective search such as a patentability search imposes the need to recognize obsolete nomenclature as well as modern nomenclature. A searcher for references to the drug gepefrine, a simple and easily named compound, would find that it was claimed in the 1944 patent shown in Figure 5 under a name that seems decidedly quaint to a chemist of the 1980s. The older the reference or the index through which it is found, the

GEPEFRINE



(+)-(S)-2-(2-AMINOPROPYL)PHENOL

Patented Oct. 31, 1944 2,361,373

UNITED STATES PATENT OFFICE

2,361,373

METAHYDROXYBENZYLMETHYLCARBINAMINES AND MEDICINAL PREPARATIONS COMPRISING THE SAME

Gordon A. Alles, Los Angeles, Calif.

No Drawing. Application July 5, 1938, Serial No. 217,584

5 Claims. (Cl. 260—570.8)

I claim:

1. A metahydroxybenzylmethylcarbinamine of the group consisting of metahydroxybenzylmethylcarbinamine and salts thereof.
2. Metahydroxybenzylmethylcarbinamine.
3. A salt of metahydroxybenzylmethylcarbinamine.
4. Metahydroxybenzylmethylcarbinamine bitartrate.
5. Metahydroxybenzylmethylcarbinamine sulfate.

GORDON A. ALLES.

Figure 5.

less likely it is to use systematic, modern nomenclature. At the same time, the older the reference, the less likely it is to have been indexed by means of a topological or fragmentation code that would facilitate recognition of references to the compound.

That these older references are less accessible and less easily recognized than more recent publications does not mean that they can be safely ignored, of course. Part of the paradox is that the emergence of computerized indexing increases, rather than decreases, the complexity of patentabilty searching. The new methods do not relieve the searcher of the need to search with the traditional manual techniques, for modern indexing systems have seldom been applied retrospectively. An exhaustive patentability search performed in 1985 incorporates virtually the whole of the manual search that would have been performed in 1965. The enormous volume of prior art published in the last 20 years must also be searched, but the number of publications has increased to the extent that it is impractical to attempt a manual search of the newer art. If a thorough search is to be completed in a realistic amount of time, the searcher must also master the computerized retrieval techniques that make it possible to search quickly through the output of the information explosion.

A patentability search does not have to be exhaustive if a reference that describes the invention is found. Once a reference has been identified that proves the invention to be unpatentable, the search is over. If no reference is found, however, the searcher is confronted with the final dilemma of the patentabilty search: how does one know when the search is finished?

The ideal of searching universally is unreachable without unlimited resources. Often, the decision to stop searching is reached when the searcher has run out of time or money with which to continue. In such cases the search is over, but it may not be complete. A patentability search is complete when all of the accessible information has been screened by means of an appropriately broad search strategy.

Although it is impossible to examine every possible source of prior disclosures of the invention, it is not difficult for a searcher who is familiar with the principal sources of information in the relevant field of technology to identify the most likely sources. Chemical innovations are likely to have been described in the literature abstracted by one or more of the major abstracting publications. If a prior disclosure existed, there is an excellent chance that it would be found among the publications abstracted by *Chemical Abstracts* or by one of the more selective secondary publications, such as *Current Abstracts of Chemistry and Index Chemicus*, Beilstein, Gmelin, *APILIT*, *Ringdoc*, or Theilheimer,[7] or in a patent from an industrially advanced country. The most useful secondary sources of chemical information are those with comprehensive indexes that can be searched rapidly online or manually. Chemical patents can be found not only in the standard abstract publications but in specialized patent databases, such as Derwent's *Central Patents Index* and IFI/Plenum's *CLAIMS Uniterm and Comprehensive*[8] databases, which index the entire disclosure of the patent in depth and employ controlled thesaurus terms in addition to uncontrolled text or title terms. Any accessible literature that is too old or too obscure to have been included in the secondary publications must be searched manually. Organizations that specialize in a field of technology often maintain libraries of texts and journals and files of patents and technical brochures in that field. These collections are invaluable sources of hard-to-find information. Literature that is not accessible must simply be omitted from the search.

No single database provides access to the whole of the technical literature on any subject. Searching at least two of the likely databases and searching all possible index terms in each database protect the searcher from mistakes in his or her own search strategy, from mistakes in the indexing applied by the database producers, and from gaps in the literature coverage of each database. If the elusive reference is indexed incompletely or incorrectly by one database, or omitted entirely, it may be retrieved through the independent indexing of another database. Each additional database or search strategy that fails to retrieve a pertinent reference decreases the probability that an undiscovered reference to the invention exists.

Having searched the most promising databases by means of the most promising strategies, the searcher must make an educated guess as to the prospects for retrieving a reference by searching one more database or one more patent classification code. The ability to decide with confidence when to stop searching develops with experience, but one can never know with absolute certainty that no pertinent prior art exists.

REFERENCES AND NOTES

(1) Title 35, United States Code, Section 102:

§ 102. Conditions for patentability; novelty and loss of right to patent

A person shall be entitled to a patent unless—

(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for patent, or

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of the application for patent in the United States, or

(c) he has abandoned the invention, or

(d) the invention was first patented or caused to be patented, or was the subject of an inventor's certificate, by the applicant or his legal representatives or assigns in a foreign country prior to the date of the application for patent in this country on an application for patent or inventor's certificate filed more than twelve months before the filing of the application in the United States, or

(e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent, or

(f) he did not himself invent the subject matter sought to be patented, or

(g) before the applicant's invention thereof the invention was made in this country by another who had not abandoned, suppressed, or concealed it. In determining priority of invention there shall be considered not only the respective dates of conception and reduction to practice of the invention, but also the reasonable diligence of one who was first to conceive and last to reduce to practice, from a time prior to conception by the other.

(2) "Beilstein Handbook of Organic Chemistry"; Springer-Verlag: Berlin, Heidelberg, and New York; 4th and 5th supplementary series. "Gmelin Handbook of Inorganic Chemistry"; Springer-Verlag: Berlin, Heidelberg, and New York; 8th ed.

(3) Title 35, United States Code, Section 103:

§ 103. Conditions for patentability; non-obvious subject matter

A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made. . . .

(4) Roth, D. L. "The Role of Subject Expertise in Searching the Chemical Literature...and Pitfalls That Await the Inexperienced Searcher". *Database* **1985**, *8* (1), 43–46.
(5) Lear Siegler, Inc. v. Aeroquip Corp. *U.S. Pat. Q.* **1984**, *221*, 1025–1034.
(6) Fisanick, W. "Requirements for a System for Storage and Search of Markush Structures". In "Computer Handling of Generic Structures"; Barnard, J. M., Ed.; Gower: Hampshire, U.K., 1984; pp 106–127.
(7) *Current Abstracts of Chemistry and Index Chemicus* is published by the Institute for Scientific Information. Philadelphia, and is available as Index Chemicus Online through Telesystems/Questel. *APILIT* is published by the American Petroleum Institute's Central Abstracting and Indexing Service and is available online through ORBIT Search Service. *Ringdoc* is a pharmaceutical literature abstracting service published by Derwent Publications Ltd., London, and is available online through ORBIT Search Service and DIALOG Information Services. *Theilheimer's Synthetic Methods of Organic Chemistry* is published by S. Karger, Basel, Munich, Paris, London, New York, Tokyo, and Sydney, as a yearbook based upon the Chemical Reactions Documentation Service, produced by Derwent Publications Ltd., London, and available online through ORBIT Search Service.
(8) *Central Patents Index* is published in print, microfilm, and magnetic tape forms by Derwent Publications Ltd., London, and is available online as a component of the World Patents Index database through ORBIT Search Service, Telesystems/Questel, and DIALOG Information Services. *CLAIMS Uniterm* is published in print and magnetic tape forms and *CLAIMS Comprehensive Database* is published in magnetic form by IFI/Plenum Data Co., Alexandria, VA, and both are available online through DIALOG Information Services.

A Numerical Index for Characterizing Data Set Separation†

DIANA HUNTER LAFEMINA and PETER C. JURS*

Chemistry Department, The Pennsylvania State University, University Park, Pennsylvania 16802

Received August 14, 1984

A method is reported for assessing the degree of separation between two clusters of points representing chemical data. The method, based on trend vectors, has been tested on both randomly generated and actual data sets. The index is intended to be a quick method for detecting the relative degree of separation between data sets (i.e., suitability of descriptors being used) in structure–activity studies or other pattern-recognition studies.

In pattern-recognition studies, a data set is represented as a set of points in a high- (greater than three) dimensional space whereby clustering, mapping and display, or discriminant development methods are commonly used to investigate the data. One of the basic assumptions underlying pattern recognition is that the distance between points is related to the similarity between points. This assumption applies to structure–activity relation (SAR) studies where the structure of organic compounds of common biological activity are encoded by numerical descriptors and represented as points in high-dimensional space. Compounds of similar activity are expected to be grouped in the same general region of the data space, where the groups may be separated or overlapped. The greater the degree of overlap, the less information can be gained from the system, which reflects the unsuitability of the descriptors being used. There is also the added problem of intuitive judgements being made as to class separability. Visual interpretation of class information is not feasible because most structure–activity studies are done in more than three dimensions. Therefore, a method to quantify the degree of class separability is needed.

This study focuses on the development and use of numerical index for characterizing data sets, analyzing both the effect of size variation and distance separation between data sets. The strategy presented can also be used to assess the degree of similarity between compounds in an assigned class, as well as between a new compound and well-defined classes.

When attempting to understand the structure of a data set that appears to be disordered, clustering routines prove to be very useful.[1] Although there are several ways of defining a cluster,[2,3] distance metrics are often used because the distance between points is a convenient method for establishing similarity between patterns in Euclidean space.[4,5–10] The distances between points can be used to include or exclude a point from a given class. Points grouped in the same class are considered to be similar, while those points that lie outside the class are different.[11]

A variety of cluster-defining techniques have been reported in the literature. Reviews of many of these techniques can be found in Sneath and Sokal,[8,9] Everitt,[12] Hartigan,[2] and Späth.[10] For example, Ling[13] proposed a specific definition of a cluster and also two indices for measuring compactness and relative isolation of these clusters. The nearest-neighbor algorithm of Sneath[14] is another formulation of the same approach. The majority of these clustering techniques calculate similarities and distances between data points so as to summarize information regarding their possible relationships,[12] while we have focused on developing a general numerical index of characterizing data set separation. The trend vector distance ratio index (the R index) yields an initial, quick indication of the relative degree of class separation/overlap (i.e., descriptor suitability) occurring in a data set.

The trend vector distance ratio index is based on the concept of trend vectors,[15] where a trend vector is defined as the Euclidean distance between class centers. The centers are representative of the average coordinates for each class point in n-dimensional space. The trend is then compared to the sum of the average radii for the two classes being studied. The R index provides a measure of class separability and class overlap.

DATA SETS AND METHODOLOGY

In the first step of the study, three data sets (each consisting of five class pairs) were generated to model idealized spherical classes. They provide well-characterized sets in which the effects of separation between class centers and class-size variations could be studied without the added concern of class shape. The data sets were generated to model the two-class situation often found in structure–activity studies. Data set 1 was used to investigate the effect of varying the distance of separation between class centers. Separation distance is defined as the Euclidean distance between two cluster centers. To observe this effect, the size of each cluster was held constant at one standard deviation unit. The ratios of separation to cluster radius were 10:1, 5:1, 3:1, 2:1, and 1:1. These ratios represent separations of one standard deviation (the radius of a cluster) to n standard deviations, where n varies with the extent of separation desired.

† This research was partially supported by Grant CHE-8202620 from the National Science Foundation. The PRIME 750 minicomputer was acquired with the partial financial support of the National Science Foundation.

0095-2338/85/1625-0386$01.50/0