

Inter-Index Patent Searching by Computer*

J. M. CATTLEY

Gulf Research and Development Company, Pittsburgh, Pennsylvania

J. E. MOORE, D. G. BANKS, and P. T. O'LEARY

Gulf Oil Corporation, Pittsburgh, Pennsylvania

Received September 15, 1965

Two indexes, the U. S. Patent Office Classification System and IFI's Uniterm Index to Chemical Patents, were combined into an inter-index searching mechanism. Various examples are given of the use of the merged indexes.

When two indices apply to essentially the same set of documents, the choice of which to use depends upon what information is sought. One question may be best answered by the first index, another by the second. Use of the indices in combination should offer even greater flexibility. All but the simplest can be expressed in terms of requiring two or more features in the same document: the basis of both coordinate searching and of the construction of many hierarchical classifications. When one index defines more closely one of the features sought, the other index another of the features sought, use of the indices in combination should provide more pertinent results, freer of documents of no or peripheral interest.

To exploit the flexibility of inter-index coordination by computer of two indices available on magnetic tape, we worked with the U. S. Patent Office Classification System, an index of hierarchical type, and the Uniterm Index to Chemical Patents, an index of coordinate type published by Information for Industry, Inc. Any other index to U. S. chemical patents could also be used in the same way.

The U. S. Patent Office started classification of patents in the 1830's. The classification systems have undergone several revisions as the number of patents increased and their subject matter widened. The classification system now in use is basically the one adopted in 1912. It consists of more than 300 main classes and about 60,000 subclasses. The chemical patents are found in 48 main classes (all of which are not entirely chemical), either as direct references or cross-references. These classes are listed in Table I, which had appeared in the *Official Gazette*, December 8, 1964. In the study being reported here we worked only with "original" references. We plan use of a similar tape containing both original and cross-references. As direct references, the total number of patents in these 48 classes is 585,000. For 1950-1963, the period we studied, the number of patents is 163,729. These are predominantly of a chemical nature, but some of the patents in these classes are of a nonchemical nature, such as mechanical shaping of material (of plastics, Class 18; of metals, Class 29).

The Uniterm Index of Chemical Patents covers patents from 1950 to date. It is prepared by the careful selection of patents from the *Official Gazette*. These are then reviewed by indexers, who select keywords or descriptors from the patents. The taped form of the index employs about 9200 descriptors. The taped index through 1963

Table I. Classification Data in Magnetic Tape

Class No.	Title	No. of Patents
8	Bleaching and Dyeing; Fluid Treatment and Chemical Modification of Textiles and Fibers..	7,321
18	Plastics	9,583
21	Preserving, Disinfecting and Sterilizing	2,030
23	Chemistry	23,246
29	Metal working	26,429
31	Dairy	1,736
34	Drying and Gas or Vapor Contact With Solids..	8,920
44	Fuel and Igniting Devices	3,873
48	Gas, Heating and Illuminating	8,214
51	Abrading	23,733
55	Gas Separation	9,516
62	Refrigeration	20,022
65	Glass Manufacturing	9,089
71	Chemistry, Fertilizers	2,645
75	Metallurgy	14,590
94	Roads and Pavements	5,732
96	Photographic Chemistry, Processes and Materials	5,850
99	Foods and Beverages	22,696
106	Compositions, Coating or Plastics	19,736
117	Coating: Processes and Miscellaneous Products..	16,231
118	Coating Apparatus	9,542
127	Sugar, Starch and Carbohydrates	2,420
131	Tobacco	9,108
134	Cleaning and Liquid Contact With Solids	5,204
136	Batteries	8,054
146	Vegetable and Meat Cutters and Comminutors..	13,038
148	Metal Treatment	6,194
149	Explosive and Thermic Compositions or Charges..	2,318
156	Adhesive Bonding and Miscellaneous Chemical Manufacture	12,835
159	Concentrating Evaporators	2,483
161	Stock Material and Miscellaneous Articles	7,839
162	Paper Making and Fiber Liberation	7,832
167	Medicines, Poisons and Cosmetics	10,782
176	Nuclear Reactions and Systems	925
195	Chemistry, Fermentation	2,818
196	Mineral Oils: Apparatus	1,654
202	Distillation	7,468
204	Chemistry, Electrical and Wave Energy	12,628
208	Mineral Oils: Processes and Products	12,590
209	Classifying, Separating, and Assorting Solids ..	21,202
210	Liquid Purification or Separation	17,296
233	Centrifugal-Bowel Separators	2,039
252	Compositions	19,351
260	Chemistry, Carbon Compounds	95,832
261	Gas and Liquid Contact Apparatus	12,060
263	Heating	8,896
264	Plastic and Non-Metallic Article Shaping or Treating: Processes	11,450
266	Metallurgical Apparatus	4,455

includes 127,993 patents. The number of descriptors used per patent ranges from 5 or 6 to 150-170, varying with the length of the patent and the amount of material in it. The average number of descriptors is 29. Table II shows the number of patents in the Uniterm Index by year from 1950 to 1963.

Table II. Number of Patents in Uniterm Index, Year by Year, 1950-1963

Year	Approx. no. of patents
1950	6,777
1951	7,273
1952	6,848
1953	6,361
1954	6,134
1955	6,065
1956	11,108
1957	8,833
1958	10,633
1959	11,532
1960	9,802
1961	10,988
1962	13,521
1963	12,270
Total	128,145 ^a

^aBased on accession numbers, a few of which were assigned to patents subsequently withdrawn or indexed exclusively to minor terms.

We are subscribers to the Uniterm Index and have used it on a routine basis for over a year and a half. Early this year we obtained from the U. S. Patent Office on magnetic tape a listing of U. S. patents with original reference to classes of chemical interest. The patents were arranged in numerical order with the original reference listed for each patent. One of our first steps was to compare the Patent Office list of patents with the Uniterm list. What we found is shown in Table III.

The patents which appear in the Uniterm file but not on the Patent Office classification tape can be found as original references in other Patent Office classes which are predominantly nonchemical. It has been estimated that about 40% of these appear as cross-references in the chemical classes. For example, if a patent describes a composition for the lubrication of an internal combustion engine and the claims allowed in the patent merely call for the process of lubricating an internal combustion engine with this composition, its original reference might be to Class 123, Subclass 1, entitled "Internal Combustion Engines-Miscellaneous." It would be cross-referred to suitable chemical classes related to such composition.

The patents which appear on the Patent Office, but not on the Uniterm, tape are probably largely of non-

chemical interest since, as mentioned previously, certain of these classes are mixtures of chemical and mechanical patents. This reflects the complexity of the Patent Office Classification System. As pointed out by J. L. Garland ("Optimizing Information Searches," a chapter in "Information Handling: First Principles," P. W. Hower-ton, Ed., Washington, D. C., 1963), "Its complexity may be inferred from a few of its decision rules. Chemical compounds are classified according to their chemical constitution; processes for making compounds are classified according to the resulting products; manufactured products are classified according to their utility; machines for making products are classified according to their mode of operation and effect." We have analyzed by U. S. Patent Office class the patents issued in 1962-1963, which appear on the Patent Office tape but not on the Uniterm tape. This analysis is shown in Table IV. A study of the nature of these classes shows that the classes that contain large numbers of patents that do not appear on the Uniterm tape are mixtures of mechanical and chemical classes. The purely chemical classes are essentially completely covered in the Uniterm Index.

The classes listed in Table IV can be subdivided into two groups, the first containing classes of chemical-mechanical nature, the second containing classes of essentially purely chemical nature. We have arbitrarily made a subdivision of this type in which each subdivision contains 24 classes. The chemical-mechanical subdivision accounts for 6388 of the patents in the U. S. Patent Office collection but not in the Uniterm Index. The chemical subdivision, on the other hand, accounts for a mere 186 of these patents. Particular examples showing the nature of this distribution are Class 29 on metal working with 1011 patents, Class 62 on refrigeration with 698 patents, and Class 51 on abrading with 594 patents, all these classes containing a high proportion of mechanical patents. Interesting chemical classes are Class 260, on the chemistry of carbon compounds, which accounts for only 18 patents (of a total of about 9000 patents in the period which bear that original classification); Class 252, on compositions, which accounts for only two patents; and Class 96, on photography, which contains no patents not appearing in the Uniterm magnetic tape index. Of the 186 "chemical class" patents not on the Uniterm magnetic tape index, some have been indexed as "minor" terms in a book form of the Uniterm Index.

Tape Layout. Our file of the Uniterm Index is arranged in serial form. Each document appears as an accession number. It is followed by the U. S. patent number, the total number of descriptors assigned to the patent, and a list of the descriptor identification numbers. We have inserted the U. S. Patent Office classification for each document immediately after the patent number. This arrangement is shown in Table V.

This is the present form of our index, which is on three reels of tape for the period 1950-1963. However, several other possibilities exist for arranging this information. A combination inverted-serial file arranged according to Patent Office main and subclasses could be made. This is exemplified in Table VI. This would present the individual patents serially under each subclass with their descriptors. This arrangement might reduce search time considerably, particularly on a computer with random access to

Table III. Comparison of Uniterm and U. S. Patent Office Chemical Files, 1950-1963

Total in Uniterm	Total in USPO	In Uniterm only	In USPO only	In Uniterm and USPO
127,993	163,729	10,494	46,230	117,499

INTER-INDEX PATENT SEARCHING BY COMPUTER

Table IV. Breakdown by Class of Patents on USPO Tape Only

Class	Title	Frequency
8	Bleaching and Dyeing; Fluid Treatment and Chemical Modification of Textiles and Fibers	6
18	Plastics	457
21	Preserving, Disinfecting and Sterilizing	24
23	Chemistry	1
29	Metal Working	1101
31	Dairy	19
34	Drying and Gas or Vapor Contact with Solids	246
44	Fuel and Igniting Devices	0
48	Gas, Heating and Illuminating	0
51	Abrading	594
55	Gas Separation	203
62	Refrigeration	698
65	Glass Manufacturing	110
71	Chemistry, Fertilizers	1
75	Metallurgy	0
94	Roads and Pavements	166
96	Photographic Chemistry, Processes and Materials	0
99	Food and Beverages	260
106	Compositions, Coating or Plastic	0
117	Coating: Processes and Miscellaneous Products	4
118	Coating Apparatus	229
127	Sugar, Starch and Carbohydrates	0
131	Tobacco	151
134	Cleaning and Liquid Contact with Solids	242
136	Batteries	97
146	Vegetable and Meat Cutters and Comminutors	219
148	Metal Treatment	0
149	Explosive and Thermic Compositions or Charges	0
156	Adhesive Bonding and Miscellaneous	
161	Chemical Manufacture	
159	Concentrating Evaporators	25
161	Stock Material and Miscellaneous Articles	77
162	Paper Making and Fiber Liberation	73
167	Medicines, Poisons and Cosmetics	0
176	Nuclear Reactions and Systems	6
195	Chemistry, Fermentation	0
196	Mineral Oils: Apparatus	0
202	Distillation	1
204	Chemistry, Electrical and Wave Energy	1
208	Mineral Oils: Processes and Products	0
209	Classifying, Separating, and Assorting Solids	377
210	Liquid Purification or Separation	307
233	Centrifugal-Bowl Separators	37
252	Compositions	2
260	Chemistry, Carbon Compounds	18
261	Gas and Liquid Contact Apparatus	142
263	Heating	177
264	Plastic and Non-Metallic Article Shaping or Treating: Processes	257
266	Metallurgical Apparatus	85

Table V. Format of Merged Index

A00001 ^a	2492948 ^b	252 ^c	417000 ^d	29 ^e
01342 ^f		01360	12221	12880
16000		16650	16780	20390
23430		24250	28610	28680
30180		32280	43790	50184
58820		59300	59340	59470
61470		65440	66180	66700
69080		73350	73860	76420
77220				

^a Accession number. ^b Patent number. ^c Class. ^d Subclass (this being Subclass 417). ^e Total number of descriptors used. ^f Descriptor number.

Table VI. Format of Possible Inverted-Serial File

252 ^a	417000 ^b
A0001 ^c	2492948 ^d 29 ^e
01342 ^f	01360 12221 12880 16000
16650	16780 20390 ... 77220
A00091	2493494 28
00148	19600 12690 12880 15272
16610	17550 20250 ... 77220
A00097	2493526 31
06980	12880 16000 16260 16580
16650	20390 25380 ... 76420

^a Class. ^b Subclass. ^c Accession number. ^d Patent number. ^e Total number of descriptors used. ^f Descriptor number.

descriptors. In the case of Class 260, which is the most heavily posted with 50,000 patents in the 1950-1964 period, we will probably make several subfiles since there are logical places where this file can be separated.

Frequency Tables. We have prepared frequency tables of the Patent Office classes and of the terms in the vocabulary of the Uniterm Index (Tables VII and VIII). These tables are useful aids in preparing search questions. A sample of the table for Class 260, Subclasses 465.1 to 503, is shown in Table VII. From the table an indication can be obtained of the periods of activity in any art, patent- and research-wise. For example, Class 260, Subclass 467, which relates to acyclic nitrates, contains 116 patents. Of these, 14 issued before 1900, six in the period 1900-1909, four in the period 1910-1919, nine between 1920 and 1929, 21 between 1930 and 1939, 27 between 1940 and 1949, and 35 between 1950 and 1964, inclusive. Looking at Subclass 465.3, which relates to synthesis of nitriles from hydrocarbons, we note that the total number of patents so classified is 151. Of these, none issued before 1940, 47 issued between 1940 and 1949, and 104 issued after 1949.

From these statistics it is evident that some subclasses contain large numbers of patents issued prior to 1950. We have retained in its original form the taped index of Patent Office classes for searching patents issued prior to 1950 in arts where we believe significant information is contained in these earlier patents.

Some statistics on the distribution by class may be of interest. The largest subclass we noticed appears in Class 260, Subclass 461, carbocyclic and acyclic esters

the index file. Testing of an arrangement of this type has been set aside for the present.

Another possible arrangement is the inverted file with accession numbers and patent numbers arranged under Uniterm descriptors and also under the main classes and subclasses.

A further possibility, and one we intend to use, is a breakdown of the entire file into subfiles on separate tapes for the various Patent Office classes or for subjects defined by combinations of Patent Office classes and Uniterm

Table VII. Subclass Frequency Listing for Main Class 260

PAGE 15

SUBCLASS	TOTAL	THRU 1899	1900 1899	1910 1909	1920 1919	1930 1929	1940 1939	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964
465100	71	C	C	C	1	14	20	2	4	1	3	C	C	5	4	4	1	C	2	1	4	5
465200	63	C	C	C	C	9	23	2	1	4	C	1	C	2	6	1	1	2	2	2	0	7
465300	151	C	C	C	C	C	47	7	4	4	6	6	3	12	15	14	2	1	3	7	4	16
465400	117	C	C	C	C	3	38	5	8	6	2	C	4	5	6	6	6	3	4	10	4	7
465500	86	C	C	C	C	2	24	3	5	1	C	C	2	5	6	6	0	5	7	5	7	8
465600	78	C	C	C	C	2	6	32	5	4	1	C	1	5	3	5	2	1	C	2	2	2
465700	55	C	C	C	C	C	27	2	1	C	1	3	1	5	3	2	3	2	C	1	1	7
465800	111	C	C	C	C	C	37	9	7	5	3	5	4	6	8	2	4	3	2	2	5	9
465900	120	C	C	C	C	4	45	7	4	3	3	11	2	6	9	7	C	3	5	1	3	7
466000	27	C	C	C	C	3	5	1	1	1	C	C	1	3	C	3	2	2	C	1	4	0
467000	116	14	6	4	9	21	27	C	2	3	2	5	1	4	C	3	1	3	6	0	4	1
468000	191	1	3	1	1	16	42	6	8	5	3	6	1	2	11	13	14	7	13	17	12	9
468500	63	C	C	C	C	8	13	C	1	4	5	1	2	12	2	5	5	2	C	3	0	0
469000	60	1	1	1	1	5	17	3	1	5	5	3	2	2	4	2	2	C	C	1	3	1
470000	58	1	1	C	2	7	15	3	9	2	9	9	2	3	C	5	1	3	3	6	6	0
471000	214	7	4	3	5	8	31	11	8	5	7	5	9	4	8	19	10	5	8	25	17	11
472000	92	C	5	2	14	9	27	1	4	3	1	1	3	C	2	4	4	4	3	2	1	2
473000	171	2	C	5	1	18	40	4	12	11	4	4	2	9	8	7	3	7	16	5	6	7
473500	42	15	11	7	C	4	C	C	C	C	1	C	C	1	C	2	C	1	C	C	C	0
473600	56	41	3	4	2	1	C	1	2	C	C	2	C	C	1	1	C	0	0	0	0	0
474000	68	9	10	7	3	16	11	C	C	1	1	1	C	1	C	1	3	C	1	2	1	0
475000	355	C	C	C	1	15	58	22	5	9	19	26	7	15	23	19	34	12	14	32	14	15
476000	113	2	2	8	2	12	20	6	6	3	2	5	C	3	1	7	8	7	2	7	8	1
477000	37	C	6	6	1	4	7	2	2	1	1	1	C	1	0	2	1	1	1	C	0	0
478000	11	C	C	C	1	C	3	C	1	0	2	1	C	C	1	C	C	C	1	1	0	0
479000	152	6	11	10	4	11	32	5	1	0	4	3	C	5	7	3	9	C	9	13	8	7
480000	42	C	5	10	4	5	3	C	C	0	C	C	C	0	1	C	0	3	2	3	2	0
481000	86	C	C	C	C	6	31	3	2	5	2	7	2	1	2	6	5	1	5	4	3	1
482000	210	2	5	5	6	9	27	16	6	3	4	4	5	7	10	12	12	11	27	23	12	4
483000	98	C	C	C	4	7	18	4	13	5	9	3	3	4	1	7	8	2	1	5	1	3
484000	199	C	2	2	7	25	80	17	13	7	3	1	5	2	6	5	9	4	2	6	2	1
485000	224	C	3	C	5	18	49	13	13	11	6	7	8	5	16	14	13	4	7	13	12	7
486000	279	C	1	C	C	23	75	8	6	5	6	6	2	9	8	14	18	10	21	29	26	8
487000	77	C	3	C	C	3	17	3	3	2	3	6	C	5	6	12	3	1	2	1	3	3
488000	211	C	2	4	18	42	82	7	10	11	7	8	3	8	15	20	16	7	11	24	12	4
489000	47	2	4	1	4	5	8	1	C	1	C	C	C	0	4	3	1	1	2	4	2	0
490000	51	C	C	C	C	8	12	3	5	1	1	4	2	1	2	3	C	C	C	6	2	1
491000	59	C	C	C	3	12	11	3	6	4	0	3	C	1	3	4	1	C	1	3	3	1
492000	11	C	C	C	1	10	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	0
493000	18	C	C	8	3	2	3	C	C	1	C	C	C	C	0	1	C	C	C	0	0	0
494000	20	C	C	C	5	3	7	1	1	C	C	C	C	C	C	C	C	C	C	1	2	0
495000	18	C	C	C	4	10	2	1	C	0	2	C	C	C	0	1	C	C	C	0	0	0
496000	20	C	C	C	1	10	1	1	C	1	2	1	C	0	C	C	C	C	1	1	1	0
497000	56	C	C	C	1	18	11	5	C	C	1	2	1	3	C	3	C	C	2	3	6	0
498000	44	C	C	3	7	11	12	1	1	C	C	1	1	2	1	C	C	1	1	1	0	1
499000	28	C	C	2	4	11	3	C	C	C	2	C	C	1	0	1	2	C	C	0	2	0
500000	136	C	9	2	4	22	34	0	4	6	5	5	2	3	15	5	4	3	3	4	3	3
501000	216	3	1	1	2	17	47	3	9	6	10	12	6	13	12	16	12	10	8	9	8	11
502000	32	C	7	1	C	1	7	C	C	1	C	C	C	0	4	5	2	C	C	1	0	3
503000	19	C	C	C	C	7	7	1	C	2	C	C	C	2	0	C	C	C	C	0	0	0

and processes of making same, phosphorus-containing acids. This contains 1128 patents, 955 of which issued after 1949. In the other direction, there are some subclasses that contain but one patent for the entire period up to 1963.

Use of Merged File. The indices in combination are searched by the coordinate procedure; that is, a search consists of specifying what descriptors or subclasses must be present or absent on patents to be retrieved. Four types of specification are used. This first type (an "and" group) is a list of Uniterms or Uniterms and a subclass, all of which must appear on each patent retrieved. The second type (an "or" group) is a list of Uniterms, subclasses, or Uniterms and subclasses, at least one of which must appear on each patent retrieved. The other two types of specification provide for selecting patents on the basis of the absence of descriptors or subclasses rather than on their presence. Thus, the third type (a "not any" group) is a list of Uniterms or subclasses, none of which may appear on any patent retrieved. The fourth (a "not all" group) is a list which rejects a patent only if every descriptor (and, if desired, a subclass) in the list appears on it.

Several ways of using the merged file are available. Questions can be asked both using Patent Office classes

and Uniterms as descriptors, merely Uniterms as descriptors or just Patent Office classes as descriptors. In considering the writing of a question, if the person preparing the question is certain of the subclass or subclasses, then he will use them as descriptors either in conjunction, disjunction, or negation. If he is uncertain of the Patent Office classification or wishes to omit it for other reasons, he can merely use the Uniterms. If he should want a listing of patents in any given subclass, he would just use the subclass as a descriptor.

Several aids to use of the U. S. Patent Office classification and the Uniterm vocabulary are available. The "Manual of Classification" is a listing of classes and subclasses in hierarchical order. Table IX shows as a typical sample of this manual, Class 260, Subclasses 410.9 to 518. The "Alphabetical Index to the Classification" guides those unfamiliar with the classification system to subclasses of interest. More detailed definitions of the subclasses are available in "The Class and Subclass Definitions," some of which are out of print but all of which can be consulted in the U. S. Patent Office. A typical page of definitions from the bulletin on Class 260 (No. 200, Revision 3, Classification Bulletin of the United States Patent Office, Class 260—Chemistry, Carbon Compounds) appears as Table X.

INTER-INDEX PATENT SEARCHING BY COMPUTER

Table VIII

DESCRIPTOR AND TEXT	FREQUENCY BY YEAR														PAGE 205	
	50	51	52	53	54	55	56	57	58	59	60	61	62	63		
45553 NICOTINYLATION	0	0	0	0	0	1	0	0	0	0	0	0	0	0		
45556 NIEUWLAND CATALYST	0	0	0	0	2	2	1	0	1	0	1	0	0	0		
45560 NIGHT	0	1	0	1	1	0	0	0	1	2	2	2	0	3		
45570 NIGRE	1	1	1	1	0	1	0	1	0	1	1	1	2	0		
45580 NIP	0	0	0	0	1	0	0	0	3	2	1	7	7	18		
45590 NIPPER	0	0	0	0	0	0	0	0	2	0	1	3	1	0		
45600 NIPPLE	1	0	3	7	0	1	12	3	0	2	2	4	7	10		
45620 NITRATION	34	34	35	27	32	11	30	33	30	58	48	80	45	40		
45630 NITRIC ACID, NITRATE	264	296	325	287	275	179	421	433	576	717	673	647	375	310		
45640 NITRIC OXIDE	8	10	7	10	11	3	30	23	21	29	24	19	32	32		
45650 NITRIDE, NITRIDING	12	10	15	11	16	7	20	41	50	37	47	42	30	30		
45660 NITRILE	115	184	144	87	95	0	0	151	152	173	198	194	80	110		
45680 NITRO, NITRO GROUP	22	41	45	67	52	39	48	173	72	158	135	88	49	88		
45690 NITRO ALCOHOL	5	0	5	0	2	0	3	1	3	3	15	21	7	3		
45700 NITROALKYLATION, NITROALKYL	0	0	0	2	0	3	0	0	2	2	6	12	0	0		
45710 NITROAMINE	0	0	4	0	4	0	0	1	0	2	1	24	2	0		
45720 NITROANILINE AND HYDROCHLORIDE	13	7	16	9	11	5	4	17	14	17	21	13	12	11		
45730 NITROARYL	2	2	2	5	0	0	3	0	0	6	12	9	1	0		
45740 NITROBACTERIA	0	0	0	0	0	0	0	0	0	0	0	1	1	0		
45750 NITROBENZENE	57	50	58	66	45	29	68	55	56	66	75	61	70	49		
45760 P- NITROBENZOIC ACID	6	5	0	3	0	0	3	1	0	0	7	4	5	0		
45764 NITROBENZYL	0	0	0	1	0	0	0	0	1	0	7	0	0	0		
45770 NITROCYCLOHEXANE	0	1	1	3	2	8	9	13	10	7	4	3	4	1		
45780 NITROETHANE	9	9	9	9	10	6	11	12	10	8	7	7	12	6		
45787 NITROGEN 15	0	0	0	0	0	3	0	2	0	0	0	0	0	0		
45790 NITROGEN, NITROGENOUS	284	397	254	181	269	210	314	319	400	530	378	427	1174	769		
45800 NITROGENATION	0	0	0	0	0	0	0	0	1	5	2	10	5	0		

The Uniterm vocabulary is listed in alphabetical order. A KWIC Index prepared by the Dow Chemical Company alphabetizes word roots so that the searcher will become aware of pertinent descriptors including those whose initial syllable would otherwise remove them from proximity to closely related descriptors. For example, "trichlorides" would be near the end of the alphabetical list, whereas "chlorides" would be near the beginning. In the KWIC Index they would both be listed under "chlorides." A page from this index is shown as Table XI.

Our present programs provide for a choice of several types of response to questions, and we plan to have at our disposal in the near future several additional types. Perhaps the simplest type of response is a list of accession numbers, their corresponding patent numbers, and the class and subclass of each patent fulfilling the specifications of the question. This type of answer or printout has the advantage of brevity, but other types, which we shall describe below, seem much more attractive to us, particularly when the answer to a question is voluminous.

In reviewing a printout consisting of a list of accession numbers (each patent in the Uniterm file receives an "accession number" consisting of a letter indicating the

year of issue and five digits), patent numbers, and class and subclasses, the decoding of the subclass numbers, that is, looking them up to see what they mean, has been rather troublesome. To avoid this, we plan in the near future to program the computer to list in the answer to a question those classes and subclasses in which patents of the answer appear. These can then be looked up once and for all, or, better yet, the computer can print out also in English the title of each class and subclass along with its numerical code. A further refinement of this would be the sorting of the patents in the answer by class and subclass so that all the patents in a particular subclass would be printed out together. This type of printout is not obtained at present by us because of the machine sort time which would be necessary.

In more elaborate printouts or responses, the accession number, patent number, and numerical code for class and subclass would be accompanied by the title in English of the class and subclass for each of the patents in the answer. This type of printout would be particularly useful in answers to questions defined in terms of Uniterms or Uniterms and subclasses. Perhaps the most complete type of printout, and one which would perhaps eliminate

Table IX. Class 260, Chemistry, Carbon Compounds

CARBOCYCLIC OR ACYCLIC		CARBOCYCLIC OR ACYCLIC	
	Fats, fatty oils, ester-type waxes or higher fatty acids	461	Esters and processes of making same
	Synthetically produced higher fatty esters	462	Phosphorus-containing acid
410.9	With acyclic monohydric alcohol	463	Boron-containing acid
412	Recovery or extraction from residues or organic material	464	Carbonic or chlorcarbonic acid
		465	Nitrile or isonitrile
412.1	From fish livers	465.1	Aromatic
412.2	From legumes, nuts or seeds	465.2	Acyclic
412.3	With acidic or basic material		From non-oxo-carbonylic compounds (through the amide)
412.4	Solvent extraction	465.3	From hydrocarbons
412.5	Foots, textile treating liquor, absorbents, sludges or industrial wastes	465.4	Non-oxo-carbonylic
		465.5	Amino
412.6	Rendering	465.6	Oxy
412.7	With acidic or basic material	465.7	Halogen
412.8	Solvent extraction	465.8	Polynitrile
413	Higher fatty acids or their salts	465.9	Unsaturated
414	Heavy metal or aluminum containing	466	Nitric or nitrous acid
415	By hydrolysis of fats, oils, waxes or esters	467	Acyclic nitrates
416	With organic sulfonic acids	468	Carboxylic acid
417	With alkali-forming metal compound	468.5	Hydrophenanthrene nucleus containing
418	With subsequent treatment	469	Aromatic
419	Purification	470	Sulfur-containing
420	Purification of fats, oils and waxes	471	Amino or nitro
421	By esterification of free fatty acids	472	Alkamine
423	By treatment with oxidizing agents	473	Oxy
424	By treatment with acid, acid anhydride or acid salt	473.5	Tannins and reaction products thereof
		473.6	With extraction from bark or vegetable material
425	By treatment with basic substance		Salicylic acid
426	With organic solvent	474	Polycarboxylic acid
427	With solid absorbent	475	Monocyclic
428	By physical treatment only	476	Alkamine
428.5	With solvent	477	Acyclic
429	Heavy metal containing	478	With phenols
429.1	Actinide series metals	479	Salicylic acid (e.g., aspirin)
429.2	Lanthanide series rare earth metals	480	Sulfur containing
429.3	Zirconium and hafnium	481	Amino
429.5	Titanium	482	Oxo
429.7	Tin	483	Oxy
429.9	Zinc	484	Polycarboxylic acid
430	Gold or silver	485	Unsaturated acid
431	Mercury	486	Halogenated acid
432	With other heavy metal	487	Unsubstituted acids of the acetic series
433	Aromatic	488	With terpenes
434	Carbonylic	489	Alkamine (e.g., choline acetate)
435	Lead	490	Esters from esters
436	Acetates	491	From alkyl sulfates
437	C-Pb (e.g., plumbanes)	492	From alkyl halides
438	Copper or chromium	493	Esters from aldehydes
439	Iron, nickel or cobalt	494	Esters by dehydrogenation of alcohols
440	Arsenic	495	Esters from ethers
441	With other heavy metal	496	Esters from olefines
442	Arsonic acids	497	Esters from acetylene
443	Carbonylic	498	Purification or recovery
444	Arseno	499	Acids or salts thereof
445	Carbonylic	500	Amine addition salts (including betaines)
446	Antimony	501	Peroxide
447	Bismuth	502	Sulfonic acids
448	Aluminum containing	503	Hydrophenanthrene nucleus containing
448.2	Silicon-containing	503.5	Sulfonation products of non-aromatic hydrocarbon mixtures
448.8	Silicon-containing-acid esters	504	Aromatic
449	Hydrogenation of carbon oxides and products thereof	505	Diaminomethylene
		506	Non-oxo-carbonylic
449.5	Methanol synthesis	507	Amino
449.6	With iron, cobalt or nickel catalyst	508	Oxy
450	With subsequent treatment	509	Polyamino
451	Partial oxidation of non-aromatic hydrocarbon mixtures and products thereof	510	Oxo
		511	Oxy
452	With subsequent treatment	512	Acyclic
453	Esters and processes of making same	513	Carboxylic acids
454	Thiocyanic acid or isothiocyanic acid	514	Hydrophenanthrene nucleus containing
455	Thiocarbonic acid or thiocarboxylic acid	514.5	Aromatic
456	Sulfoxy acid	515	Bivalent sulfur containing
457	Sulfuric acid	516	Diarylketones
458	Acyclic	517	Amino
459	Unsubstituted alkyl	518	
460	From olefines		

491

INTER-INDEX PATENT SEARCHING BY COMPUTER

Table X. Classification Definitions

Sub. 239.5.

(9) Note. See this class, subclass 705, for extraction processes of general application in the field of carbon compounds.

236.6. Extracts under subclass 236.5 which are essential oils, including reaction products thereof, which are not more specifically provided for below. Essential oil reaction products which are esters, alcohols, ethers, aldehydes, ketones, and acids are under those respective headings in this class. See, for example, subclasses 489, 587+, 631 and 631.5.

(1) Note. This subclass includes the removal of odoriferous oils from plants, flowers, etc., enfleurage.

SEARCH THIS CLASS, SUBCLASS-

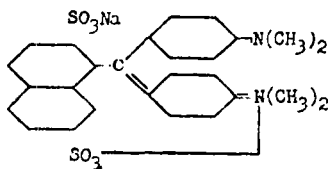
631.5, for pine oils per se and their purification treatments.

675.5, for turpentine per se and their purification treatments.

SEARCH CLASS-

202, Distillation, subclass 46+, for the removal of essential oils from natural sources by distillation.

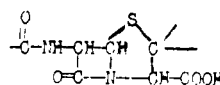
239. Carbon compounds which contain a ring composed of carbon and at least one element from the group consisting of nitrogen, sulfur, selenium, tellurium, and oxygen. Excluded herefrom are those carbon compounds wherein the only heterocyclic nucleus present is produced by salt formation between amino and acid groups, e.g., betaines, which are placed with the corresponding open chain compounds, particularly subclass 501, and



which is found in subclass 388.

the need to review patent claims, would contain, in addition to the accession number, patent number, class, and its English title, a list of the Uniterms used in characterizing the patent. This would contain all the information

239.1. Compounds under subclass 239 containing the grouping:



and derivatives thereof.

(1) Note. This subclass includes, for example, the various penicillins and their salts, addition products, esters, amides and substitution products.

SEARCH CLASS-

167, Medicines, Poisons, and Cosmetics, subclass 65, for medicinal compositions containing penicillins and their derivatives

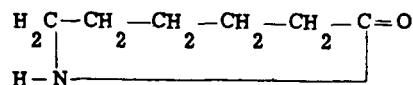
195, Chemistry, Fermentation, subclass 36, for processes of producing penicillin by mold fermentation of carbohydrates.

239.3. Compounds under subclass 239 which contain in the ring the H X group



(X=O or S) and in which the ring contains at least seven members.

(1) Note. This subclass contains for example:



SEARCH THIS CLASS, SUBCLASS:

294.7, for delta lactams, e.g., alpha piperidone.

326.5, for gamma lactams, e.g., alpha pyrrolidone.

239.5. Compounds under subclass 239 which contain the cyclopentano-phenanthrene nucleus.

SEARCH THIS CLASS, SUBCLASS-

397+, for non-heterocyclic cyclopentano-phenanthrene compounds.

available on the index tape plus the English titles corresponding to Uniterm numbers and subclass numbers, this English being supplied from auxiliary tapes. Ultimately, we shall print out in addition to this information

Table XI

KNIC INDEX OF IFI VOCABULARY	GR+DC	DECEMBER 18, 1964	PAGE	79
POTASSIUM CHLORIDE			55400	
PHOSPHORUS OXY CHLORIDE			52410	
PHOSPHORUS PENTA CHLORIDE			52430	
PHOSPHORUS TRI CHLORIDE			52450	
PLATINUM CHLORIDE			53695	
OXALYL CHLORIDE			49300	
OXY CHLORIDE			49510	
PALLADIUM CHLORIDE			49810	
PENTA CHLORIDE			50620	
NICKEL CHLORIDE			45430	
NICKELOUS CHLORIDE			45495	
NITROSYL CHLORIDE			45930	
MOLYBDENUM PENTA CHLORIDE			44040	
MERCURIC CHLORIDE			41830	
METAL CHLORIDE			41985	
METHACRYLYL CHLORIDE			42204	
METHALLYL CHLORIDE			42230	
METHANE PHOSPHONIC DI CHLORIDE			42255	
METHANE PHOSPHONYL DI CHLORIDE			42257	
METHANE SULFONYL CHLORIDE			42270	
LITHIUM CHLORIDE			39840	
MAGNESIUM CHLORIDE			40650	
MANGANESE CHLORIDE			41090	
MANGANOUS CHLORIDE			41170	
LEAD CHLORIDE			38870	
IODINE CHLORIDE			36941	
IRON CHLORIDE			37110	
HYDRAZINIUM CHLORIDE			33810	
HYDROGEN CHLORIDE			34230	
HEXA CHLORIDE			32703	
GOLD CHLORIDE			31020	
FERRIC CHLORIDE			27600	
FERROUS CHLORIDE			27780	
ETHYL ALUMINUM DI CHLORIDE			26420	
ETHYL ALUMINUM SESQUI CHLORIDE			26430	
DI ETHYL ALUMINUM CHLORIDE			22040	
DI ETHYL AMINO ETHYL CHLORIDE AND HYDROCHLORIDE			22100	
ETA-DIMETHYLAMINOETHYL CHLORIDE			22780	
CUPRIC CHLORIDE			18200	
CUPROUS CHLORIDE			18260	
CYANOGEN CHLORIDE			18560	
CYANURIC CHLORIDE			18500	
COBALTOUS CHLORIDE			15410	
COLUMBIUM PENTA CHLORIDE			15350	
COPPER CHLORIDE			15470	

the assignee to whom each of the patents has been assigned. A printout of this type is illustrated in Table XI-A. We feel that printouts of this type should make it possible for the questioner to select the very documents that he is most interested in reading and eliminate the need for his sifting or having someone sift for him the computer output from books of patent claims. A careful reading of Table XI-A bears this out.

As mentioned before, we maintain the Patent Office classification as a separate tape and use it to get a list of the U. S. patents in any chemical subclass prior to 1950, if desired, for anyone wishing to extend a search prior to 1950.

Example Questions. Typical questions using Uniterm descriptors, using subclasses as descriptors, and using both Uniterms and subclasses as descriptors are shown in Tables XII, XIII and XIV.

The subject of each of these questions was the use of tetramethyllead as a gasoline additive. In the first question, therefore (Table XII), we require that any document retrieved bear the descriptor 69605, tetramethyllead. The tetramethyllead must, however, to be of significance in the answer to our question, be employed in gasoline or motor fuel of some type, and, accordingly, we also require that the document have at least one of the descriptors: 30230, gasoline; 17525, cracked gasoline; 44316, motor fuel; 44790, naphtha; 39400, light naphtha; 32360, heavy naphtha; 17528, cracked naphtha, cracked oil; or 75380, virgin naphtha, virgin oil.

Consider next the second question (Table XIII) with only Patent Office subclasses. In the Patent Office classification, Classes 44—Fuel and Igniting Devices, 252—Compositions, and 260—Chemistry, Carbon Compounds, are of interest. In Class 44, Subclass 69, Artificial Fuel-Liquid-Heavy Metal or Aluminum-Containing-Organic Compound-Lead Containing, seems most likely to contain

Table XI-A. One Possible Reply or Printout Type

(OR) 252	428000	Compositions-catalysts or solid adsorbents-organic compound containing-with inorganic material (except water)
(XR) 252	475000	Compositions-catalysts or solid adsorbents-metal, metal oxide, or hydroxide containing-Group II metal
(XR) 260	498000	Chemistry, carbon compounds-carbocyclic or acyclic-esters and processes of making same-carbocyclic acid-acyclic-unsubstituted acids of the acetic series-esters from acetylene
	00255	Acetic acid, acetate
	00400	Acetylene
	00750	Activated carbon, activated charcoal
	12221	Carbon, carbonaceous
	12890	Catalyst-production
	26170	Ester
	27400	Fatty acid
	43790	Mixture
	46190	Noncaking, nonagglomerating, nonclumping
	49380	Oxide
	67950	Support
	74460	Vapor
	74860	Vinyl acetate
	75020	Vinyl ester
	76890	Zinc acetate
	76940	Zinc compounds, zinc salts
	77030	Zinc oxide

Subject of patent as deduced from subclass titles and descriptors: A new catalyst has been developed for making vinyl acetate (or other vinyl esters) by reacting acetic acid (or another carboxylic acid) with acetylene. The catalyst is a zinc compound on a support, such as activated carbon.

Claim 1, for comparison with description deduced from subclass titles and descriptors: The method of preparation of a catalyst for the manufacture of a vinyl ester of an aliphatic monocarboxylic acid containing 1 to 4 carbon atoms from acetylene and the vapor of said acid, comprising dry-mixing zinc oxide powder and active carbon, contacting the mixture with at least 2 moles of said acid in liquid form per mole of zinc oxide and less than 8 parts of water for each part of the zinc oxide, and drying the catalyst.

Table XII. IFI Search for Tetramethyllead in Gasoline

U. S. Chemical Patents File—IFI Uniterm
Index Search Request MEPB 1

Each Document Must Have All of the Following:

69605 Tetramethyllead

Each Document Must Have at Least One Term From:

30230 Gasoline

17525 Cracked Gasoline

44316 Motor Fuel

44790 Naphtha

39400 Light Naphtha

32360 Heavy Naphtha

17528 Cracked Naphtha, Cracked Oil

75380 Virgin Naphtha, Virgin Oil

the information we want. However, our experience in using the U. S. Patent Office classification has taught us that Subclass 76, Artificial Fuel-Liquid-Sulfur, Seleni-

Table XIII. USPO Class Search for Tetramethyllead in Gasoline

U. S. Chemical Patents File—IFI Uniterm
Index Search Request MEPB 2

Each Document Must Have at Least One Term From:

C044
D069000
D076000
C252
D386000
C260
D437000

um, Tellurium, Silicon, Phosphorus or Boron-Containing, also contains patents describing use of organic lead additives in motor fuels in combination with the sulfur-, selenium-, tellurium-, silicon-, phosphorus-, or boron-containing additives. In Class 252, Subclass 386, Preservative Agents—Antiknock or Antiexplosion, may also contain a reference to the use of tetramethyllead in motor fuel. In Class 260, Subclass 437, Carbocyclic or Acyclic—Heavy Metal Containing—Lead—C—Pb (*e.g.*, Plumbanes), relating to the chemistry or organolead compounds containing a lead-carbon bond, may also include a patent which refers to the use of tetramethyllead in gasoline. Hence, all these subclasses have been specified in an “or” group in the questions.

Finally, Table XIV shows a question phrased with both class and subclass descriptors and with a Uniterm descriptor. The same subclasses have been used here as in the previous question. However, by use, in addition, of the Uniterm descriptor 69605, tetramethyllead, we will obtain in the reply to this question only patents which mention specifically tetramethyllead, and we will eliminate those which mention other organolead derivatives but not tetramethyllead.

Table XIV. Inter-Index Search for Tetramethyllead in Gasoline

U. S. Chemical Patents File—IFI Uniterm
Index Search Request MEPB 3

Each Document Must Have at Least One Term From:

C044
D069000
D076000
C252
D38600
C260
D437000Each Document Must Have All of the Following:
69605 Tetramethyllead

A word of explanation is due on the class and subclass symbols. The letter “C” preceding three digits signifies a main class. When the class number contains only one or two digits, the appropriate number of zeros is filled in to the left to make a three-digit number, as shown in Tables XIII and XIV for Class 44.

The letter “D” signifies the subclass. Subclasses come in decimal form with up to three digits to the left of the decimal point and, as far as we can determine, up to two digits to the right of the decimal point. The Patent

Office, however, has allowed for three digits to the right of the decimal point, and, accordingly, we employ a six-digit number to represent subclasses, the first three digits representing the digits to the left of the decimal point and the last three digits to the right of the decimal point. With this distinction between classes and subclasses, our programmers have eliminated the necessity of rewriting the class each time a subclass belonging to the same main class is used in a question. Thus, in the examples shown in Tables XIII and XIV, D06900 and D07600 both designate subclasses of Class 44 even though C044 is written but once. This saves time both in writing and keypunching, particularly when a large number of subclasses belonging to one main class is used in phrasing a question.

When a list of cross-references in tape form has been merged with the Uniterm Index, a project now underway at our laboratory, it will not be necessary to list so many subclasses as we have had to in the examples discussed above. It should be emphasized, however, that by using a combination of class and subclasses and Uniterms, the burden of reviewing all the patents in each of the subclasses has been lifted from the searcher by the selective function performed by the Uniterms.

Performance of the Indices. Some idea of how the indices in combination perform in comparison with their separate performance can be had from Table XV. This lists seven searches, each on a different and unrelated subject and each performed in three ways: by using Uniterm descriptors alone, by using U. S. Patent Office subclasses alone, and by using a combination of Uniterms and subclasses. Searches A, E, and F are novelty searches. Searches B, C, D, and G were made to find out the state of the art. For each search made in each way, we have tabulated the total number of patent numbers retrieved and for those which we carefully reviewed the number of patents in that total which were pertinent to the subject. In the righthand-most column is tabulated the total number of pertinent patents we are aware of. This represents a minimum of pertinent patents. There may be others we have not learned about.

How use of indices in combination can increase selectivity is shown by search A. Here, there were four pertinent patents in a total of 13 retrieved in the search of the indices in combination. In using Uniterm descriptors

Table XV.^a Inter-Index vs. Single-Index Searching

Search and period	Number of patents retrieved						
	IFI		USPO		IFI and USPO		Minimum pertinent
	Total	Pertinent	Total	Pertinent	Total	Pertinent	
A 62-63	123	3	187	4	13	4	4
B 50-63	6	6	88	4	4	4	6
C 50-63	14	0	954	^b	7	0	^b
D 50-63	782	^b	32	^b	16	14	14
E 62-63	49	11	86	6	7	6	14
F 50-63	41	1	201	^b	1	1	1
G 50-63	354	^b	795	^b	107	^b	^b

^a This table summarizes the character of the results of seven unrelated searches performed in three ways. The “Minimum pertinent” column tabulates the number of pertinent patents we were aware of. There may be others we have not learned about.

^b Not determined.

alone, however, the total output was 123, which would have had to be screened to locate three of the pertinent patents. With U. S. Patent Office classes it would have been necessary to screen 187 to find four pertinent patents.

In some instances the performance of the combination is hardly superior to that of one of the indices alone. Search B illustrates that nicely, for in the Uniterm search a total of six out of six pertinent patents were retrieved with no patents not relating to the subject searched. The indices in combination yielded a selective output, but two patents of interest were not retrieved. These had been cross-indexed to the subclass specified in this search. Therefore, had we at the time this search was made the subclasses available on magnetic tape, a total of six pertinent patents would have been retrieved by that combination of indices. In this instance the use of subclasses alone, however, would have required the review of 88 patents to find four (or with cross-references, six out of more than 88) which were pertinent.

In search C the absence of pertinent patents may seem surprising in a state-of-the-art search. In explanation, the search was for producing a type of ester by a certain reaction. We had learned from people familiar with this art that it would be unlikely that this reaction had been carried out successfully. The results of the search confirmed their prediction.

In search D the indices in combination performed very well. We found in the output enough information to satisfy the questioner, and hence the outputs from the indices used separately were not reviewed.

In search E, we may not have phrased our questions as well as we should have. Interestingly, we found from a review of the patents retrieved in each of the three types of searches that there were at least 14 pertinent patents. In the Uniterm search, 11 of these were retrieved out of a total of 49 in the Uniterm output. Our use of the subclasses resulted in our retrieving only six pertinent patents out of a total of 86 printed in the output. By using the indices in combination, we retrieved the same six and only one nonpertinent patent.

In search F the use of the combination of indices retrieved one patent, which was adequate for our requirements. Accordingly, we did not review the voluminous output from use of the Patent Office subclasses. We did review the somewhat smaller Uniterm output to find out that only one of the total of 41 printed out was pertinent. That patent was the same one which we retrieved by the use of the indices in combination. It was, of course, also among the 201 patents retrieved by the use of subclasses alone.

Search G is an example of a broad state-of-the-art search. It shows a limitation of output by the use of the indices in combination. We have not, however, reviewed the outputs of the three searches and so cannot comment further on their quality.

This table may be looked on as an indication of what can be done by using combinations of indices. It does not, and should not be interpreted to, show the ultimate which can be achieved in this way. The performance of one or the other index or the combination of indices depends upon how they are used and the questioner's understanding of the indexing philosophy. As experience in their use accumulates, the effectiveness of the indices

should improve. In addition, when cross-references to U. S. Patent Office classes have been incorporated in the index tape, the use of the indices in combination should improve considerably.

Index Term Correlation. While our primary purpose in the merging of these indices was to test their performance in combination for searching, this project has given rise to interesting by-products. For example, the merged index has allowed us to analyze the information content of subclasses in terms of Uniterms or the information content of Uniterms in terms of subclasses. We shall present a few examples below.

Term Tallies. As we mentioned above, our search program allows us to segregate on a separate tape (we call a save tape) those patents which are defined by logical combinations of Uniterms or subclasses or Uniterms and subclasses. Another program allows us to list the subset of the vocabulary used in indexing the patents on a save tape and to tabulate the number of patents in the save tape to which each of the descriptors, Uniterm or subclass, in that subset has been applied. A tabulation of this type we call a "term tally."

One example of this type of procedure is illustrated by the obtaining of a save tape on "pharmaceuticals." This was defined by requiring that all patents retrieved for the save tape be in Class 167, Subclass 65, on organic medicines, or that it bear one of the Uniterms "pharmaceuticals" or "pharmaceuticals-synthesis." The resulting save tape contained 12,642 patents. A term tally showed that 7131 Uniterms had been used in indexing these patents. A subclass tally, analogous to a term tally, showed that 953 subclasses were represented by the patents on the save tape. This tally also listed the number of save-tape patents in each of the subclasses.

From the subclass tally, which listed the subclass and the number of patents in it, a tabulation was made of the number of subclasses containing only one patent, the number of subclasses containing only two save-tape patents, etc. This is shown in Table XVI. It is interesting to note from this tabulation what a large number, *viz.* 953, of subclasses is represented by this subject of pharmaceuticals.

An examination of the definitions of the 953 subclasses showed that many of these represented subjects remote from the pharmaceutical art. It seemed desirable to eliminate from the save tape the patents represented by these

Table XVI. Distribution of "Pharmaceutical" Patents among Subclasses

Number of "pharmaceutical" patents	Number of sub- classes containing that many patents
1	402
2	105
3	60
4	37
5	32
6-10	91
11-20	87
21-50	80
51-100	36
>100	32
Total 12642	Total 953

subclasses, and this was subsequently done to obtain a new, refined save tape more directly pertinent to the subject of inquiry. This resulted from elimination of 479 subclasses to leave a save tape with 474 subclasses represented. The number of documents on the refined save tape was 10,759, and a term tally of the refined save tape showed that a total of 6201 Uniterms had been used to characterize the refined save-tape patents. A sample of a page from the term tally appears as Table XVII.

Table XVII. Term Tally for "Pharmaceuticals" Save Tape

Uniterm number	Uniterm	Frequency of application to patent
04460	Anthelmintics, Nematocide	88
04470	Anthracene, Anthracene Oil	14
04490	Anthranilic Acid	20
04500	Anthraquinone	13
04510	Antiacetylcholine	7
04520	Antiadhesive, Antisticking	5
04530	Antiaging	4
04550	Antialdosterone	16
04560	Antiallergenic	53
04575	Antianabolism	1
04580	Antiandrogen	46
04590	Antianemia	13
04600	Antiarthritic	148
04610	Antibacterial, Antimicrobial	114
04615	Antibarbiturate	2
04620	Antibiotics	1058
04640	Antibody	16
04650	Anticholesterogenic	27
04660	Anticholinesterase	19
04670	Anticoagulant	78
04680	Anticonvulsant, Spasmolytic	444
04700	Anticorrosion, Antifouling	20
04710	Antidepressant	22
04720	Antidesoxycorticosterone	4
04740	Antidiabetic	14
04750	Antidote	9
04760	Antiedemic	7
04770	Antiemetic	86
04780	Antienzyme	37
04790	Antiepilepsy	32
04800	Antiestrogen	106
04825	Antifertility	5
04830	Antifibrillatory	30
04840	Antifibromatogenic	1
04850	Antifoaming	30
04860	Antifogging	2
04865	Antifolliculoid	5
04870	Antifreeze	4
04890	Antigen	37
04897	Antiglucocorticoid	2

Aside from its statistical function, the term tally is an alphabetical index to the patent art represented on the save tape. Thus, people interested in the subject defined by the save tape may be provided with copies of the term tally, which they may use in framing questions for the save tape; for the save tape can be searched in the same way as can the entire file. Advantageously, however, the searching of the save tape takes considerably less computer time.

For small save tapes containing few documents, say 1000, a dual dictionary can be compiled by the computer for the save tape, and copies of this dictionary may be supplied to those interested in the subject on the save tape. They may then clerically search *via* that dual dictionary the subject represented by the save tape.

Term Class Correlation. In a hierarchical classification it is sometimes difficult to know precisely where to look for the information one wants. To some extent, cross-referencing serves as a guide here. However, we have found that an excellent insight into the contents of subclasses and the classification philosophy can be provided by counting the number of members of the intersection of the sets represented by a subclass and by a Uniterm. In fact, our programmers have provided for tabulating arrays of such intersections in a matrix form so that classification patterns become evident. An example may make this clearer.

U. S. Patent Office Class 44 is entitled "Fuel and Igniting Devices." Subclasses 50 to 80 on "Liquid Fuels" are of interest to petroleum chemists. These subclasses are defined in terms of the composition of fuel additives, not in terms of the type of fuel, that is, the nature of the petroleum fraction or other material. It would be of interest to find out which of the types of additive represented by the subclasses are designed for gasolines, for example, which for diesel fuels, which for jet fuels, etc. To get an answer to these questions, we tabulate the number of patents in intersections of these subclasses with Uniterms designating types of fuel, for example, gasoline, diesel fuel, jet fuel, kerosene, fuel oil, residual fuel, etc. Our present program allows for 25 subclasses in a matrix of these intersections and up to 100 Uniterms. In addition, if we desire, we can tabulate similarly as a matrix the number of patents in the intersections of Uniterms. In this instance, the matrix can have a size up to 75 by 100. The combination of two matrices, 75 Uniterms by 100 Uniterms, and 25 subclasses by the same 100 Uniterms, can be compiled in the same computer run. An example, in fact that which we discussed on Class 44, appears as Table XVIII. This table shows the

Table XVIII.^a Subclass-Uniterm Correlation for Selected Subclasses in Class 44, Fuels and Igniting Devices

Subclass	Artificial fuel-liquid	Gasoline	Fuel oil	Diesel fuel	Residual oil	Jet	Anti-oxidant	Antiknock	Anti-stalling	Detergent
57	With ignition promoter	1	4	17	0	4	0	1	0	0
63	Heterocyclic carbon compound	72	20	15	1	13	14	17	5	2
66	Fat, fatty oil, fatty oil acid, or derivative	41	22	9	2	6	1	8	10	4
69	Lead-containing organic compound	108	3	4	1	2	13	91	1	1
76	S, Se, Te, Si, P, or B containing	15	7	4	0	6	4	4	2	3

^aThis table shows the number of patents in each subclass with the descriptors shown. For example, in Subclass 57, one patent bears the descriptor, "gasoline," four patents the descriptor, "fuel oil," etc.

distribution among 25 of the 31 subclasses between subclass 50 and subclass 80 of patents on the various types of fuel, such as gasoline, diesel fuel, kerosene, jet fuel, etc., and also the distribution among these subclasses of patents on additive functions.

A similar company-class or company-Uniterm correlation will show the areas of activity of any companies

selected or which companies have been active in the area represented by a particular subclass or group of related subclasses. The same can be done by company-Uniterm correlations. By company, we mean, of course, the assignee recorded for the patents in the U. S. Patent Office and Uniterm files. These will be available shortly on magnetic tape.

The Indexing of Technical Books*

J. FREDERIC WALKER
Box 27, Haddam, Connecticut

and ROBERT F. SCHIRMER
E. I. du Pont de Nemours and Company, Inc., Information Systems
Division, Secretary's Department, Wilmington, Delaware

Received September 7, 1965

The indexing system designed for the ACS monograph "Formaldehyde" (3rd edition) is described. The system, a modified concept coordination, uses a limited number of roles identified by mnemonics.

In spite of recent advances in the technology of retrieving scientific information, the method of indexing chemical texts has remained relatively static. Accordingly, in the preparation of the third edition of the ACS monograph, "Formaldehyde" ("Formaldehyde" III) (7), it was decided to employ some of the techniques used in concept-coordination information systems with the object of obtaining a more informative and compact index. At the same time, data-processing methods were employed to simplify and reduce the work involved.

In the following discussion, the index resulting from this work will be described and compared with the conventional index used in the second edition of "Formaldehyde" ("Formaldehyde" II) (6). Methods of index preparation including an improved method of assembly developed by Professor L. F. Fieser (2) will be reviewed.

"FORMALDEHYDE" III INDEX

A major innovation in this index is the use of a limited number of letters to indicate the function or role of the chemicals, reagents, or materials indexed in specific page references. These function or role indicators (5) serve in the place of wordy subentries to designate pages containing information on the material indexed as a product, reactant, catalyst, etc. In addition, small separate indexes were prepared for special chemical reactions, property data

on the various forms of formaldehyde, and end use applications.

Before proceeding further, it should be noted that some use of symbols to designate the type of information covered by page references has already appeared in technical books. The use of boldface type to indicate a major page reference is old and well known. The employment of special symbols is noted in Fieser and Fieser's "Organic Chemistry" (3rd ed) (3) in which an asterisk is used to denote pages containing physical constants and a dagger sign to designate pages with descriptions of methods of synthesis. Use of multi-part indexes is found in the collective volumes of "Organic Syntheses" (4).

The composition of the index of the 660-page text of "Formaldehyde" III is summarized in Table I. The Author Index lists all authors in alphabetical order followed by page numbers in numerical order. Chapter reference numbers, if any, precede the page numbers and are in parentheses. This author-reference number-page number system is standard in many texts.

Table I

Section	Items	Printed pages
Author Index	2955	22
Subject Index		
Explanatory preface		1
(1) Chemicals	1600	15
(2) Reactions	36	0.5
(3) Properties	37	1
(4) Applications	125	1.5

* Presented before the Division of Chemical Literature, 150th National Meeting of the American Chemical Society, Atlantic City, N. J., Sept. 13, 1965.