to decide whether other searches are needed. We can turn our IFI patent search requests around in 24 hours if that is needed. The database is used for all levels of searching: "quick and dirty", "find some background information", novelty search, validity search, infringement search, and state-of-the-art search.

We have been subscribers to this database since its creation in 1971. Prior to that time we were subscribers to the Uniterm tape and the printed Uniterm Index. We have had the advantage of growing with the file. Our decision to purchase the Comprehensive Database in 1972 was based on our need to have superior patent coverage in the fields of polymer chemistry and synthetic organic chemistry and also to permit more cost-effective use of professional time in providing search services. We also believed that the price of the merged database was an information bargain. Our experiences of the past five years have not altered our original position. The present annual subscription cost is high, and we do monitor our costs of operation each year to evaluate the cost effectiveness of this database to our total information service. However, intangible benefits are realized in having any database in-house; i.e., the search work is held confidential, the searcher is very familiar with the information needs of the client, and the searches can be restructured or up-dated at a reasonable cost. These factors also are considered in maintaining this database in-house.

In conclusion, we believe the IFI Comprehensive Database is a cost effective means of providing our R&D staff with a fast, comprehensive in-house search service for the United States patent literature.

LITERATURE CITED

(1) J. M. Cattley, et al., "Retrieving Patents by Weighted-term Searching", *Chem. Eng. Prog.*, **62**, 91–96 (Oct 1966).
(2) P. T. O'Leary et al., "Computer Searching of Chemical Patents", *J. Chem. Doc.*, **5**, 233–238 (1965).
(3) L. E. Rasmussen and J. G. Van Oot, "Operation of DuPont's Central Patent Index", *J. Chem. Doc.*, **9**, 201–206 (1969).
(4) M. Z. Balent and J. M. Emberger, "A Unique Fragmentation System for Indexing Patent Literature", *J. Chem. Inf. Comput. Sci.*, **15**, 100–104 (1975).
(5) Private communication, John W. Lotz, IFI/Plenum Data Co.
(6) Private communication, Patrick T. O'Leary, Gulf Research and Development Co.

A User's Experience with the Derwent Patent Files†

STUART M. KABACK

Information Research and Analysis Unit, Research Corporate Services, Exxon Research and Engineering Company, Linden, New Jersey 07036

Received April 20, 1977

The Central Patents Index-World Patents Index System, produced by Derwent Publications, Ltd., provides a wide range of products and services, ranging from alerting and in-depth abstracting through retrospective information retrieval. All chemically related patents have been covered since 1970, and nonchemical patents since 1974, but portions of the file go back as far as 1963. Multiple tools using a variety of parameters are available for information retrieval, and the recent availability of on-line search capability has increased substantially the utility of the Derwent files. In summary, the CPI and WPI constitute a powerful information resource which works superbly in some areas, less effectively in others.

The family of Derwent's patent information services forms one of the most complex information systems offered to the public today. There are alerting services in the form of expanded titles, short abstracts, and long abstracts, multiply packaged and repackaged. There is retrieval via classified browsing files, various printed indexes, punch card sorting, computer tapes, and, since early 1976, on-line interactive searching. In the on-line files alone the subject retrieval parameters include Derwent classes, manual code classes, international patent classes, multipunch codes, and title keywords. Wherever you approach the Derwent system there are generally multiple options—options that may tend to confuse even the most experienced Derwent user, let alone the uninitiated.

Derwent's Central Patents Index and World Patents Index cover virtually all of the world's chemically related patent publications, as well as a large percentage of nonchemical patents. Table I shows something of the volume and character of this patent coverage. Derwent's estimates for the current year put the annual total, chemical and nonchemical, at about 500 000 patent documents. Nearly half of these are "basics"—patent cases new to the data base. Most of the remainder are equivalent to references already in the system, although a small proportion are patents from so-called "minor" countries which are given partial coverage. About half of all

Table I. Derwent Patent Coverage, 1977 (est)

	CPI	WPI only	Total
Total basics	104 000	127 500	231 500
Total equivalents	147 000	100 500	247 500
"Minor" country patents not fully in system			16 000
Grand total	251 000	228 000	495 000

Table II. CPI-WPI Country Coverage

Full Coverage					
Belgium	(BE)	West Germany	(DT)	Soviet Union	(SU)
Canada	(CA)	France	(FR)	United States	(US)
Switzerland	(CH)	United Kingdom	(GB)	South Africa	(ZA)
East Germany	(DL)	Netherlands	(NL)		
Full Chemical Coverage					
Japan	(JA)				
Partial Coverage					
Brazil	(BR)	Israel	(IL)	Rumania	(RU)
Czechoslovakia	(CS)	Norway	(NO)	Finland	(SF)
Denmark	(DK)	Austria	(OE)	Sweden	(SW)
Hungary	(HU)	Portugal	(PT)		

† Presented in the symposium on "Meeting the Challenges of the Changing Patent Literature", Division of Chemical Information, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 21, 1977, and 11th Middle Atlantic Regional Meeting of the American Chemical Society, Newark, Del., April 20, 1977.

the patents involved in the system have some chemically related aspect and are included in the Central Patents Index, or CPI. Nonchemical patents go into the World Patents Index, WPI, but even here there is a nomenclature complication, because the WPI designation is also used for the combined on-line file, chemical and nonchemical. Further, as we shall see shortly, there is a WPI alerting gazette covering the chemical area. In general, though, the term CPI will be used here to refer to the chemical database, and WPI the nonchemical database.

Table II shows country coverage in the CPI–WPI system. All patents, chemical and nonchemical, are fully covered for 11 "major" countries. Japan issues approximately 200 000 examined patents and unexamined Kokai per year, so many that only the chemical ones are covered. For 11 additional countries, titles and bibliographic details are noted for alerting purposes, but the patents are not abstracted and do not enter the retrieval system unless they are equivalent to cases that have appeared in one of the "major" countries. Note that Derwent abstracts and indexes do not use the full names of the countries involved, but rather the two-letter standard ICIREPAT abbreviations. Many of these abbreviations are self-evident, but some, such as SF or CH, may not be. This is one of the many little challenges of the Derwent system.

In any event, the volume of information involved is large, very large. It's offered in a variety of forms, and accessed in a variety of ways. In using any tool it is essential to know just what that tool's characteristics and capabilities are. This is certainly true for information tools and, in the case of the CPI–WPI system with its manifold complexities, mastery does not come easily. The author has years of working with Derwent products and still looks forward to the day when he will finally feel that he has tamed the tiger, but until that day he will still find this to be an invaluable information source.

A brief look at the history of the Derwent organization and its products is in order, because the present character of Derwent products is highly dependent on that history. Derwent got its start in 1951 by recognizing a need for rapidly produced alerting abstracts of British patents. Over the next dozen years or so, similar products were begun for a number of other countries. Up to this point, alerting was the only focus; no means were provided for retrospective retrieval.

The jump into the world of retrieval took place in 1963, with the start of Farmdoc, which covered pharmaceutical patents. In 1965 there was Agdoc, on pesticide and fertilizer patents, and in 1966 Plasdoc, on polymers. Literature services born in the 60's, most notably Ringdoc, are outside of the scope of this paper.

Farmdoc, Agdoc, and Plasdoc had two main methods for retrieving subject information: manual codes and multipunch codes. Manual codes were developed to form classified files of abstracts to be searched by browsing. The manual code systems have been subject to various modifications and subdivisions over the years, to adapt them to changing and growing technology.

The multipunch codes provided a deep indexing system, based on information in the full patent specification, originally for use on punch card sorters and later adapted to computer tape searching. The Farmdoc–Agdoc code is primarily a structural fragmentation code, coupled with terms describing compound properties and uses. The Plasdoc multipunch code covers all aspects of polymer information. These codes, too, have been modified as needed over the years.

In 1970, Derwent expanded its documentation services to cover the entire range of chemistry and chemical technology. Farmdoc, Agdoc, and Plasdoc became three of twelve sections of the new Central Patents Index. Manual code retrieval was developed for each of the new CPI sections, and the Farmdoc–Agdoc multipunch code was adapted to Chemdoc, the general chemical section of CPI. No comparable retrieval system was devised for the remaining eight CPI sections. When the nonchemical WPI services were started in 1974, only international patent classification was used for subject retrieval.

Table III. CPI Sections

	Basics, 1977 (est)
A: PLASDOC (POLYMERS)	32 600
B: FARMDOC (PHARMACEUTICALS)	9 000
C: AGDOC (AGRICULTURALS)	5 100
D: FOOD, FERMENTATION, DETERGENTS, COSMETICS, LEATHER, WATER TREATMENT, ETC.	10 800
E: CHEMDOC (GENERAL CHEMICAL)	19 400
F: TEXTILES, PAPER, CELLULOSE	9 600
G: PRINTING, COATING, PHOTOGRAPHIC	7 800
H: PETROLEUM	5 100
J: CHEMICAL ENGINEERING	10 400
K: NUCLEAR, EXPLOSIVES, PROTECTION	2 500
L: GLASS, REFRACTORIES, CERAMICS, ELECTROCHEMISTRY	12 500
M: METALLURGY	18 000

Table IV. WPI Sections

P1-P3	HUMAN NECESSITIES
P4-P7	PERFORMING OPERATIONS
Q1-Q4	TRANSPORT, CONSTRUCTION
Q5-Q7	MECHANICAL ENGINEERING
R1-R3	INSTRUMENTATION
R4-R5	ELECTRICAL ENGINEERING

As a result of this history, the various sections of CPI and WPI have sharply differing retrieval capabilities. The reason for this disparity is basically economic: the most elaborate retrieval techniques have been developed in areas in which there were sufficient subscribers, prepared to pay sufficient money, to finance the system. Derwent is an entrepreneurial venture, in business to make a profit. It has provided subscribers with just what they have been willing or able to pay for. At times one misses the refinements that might come from the support of government or of a prestigious scientific society, but, on the other hand, neither of these produced a Central Patents Index. It took Derwent, the entrepreneur, to do that.

Let us look now more closely at the constituents of the CPI and WPI, how they work, where they work best, and where they might be improved. As already noted, the CPI covers the full range of chemistry and chemical technology. The volume of patents covered is very large, and the cost for handling this large volume of material is correspondingly high, particularly if there is to be any great sophistication involved in the retrieval system. The only viable way to handle this was by offering CPI in a series of independent sections, since few organizations are large enough, or have broad enough interests, to be able to justify acquiring the whole package. Table III shows the twelve CPI sections, along with an indication of their relative sizes, based on Derwent projections for the current year. The size varies widely, from a high of over 30 000 for Plasdoc down to about 2500 for section K, a hybrid section including nuclear technology, explosives, and protection technology. The estimate for total basics, 104 000, is somewhat greater than the 98 000 covered in 1976, the previous high water mark.

Besides K there are three other hybrid sections: D, G, and L. One could suggest alternative ways in which some of these sections might have been combined, but the fact is that they are as they are. Subscribers can purchase portions of the file without having to buy the whole thing and, if the portions offered do not necessarily correspond to one's exact needs, it is simply impossible to please everybody.

The arrangement of the nonchemical material in the WPI is shown in Table IV. The three main sections, for general,

mechanical, and electrical patents, are each split into two portions for the purpose of issuing abstract booklets. Thus, the nonchemical WPI consists essentially of six subsections.

Patents received by Derwent are checked to see whether they are new to the system or equivalent to known patents. The checking is done primarily on the basis of priority dates and application numbers. Non-convention patents are also checked, and Derwent frequently succeeds in identifying non-convention equivalents. Derwent's patent concordances are quite excellent, and the on-line file is especially fine for the retrieval of families of patents, since it can be entered with any member of a patent family, or the priority application number, to retrieve directly the family as it exists in nearly two dozen countries.

Coverage of chemically related patents is, to all intents and purposes, complete. Before the start of the WPI there were always some instances of missed patents, especially in the early days of the system. Now, since nonchemical patents do go into the WPI, there is added insurance that—if the patent exists—it is in the system somewhere. If you follow CPI and WPI assiduously to keep up with new developments, you can be quite sure that they will all be covered.

New patents are classified into one or more broad Derwent classes, which are used in allocating each patent to one or more sections of the CPI–WPI system. CPI classification is done by Derwent personnel. WPI classification, on the other hand, is based on international patent classes, as assigned by issuing patent offices. A given patent may appear in as many as four sections of the CPI. The average is about 1.5 sections per patent. There are definitely times when one wishes for a more generous multiple classification. This seems to happen most often with inventions that might go into both Chemdoc and one of the other sections.

For example, new compounds claimed for pharmaceutical or agricultural utility are placed in Farmdoc and/or Agdoc, but not in Chemdoc. They will be missed by the Chemdoc user who might be interested in them for other purposes. A patent on a pinene derivative made by hydroformylation appeared only in Farmdoc and was missed by someone following oxo process patents in Chemdoc. In general there is a bias in CPI to concentrating on end products, rather than starting materials or processes, a bias that clearly traces to the fact that chemical retrieval was built on a system designed for pharmaceuticals and pesticides. For another example in the process vein, processes involving a single hydrocarbon feed or product go just into Chemdoc, while similar processes using mixed hydrocarbon feeds go in the Petroleum section. An example is the hydrodealkylation of toluene and of mixed aromatics. There are obvious economies for Derwent in this restricted classification, but also obvious potential pitfalls for information users.

In short, coverage of chemical patents is quite complete, but the user must check carefully where he looks for his information. It may turn up where he did not expect it. Derwent does publish coverage guidelines, and users are advised to study these well.

Completeness is one of the very great virtues of CPI and WPI. A second great virtue is timeliness—consistent timeliness. Within 5 to 6 weeks of their publication, all patents from most key countries are covered in a WPI gazette—one for each of the general, mechanical, and electrical sections, and one that corresponds to the full chemical content of the CPI. It is very valuable to have all patents for each country covered at the same time, rather than having some taken rapidly and the rest strung out over an extended period.

Listings in the WPI gazettes appear in two sections: one organized by patentee and one by international patent class. Thus, they can be scanned for key companies and for subject

```
SLMB *                                  83684X/45 *FR 2296-858
Dipmeter measurements in boreholes - for determination of angle of dip
and azimuth of strata traversed by borehole (BR240876)
   SOC PROSP SCHLU  30.12.74 US-537998 (30.12.74 US-537310)
   H01 R11 +R19  (03.09.76) G01c-09 G01v-03/18
```

Actual Patent Title: Method and Apparatus for Dip Measurement.

Actual Patentee Name: Societe de prospection electrique Schlumberger

Figure 1. Sample WPI heading.

```
93402X/50                H06                    MITP 23.04.75
MITSUBISHI OIL KK (TOKM-)                       *J5 1124-106
   23.04.75 JA-043703 (27.10.76) C10L-01/32
Stable water in oil emulsion fuel - with high water content dispersed
uniformly as small particles
```

TOKM- Means patent joint with TOKYO MACHINERIES

Figure 2. Joint patentee ambiguity.

J7 6012-345 = JAPAN EXAMINED 76 12345

J5 1123-456 = JAPAN KOKAI 76 123456

Figure 3. Japanese patent numbering.

matter concepts. A sample WPI listing shown in Figure 1 has been selected carefully to show a number of features.

On the first line there are a patentee code, an accession number, and the French patent number. The stars show that this is a basic, not an equivalent.

Next comes the title, which is a great deal more informative than the original patent title, as you can see. The title is informative enough to form a mini-abstract. After the title is the rather cryptic indication that the earliest publication of this patent was in Brazil, on August 24, 1976.

The next line gives an overabbreviated listing of the patentee name, followed by the earliest and latest priorities for this patent. If there were any additional priorities they would not appear here. The truncated name is an obvious weak point, and one wishes that Derwent allotted more space for this data element.

The final line gives Derwent classes, the publication date in France, and the international patent classes.

A great deal of information is crammed into a very small space, accomplished in part by the use of a highly telegraphic style that can be confusing. Other confusing things can also appear in the headings. Thus, in Figure 2, it is hardly apparent that the TOKM- following the patentee refers to a joint patentee, Tokyo Machineries. Or consider Figure 3, which shows how Derwent distinguishes between examined and unexamined Japanese publication numbers. Examined numbers are presented in the Christian year format, unexamined in the Japanese Imperial year format, with the numbers field displaced into the country field to account for the fact that more than 100 000 Kokai issue per year. One shudders to think of what will happen when Japan has a new Emperor and the Imperial year numbering changes.

By and large, the telegraphic style does not really present problems for information specialists, who will take the trouble to learn what all of the data elements mean. The amount of information conveyed is gratifying and, indeed, there is one valuable item that really should be added: the names of inventors, where available. But for the bench chemist or engineer, or the patent attorney, these data elements seem to create a barrier. Derwent does publish in each booklet guides to the interpretation of the headings, but our experience at Exxon indicates that few users seem to consult these. The average user would be greatly helped by more explicit presentation of the information in the headings.

The basic material from the WPI gazette is then used, in slightly rearranged form, as the heading for a series of abstract publications, each of which appears one week after its predecessor: relatively short alerting abstracts, arranged first by country and then in subject groupings by Derwent class; longer basic, documentation abstracts arranged by country;

Table V. Available Alerting Formats

WPI Gazette	Heading only
Alerting Abstracts, Country Order	Short abstracts, arranged by country; equivalents reabstracted
Alerting Abstracts, Classified	Same short abstracts, in broad subject groupings; only examined equivalents reabstracted
Basic Abstract Journal	Documentation abstracts, arranged by country; manual code index included
Basic Abstract Profiles	Subgroupings of large sections, mainly from Plasdoc and Chemdoc, also catalysis

CUST- H09 *84719X/45 *US 3988-121
Fire-lighters of cellulose powder and wax compsn. - with partial surface coating of mixt. of cellulose powder and long-chain diacyl peroxide
CUSTOM RES & DEV. 09.01.76-US-647948 (19.02.74-US-443565)
E14 (26.10.76) C10l-05/02 C10l-11
A firelighter tablet consists of (a) a mixt. contg. a wax with m.pt. 115-200°F (115-180°F) and a cellulosic wood powder which forms 40-80 wt.% of the mixt., and (b) on the surface of the tablet, 0.02-1 (0.05-0.5) esp. 0.05-0.3 wt.% w.r.t. the wt. of the tablet, of a mixt. contg. 25-90 wt.% of a <20C diacyl peroxide, and a cellulosic wood powder. The tablets ignite easily and burn for a sufficient time to kindle logs or synthetic or natural solid fuel. The tablet retains its structure during packaging and storage. 9.1.76 as 647948 (5pp).

Figure 4. Sample alerting abstract.

and, in polymers and some other high interest areas, basic abstracts arranged in subject profile groupings. Table V lists the five formats which can be used for alerting purposes, and Figures 4 and 5 show the comparison between a typical alerting abstract and the documentation abstract of the same patent. As you can see, the documentation abstract adds a great deal of useful information. Unfortunately the volume of Japanese documents, coupled with language problems, leads to a much lower abstract quality for Japanese publications. Economics seem to make this nearly unavoidable, at least for the foreseeable future.

Given these multiple formats, which should one use for alerting purposes? It is hard to give cut and dried answers, because each has its advantages. The WPI gazette format, with heading only, is quite useful for picking up red-hot items at the earliest possible date. We do not consider it optimum as a general alerting tool, however, since the expanded titles, informative as they are, merely whet the appetite for more.

Country order alerting booklets are aimed mainly at patent attorneys, who may want to keep up not only with the issuance of various patent cases, but the countries in which they appear. The chemist or engineer is better off with the classified alerting booklet, which eliminates duplicate abstracts for equivalents and arranges things in subject groupings. Basic abstract journals are clearly more informative and are strongly recommended, but many users will find their volume excessive, particularly in Plasdoc or Chemdoc. They are further handicapped by not being in subject order. Users who can narrow their interests to one or two profile booklets can get around these problems. For those organizations that are prepared to take the effort, custom repackaging of Derwent basic abstracts can be highly effective in meeting their alerting information needs.

An alternative alerting method is, of course, computerized SDI. Derwent does provide computer tapes for in-house SDI, but our organization has no experience with these. Undoubtedly an on-line SDI facility will be developed before too long, but at present economics have prevented the start of such a service.

Summing up the alerting situation, completeness, timeliness, and excellent titles and abstracts, especially the basic abstracts, make the CPI especially strong here. The telegraphic heading style can cause problems for the nonspecialist user, and the completeness (which is, of course, a great virtue) can also cause problems by making some sections large enough to scare off some users. The multiplicity of packaging formats may also be confusing, making it important for information specialists to do an effective job of informing their user communities of just what is available, and how it can best be used. It should be noted that decisions taken at the 1977 Derwent subscribers' meeting should help correct some of the problem areas noted here.

Let us turn now to the subject of retrospective retrieval of information from CPI and WPI. The manual codes are available across the entire CPI. They are used to provide files of basic abstracts to search through, and the quality of the basic abstracts is one of their greatest strengths. A collection of clear, detailed summaries of related technology may not be the most modern method of information retrieval, but it is still a highly effective one. The browsing human mind can spot concepts that could never be articulated to your friendly computer and, even if they could, chances are that they might not have been indexed. There are problems, though, associated with the manual codes. Efforts have been made to subdivide codes that produce large numbers of items, but nevertheless some classes are exceedingly large and take a long time to

84719X/45 H09 E14 CUST- 19.02.74
CUSTOM RES & DEV *US 3988-121
09.01.76-US-647948 (+443565) (26.10.76) C10l-05/02 C10l-11
Fire-lighters of cellulose powder and wax compsn. - with partial surface coating of mixt. of cellulose powder and long-chain diacyl peroxide

Fire kindling tablets are claimed comprising (a) a mixt. of a wax of m.pt. 115-200°F (115-180°F) with 40-80 wt.% of cellulosic powder, and (b) on the surface of the tablet, a coating, in an amount of 0.05-0.5 wt.% (w.r.t. the wt. of the tablet), and containing a mixt. of cellulosic powder with 25-90 wt.% of a <20C diacyl peroxide.

ADVANTAGES

The tablets ignite easily and burn for a sufficient time to kindle logs or other natural synthetic solid fuels. The tablet retains its structure during packaging and storage.

DETAILS

The cellulosic powder may be from wood (prefd.), leaves, straw, nuthulls, shells or rice hulls, in a particle size of 50-500 (50-200) US mesh. The prefd. amt. of cellulosic powder in mixt. (a) is partic. 50-75 wt.%. The wax may be natural animal or mineral wax, petroleum wax (prefd.) or synthetic wax. Paraffin wax and/or microcrystalline wax are partic. prefd., the wax contg. < 9.5 wt.% oil. The peroxides are generally 20-40C (20-28C), free from aromatic unsatn. and pref. free from aliphatic unsatn. Lauroyl peroxide is specifically claimed. The mixt. (b) pref. contains 50-85 (65-80) wt.% peroxide. In the prepn. of the tablets, the mixt. (a) is compressed to give the required shape. Mixt. (b) may be placed in the tabletting machine before and/or after mixt. (a), forming a tablet with peroxide at one or both ends. Alternatively a rod of mixt (a) may be cut into lengths, and painted or sprayed with, or rolled in mixt. (b), or the tablet may have slots into which mixt. (b) is compressed. The size of the tablet is esp. 0.5-4 inches, and the tablet is partic. cylindrical.

F(10-A4) H(8-A, 9-F).

EXAMPLE

A 1:1 mixt. of 100 mesh wood powder and 138°F melting paraffin wax was compressed at 700 psi for 5 seconds to form a tablet with dia. 1.25 inches and length 0.5 inch. A mixt. of lauroyl peroxide and wood powder was introduced before and after the wood powder-wax mixt. The ratio of peroxide to wood powder was varied and the time to ignition, and the time to burn over the top of the tablet were determined. The results were as follows:

Peroxide (g)	wood powder (g)	time to ignition (seconds)	time to burn over the top (seconds)
1	1	5	45
1.5	0.5	1	25
0.5	1.5	15	70
2	-	1	43
-	2	24	235

(5 pp.).

US 3988121/95

Figure 5. Documentation abstract of same patent.

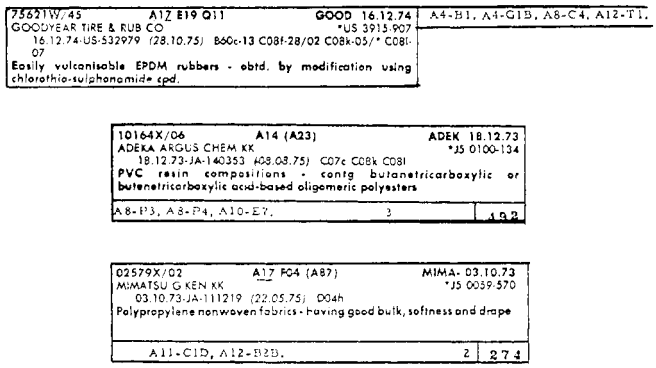


Figure 6. Manual code "underclassification".

Table VI. Retrieval of Polypropylene Preparation, 1975

Total Relevant Patents: 92

	IPC C08F-10/06	Manual codes A4-G3A	Manual codes A4-G1A	Manual codes Combined	Multipunch 050/A +347/A
Total retrieved	13	34	234	268	115
Hits	10	24	61	85	86
Relevance, %	77	71	26	32	74
Recall, %	11	26	66	92	93

search. More of a problem, however, is underclassification. Consider the three items in Figure 6. The first, on vulcanizable EPDM rubber compositions, would appear to belong in the class for ethylene–olefin copolymer compositions, but in fact appears just in the generic class for olefin polymer compounding. Users are cautioned to be careful to check generic manual code classes, as well as the most specific ones.

The next two examples, though, do not have an easy solution: a PVC composition not in the class on PVC compositions, and a polypropylene use not in the class for polypropylene uses. Neither is placed in any class based on composition, and both will be missed in searches for these polymers. Derwent would help users considerably by being more liberal in the use of manual codes.

Manual code cards take time to file and take space to store. We find that they are well worth the time and space. They can be obtained on microfilm, but the cards themselves should be used if at all possible. It is far easier and quicker to search cards than to search on microfilm.

We had no experience with multipunch code retrieval until the advent of the on-line file, but have used multipunches extensively since then. The on-line file has greatly enhanced retrieval from Derwent's database. There are many searchable parameters in this file. By all means, familiarize yourselves with them and use them, because some searches respond well to some retrieval methods, others to an entirely different set of methods. Here are some concrete examples.

In one test, we retrieved patents from 1975 on the preparation of solid polypropylene homopolymer. Results are summarized in Table VI.

Title keywords were not used since experience suggested that they would not work too well. International patent class was a failure, with poor recall. Both manual codes and multipunch codes gave excellent recall, the latter with very high relevance. Each retrieved some items not found by the other. Thus, the multipunch code found several patents coded only for the polymerization catalyst in the manual code system. The manual code search picked up several items whose abstracts were not explicit enough to multipunch code them for propylene as a monomer, but which seemed to a browsing searcher to be relevant to polypropylene. Thinking humans can do things a computer cannot. Note again the need to search specific *and* generic classes in the manual code.

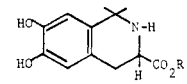
Table VII. Retrieval of Oxo Processes, 1975–76

	IPC	Title words + manual codes	Multipunch	All 3
1975: 64 Relevant Patents				
Total retrieved	40	48	104	
Hits	31	37	36	59
Relevance, %	78	77	35	
Recall, %	48	58	56	92
1976: 69 Relevant Patents				
Total retrieved	50	57	86	
Hits	32	47	31	58
Relevance, %	64	82	36	
Recall, %	46	68	45	84

IPC terms: C07c-045/08, 10, 12; C07c-027/20, 22, 24; C07c-029/16, 18

Title words: Hydroformylat:; Oxonat:; Carbonylat:; Oxo + (E10-D01: or E10-E04: or Reaction # or Synth:)

Multipunch: 55&/E3 (forming C-C bonds) + 480/E3 (aliphatic aldehyde)



E3 MULTIPUNCH POSITIONS:

195 Hydrogenated isoquinoline
493 Aromatic multihydroxy
51& Heterocyclic ester
59- One fused heterocyclic
60- No monocyclic heterocyclic
62- No carbocyclic aromatic
620 No alicyclic
713 Fused heterocyclic compound

Figure 7. Complex compound multipunch retrieval.

Look now at an entirely different situation: 1975 and 1976 patents on a chemical process, the hydroformylation of olefins, shown in Table VII. Here direct manual code searching was not tried, because previous experience had shown a great deal of manual code scatter for this subject. Three methods were used again: international patent class, a computer combination of title keywords with manual codes, and multipunch codes. This time international patent class was considerably more useful then before. Title keywords showed up best for this search, and even the multipunch code gave useful recall, though with lower relevance—this in a process area that is not its great strength. Each retrieval method found some items not found by any other. All three in combination retrieved most, but not all, of known references.

As already mentioned, the Farmdoc–Agdoc–Chemdoc multipunch code is geared to the retrieval of chemical compounds, especially the fairly complicated compounds that one finds in pharmaceuticals and agriculturals. Thus, a test retrieval of a dihydroxy tetrahydroisoquinoline ester in a known patent led at a first pass to just eight references, including the desired one, and computer elimination of unwanted functional groups cut this further to three references, again including the correct patent. This is shown in Figure 7. On the other hand, a test run for allene, shown in Figure 8, retrieved nearly 300 references for a $4^1/_2$-year period, and a spot check showed that very few of these were actual hits. Allene was selected for this test because its two unconjugated double bonds in a three-carbon molecule promised to make it a relatively good simple compound for the system, but it turned out to give unac-

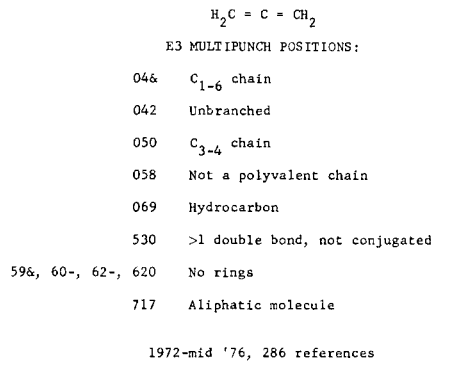


Figure 8. Simple compound multipunch retrieval.

ceptably low relevance. The multipunch code may be very effective for complicated molecules, but it is relatively ineffective for simple ones.

Other retrieval problems stem from the fact that products are indexed, but not starting materials, again reflecting the interests of the end-product-oriented pharmaceutical and agricultural industries. It is very difficult to retrieve information on the utilization of a given starting material. Catalyst retrieval, too, is sketchy at the moment. For example, an attempt to isolate rhodium catalysis in the oxo test search gave only middling results. Hopefully the new catalysis manual codes introduced this year will help.

We saw how effective title keywords were for retrieval in the oxo example. They would have been even more effective had it not been for a number of typographical errors: hydroformalation, hydroformulation, and hydro-formylation. The misspelling rate in Derwent titles is far too high, and the unstandardized use of various punctuation marks causes one to miss relevant items. New hyphenation rules promise to help the punctuation problem. Users should certainly try to truncate terms wherever possible, to allow for variant endings, misspellings near the end of the word, and hyphenated combinations. Use the "Neighbor" command liberally, too. But of course misspellings at the start of words are going to be missed no matter what you do, and there is a great need for Derwent to tighten up on the proofreading of their abstract headings. One more point for searchers: remember to use both American and British spellings.

Tests are underway on the addition of indexing keywords to Plasdoc, and, while this is welcome, it would seem more important to add this feature in other sections first, since Plasdoc retrieval facilities are strong in comparison with those for most other sections. In particular, Chemdoc needs better retrieval of starting materials, processes, and simple compounds.

One of the problems in retrieval stems from the fact that changes in indexing have occurred over the years. Improved indexing is, of course, beneficial to users, but the searcher often has to devise different search strategies for different time periods.

Given a system as complex as the CPI, one of the greatest needs for users is detailed instructional materials. In fact the telegraphic style of Derwent headings extends to their instruction manuals, and leaves much unexplained. Recent communiques from Derwent have acknowledged that fact and promised more detailed manuals for the future. These will be very welcome, as will a promised listing of past multipunch codings, which will be an invaluable aid in devising multipunch search strategy.

Retrieval, then, presents more problems than does alerting. Nevertheless, there is considerable retrieval capability, and it has been greatly enhanced by the advent of the on-line file. There is evidence that titles, always a Derwent strong point, have improved and will continue to be improved. Hopefully the keywording experiment will succeed and will be extended to additional sections of CPI. In the meantime, CPI searchers are advised strongly to make use of all of the available, complementary, search parameters.

This paper has not attempted to touch on every aspect of the Derwent products, much less explore them all in detail. Its aim has been to present a balanced picture of the most important strengths and limitations of the system. If there has been a stress on some of those limitations, it has been in the spirit of striving for improvements in a powerful but imperfect information resource. Let there be no misunderstanding: based on extensive experience we regard the Derwent CPI–WPI system as an invaluable information asset, indispensible for anyone who needs to know about the new technology of the present and the recent past.

On-Line Retrieval of Chemical Patent Information. An Overview and a Brief Comparison of Three Major Files†

ROGER GRANT SMITH,* LOUISE P. ANDERSON, and SUSAN K. JACKSON
Merck & Co., Inc., Rahway, New Jersey 07065

Received April 21, 1977

Databases which contain patent information and are available to the public via on-line computer networks are listed. A comparison of document and retrieval parameters is made for those on-line databases which provide comprehensive coverage of the chemical patent literature, i.e., CHEMCON, CLAIMS™, and WPI. Examples of representative searches run in each of these three databases are presented and discussed. The conclusion is reached that no one database may be relied upon to provide all possible relevant answers.

INTRODUCTION

The ability to retrieve patents by means of on-line access to computers is a relatively recent phenomenon. Producers of machine-readable patent databases have been slow to capitalize on the advantages of on-line retrieval for obvious economic reasons: comprehensive patent databases are large and therefore expensive to maintain on-line, while their appeal has always been limited.

Considering the cost factors, it would seem unlikely that many organizations would care to sustain the entire cost of operating an in-house on-line retrieval system for a large file of patents. One company that did perform pioneering work

† Presented in the symposium on "Meeting the Challenges of the Changing Patent Literature", Division of Chemical Information, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 21, 1977, and 11th Middle Atlantic Regional Meeting of the American Chemical Society, Newark, Del., April 21, 1977.

*Author to whom inquires should be sent.