Chemical Patents Information: The Challenge of Change†

MONTAGU HYAMS

Derwent Publications Ltd., Rochdale House, London WC1X OBP, U.K.

Received September 25, 1984

The greatest challenge to the efficient handling of chemical patents for information purposes has been the decision by all the major patent offices, with the exception of the U.S. Patent Office, to publish without examination. Early disclosure has enhanced the value of patent specifications for current-awareness purposes, but the issuance of large numbers of unedited, badly printed high-cost documents of doubtful legal significance has proved to be a nightmare, not only for the research chemist but also for those involved in licensing. The problem has been made all the more difficult due to the long delay in examination and the unreliability of the examination process itself due to bad classification, incomplete search files, and the restriction of Japanese and Soviet references to the PCT minimum documentation abstracts. The availability of on-line access has posed problems for publishers such as Derwent, who, faced with the possibility of loss of print-product income, have been forced to maintain up-front subscription charges. Political considerations have so far prevented the INPADOC, EPO, and JAPATIC tapes from being mounted on on-line hosts in the U.S., so that cross-file searching with other patents files is not possible. The current changes being contemplated in the provision of chemical patents information are the facility for the topological searching of Markush structures, statistical analysis for monitoring technological changes, automatic translation, at least of keywords, and the provision on-line of drawings and chemical formulas with optical discs.

I would first of all like to consider the changes in the source documents from which we have to work, namely, the patent specifications. In the early 1960s, in most countries one could not claim for chemical compounds per se, but only for methods of producing them. As a result, the chemical patent literature was unnecessarily complex and voluminous, as inventors thought up more and more ways of preparing a valuable compound—many of the methods being quite uncommercial, but merely to block out the competition. And competitors in their turn rushed to think up novel processes for getting round the invention. It was in such circumstances that the early disclosures of Belgian, Australian, South African, and Indian patents became so important—it gave those in the know an important start in the race to think up alternative synthetic routes.

But all this changed by the early 1970s, by which time most countries did allow new compounds to be claimed per se. So for a while the chemical patent literature was precise, and the patent specifications were highly respected as powerful legal tools. However, the big drawback was that, in most cases, the information when published was so out of date that it only had archival value.

But around the same time a far-reaching change began to appear in the patent laws, which undid all this good work and has now left the patent information world in complete disarray. I refer of course to the almost universal practice of early publication without examination. I congratulate the U.S. Patent Office as being the only major office not to do this, mainly I suspect because of the powerful lobbying through self-interest of your patent attorneys.

As a result, what used to be clearly defined, well-edited, typeset, and for the most part legally enforceable patent documents has been replaced by abstruse, unedited, badly offset litho printed disclosures of doubtful legal significance. Yet we have had to apply the same high standards of indexing and coding for retrieval to these pieces of literature as had to be applied to the meaningful granted patents of the past. It is mainly for this reason that patent searches today yield so many false drops, and such hits as are obtained have then to be evaluated for likely validity.

One of the main advantages we had envisaged arising from the so-called "18 month from first priority date" publications was that all family members would issue at or around the same time, and hence, we could select the English language version as basic document, or failing this maybe the French or German. As you can see from Figure 1 (all statistics used in Figures 1–17 were taken from Derwent sources), this has not proved to be the case. It is impossible to predict the order of publication, or the likely delay. The chances are that a Japanese equivalent will issue before a British. If, as is desirable, we want to make the European patent our basic or first issue member, we would have to delay processing all other countries for up to 2 months. If only the major offices would in fact publish in or around the prescribed 18 months, there would not be the need to constantly modify the family member line of our online file, and we could produce our abstracts and make the basic microfilm from the most easily understood language version.

One good thing has resulted from publication without examination. It has increased the value of the patent literature as a current awareness tool for academics and technicians, and for this of course we are most grateful. However, it has not relieved the examiners of their burden—only put off the evil day. In other words, issuance of the document that really matters—the granted patents—is still just as late as ever.

In this respect the U.S. is not too bad, as Figure 2 shows. Most granted U.S. patents are published within a 2-year period, as compared with Japan where the period is some 4–5 years on average. The delay is usually much longer for foreign applicants, who allege that the Japanese purposely hold up grant for commercial reasons.

Another explanation is evident from Figure 3. Due to changes in the law, the number of Japanese applications to be examined is nearly 5 times as great in Japan as it is in the U.S., yet there are only about half the number of examiners. In Japan, an examiner deals with one case per day—in the U.S. with only 1.5 per week. In the European Patent Office (EPO) there has been some attempt, at least in the chemical area, to speed up the preexamination process by the online searching of commercially available files, but current rate of access is a mere 1200 h for the whole of the EPO.

Figure 4 is intended to highlight some of the problems now facing those of us wishing to study patent specifications. Belgian and Canadian documents can now only be consulted

† Presented at the 1984 Annual Meeting of the American Chemical Society.

0095-2338/85/1625-0365$01.50/0

	1-7	8-14	15-30	31-60	More
GB	2%	29%	37%	18%	14%
NL	-	-	100%	-	-
DE	75%	18%	2%	3%	2%
EP	6%	27%	21%	40%	7%
FR	99%	-	-	-	1%
JP	6%	60%	14%	10%	10%

Figure 1. Days delay in publication compared with due date of 18 months from first priority.

Less than 12 months	3%
12 - 18 months	20%
18 - 24 months	40%
More than 2 years	37%
Can be up to 10 years	

Figure 2. Delay in publication of U.S. patents from filing date.

JAPAN	460,000 applications/year
	1800 examiners
	260 applications/year/examiner
USA	100,000 applications/year
	1300 examiners
	77 applications/year/examiner

Figure 3. Examiners output.

as microfiche. Hard copies if requested are made from the fiche and not from the typewritten copy; hence, they are of very low quality.

Since the German Offenlegungsschriften, or unexamined patent applications, are not numbered consecutively, it is impossible to check any consignment for completeness, especially since some documents listed in the Patentblatt are in fact subsequently withdrawn.

The same applies to Soviet patents, which appear in random-number order some 10–23 weeks after the *Claim 1 Gazette* disclosure and which bear no publication date—merely the data sent for printing—so that the date of actual disclosure remains a mystery.

Trying to follow the progress of the 4000 Japanese applications published each week is virtually impossible, since the numbering changes completely and in unrelated fashion as the document passes through four different stages—filing, then publication first as unexamined and then as examined, and finally as the granted patent specification.

During the periods October 1983 through April 1984, the numbers of U.S. patents missing from each weekly consignment was very high, often approaching 50%. Currently on average, around 50 U.S. patents issue 1 week late, and 10 are 2 weeks late—and this delay holds up the whole of our processing. Moreover, since the *U.S. Gazette* is invariably published a few days late, the true full disclosure data of many U.S. patents is not really the one printed on them.

To illustrate still further the sort of problems with which we are faced in our attempts to be current and strictly sequential with our information, our shipment of U.S. patents bearing July 31, 1984, publication date was inadvertently diverted to Saudi Arabia, and has yet to be recovered. Earlier this year a shipment turned up in Bombay.

BELGIUM, CANADA

Microfiche only
Printed copies from fiche

GERMAN OLS

Non-consecutive numbering
Some in Gazette withdrawn

JAPANESE

Different numbers as passes through stages of filed, unexamined, examined and sealed

SOVIET

Non-consecutive numbering
10-23 weeks after Gazette
No actual publication date

Figure 4. Source documents.

Certain 7H-pyrrolo[3,2-f]

quinazoline-1,3-diamines

are useful as pesticides.

Figure 5. Abstract of U.S. 4451466.

These problems, though they may seem relatively trivial on an individual country basis, become serious when dealing with 11 000 documents each week from 26 different countries, all presenting their own particular problems. And remembering also that we need to report at the same time upon *all* the patents issued by a given country in a given week.

An improvement introduced by most Patent Offices during the last 5 years has been the introduction of front-page information with INID codes—Internationally agreed Numbers for the Identification of Data. However, the abstracts provided on these front pages leave much to be desired, since they are provided by the applicant. Typical is the abstract shown in Figure 5. The main claim alone is of quite a complex nature and for the compounds per se—not merely for their use as pesticides. In general, corporations tend to give as little information away as possible in their abstracts. Proud or license-seeking private inventors tend to do just the opposite.

Harder still to cope with is the abstract of an unexamined European patent, since it is only in the language of filing—English, French, or German. For the granted European patent, even though all the claims are in all three official languages, the specification itself is only in the language of the filing. Thus, an Englishman, or an American for that matter, wishing to know if his product is likely to infringe a granted British patent, now has quite a problem. He has to search not only through British patents but also through all those granted European patents designating Britian. And half of these—but for the claims—are entirely either in French or German.

Figure 6 shows the cost of specifications from different patent offices. Except for British patents, the rise in cost has been roughly in line with inflation.

However, if one looks at the extreme right-hand column, where the 1984 prices are all expressed in U.S. dollars, it can be seen that with the exception of Japan all the costs are extremely high compared with other technical literature, bearing in mind also the large numbers of documents involved.

There can be no doubt that this high cost, together with the linguistic and postage problems, is the main reason for the lack of popularity of patent documents as a technical information source.

	1974	1980	1984
GB - p	25	125	185 = $2.42
US - $	0.5	0.5	1.0 = $1.00
FR - FFr	2.5	5.0	10 = $1.13
JA - Yen	-	20	24 = $0.10
PCT- SwFr	-	8	10 = $4.14
EPA- p	-	140	160 = $2.10
EPB- p	-	270	330 = $4.32

Figure 6. Cost of a patent specification.

In Japan, where patent specifications are so cheap by comparison, interest in patents is extremely widespread, as evidenced by the high volume of filings. As a consequence, awareness of the latest technological developments is widespread, and the Japanese economy fluorishes.

The same is true of the Soviet Union, where the number of filings has increased 5-fold in the last decade to over 1800 per week. Far more copies of the *U.S. Official Gazette* translated into Russian are distributed in the Soviet Union than there are copies of the original *Gazette* sold in the U.S. The reason is that by law every factory above a certain size is required to purchase a copy, so as to promote technological awareness.

Would not it be useful if every U.S. corporation above a certain size was obliged to purchase all new U.S. patents in classes corresponding to its main lines of business. The price of patent copies would come right down; and interest in patent information would go right up.

Finally, in connection with the source documents, some of the changes in nomenclature imposed by the patent offices have caused us great problems in the computer handling of a multinational patents file.

Figure 7 shows some of the changes in two-letter country codes introduced in 1978 through a dispute between WIPO—who had used Icirepat codes—and an international maintenance agency who insisted on a change to standard ISO 3166 codes. By 1981 we were forced to follow suit; otherwise, our data would no longer be standard and crossfile searchable. Thus at great expense we had to change completely all our internal backfiles and to reload completely our online files.

Some 15 years ago, the IPC group increased from two to three digits. Ten years later, the IPC subgroup likewise was increased from two to three digits. In each case we had to alter all our programs to cater for these changes. Now that some subgroups in the fourth edition of the IPC are going to have four digits—heaven knows why—and there will be two types of IPC, we shall have to make even more costly program and file changes.

Likewise, our numbering system needed rescheduling when the number of Japanese filings exceeded 99 999 in the year and when the number of Soviet patents reached one million. If we had not done this, crossfile searching with other online databases would not have been possible.

The next change I want to talk about is the deterioration in the standards of classification and examination of patent specifications, which makes the whole process of patenting less attractive than it used to be, thereby putting at risk the livelihood of those of us in the patent information business. There was a time when you could feel pretty confident that a U.S. patent, when granted, would be reasonably watertight and legally enforceable. Alas, this is no longer the case, and the main reasons for this are so obvious that I am amazed that appropriate action to rectify the situation has neither been taken, nor is it contemplated for the foreseeable future.

Country	Old	New
E.Germany	DL	DD
W.Germany	DT	DE
Japan	JA	JP
Sweden	SW	SE
Romania	RU	RO
Eire	EI	IE

Figure 7. Country code changes for 1978 of Icirepat to ISO.

First, let us consider the patent documents available to the examiners for their search. Up to 10% of the U.S. patents in the classified files are missing. Pending the doubtful achievement of a so-called paperless patent office in the not too distant future, would it not make more sense meanwhile to devote more of the several hundred million dollars available to filling the gaps in the examiners files through recourse to the computerized lists of U.S. patents by class?

There is the question of documents foreign to the examining Patent Office. For Japanese and Soviet patents, only the abstracts need to be searched under PCT minimum documentation rules. But these now constitute 40% of total current filings, and many of the Japanese abstracts in particular are of a very unsatisfactory nature. Moreover, many of the U.S. examiners are unfamiliar with the German and French languages, so they tend to rely upon the Derwent English language abstracts attached to the front pages of the documents.

These problems were not so important in the past, since the numbers of Japanese and Soviet documents were not so high; and such foreign language documents that had to be dealt with were well printed, well edited, and clearly defined examined patent specifications, which, for these reasons, were far easier for the examiners to interpret. Moreover, it is becoming increasingly difficult to cover adequately the growing mass of nonpatent literature for searching purposes, especially since the so-called PAL project did not receive adequate financial support. Although online access to supermarket databases such as on Dialog should ease this situation, it is foolish to imagine that the nonpatent literature can be covered adequately. It is surprising, however, the depths to which some examiners will go—for example, an article in a children's comic called *The Beano* was cited against British 2 117 179.

Having regard to the difficulty of interpreting many of the source documents at their disposal, it is not surprising that the examiners do not classify them with any high degree of consistency. Figure 8 has not been specially selected; it is a typical example of how the IPC was applied differently by three different national patent offices to the same invention. Even at the broad first four character or subclass level, in only one instance—CO9H—is there any agreement. The U.S. office has in fact chosen incorrectly to ignore the main sections A and F altogether.

With such inconsistency and lack of precision at the broad level, it is surely questionable whether the provision of over 50 000 subgroups in the IPC can be justified in practice. From Figure 9, for example, it can be seen that 34 subgroups are provided to handle the 800 new inventions per year on cos-

US 4,412,947	1 NOV 83	C07G-007/00
		C08H-001/06
		C08L-089/04
DE 3,315,678	1 DEC 83	A61K-009/70
		A61K-037/12
		A61L-015/01
		C09H-009/02
FR 2,527,621	2 DEC 83	C08J-005/02
		C08J-009/28
		C09H-001/04
		C14C-001/08
		F26B-005/06

Figure 8. IPCs applied to same family members.

IPC A61K-07

Cosmetics, toilet preparations

1981-1983 2491 new inventions

Over 50% in 07/00, 07/06, 07/46

Rest in other 31 sub-groups

8 sub-groups less than 3/year

Figure 9.

IPC A44B-01

Buttons

35 inventions per year

23 sub-groups

01/04 Buttons - ornamental
(one per year)

Figure 10.

metics and toilet preparations but that 31 of these subgroups only average 10 documents per year each with eight of them yielding less than three per year.

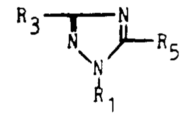
According to Figure 10, there are only some 35 inventions each year to do with buttons, yet 23 subgroups are provided to deal with them. For example, if one is only interested in buttons if they are ornamental, there will on average be just one document per year, assuming of course that the examiner has done his work properly, bearing in mind the constraints placed upon him by the examining office of limiting the number of IPCs applied per document to an average of less than two.

It is this limitation rule, linked to the practice of having to file as many hard copies as there are assigned IPCs in the appropriate shoes, that leads to further failings in the search process. For example, the compounds shown in Figure 11 are characterized by having both a pyridine and a triazine ring. Yet through the priority rules, the document is only classified under pyridine in the U.S. system and under triazine in the IPC. Searching under the Derwent manual code, or simply under the title terms, will give hits under either concept.

Figure 12 is a further example. The dibenz-azepine concept has not been provided for, whereas this would have given a far more specific search. This slide is surely a warning that, in the chemical area at least, you cannot rely upon ordering

US 3,963,731

Pyridyl-triazoles of formula



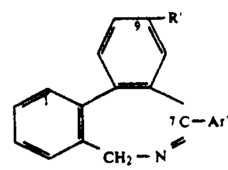
R_1 = benzenesulphonyl
R_3, R_5 = opt. substd. pyridyl

US 260-294.8 S-containing. pyridine
IPC C07d - 249/08 1,2,4-triazole
MC B7-D4 = Pyridine B7-D13 B7-D13 = Triazine

Figure 11.

US 4,434,100

7(2-Thienyl)dibenz[c,e]azepines



Ar' = thienyl
R = F, Cl, Br, CF_3

US Class 260/330.3 = S-contg. ring +
N-contg. ring

Figure 12.

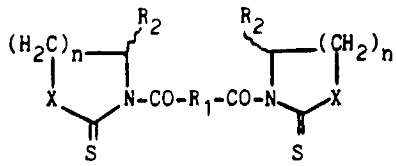
EP-23435	US-4249747
US-3062925	US-2310048
-3220738	-2820853
-3925628	-3918723
-3881073	-4011417
-3383474	-4136884

Figure 13. Citations on same family members.

U.S. patents by class alone to give you a very satisfactory result. Even a simple word search on the titles is often more effective.

Presumably, once a document has been misclassified, or through the need to restrict the number of classifications applied has been omitted from an examined classified collection, there is no choice of it turning up in a search. This is probably one of the reasons why the citations given by different offices against the same invention are so variable. Figure 13 is fairly typical—it can be seen that not one of the five U.S. patents cited by the European office appears in the list cited by the U.S. office, and vice versa.

In fact, anyone carrying out a search with the Derwent files will in general not find it difficult to locate telling prior art against a large number of recently granted U.S. patents, especially in the chemical area. But presumably this will all change eventually when PRC and Chemical Abstracts Service have mechanized the U.S. office—or will it? Let us hope that in the intervening period—and this may be a lot longer than



(I)

R_1 = An acyclic or cyclic divalent gp. contg. at least 1C which, on substitution of a heterocyclic gp. in (I) with a nucleophilic reagent, will stand as an asymmetric centre in the substitution prod.;

Figure 14.

you think—U.S. corporations will not become so dissatisfied with the patent system that they cease filing altogether, or cut down very severely.

Now a few words about chemical coding and retrieval. Chemists are becoming less willing to learn complicated coding systems, or less competent to tackle this problem, so we are now being asked to provide a graphics retrieval system in place of our current fragmentation code. There is also a demand for the ability to carry out searches at different levels of specificity, in order to render easier the task of screening search results, especially where documents are based on wide-claim unexamined applications. They want the thoroughness or degree of specificity to reflect the nature or importance of the search being undertaken—thus a simple research chemist request need not be as broad as an important infringement or prior art search.

The last few years have seen a dramatic increase in the usage of chemical information systems that employ graphics techniques—for example, CAS Online and DARC. These systems encode chemical structures topologically, in other words by determining the shape of the molecule and the interconnections between the atoms. Such systems allow you to input search statements and receive answers in the form of the chemical structure diagram—which is of course the language of normal use of all chemists.

This approach will not however solve the problem of encoding Markush formulas—the generic formulas found in most chemical patents. Topological coding systems such as DARC and CAS Online are designed to encode single compounds into the database. You can search with generic search statements, but the database you are searching comprises single compounds—in the case of both DARC and CAS Online you are searching the CAS Registry collection of somewhat over 6 million compounds. However, a single claim in a single chemical patent can easily encompass 6 million compounds—indeed, there are many Markush formulas that can cover an infinite number of compounds.

Later on, Professor Lynch is going to talk about the problems of graphics input and retrieval of Markush formulas. But I merely want to give you here a brief overview of the current problem as I see it by reference to a typical example.

Figure 14 shows the first part of the main claim of European 81817. R_1 is defined as "an acyclic or cyclic divalent gp. contg. at least 1C which, on substitution of a heterocyclic gp. in (I) with a nucleophilic reagent, will stand as an asymmetric centre in the substitution prod." This definition is very wide, difficult to interpret, and probably invalid as it stands. Presumably, it will be whittled right down at the granted stage. Accordingly, it only needs to be given the full treatment for anyone carrying out an important novelty search. It is unlikely, as it stands, to affect an infringement search and is of very little interest to the ordinary research chemist.

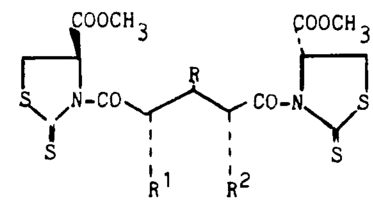
The same is true for the wide definition of R_2 shown in Figure 15: "R_2 = *a substituent of configuration and molecular size such that the* regioselectivity in the substitution reaction

R_2 = A substituent of configuration and molecular size such that the regioselectivity in the substitution reaction of the heterocyclic gp. with the nucleophilic reagent can be determined by a stereochemical interaction with R_1, the steric configuration of the 2 asymmetric C atoms to which the R_2 substituents are attached being identical;

or, preferably,

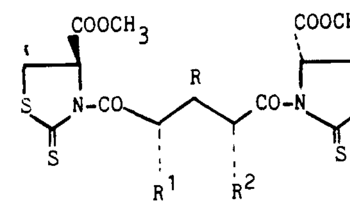
R_2 = hydroxyalkyl, alkoxyalkyl, COOH, alkoxycarbonyl, substd. carbamoyl or alkyl.

Figure 15.



R	R_1	R_2
Me	H	H
MeCONH	H	H
$MeOC_2H_4OCH_2O$	H	H
H	Me	Me

Figure 16.



R = Me, MeCONH, $MeOC_2H_4OCH_2O$ or H

R_1 = H or Me

R_2 = H or Me

Figure 17.

of the heterocyclic gp. with the nucleophilic reagent can be determined by a stereochemical interaction with R_1, the steric configuration of the 2 asymmetric C atoms to which the R_2 substituents are attached being identical."

However, the preferred value of R_2, taken from a subsidiary claim, namely, "R_2 = hydroxyalkyl, alkoxyalkyl, COOH, alkoxycarbonyl substd. carbamoyl or alkyl," is more readily understood and more likely to be relevant. Accordingly, it would be nice, where appropriate, to search only down to this level and ignore any wider disclosures for the particular type of search involved. In point of fact, as shown in Figure 16, only four compounds have been exemplified in the patent, and these are of particular interest to the research chemist, who may be happy to limit his search to these, as per CAS Online.

However, the patent chemist is interested in closely related compounds, and so it could be of advantage to search the Markush structure shown in Figure 17, and developed from all the variables of Figure 16. In other words, from the four specific examples a Markush has been developed covering not only these compounds but 12 others that are closely related.

Apart from retrieving further valuable information, there is a big processing advantage in using this developed Markush technique—which I have called level three, the new specific examples being designated as level four. Levels one and two are reserved for the generic or nonspecific claimed disclosures. Instead of having to input four separate structures, only one graphics input is required.

This may appear to be relatively insignificant in the present case, but in general—for patents involving new compounds—there are around 120 documents each week averaging 35 exemplified compounds each. Some patents even have several hundred examples. Thus the task of separately inputting some 4000 new specific compounds each week is reduced to representing just 120 structures. Storage costs are significantly reduced also.

As I have said, Professor Lynch will be talking generally about graphics Markush later, but as far as Derwent is concerned, I can tell you that our plans are as follows: We shall start creating on a current basis a file ready for searching early next year topologically at what I have described at level three. It will cover all the specifics exemplified in two types of patent—not only those involving new chemical compounds but also those in which known compounds are covered by a Markush.

There will be a total of some 650 documents involved each week. We also plan to go back at least 10 years for patents involving new compounds. We plan eventually to apply graphics to cover completely the main claims also when—as is expected—a fully generic graphics capability is developed by mid-1986.

Meanwhile, we shall continue to code the generics by our fragmentation code, and in any case, this code must still be applied in searching our backlog. To make this code easier to understand and apply, a microcomputer-driven search menu has been developed, and inhouse instruction in the use of this menu will be taking place at subscriber locations throughout the U.S. in the fall of 1985.

Without doubt, the greatest change we have seen in the provision of chemical patents information is the ability of database producers to provide subscriber access online through host computers such as SDC and Dialog. But, here again, the change has created a further challenge to the database producer, anxious to preserve the profit margins essential for the continued creation of a high-quality product.

Subscribers are beginning to ask why should I subscribe to an expensive upfront service such as CPI or printed product such as *Chemical Abstracts* when for a lot less money I can access them online through SDC, for example, along with a lot of other databases of possible interest? Moreover, by using high-speed terminals, or downloading the information, costs can be reduced even further.

There is a still further problem. Let us suppose a company that would have spent $1000 in 1984 on conventional Derwent products decided to allocate half of its available funds to online. And let us suppose Derwent only gets 40% of the online income from SDC. Then Derwent will only receive $700 from the company instead of $1000.

One solution of course would be to increase the online charge. However, Derwent spends the equivalent of $15 million per year in producing its patents services. If we were to rely upon online income alone and even if we were able to get 50% return on income from the online spinners, at the present rate of 40 000 h per year we would have to charge $750 per hour access just to break even, and would still stand the risk of lower usage through downloading and/or higher speed terminals. Moreover, it would not be appropriate to seek to increase revenue by charging for the number of online hits, for the most valuable results of a patent search are often when no hits are obtained whatsoever.

So we at Derwent are convinced that our policy of upfront subscription fees is the only viable one in the long term, no matter how unpopular it may seem with some people at the moment. I feel certain that other quality database producers will be forced to follow suit, by imposing some form of subscription fee in addition to the hourly access charge.

In any case, I am confident that, with the increasing power of microcomputers, we shall soon be back to the pre-online scenario, where it will become more cost effective and convenient for the larger users to access the more important databases such as WPI and *Chemical Abstracts* inhouse. However, instead of using magnetic tape on main frames as hitherto, they will use digital disc storage on minis or micros with the convenience of inhouse graphics output, and no telecommunications charges, failures, or bottlenecks.

Also looking into the future, I foresee the use of patents information for the forecasting of technology trends becoming a very powerful and status-improvement tool in the hands of the patents information manager, through improvements in computerized statistical analysis techniques. Derwent has developed a program based on dBASE II (now III) for carrying out such work on a microcomputer, from answers downloaded onto the disc from an online search. We also plan to have ready by the end of the year a system being developed jointly with Telesystems using a QUESTEL/SAS or pSTAT interface, whereby the analyses can be carried out completely online.

Using such techniques in conjunction with substructure searching, it should be possible to identify which substructures are being actively researched by whom and where. Even within a given patent document, where a large number of specific compounds has been encoded, it should be possible to search the various members to determine which are most relevant to the invention and in particular to establish new emerging structure–activity relationships. Mike Dixon will be telling you more about this.

What I do think we shall see in the future is the requirement, or the invitation, to submit patent applications on floppy discs, from which printed copies will only be produced on demand. Even further into the future, I think that all patent applications will be converted through an automatic translation program into a standardized language—probably American English—before being placed online for international access.

Finally, I would like to describe just some of the steps being taken by the major national patent offices to cope with the online situation, since the present plans are certainly not in the best interest of you the users and would appear to be particularly biased against users in the U.S.

Did you know that the U.S. Patent and Trademark Office, along with other offices, supplies tapes to INPADOC in Austria to create a patents family tape that can only be accessed at very high costs from the INPADOC computer in Vienna or from Pergamon-InfoLine in London? Did you know that the tapes of the European Patent Office can only be accessed publicly from a computer in Europe, and even then only through the intervention of INPADOC, and that the signs are that the English language abstracts of the Japanese Patent Office will only be made available from a computer in Japan, whereas there are absolutely no restrictions against mounting of the USPTO full-text tapes anywhere in the world?

It would also have been rather nice had Chemical Abstracts Service and Derwent been able to agree on closer cooperation. Unfortunately however, no satisfactory formula has been found at the moment because of the apparent competitive nature of our two services, but I am confident that with good will on both sides some means of preventing duplication will be worked out between us in the not too distant future for the mutual benefit of all concerned.

In order to give widest possible access to its patents files, Derwent is going to the expense of putting them up on three different online hosts, two of them in the U.S. It is regrettable that neither U.S. host will be allowed also to mount the INPADOC, EPO, or JAPATIC patents files—nor will they be allowed to mount the CAS abstracts. Unless these barriers can be overcome, the dream of being able to interrogate the world's major patents databases at one and the same time on a single host computer will not be possible.

My hope is that, if ever I am privileged to talk to the Chemical Information Division of the ACS again, I shall be able to report that this dream has come true—and that you in this Division have played an important part in its achievement.

Polymer Patent Information Systems Could Be Even Better![†]

STUART M. KABACK

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

Received December 17, 1984

Patents that involve polymers are a very substantial portion of the patent literature. Both Derwent Publications and IFI/Plenum have highly developed systems for indexing polymer patents, and significant features relating to polymers (though not restricted to patents) are found in other databases, most notably *Chemical Abstracts* and the CAS Registry. This paper surveys some strengths and limitations of today's systems and looks toward possible future advances.

INTRODUCTION

Patents that involve polymers constitute a very substantial portion of the patent literature. *Plasdoc*, the section devoted to polymers, is by far the largest component of the chemically related patents in Derwent's *Central Patents Index* or CPI. Since 1981, nearly one-third of the patents in the CPI have been assigned to *Plasdoc*, averaging nearly 850 basic patents (first-abstracted members of patent families) per week during 1984.

Polymers have a huge economic importance. In the U.S., sales of plastics alone exceeded 18 million metric tons in 1983; the figure does not include elastomers and other types of polymer. It is essential to provide this enterprise with timely and accurate documentation of polymer patents, for both current awareness and retrospective information retrieval.

Two information services are especially important for current awareness on polymer patents: the aforementioned Plasdoc and *Chemical Abstracts*. For retrospective retrieval these two are again of prime importance, along with the CLAIMS family from IFI/Plenum. Other files also contain polymer information but lack a specialized polymer retrieval system or are too limited in scope to be of general interest.

What must a polymer patent information service deal with? Figure 1 shows something of the scope that is needed. The range is enormous, involving not only chemistry and chemical engineering but also mechanical engineering, electrical engineering, physics, and other disciplines. And the complexity of the technology can be prodigious. A satisfactory information system must be able to describe the significant points of individual patents so that they can be retrieved reliably, with as much specificity as possible. With 40 000 or so new inventions involving polymers being added annually to what is already a massive heap, discrimination is essential to effective information retrieval.

Until about the mid-1960s there were no specialized information systems for polymer patents. *Chemical Abstracts* covered many of them, with an emphasis on the chemistry, and IFI's *Uniterm Index* indexed U.S. polymer patents. Then a lot of things began to happen. Chemical Abstracts Service (CAS) developed its Registry system for chemical substances and applied it to many polymer systems. CAS also set out to encompass more of the technology of polymers through two special publications: *POST-J* for journal literature, *POST-P* for patents. Meanwhile, Derwent, which had recently started documentation services on pharmaceutical and agricultural chemical patents, began *Plasdoc* in 1966. Finally, du Pont started a highly sophisticated in-house patent indexing system in 1964, which was later modified and merged with IFI's *Uniterm* system to form the basis of IFI's *Comprehensive Data Base*, or *CLAIMS-CDB* as it is known in the online form.

CURRENT AWARENESS SERVICES

Before considering the characteristics of the retrieval systems, let us look at the current awareness situation, with respect to timeliness and also abstract content. This is also an appropriate place to consider differences in coverage.

Derwent has always aimed at providing information as rapidly as possible. Derwent alerting abstracts for the most important patent-issuing bodies are normally in the hands of subscribers slightly less than 2 months after the publication dates of the patents. *Chemical Abstracts* was considerably slower, but has improved its timeliness to where it will cover these major countries about 1–2 months later than Derwent. A significant difference between the two, however, is that Derwent covers all of any country's output for a given week in a single week. There is no time distribution for any country in Derwent coverage. CA, on the other hand, does have a time distribution, so that, while the bulk of coverage in major countries is taken care of within about 4 months, there is inevitably a tail to the distribution that brings in some abstracts considerably later. Thus, one never knows when CA might have finished covering the patents issued on a given date.

For the Soviet Union the timeliness picture is reversed. Here, CA takes its abstracts from brief descriptions published by the Soviets (Otkrytiya, Izobreteniya), while Derwent waits to obtain full copies of Soviet patents to work from. As a result, CA's Soviet coverage runs some months ahead of Derwent's. Note, too, that CA enters listings for equivalent patents in its printed index very rapidly, more rapidly than Derwent, because of the input that it gets from INPADOC, the International Patent Documentation Center. CA does not

[†] Presented at the Herman Skolnik Award Symposium, Division of Chemical Information, American Chemical Society, Philadelphia, PA, Aug 28, 1984.

0095-2338/85/1625-0371$01.50/0