data to assume such a fate or to be unworthy of the Herculean effort required for retrieval calls to question the very purpose of obtaining the data to begin with. The database now exists to ensure that data are not lost, and that they are easily accessible and in a useful format.

ACKNOWLEDGMENT

The encouragement and support of R. G. Laughlin (The Procter and Gamble Company) and M. Chase, P. Fagan, and D. Bickham (National Institute of Standards and Technology [NIST]) throughout this project is gratefully acknowledged. Additional thanks are extended to all those who contributed comments, data, and reprints to the database. The project was funded in part through a grant for Critical Compilations of Physical, Chemical and Materials Data through NIST and a grant from The National Institutes of Health (Grant DK36849) to M.C.

REFERENCES AND NOTES

(1) Silvius, J. In *Lipid-Protein Interactions*; Jost, P., Griffith, G., Eds.; Wiley and Sons: New York, 1982; pp 239-281.
(2) Caffrey, M. In *Metabolism and Dry Organisms*; Leopold, A. C., Ed.; Cornell University Press: New York, 1986; pp 242-258.
(3) IUPAC-IUB Commission on Biochemical Nomenclature *J. Biol. Chem.* **1967**, *242*, 4845-4849.
(4) IUPAC-IUB Commission on Biochemical Nomenclature *Eur. J. Biochem.* **1977**, *79*, 1-9.
(5) IUPAC-IUB Commission on Biochemical Nomenclature *Eur. J. Biochem.* **1977**, *79*, 11-21.
(6) Luzzati, V. In *Biological Membranes, Physical Fact and Function*; Chapman, D., Ed.; Academic Press: New York, 1968; Vol. 1, pp 71-123.
(7) Fontell, K.; Mandell, L.; Ekwall, P. *Acta Chem. Scand.* **1968**, *22*, 3209-3223.
(8) Ekwall, P. *Advances in Liquid Crystallography*; Brown, G. H., Ed.; Academic Press: New York, 1971; Vol. 1, Chapter 1.
(9) Windsor, P. A. *Chem. Rev.* **1968**, *68*, 1-40.
(10) Reed, R. A.; Shipley, G. G. *Biochim. Biophys. Acta* **1987**, *896*, 153-164.
(11) Waldrop, M. M. *Science* **1990**, *248*, 674-675.

Structure Searches in Patent Literature: A Comparison Study between IDC GREMAS and Derwent Chemical Code

KARL HEINZ FRANZREB, PIA HORNBACH, CLAUDIA PAHDE, GOTTFRIED PLOSS, and JÜRGEN SANDER*

Hoechst Aktiengesellschaft, D-6230 Frankfurt/Main 80, Federal Republic of Germany

Received October 24, 1990

Patent searches by fragment codes in the Derwent WPI/L and IDC GREMAS Files as well as by closed substructures in the CAS Registry and the Derwent WPIM Files are compared. The most relevant answers are found in the IDC GREMAS File; by the Derwent fragment code the IDC result is almost reached. On the contrary, the search results obtained from the CAS Registry File are so incomplete that they are ruled out as an alternative. Derwent WPIM File as a successor to Derwent Chemical Code cannot serve as a replacement for the GREMAS File at the present time because of lack of completeness. Higher costs at IDC GREMAS compared with Derwent WPI/L for database use contrast with lower costs for relevance checking. On the whole, the IDC GREMAS File offers the best cost-benefit ratio for Hoechst AG.

INTRODUCTION

Particularly when investigating a patent status, it is important for industry to be able to have full information provided with as few irrelevant answers as possible. The Internationale Dokumentationsgesellschaft für Chemie mbH (IDC) was founded by chemical industries with this objective. Its task is to create and update files of information in the areas of primary interest to the shareholders; this also includes the maintenance and further development of the input and retrieval systems.

The GREMAS code is used for encoding full structures and generic structures in the field of organic chemistry. An efficient retrieval method developed for searching in this GREMAS File enables us to search with a high recall and precision.[1]

The IDC File is directed particularly to the needs of the shareholders and accordingly only covers the patent areas important to them. On the contrary, the WPI/L[7] File of Derwent covers all patent areas.[2] The GREMAS code on which the IDC Files are based is more precise than the chemical code used by Derwent for encoding structures; as a result IDC requires more time and effort for indexing but consequently yields more precise search results. This qualitative statement is known and has also been described in the literature.[3]

The aim of this study was to make a cost-benefit analysis between GREMAS and WPI/L searches. In addition, searches in the Registry and CA Files of the Chemical Abstracts Service (CAS) were included. On the one hand, we were concerned to make a quantitative analysis as to whether the higher indexing costs for the IDC could be compensated by cost savings in personnel for the relevance checking of a lower IDC proportion of false drops. On the other hand, we were concerned to make a value statement regarding the IDC File in comparison to the WPI/L and/or CAS Files on the basis of the number of relevant hits found.

Derwent will be suspending the indexing by chemical code in favor of the topological encoding by Markush DARC. The WPIM File based on Markush DARC was therefore also taken into account in this test; however, only for the small period of time hitherto available from Derwent week 01/87.

FILE CONTENTS

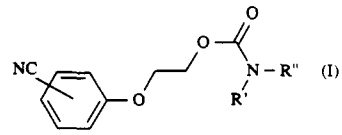
In principle the period from 1963 onward may be used for comparing the IDC GREMAS and the Derwent WPI/L Files. Since 1970, however, the CPI abstracts from Derwent have been used as a common basis for both files. They are indexed by Derwent using the chemical code and in parallel by IDC using the GREMAS code.

In the CA File from CAS also included in the test, patent literature is indexed in addition to journal literature. Searching of these patents using structure formulas can however only be made using specific patent examples filed in the Registry. Generic parts such as, for example, alkyl, are permitted for the structures of the query formulation in CAS, but the CAS Registry contains only specific compounds. The references

0095-2338/91/1631-0284$02.50/0

Query

Cyanophenoxyethyl carbamates of the formula (I) (where R', R" = H, (1-4 C) alkyl or NR'R" forms a saturated 5-, 6- or 7-membered ring)



(I)

Interpretation for the test:

1) Alkyl residues are interpreted as being unsubstituted
2) The saturated 5, 6 or 7 membered heterocycle is interpreted as being isolated and unsubstituted and containing any further heteroatoms, which themselves can be substituted by unsubstituted alkyl or phenyl

Results

1978-1988	WPI/L	GREMAS	CA
Total	341	14	0
Relevant	1	1	0
	0D85-197713	0D85-197713	- (CA103:191470[8])
Non relevant	340	13	0

1987-1988	WPIL	GREMAS	WPIM
Total	113	4	5
Relevant	0	0	0
Non relevant	113	4	5

Figure 1. Test 1.

associated with these structures can be found in the CA File. The patent claims described in the abstracts of the CA File as Markush structures are not searchable.

The Derwent WPIM File contains the structures from the complete patent specification; the associated bibliography and the abstracts can be found in the WPIL File.

GENERAL TEST CONDITIONS AND TEST QUERIES

The comparison relates to the pure structure search type without further restrictions such as nonstructural features. Searches for patented reactions were also excluded, since these can only be searched for in the IDC files.

Journals are only included in the GREMAS and CAS Files. They were therefore not taken into account in the comparison.

We have used queries from a comparison study in 1979.[4] Originally in this test, files which are encoded using the Derwent Ring and Farmdoc Codes were compared with one another in the period from 1970 to 1978. Supplementarily the IDC GREMAS File was also tested. The seven queries used for this comparison were selected by the AIOPI, a working group of the English pharmaceutical industry. These queries thus favor none of the files to be compared. We took over these queries for a new test for the period from 1978 to 1988.

The queries with their exact specifications are shown in Figures 1–7.

The GREMAS search strategies were defined by employees of the Scientific Information Department of Hoechst AG after consultation with IDC. The queries were formulated in such a way that only the required hierarchical levels were addressed. The searches were carried out in the serial GREMAS File of the IDC.

The WPI/L and WPIM searches were carried out by Derwent Publication Limited on the host Telesystemes Questel; Derwent was informed of the background of the searches.

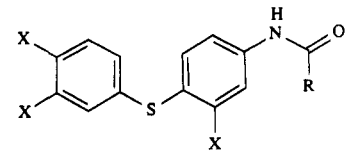
The CAS searches were carried out by the Scientific Information Department of Hoechst AG in the Registry and CA Files on the host STN International.

SEARCH RESULTS

The relevance checking of the answers was carried out by the Scientific Information Department of Hoechst AG. Answers which only matched a higher hierarchical level than the one required were considered neither as hits nor as false drops, since these are only important if no specific, more obvious references describe the prior art. If necessary, answers of this type can be obtained from GREMAS searches selecting specifically the hierarchically higher level.

Query

Diphenylthioethers of the formula (II) (where X = H, halogen, (1-5 C) alkyl; R = (1-5 C) alkyl)



(II)

Interpretation for the test:

Alkyl residues are interpreted as being unsubstituted

Results

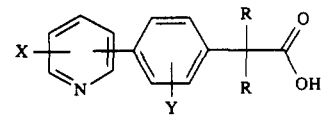
1978-1988	WPI/L	GREMAS	CA
Total	92	36	1
Relevant	4	4	1
	0D84-070356	0D84-070356	CA101:66008
	0D86-132499	0D86-132499	- (CA106:18377[8])
	0D86-273555	0D86-273555	- (CA108:113198[8])
	0D86-319057	0D86-319057	- (CA106:176188[8])
Not taken into account	2	0	0
	0D78-035559[11]	-	-
	0D85-203219[12]	-	-
Non relevant	86	32	0

1987-1988	WPIL	GREMAS	WPIM
Total	41	10	13
Relevant	0	0	0
Non relevant	41	10	13

Figure 2. Test 2.

Query

Pyridylphenylacetic acids of the formula (III) (where X = Y = H, halogen, alkyl, allyl, alkoxy, NO2, NH2, polyhalogen (e.g. CF3); R = H or alkyl)



(III)

Interpretation for the test:

1) Residues alkyl, allyl and alkoxy are interpreted as being unsubstituted
2) Polyhalogen (e.g. CF3) is interpreted as being polyhaloalkyl

Results

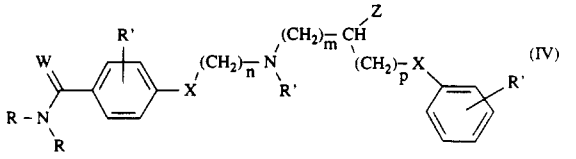
1978-1988	WPI/L	GREMAS	CA
Total	166	15	0
Relevant	5	7	0
	0D81-048176	0D81-048176	- (CA96:19840[9])
	0D82-098287	0D82-098287	- (CA98:197803[8])
	0D83-K26680	0D83-K26680	- (CA98:178970[9])
	-	0D83-K26681	- (CA98:215332[8])
	0D83-847177	0D83-847177	- (CA100:120696[8])
	0D85-141060	0D85-141060	- (CA96:19840[9])
	-	0D88-228683	- (CA110:172854[8])
Non relevant	161	8	0

1987-1988	WPIL	GREMAS	WPIM
Total	71	5	0
Relevant	0	1	0
	-	0D88-228683	.[10]
Non relevant	71	4	0

Figure 3. Test 3.

Query

Phenoxyalkanolamines of the formula (IV) (where W, X = O or S; Z = OH or Cl; R = H, alkyl or N(R2) forms a saturated ring; R' = H or lower alkyl; m, n, p = the same or different 1, 2 or 3)



(IV)

Interpretation for the test:

1) All alkyl residues are interpreted as being unsubstituted
2) The saturated ring formed by N(R)2 is interpreted as being isolated and unsubstituted (only alkyl substituents are permitted) and containing any further hetero atoms which in turn can be substituted by unsubstituted alkyl or phenyl

Results

1978-1988	WPI/L	GREMAS	CA
Total	591	27	4
Relevant	5	10	4
	-	0D78-008283	- (CA89:12153[8,13])
	-	0D79-056367	- (CA91:181467[8,13])
	-	0D81-062018	CA96:40926
	-	0D81-075594	- (CA95:209678[8,13])
	0D81-094874	0D81-094874	CA96:57785
	-	0D83-815050	- (CA100:12674[8,13])
	0D84-075350	0D84-075350	CA101:90757
	0D84-108792	0D84-108792	CA101:60142
	0D87-177921	0D87-177921	- (CA108:94211[9,13])
	0D87-177922	0D87-177922	- (CA108:94212[9,13])
Not taken into account	3	0	0
	0D87-213788[11]	-	-
	0D87-258043[11]	-	-
	0D87-298883[11]	-	-
Non relevant	583	17	0
1987-1988	WPIL	GREMAS	WPIM
Total	178	9	.[14]
Relevant	2	2	.[14]
	0D87-177921	0D87-177921	-
	0D87-177922	0D87-177922	-
Not taken into account	3	0	0
	0D87-213788[11]	-	-
	0D87-258043[11]	-	-
	0D87-298883[11]	-	-
Non relevant	173	7	.[14]

Figure 4. Test 4.

Only 10 matching hits were found with the original WPI/L query formulation for all queries together. The query formulations were checked after presentation of the IDC results, and coding errors in the query formulations were discovered in queries 2, 3, and 5. After correcting the errors, the WPI/L searches were repeated. The results presented here relate to this corrected version.

The individual results are shown in Figures 1–7 and the overall results in Figures 8–11.

DISCUSSION OF THE RESULTS

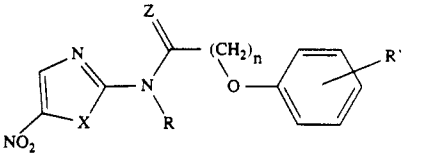
A total of 23 relevant answers were found in the GREMAS searches; apart from one exception, this set contained all the relevant hits found in other systems. In contrast, seven relevant answers were found only by the IDC, but not by other systems (Figures 8 and 9).

The one relevant reference which was not found by IDC was present in the GREMAS File. However, it could not be obtained with the query formulation since, as the result of an error, the specific and the general formulations were not coded together as alternatives, but were divided into two formulas.

Out of the total number of 96 answers found in the IDC File, 73 were nonrelevant. The ratio of 1 hit per 3 false drops permits rapid relevance checking. Processing is additionally

Query

Nitroimidazoles, nitrothiazoles and nitrooxazoles of the formula (V) (where X = N, O or S; Z = O, S or NH; R = H or lower alkyl; R' = halogen, lower alkyl, alkoxy or acyloxy; n = 1, 2 or 3)



(V)

Interpretation for the test:

1) Residues alkyl and alkoxy are interpreted as being unsubstituted
2) Residue acyloxy is interpreted as being unsubstituted alkanoyloxy

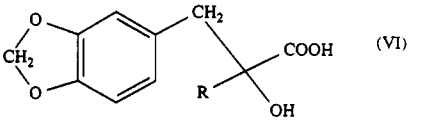
Results

1978-1988	WPI/L	GREMAS	CA
Total	145	4	0
Relevant	2	1	0
	0D82-072488	0D82-072488	- (0D97:182225[8])
	0D88-199351	.[15]	- (0D110:23890[8])
Non relevant	143	3	0
1987-1988	WPIL	GREMAS	WPIM
Total	60	1	0
Relevant	1	0	0
	0D88-199351	.[15]	.[10]
Non relevant	59	1	0

Figure 5. Test 5.

Query

Benzodioxole derivatives of the formula (VI) (where R = H or alkyl)



(VI)

Interpretation for the test:

1) Ring system is interpreted as being not substituted by further residues
2) Residue R is interpreted as being H or any unsubstituted alkyl
3) An undefined position of the side chain on the aromatic ring is permitted, but not another defined position

Results

1978-1988	WPI/L	GREMAS	CA
Total	2	0	0
Relevant	0	0	0
Non relevant	2	0	0
1987-1988	WPIL	GREMAS	WPIM
Total	1	0	2
Relevant	0	0	0
Non relevant	1	0	2

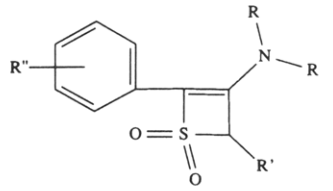
Figure 6. Test 6.

further facilitated: The Derwent references are supplemented by the IDC by formula sheets in which the structure searched for can be detected considerably more quickly than in the original text. All numbers of the formula schemes which hit the query are stated in the answer list.

In the Derwent WPI/L searches, after correction of the query formulations, a total of 17 relevant answers were found, one of these only with WPI/L. Seven relevant answers were missing, five of these for query 4. The low number of hits for query 4 can be traced to the limitation of the query using "manual codes". This restriction was made to reduce the number of answers.

Query

Thietane derivatives of the formula (VII) (where R = lower alkyl or NR2 forms a saturated or unsaturated 5- or 6-membered ring; R' = lower alkyl; R" = H, halogen, NO2, NH2, alkyl, alkoxy, thioalkoxy, carboxy, aryloxy or cycloalkoxy)



(VII)

Interpretation for the test:

1) Residues alkyl and alkoxy are interpreted as being unsubstituted
2) Thioalkoxy is interpreted as S-alkyl (unsubstituted)
3) The 5- or 6-membered ring is interpreted as being isolated and optionally containing any further heteroatoms and being optionally substituted by any residue

Results

1978-1988	WPI/L	GREMAS	CA
Total	5	0	0
Relevant	0	0	0
Non relevant	5	0	0

1987-1988	WPIL	GREMAS	WPIM
Total	1	0	1
Relevant	0	0	0
Non relevant	1	0	1

Figure 7. Test 7.

A search subsequently carried out for query 4 without this restriction supplied 9 of the 10 relevant answers; however, the total number of answers to be considered increased with this query formulation from 583 to 802.

Two answers for query 3 and one answer for query 4 could not be found as a result of coding errors in the WPI/L File.

In the WPI/L File, one further answer was found for query 2, and three further answers were found for query 4, which match a higher hierarchical level than the one required. For instance "linking groups" are claimed in the patent, but without specific details being given of these groups. These answers were not considered as hits since it should be possible to define the level up to which answers are of interest in a patent structure searching system. A further reference found for query 2 was equivalent to another reference in this query; this was not considered as an additional hit.

The total number of Derwent WPI/L was 1342 answers; of these 1320 were not relevant. The ratio of hits to false drops is 1:78. The relevance checking thus requires a high effort. There is the risk that relevant hits are overlooked in such a large amount of false drops. The search results of query 1 and query 4 contained so many false drops that we would not be able to process them in our day-to-day operations for reasons of time.

On the basis of query 4, it is clear that, although it is possible to reduce the number of answers by modifying the search strategy, the number of relevant hits is also reduced. This is equivalent to a value reduction of the search result.

The searches performed in the CAS Files delivered five answers which were all relevant. The GREMAS and WPI/L answer sets also contained these relevant hits. However, 19 relevant answers were not retrieved in the CAS Files. The result demonstrates that, although it is possible to search with very little noise in the CAS Registry, which contains only patent examples of the patents, a complete result cannot be expected.[5]

A noteworthy fact is that all 24 relevant references are contained in the CA File. However, only in five cases one of the patent examples hit the query formulation. With query

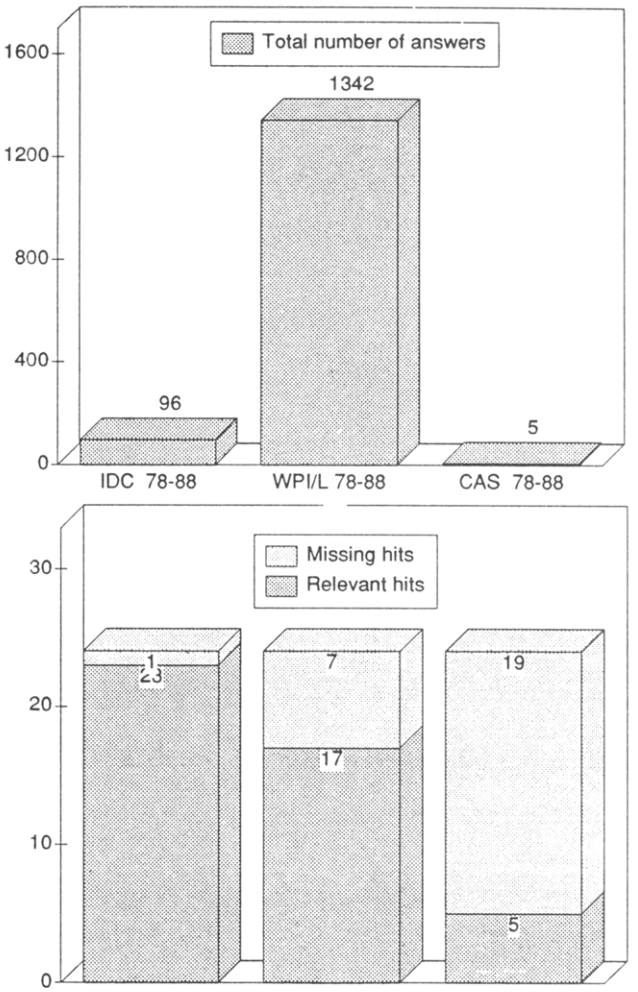


Figure 8. Total number of answers and relevant and missing hits of the GREMAS, WPI/L, and CAS searches (period 1978–1988).

4, several relevant hits were not found, although specific compounds matching the query are stated in the original patent; these were not indexed by CAS however.

The Markush formulas of the patents are given in the abstracts of the CA File. If it was possible to search these Markush structures, five further references would have been found. With the remaining 14 references, the patents were not indexed in sufficient detail. Thus, the CAS abstract frequently only refers to "optionally substituted", although the substituents are specifically defined in the original patent.

A result comparable to the GREMAS or WPI/L result can only be expected in the CAS Files if the possibility of accessing the Markush structures of the patent claims is created, and if the indexing depth for the patents is increased. This approach is being pursued by CAS with the new MARPAT File. But this file contains only Markush structures of patent literature of the CA File from 1988 onward. Because of the small temporal overlap, MARPAT was not supplementarily included in the comparison.

In the Derwent WPIM File it was virtually impossible to carry out query 4 since a processable result could only have been achieved by splitting it into a number of subqueries with the concomitant large expenditure for time and effort. The remaining six queries delivered 21 answers, none of which matched (Figures 10 and 11). The comparison figures for the IDC File are 19 nonrelevant answers and one relevant hit, for the WPI/L File 286 nonrelevant answers and one additional relevant hit. The reason for missing the two relevant answers with WPIM is that both references were not indexed.[6]

Although the number of answers is insufficient to make a conclusive statement, the following trends are apparent in

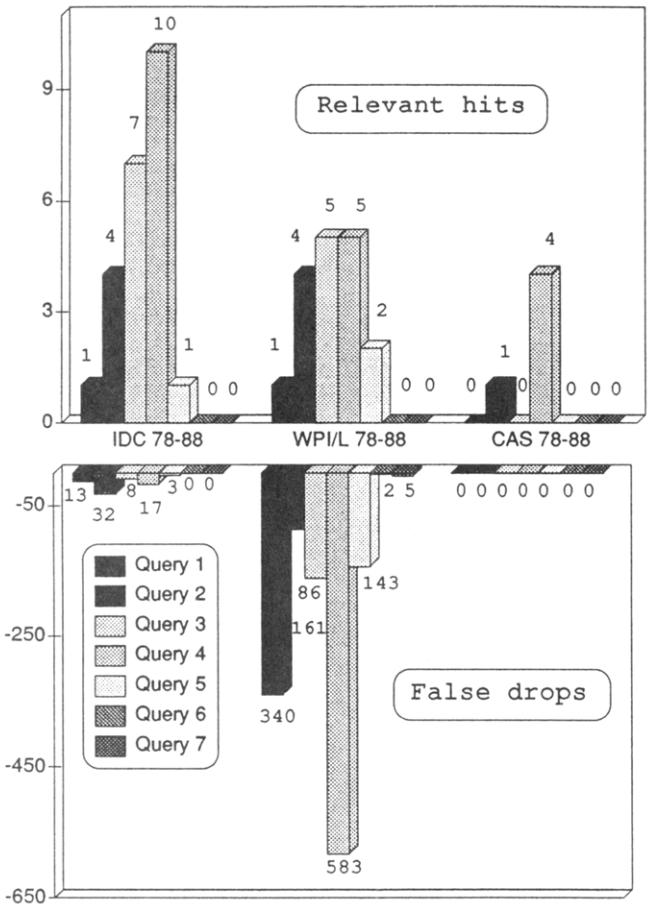


Figure 9. Relevant hits and false drops of the GREMAS, WPI/L, and CAS searches, broken down for queries 1–7 (period 1978–1988).

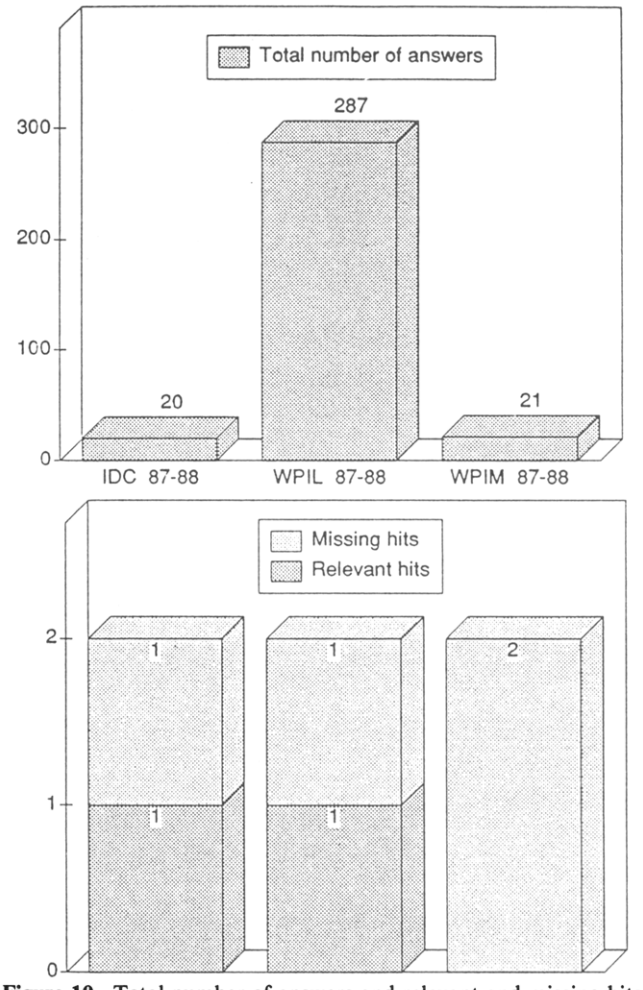


Figure 10. Total number of answers and relevant and missing hits of the GREMAS, WPIL, and WPIM searches (period 1987–1988; queries 1–3, 5–7).

relation to WPIM: considerably less noise is obtained than with Derwent Chemical Code and the GREMAS result is almost achieved. However, a degree of completeness comparable to the WPI/L or IDC File has not (yet) been reached.

A weakness of WPIM is that generic groups in the query, like alkyl, cannot also address specific structure parts in the file such as methyl, as it is possible already in the CAS Registry. For if one wishes to take into account all specific groups, then these must be formulated as alternatives, which in practice is not possible.

Also with respect to noise reduction, further improvements are necessary; for instance variable points of attachment for substituents should be possible.

CONCLUSIONS

The crucial criterion for the quality of a patent documentation system is the completeness of the answer set for the search. The comparison shows that files containing only patent examples, like the CAS Registry, do not meet this requirement. WPIM contains only patents from Derwent week 01/87 onward; another database must be used for searches in previous years. However, even from 1987 onward not all relevant patents are filed in the WPIM File and not all queries can be formulated, so that WPIM likewise cannot fulfill this quality criterion at the present time.

In contrast, GREMAS (except for one missing answer) provides a complete result. Even if the Derwent Chemical Code does not achieve the GREMAS result, almost all relevant answers can however also be found in WPI/L.

A further criterion important in practice is the total number of answers from which the relevant hits have to be sorted out. With its hits to false drops ratio of 1:3, GREMAS here offers decisive advantages in comparison to WPI/L with a ratio of 1:78. In principle, the search strategy can be formulated more precisely in GREMAS; in addition it can be designed to be more flexible and considerably more efficient by the specific addressing of different hierarchical levels.

COST–BENEFIT ANALYSIS

When equivalent files are available, the costs become the decisive factor for their use. We have therefore drawn up the estimate shown below for Derwent WPI/L and IDC GREMAS:

Each Derwent WPI/L search generates costs for use of the database; these costs are the aggregate of the costs for online-time and for printing out the answers.

With the IDC, individual searches are not invoiced, but rather a membership fee, which is proportional to the number of chemists and pharmacists employed in the company, must be paid. If these fees, which are paid for creating the patent file, are distributed over the annual patent searches actually carried out, costs per search, which are 11 times higher than the Derwent WPI/L costs, are incurred by Hoechst AG.

At first glance it would seem that Derwent WPI/L searches offer clear cost advantages. However, this view does not take into account the personnel costs of the searchers.

On an average 192 answers per search were received from WPI/L and 14 answers from GREMAS. The test covered the period 1978–1988. IDC and WPI Files go back to the beginning of the sixties. In the case of routine searches in the entire file, therefore, there would be about twice as many references to be tested for relevance; therefore 28 answers with GREMAS and 384 with WPI/L have to be checked for

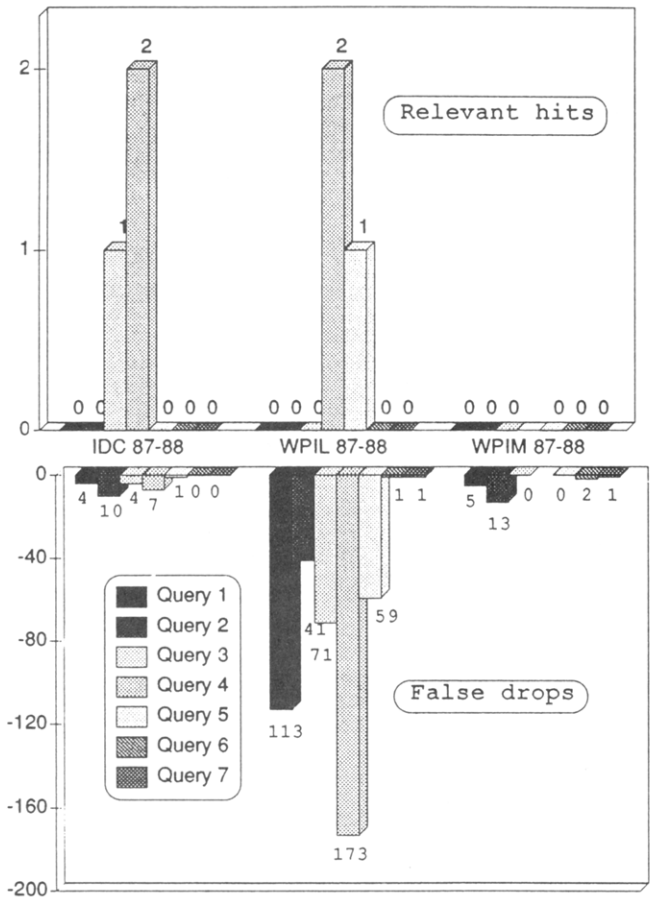


Figure 11. Relevant hits and false drops of the GREMAS, WPIL, and WPIM searches, broken down for queries 1–7 (period 1987–1988).

relevance, that is 356 answers more with WPI/L than with GREMAS.

Assuming 5 min per answer for relevance checking, an additional effort taking 1780 min or 30 h per search will be required. Thus, for example 1000 patent searches of this kind per year would result in an additional effort of 30000 h. Thus a company which uses Derwent Chemical Code and wants to thoroughly check 1000 searches per year for relevance needs 18 additional chemists.

The total costs for a search consist of the costs for database use and the personnel costs. GREMAS searches require a very high fixed-cost proportion in the form of IDC membership and only relatively low variable costs; in contrast the costs for Derwent WPI/L are proportional to the number of searches. Depending on the number of patent searches and, secondarily, on the number of potential users in the company, the IDC membership or the direct search in Derwent WPI/L will be more cost effective.

For Hoechst AG the total costs per search are distinctly lower for GREMAS than for WPI/L. As the number of searches increases, the cost advantages of GREMAS increase considerably in comparison with WPI/L. On the other hand, with a decline in the number of searches by one quarter, IDC would be more expensive than WPI/L for Hoechst AG.

SUMMARY

The comparison demonstrates that the IDC GREMAS File is superior to the Derwent WPI/L File with respect to completeness and particularly with respect to precision. The results in the CAS Registry and in the Derwent WPIM File are so incomplete that they are ruled out as an alternative for WPI/L or GREMAS.

The total costs for the searches with IDC were significantly less than those of Derwent WPI/L; the IDC GREMAS File thus offers the best price/performance ratio for Hoechst AG. The additional effort for a more precise input and hence less false drops is worthwhile in the long term. This should also be taken into account in designing future patent systems on a topological basis.

Only pure structure search could be compared in this test; however, the structures represent only one arm of the GREMAS system. In addition there are a hierarchical thesaurus for searching for nonstructural features and a patent reactions database as integral components of the GREMAS system. These two parts are not offered by Derwent or CAS. In many cases, however, the full power of the total system only comes into its own by the combination of the three parts of the GREMAS database—structures, reactions, and nonstructural features.

ACKNOWLEDGMENT

We thank B. Stockdale, Derwent Publications Ltd., for carrying out the WPI/L and WPIM searches and for stimulating ideas. We are grateful to IDC, especially G. Berger and M. Koehne for their constructive contributions and helpful discussions. We are thankful to G. Fischer-Huettelmaier, Hoechst AG, for her support. We also thank all members of Hoechst AG Scientific Documentation Team for participating in this test and for their encouragement.

REFERENCES AND NOTES

(1) (a) Fugmann, R.; Braun, W.; Vaupel, W. GREMAS—ein Weg zur Klassifikation und Dokumentation in der organischen Chemie. *Nachr. Dok.* **1963**, *4*, 179–190. (b) Meyer, E. The IDC system for chemical documentation. *J. Chem. Doc.* **1969**, *9*, 109–113. (c) Roessler, S.; Kolb, A. The GREMAS system, an integral part of the IDC system for chemical documentation. *J. Chem. Doc.* **1970**, *10*, 128–134. (d) Fugmann, R. The IDC System. In *Chemical Information Systems*; Ash, J. E., Hyde, E., Eds.; Ellis Horwood Ltd: Chichester, 1975; pp 195–226. (e) Fricke, C.; Nickelsen, I.; Fugmann, R.; Sander, J. GREDIA: A new access to GREMAS databases. *Tetrahedron Comput. Methodol.* **1990**, *2*, 167–175.

(2) (a) Kaback, S. M. Chemical Structure Searching in Derwent's World Patent Index. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 1–6. (b) Simmons, E. S. Central Patents Index Chemical Code: A User's Viewpoint. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 10–15.

(3) (a) Harsdorf, E. V.; Dethlefsen, W.; Suhr, C. Derwent's CPI and IDC's GREMAS: Remarks on their Relative Retrieval Powers with Regard to Markush Structures. In *Computer Handling of Generic Chemical Structures*; Banard, J. M., Ed.; Gower: Aldershot, U.K., 1984; Chapter 10, pp 96–105. (b) Meyer, E.; Schilling, P.; Sens, E. Experiences with input, translation and search files containing Markush formulae. In *Computer Handling of Generic Chemical Structures*; Banard, J. M., Ed.; Gower: Aldershot, U.K., 1984; Chapter 9, pp 82–95.

(4) (a) Nineham, A. W.; Phillips, E. Comparison of Ring and Farmdoc Codes. In *Derwent International Patents Conference Proceedings*; 1978; 340–358B. (b) Silk, J. A. Present and Future Prospects for Structural Searchings of the Journal and Patent Literature. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 195–198.

(5) Watermann, J. R. Using CAS ONLINE to search for patents. In *Computer Handling of Generic Chemical Structures*; Barnard, J. M., Ed.; Gower: Aldershot, U.K., 1984; Chapter 8, pp 76–81.

(6) Schoch-Grübler, U. (Sub)Structure Searches in Databases containing Generic Chemical Structure Representations. *Online Rev.* **1990**, *14*, 97–108.

(7) WPI covers new basic patents from 1963 to 1980, with new equivalents added regardless of date. WPIL contains basic patents added from 1981 onward and their equivalents. The combination of both files is called WPI/L.

(8) An equivalent reference is present, but none of the filed examples matches; the Markush formula in the abstract does not cover the structure required.

(9) An equivalent reference is present, but none of the filed examples matches; the Markush formula in the abstract covers the structure required.

(10) The reference was not indexed.

(11) The answer only matches a higher hierarchical level than the one required.

(12) Equivalent patent to 0D84-070356.

(13) The original patent contains a full structure which meets the query.

(14) The search could not be carried out.

(15) The reference is present; as a result of an error the specific and the general formulations were not coded together as alternatives but divided into two parts.