Polymer Searching in Different Databases

M. HERZ

CIBA-GEIGY AG, Scientific Information Centre, CH-4002 Basel, Switzerland

Received June 1, 1991

Polymer indexing methods are outlined for five major online databases, with respect to addition polymers, polycondensates, and block and graft polymers. Search strategies are compared that depend on designating polymers in terms of their starting materials, structural repeating units, and names, especially in the CAS and Derwent Files. The use of nonstructural terms for locating syntheses, properties, uses, etc. is discussed.

INTRODUCTION

The importance of synthetic polymers in science and technology has increased greatly, and the number of patent documents in this field has increased correspondingly. The most important databases for polymers are listed, along with their producers, in Table I: the bibliographic Chemical Abstracts (CA) File, and its accompanying Registry File for chemical structures,[1,2] produced by the Chemical Abstracts Service (CAS); the Plasdoc component of Derwent's Chemical Patents Index (CPI),[3] which is itself a subset of the World Patents Index (WPI) database; IFI/Plenum's family of CLAIMS databases,[4,5] including both the Uniterm and the Comprehensive Data Base (CDB) retrieval systems; RAPRA with both its bibliographic RAPRA File and the accompanying RAPTN trade name File;[6] and the KKF File of the German Plastics Institute,[7] which I shall refer to by its German abbreviation of DKI.

Table II shows the years of patent and literature coverage for these files online. Derwent and IFI contain patents only, and KKF contains literature only. CAS and RAPRA contain both, although RAPRA's patent component is quite limited. CAS and Derwent both include patent documents worldwide[8] while IFI includes only U.S. patents. RAPRA covered patents only during 1978-1980, for just the U.K. and U.S.

The various hosts in which the databases can be accessed are listed in Table III. As this table shows, CAS is available on all the hosts listed. Derwent and IFI are on three hosts, while RAPRA and DKI are each available on only one host.

Table IV provides information on the number of polymer documents found in these databases as of early 1991. The comparison shows that CA with over 1 050 000 polymer documents is the most comprehensive polymer database, but Derwent with over 886 000 polymer patents is the largest polymer patent database.

Derwent introduced its Plasdoc information service in 1966, equipped with the Plasdoc code for systematic coding and retrieval. That code has been modified a number of times over the years, primarily in 1968, 1972, 1977, 1978, 1982, and 1984. Key term serial numbers (KS), many of them precoordinated combinations of earlier codes, were introduced in 1978 in an attempt to facilitate more systematic searching. In many respects, however, this system no longer meets current needs for precision searching.

In 1967, CAS expanded the Registry System developed originally for low-molecular weight chemical compounds to include polymers. Most polymeric structures are registered in terms of their starting monomeric components. In addition, some substances—primarily condensation polymers—are registered in terms of their structural repeating unit (SRU). It is possible for polymers to have Registry Numbers based both on starting materials and SRU, and work at CAS is currently underway to link those numbers.

In comparing search techniques in the various databases, it is important to be aware of the indexing criteria of each file.

Table I. Polymer Databases

company	file
Derwent	Plasdoc (CPI/WPI)
CAS	CA, Registry
IFI/Plenum	CLAIMS Uniterm, CDB
RAPRA	RAPRA, RAPTN
DKI	KKF

Table II. Databases Content

	patents	literature	online
CAS	yes	yes	1967-
Derwent	yes	no	1966-
IFI	yes[a]	no	1950-
DKI	no	yes	1973-
RAPRA	some[b]	yes	1972-

[a] U.S. only. [b] U.S. and U.K. only, ca. 14 000 patents covered during 1978-1980.

Table III. Hosts

	STN	Dialog	Orbit	Questel	Datastar	BRS
CAS	X	X	X	X	X	X
Derwent		X	X	X		
IFI	X	X	X			
DKI	X					
RAPRA			X			

Table IV. Polymer Documents in Different Data Bases

	total	patents
CA Sections 35-46, 1967-	1 050 823	461 219
Derwent Section A, 1966-	886 494	886 494
IFI, 1950-	149 042	149 042
DKI (KKF), 1973-	211 000	
RAPRA, 1972-	328 554	14 089[a]

[a] 1978-1980.

Thus, for example, CAS has traditionally indexed from literature and patents only those substances, whether of high- or low-molecular weight, which are identified by physical data. For patents that means primarily compounds described in the examples, and then only those examples with hard data. However, beginning about 1980 CAS did begin to index substances claimed specifically, even when they were not backed by hard examples. The Derwent and IFI databases, on the other hand, are in general more liberal in their indexing policies. They aim to index all substances claimed, whether generic or specific, and in some cases also significant disclosures from the specification.

All search examples have been carried out in ORBIT for Plasdoc and RAPRA, and in STN for CAS, IFI, and DKI.

TYPES OF POLYMER

Polymer searching systems must deal with polymer types differing according to their manufacture and structure, including addition polymers, condensation polymers, block

0095-2338/91/1631-0469$02.50/0

Table V. Description of Polymers in the Literature

monomer-based polymers

e.g., HO—$(CH_2)_4$—OH+HOOC—⟨⟩—COOH

structural repeating units

e.g., —(—OC—⟨⟩—CO—O—$(CH_2)_4$—O—)$_n$—

trade names
e.g., Epon 828
polymer classes
e.g., polyester urethanes

polymers, graft polymers, and posttreated polymers. These principle types of polymer may be described in the literature in terms of their monomeric starting materials, structural repeating units, trade names, or polymer classes. Examples of each of these are given in Table V.

Whereas an author of an original publication can use structures, IUPAC names, generic names, molecular formulas, trade names, or abbreviations to describe polymers, each database has to index the polymer structures in accordance with its own guidelines. This indexing may differ from the way in which an author has described the polymers.

ADDITION POLYMERS

The treatment of addition polymers will be discussed first, using poly(methyl methacrylate) as an example. As noted above, this might be described by an author in various ways. Let us consider how it would be handled by Derwent and CAS.

Derwent Plasdoc Code. Coding by Derwent's staff is based on examination of the full patent specification, with some exceptions, most notably Japanese documents. As can be seen from Table VI, the homopolymer of methyl methacrylate has been coded and searchable in the Derwent Plasdoc code since 1966, irrespective of the manner in which the polymer was described by the patentee. Table VI shows the meaning of the codes and a search strategy for ORBIT.

Unequivocal indexing, which means accurate searching, for methyl methacrylate homopolymer has only been possible since 1982 with the advent of the single KS 3011. Before that time, the acid component (methacrylic acid) and alcohol component (methanol) were coded separately. Indeed, even the homopolymer concept was coded separately before 1978. For this reason, false drops can occur in older references, caused by overlaps with other acrylic acids and alcohols. This can happen, for example, when the patent contains simultaneously claimed ethyl *methacrylate homopolymer* and acrylic acid *methanol ester homopolymer*.

In addition to coding, the methyl methacrylate homopolymer is searchable by name in the title and alerting abstract of the online WPI File. The strategy shown in Table VI includes SS5, which shows one way of limiting results to patents in which the polymer is sufficiently important to be named in the title or abstract of the patent. One must of course consider possible variants in the patentee's or abstractor's description of the polymer and combine those variants using the Boolean operator OR. Such a limiting strategy can of course lose significant references. As used here it actually overrides all of the code retrieval. On the other hand, had the free language terms been *OR*ed with the code terms, the effect would have been to *augment* the set retrieved by codes. This can be useful, for example, in retrieving newer references where coding has not yet been completed. The use of free language can be a highly useful adjunct to searching with the Plasdoc code.

Chemical Abstracts Databases: The Registry File. Chemical Abstracts has a separate database containing all structures, the CAS Registry File. Each low- or high-molecular weight structure is normally indexed in the Registry File only once and is identified by its Registry Number.

Methyl methacrylate homopolymer is unequivocally indexed in the Registry System by its CAS Registry Number (RN), CAS index name and synonyms, the molecular formula of starting material and product, as well as their connection tables and fragment screens. The main search methods used for structure searches in STN are listed in Table VII. These include:

1. CA nomenclature or synonyms
2. molecular formula, polymer class identifier, and number of components
3. CA component RN, polymer class identifier, and number of components
4. structural formula and screen fragments

Methyl methacrylate homopolymer can be located on the STN host with any of the search strategies listed in Table VII. The CAS-RN 9011-14-7 can be obtained with each search method, perhaps most simply with the name search for methyl methacrylate homopolymer/CN. Depending on the search strategy used, some nonhits may be expected.

The most-used search strategy in the Registry File is the structural formula search, especially for starting monomers which carry varying substituents. Characteristic screen numbers for polymers can be used to limit the search in the Registry File to polymers (screen 2043) and can further be limiting (e.g., screen 2127, two or more components, for binary or higher copolymers).

Different methods can be used to construct the structural formula for the search. These methods include graphic input or text entry "drawing" with type-3 terminals, using, for example, STN Express for searches on the STN host and DARC on the Questel host.

CA File. The result of a Registry File search can be simply transferred to the CA File with Messenger software on the

Table VI. Derwent Strategy for Searching

	Methyl Methacrylate Homopolymer Plasdoc codes with description		
	077	methacrylic acid esters general	1966 to date
	082	hydroxy component: methyl alcohol	1966 to date
	688	homopolymers general	1966 to date
	0500	methacrylate homopolymers general	1978 to date
	0535	methyl alcohol cont. methacrylate	1978 to date
	3011	methyl methacrylate homopolymer	1982 to date
file	Plasdoc Search Strategy (1966 to date) WPAT; SUBS AM, KS; SYN + FOR AND		
SS1	23786	077 link 082 link 688	
SS2	9829	0500 + 0535	
SS3	7427	3011	
SS4	17569	1+01& or 1 +01- or 1+012 or 1+010 or 2+011 or 3	
SS5	5640	4+(methyl methacrylate: or polymethyl (W) methacrylate: or ...)	

Table VII. Polymer Structure Search Methods

CA nomenclature and synonyms
molecular formula + PMS/CI + (e.g., NC = 1)
CA component Reg No./CRN + PMS/CI + (e.g., NC = 1)
structure + screen fragments

Table VIII. Methyl Methacrylate Homopolymer Search Results Comparison in Different Data Bases

	search with	documents	patents only
CAS			
Registry File	name or Reg No.	1	
CA File	Reg no.	23491	8531
Derwent	Plasdoc codes	17569	17569
Section A		5640	5640
IFI (CDB)	Uniterms	6161	6161
KKF	names	9217	
RAPRA	codes	4154	36
	codes or Names	8262	108

STN host, by crossover of the Registry L-number to the CA File. All abstracts which contain the CAS-RN sought (e.g., the 23 491 abstracts for methyl methacrylate homopolymer with CAS-RN 9011-14-7) are searched, and the CAS-RN is highlighted when printing out the index field using the command DISPLAY HIT. In searching the CA File, polymer classes (e.g., polyacrylate, acrylic polymer), chemical names, and index terms (e.g., acrylic-terminated epoxy resins) may be used, as well as CAS-RN, with varying degrees of success.

Results. Table VIII shows a comparison of search results for methyl methacrylate homopolymer obtained in all five databases under consideration here, over their full time spans. The search was carried out in these databases with codes as well as name terms. Note that the largest number of hits came from the CA File, which includes both literature and patents. Variations among files are the result of many factors, including indexing policies, time span covered, and the number of false drops which might be involved.

CONDENSATION POLYMERS

As in the case of addition polymers, condensation polymers can be described in the original literature in terms of their starting monomers, SRU, polymer class, and the like. But the treatment of these materials by Derwent and CAS differs from that of addition polymers.

Plasdoc Code. In the Plasdoc code, polycondensates are coded by polymer type and starting monomers. For example, urea/formaldehyde resins have been coded since 1966, as illustrated in Table IX. The original coding involved four linked terms. Since 1978 three (unlinked) KS terms have been required: 1276 for the polycondensate type aminoplast, 1731 for urea, and 1517 for formaldehyde.

Chemical Abstracts. For monomer-based indexing, the CAS Registry System uses only the starting monomers and a general polymer class term (PMS/CI) or polymer screen (2043), but *not* the polymer type aminoplast. In the subsequent crossover of the Registry File to the bibliographic CA File, the CAS-RNs obtained are searched in the CA File to give 7473 abstracts, as shown in Table IX. In the other databases, codes or name terms are used. Table IX summarizes the results of the urea/formaldehyde resin search and shows substantial differences among the databases.

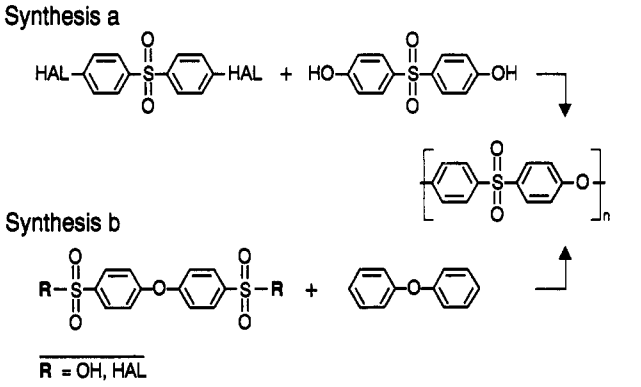
Differences between CAS and Derwent. The differences between CAS and Derwent will be discussed only briefly. The search strategies shown in Table IX for Plasdoc and for the Registry File lead to *all* aminoplasts dervied from formaldehyde and urea, regardless of whether or not they contain additional condensants such as melamine. When searching in the CAS Registry File, a restriction can be made to those polycondensates consisting of just two components, by entering

Table IX. Urea–Formaldehyde Resins

Plasdoc Coding Since 1966		
type of polymers	condensant	condensant
aminoplast	urea	formaldehyde
139/AM	185/AM link 186/AM	180/AM (1966–)
1276/KS	1731/KS	1517/KS (1978–)
Registry Representation Since 1967		
polymer term	condensant	condensant
	urea	formaldehyde
PMS/CI	57-13-6/crn	50-00-0/crn

Search Results Comparison in Different Data Bases			
	search with	documents	patents only
CAS			
Registry File	names or Reg Nos.	762	
CA File	Reg Nos.	7473	4859
Derwent	Plasdoc codes	10144	10144
Section A	codes and names	10277	10277
IFI (CDB)	Uniterm codes	4267	4267
	Uniterms or names	4357	4357
KKF	names	915	
RAPRA	codes	439	39
	codes or names	1041	105

Table X. Polyether Sulfone Syntheses



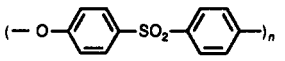
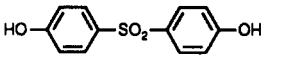
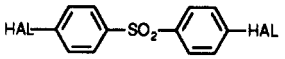
the additional condition "number of components equal 2" (NC = 2). But this restriction cannot be made in the Plasdoc code.

Another difference between the CAS Registry System and the Plasdoc code is that all monomer structures can be indexed and searched accurately in the Registry File. When searching Plasdoc, however, "others-codes" must be used for all but the most frequently occurring monomers, which have separate code positions. A total of about 750 "others-codes" are used in Plasdoc for structural terms. For example, 2,2-dimethyl-1,3-propanediol can only be indexed and searched in the diol condensant field by using the "other-code" KS 1329, which means other aliphatic diols. This code position 1329 is contained in 6267 Derwent abstracts at this writing and so is not at all selective. But this compound can be indexed accurately in the Registry File.

The essential differences between CAS and Derwent as regarding the indexing and retrieval of polycondensates will now be illustrated by a discussion of procedures for polyether sulfones.

Plasdoc Code. Table X illustrates two synthesis routes which, although they lead to the same type of polyether sulfone, are coded differently in Plasdoc. The reason is that synthesis a is a polyetherification, whereas synthesis b is a sulfonation at the aromatic nucleus. Because the polyether sulfone obtained by method a is synthesized by *polyetherification*, the polymer type is coded as POLYETHER CONTAINING SULFUR with KS code 1279 for polyethers and KS code 0203 for SULFUR CONTAINING POLYMERS. On the other hand, the polyether sulfone obtained by *polysulfonation* in method b is coded as POLYSULFONE

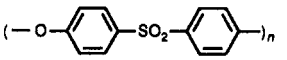
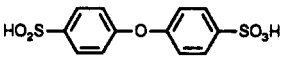
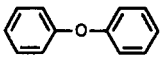
Table XI. Synthesis a

	CA Registry			Derwent	
	2043	polymer	(L3)	0016	ring in backbone
	2127	≥2 comp	(L4)	1279	polyether
	2077	≥3 comp	(L5)	0203	other S-contg polym
		structure L1		1377	other bisphenols
				3079	dihydroxydiphenyl sulfone (1982 to date)
				1962	other S-contg condens
		structure L2		1920	other aromat condens
				1962	other S-contg condens

Search Strategy

CA Registry
L1 and L2 and L3 and L4 not L5
Derwent
KS 0016 and 1279 and 0203 and 1920 and 1962 and (1377 or 3079)

Table XII. Synthesis b

	CA Registry			Derwent	
	2043	polymer	(L3)	0016	ring in backbone
	2127	≥2 comp	(L4)	1309	polysulfone
	2077	≥3 comp	(L5)	1279	polyether
		structure L1		1920	other aromat cond
				1962	other S-contg cond
		structure L2		1920	other aromat cond

Search Strategy

CA Registry
L1 and L2 and L3 and L4 not L5
Derwent
KS 0016 and 1279 and 1309 and 1920 and 1962

with KS code 1309 and POLYETHER with KS code 1279.

The starting condensants for each of the polyether sulfones can only be coded in Plasdoc under the respective "other monomer" KS codes, namely

1. 1920/KS other aromatic condensants
2. 1962/KS other sulfur-containing condensants
3. 1377/KS other bisphenols

Dihydroxydiphenyl sulfone has only had its own KS code 3079 since 1982.

CAS Files. Searching in the CAS Files requires a different strategy. The polycondensate type cannot be indexed in the CAS Registry File, but an accurate indexing of all condensants (starting components) is entirely possible. A systematic search for synthesis a and synthesis b is therefore possible on the basis of the monomers. The polyether sulfones can be searched as POLYSULFONES from the 10th Collective Index on, and as SULFONES before the 10th Collective Index, after crossover from the Registry File to the CA File.

Tables XI and XII show the difference between the indexing and search strategy for Plasdoc and CAS Registry for syntheses a and b.

STRUCTURAL REPEATING UNITS

Authors often use the smallest structural repeating unit (SRU) to describe a polymer, especially if the starting monomers are not known or if the structure of the SRU is established by analysis. Polybutylene terephthalate is shown as an example of an SRU in Table XIII. Only selected polycondensates are indexed as repeating structural units in the CAS Registry System, and no polycondensates can be indexed as SRUs in Plasdoc.

Table XIII shows the types of polymers indexed as SRUs in the CAS Registry System. In general CAS uses SRU indexing for polycondensates to supplement monomer-based indexing or where the starting materials are not indicated. In special cases, for example nylon 6, CAS uses only the SRU.

TRADE NAMES

A large number of polymers are indexed and searchable by their trade names, especially in the CAS Registry File and in the RAPRA RAPTN File. The pertinent bibliographic data are contained and searchable in the complementary CA File and RAPRA File databases. The essential difference between RAPTN and the Registry File is that the latter contains nomenclature, molecular formula, and structure where known, whereas RAPTN contains only the trade name. The indexing of trade names in the WPI, CLAIMS, KKF (DKI), and CA databases is not exhaustive, so that trade name searches are inevitably incomplete.

BLOCK POLYMERS

Turning now to block polymers, we find that the indexing of these proliferating polymers differs from one database to another, especially in CAS and Derwent, and depending on whether they are addition polymers or polycondensates.

Plasdoc Coding. In Plasdoc, only addition polymers can be indexed or searched with the KS code 0002 for block copolymers. Block condensates are coded with the KS code 0004 for co-condensates and the KS code 0005 for ordered co-condensates. A more selective, but incomplete, result can also be obtained when searching for polycondensates by using the natural language word BLOCK, if that term has been used in the title or abstract in the WPI File.

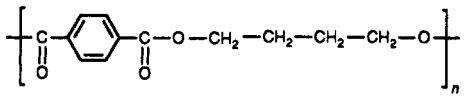
CAS Databases. Since 1987, BLOCK COPOLYMERS and BLOCK POLYCONDENSATES have been indexed in the Registry File under their own CAS-RN. After the chemical structure search, the keyword BLOCK can be used to make a selective restriction to block polymers and block polycondensates in the Registry File. This is an important difference in coding as compared with Plasdoc.

Prior to 1987, block polymers and block polycondensates had no separate CAS-RN. The indexing and searching strategy was the same as for any copolymers and co-conden-

Table XIII. CAS Registry System

Section A

type of structural repeating units for polymers
for polycondensates only
containing one or two monomers only
symmetrical monomers only
supplement for monomer-based entry



Section B

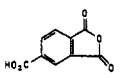
STN INTERNATIONAL®

REGISTRY FILE SEARCH RESULTS - P094150S 04 APR 91 05:18:35 PAGE 4

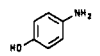
L1 ANSWER 1 OF 1

RN 125223-67-8
IN 1,3-Benzenedicarboxylic acid, polymer with 4-aminophenol, 1,4-benzenedicarboxylic acid, [1,1'-biphenyl]-4,4'-diol, 1,3-dihydro-1,3-dioxo-5-isobenzofurancarboxylic acid and 4-hydroxybenzoic acid, block (9CI)
MF $(C_{12}H_{10}O_2.C_8H_4O_5.C_8H_6O_4.C_6H_6O_2.C_6H_7NO)_x$
STE 8:PM,BLOCK

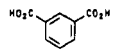
CM 1



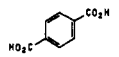
CM 2



CM 3



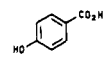
CM 4



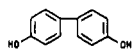
CM 5

STN INTERNATIONAL®

REGISTRY FILE SEARCH RESULTS - P094150S 04 APR 91 05:18:35 PAGE 5



CM 6



STN INTERNATIONAL®

CA FILE SEARCH RESULTS - P094151T 04 APR 91 05:18:42 PAGE 4

L2 ANSWER 1 OF 1

CA112(10):78244h Preparation of polyimide esters) with good mechanical strength. Tagami, Sanae; Takeya, Tetsuro; Fujiwara, Kenichi (Idemitsu Kosan Co., Ltd., Japan). Jpn. Kokai Tokkyo Koho JP 01197527 A2 9 Aug 1989 Heisei, 9 pp. (Japan). CODEN: JKXXAF. CLASS: ICM: C08G073-16. APPLICATION: JP 88-19561 1 Feb 1988.

AN CA112(10):78244h

AB Polyamic acids prepd. from trimellitic anhydride (I) and aminophenols are mixed with condensing agents or their mixts. contg. ≥1 of arom. hydroxy carboxylic acids or their derivs., arom. dihydroxy compds. or their derivs., arom. dicarboxylic acids, and poly(alkylene terephthalates), cyclized, and melt polymd. to give title polymers with high mol. wt., useful for elec. or mech. parts, films, fibers, etc. Thus, 154 g I in acetone was treated with 87 g *p*-aminophenol to give a slurry, which was stirred with 122 g Ac_2O and 28 g *p*-hydroxybenzoic acid at 300° and then at 320° and 1 mmHg to give a title polymer showing reduced viscosity (0.2 g/dL soln. in C_6F_5OH at 60°) 0.7 dL/g. The polymer was spun at 370° to give 30-μm fiber showing tensile strength 0.14 GPa and tensile modulus 12 GPa.

IT 125223-62-3P, *p*-Aminophenol-*p*-hydroxybenzoic acid-trimellitic anhydride block copolymer 125223-63-4P 125223-64-5P, *p*-Acetoxybenzoic acid-*p*-aminophenol-4,4'-diacetoxybiphenyl-isophthalic acid-terephthalic acid-trimellitic anhydride block copolymer 125223-65-6P 125223-66-7P ***125223-67-8P***
(prepn. of, for fibers and films, with good mech. strength and high mol. wt.)

Table XIV. Graft Polymerization

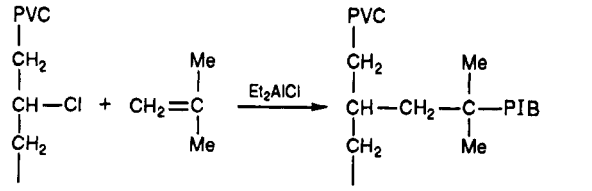


Table XV. Derwent Search Strategy for Vinyl Chloride–Isobutylene Graft Polymers

0003/KS	graft copolymer
0761/KS	CH_2=CH—Cl binary copolymer
0264/KS	CH_2=$C(Me)_2$ binary copolymer

File WPAT; subs KS

SS1	48	0003 and 0761 and 0264 (since 1978)

Table XVI. CAS Search Strategy for Vinyl Chloride–Isobutylene Graft Polymers

		Registry File	
L1	1 s	vinyl chloride/cn	(75-01-4)
L2	1 s	dimethyl ethene/cn	(115-11-7)
L3	2 s	75-01-4/CRN and 115-11-7/CRN and PMS/Cl and NC = 2	
L4	1 s	L3 and graft?	(1987 to date)
L5	1 s	L3 not L4	(since 9187)
		CA File	
L6	3 s	L4	(1987 to date)
L7	24 s	(L5/D or L5) (L) Graft?	(since 1987)
L8	11 s	(L6 or L7) RAN = (1978,)	(1978 to date)
L9	3 s	L8 and p/dt	(1978 to date)

sates made of the same monomer units. The corresponding block polymers and block polycondensates can only be searched specifically with the link operator in the IT field after the crossover of CAS-RN from the Registry File to the CA File. By way of example, styrene–butadiene block copolymer can be searched by linking the RN 9003-55-8 with the term "block".

Thus, different strategies must be used for the period before and after 1987. But in contrast to Plasdoc, the term block is utilized for all types of polymers, not just addition polymers. One shortcoming in the Registry File is that the arrangement of the sequences in block copolymers cannot be identified. For example, since 1987 the block condensate with the RN 125223-67-8 is indexed in the STE field of the Registry File with BLOCK and is searchable. But it is impossible to figure out the construction of the sequences of this block co-condensate from either the Registry File entry or from the CAS abstract 112:78244*h* (Table XIII, section B).

GRAFT POLYMERS

Graft polymers are formed by initiating the polymerization of further monomer(s) at sites along the backbone of a substrate polymer. Table XIV illustrates the graft polymerization of isobutylene onto PVC, using an alkylaluminum chloride as initiator. The following tables show that, for the period from 1978, one search strategy is needed for Derwent, whereas two are needed for CAS.

Plasdoc Code. Plasdoc coding is illustrated in Table XV, the strategy being the same as for random copolymers, but using the additional code for GRAFT. For the period since 1978

1. 0003/KS for graft copolymer
2. 0761/KS for binary copolymer of vinyl chloride
3. 0264/KS for isobutylene binary copolymer

CAS Indexing. In contrast, since 1987 indexing in the Registry File and the CA File has differed from indexing prior

Table XVII. Vinyl Chloride–Isobutylene Graft Polymers Database Comparison 1978–1991

	search with	documents	patents
CAS			
Registry File	structures	2	
CA File	Reg Nos.	11	3
Derwent			
Section A	codes	48	48

Table XVIII. CAS Search Strategy[a]

			Registry File	
L1	1	S	trimellitic anhydride/cn	(552-30-7)
L2	1	S	4,4′-diaminodiphenylmethane/cn	(101-77-9)
L3	2	S	isophthalic acid/cn	(121-91-5)
L4	107	S	552-30-7/crn and 101-77-9/crn and 121-91-5/crn and PMS/CI	
			CA File	
L5	20	S	L4/P	(P = synthesis)
L6	20	S	L5 and P/DT	(P/DT – patents)
L7	23		select L5 HIT 1-20 RN	(23 HIT-Reg Nos.)
			Registry File	
L8	23	S	E1-E23	(23 HIT-Reg Nos.)

[a] Synthesis of polyamideimides from trimellitic anhydride and isophthalic acid and 4,4′-diaminodiphenylmethane.

to that date. As Table XVI shows, two different search strategies are needed:

1. Since 1987 in the Registry File, direct linkage of component terms to GRAFT, and crossover of the result in L4 to the CA File.
2. Before 1987 in the Registry File component terms are searched without use of the term GRAFT, crossover to CA and then linkage to GRAFT?.

Table XVII gives a comparison of results in CAS and Derwent databses, based on searches at CIBA-GEIGY.

IMPROVEMENT IN SELECTIVITY BY USING NAME TERMS

When making polymer searches, the results can be enhanced by using nonstructural text terms or role indicators in all databases, especially in regard to synthesis methods, properties, uses, and the like.

Synthesis Methods. The synthesis of polyamideimide cocondensates of trimellitic anhydride, isophthalic acid, and 4,4′-diaminodiphenylmethane will be used to illustrate the different search strategies in CAS, Derwent, and RAPRA; and the results obtained in the period since 1978 will be compared. As may be seen from the search strategy shown in Table XVIII, a structure query with component RNs in the CAS Registry File, followed by crossover of the resulting RNs in L4 to the CA File with the role "P", leads only to those abstracts in which the preparation of the RNs found in the Registry File is actually described. With the "SELECT HIT" command it is possible to locate the 23 prepared RNs of the 107 indexed ones and to search by repeated crossover to the Registry File. The result obtained with the CAS system produces only hit-RNs and hit-abstracts.

Table XIX shows the Plasdoc codes with the appropriate search strategy. Use of the KS code 2148 for cocondensation and the KS code 2149 for ordered cocondensation is essential for limiting the result to syntheses only.

Table XX shows that, in searching the RAPRA File for polyamideimides, one must use not only the relevant codes but also free language terms for polyamideimides. Similarly, the concept "synthesis" must be searched with free words as well as with codes. To limit the search, only the principal starting monomer trimellitic anhydride was used. Table XXI shows the comparison of results obtained with CAS, Derwent, and RAPRA. Only the CAS result produces exclusively hits; the other databases merely come up with candidates.

Uses and Properties. Polyamideimides obtained from trimellitic anhydride and 4,4′-diaminodiphenylmethane can be used as coatings for electrical wires or cables. Two searches must be made in the CAS Registry File:

1. (1) with the SRU molecular formula $(C_{22}H_{14}N_2O_3)_n$, giving 2 hit RNs
2. (2) with the CRN of the monomers and NC = 2, giving 1 hit RN

The total of three hit RNs results in 85 abstracts in the CA File in the period since 1978. A restriction using the keywords ELEC or ELECTRIC in conjunction with WIRE or CABLES, and also the CA subsection 42-10 (coatings for wires and cables), produces the result shown in Table XXII.

In the Derwent WPI File, 288 abstracts were obtained with Plasdoc structure coding, and further restricted as follows:

1. (1) KS terms for electrical properties in conjunction with KS terms for coatings on wire, cable, or wire nets
2. (2) Natural language words ELECTRIC in conjunction with WIRE or CABLE
3. (3) Manual code A12-EO2: for cable and wire insulation or coating.

The result in CAS, WPI, and RAPRA is shown in Table XXII. It is conspicuous that uses and properties coded in Plasdoc are not always described in the altering abstract but often appear in Derwent's detailed documentation abstracts.

Table XIX. Derwent Search Strategy[a]

Plasdoc codes		search strategy			
1283	polyamide			FILE WPAT; SUBS AM, KS; SYN + FOR AND	
1285	polyimide				
1487	other aromatic polycarboxylic acids	SS1	118	1283 + 1285 + (1487 or 3107) + (1717 or 3117) + 0016 + 1458 + (0004 or 0005)	
3107	trimellitic acid (since 1982)				
1717	other diamines contg aromatic rings				
3117	diaminodiphenylmethane (since 1982)				
0016	condensat. polymer contg ring in the backbone				
1458	isophthalic acid	SS2	53	1 + 3107 + 3117	
0004	cocondensate	SS3	53	2 + 011 or 3	
0005	cocondensate ordered	SS4	38	3 + (2148 or 2149)	
2148	cocondensation				
2149	cocondensation ordered				

[a] Synthesis of polyamideimides from trimellitic anhydride and isophthalic acid and 4,4′-diaminodiphenylmethane.

Table XX. RAPRA Search Strategy[a]

		RAPRA File
SS1	474	polyamid# # (W) imid: or copolyamid# # (W) imide: polyamideimide: or polyimidamide: or polyimideamid: or polyimidamide: or polyimideamid: or polyimid# # (W) polyamide: or polyamid# # (W) polyimid: or polyimid# # (W) amide#
SS2	16	43C4/CL and 43C318/CL
SS3	182	(1 or 2) and (723/CL or 724:/CL or synthesis or syntheses or copolymerisatio: or preparation# or manuf: or prepared or prep or prepn)
SS4	145	3 and greater than 77
SS5	28	4 and trimellit:
SS6	1	5 and p/dt

[a] Synthesis of polyamideimides from trimellitic anhydride and isophthalic acid and 4,4'-diaminodiphenylmethane.

Table XXI. Synthesis of Polyamideimides Database Comparison 1978–1991

	search with	documents	patents
CAS			
Registry File	structures	23	
CA File	Reg Nos./P	20	20
Derwent	Plasdoc		
Section A	codes	48	48
RAPRA	codes or terms	28	1

Table XXII. Application of Polyamideimides Database Comparison 1978–1991

	search with	abstracts	patents
CAS			
Registry File	structures	3	
CA File	Reg Nos. and terms	28	27
Derwent			
Section A	Plasdoc and terms manual codes	60	60
RAPRA	codes or terms	12	2

Coding is, as already noted, based on the full patent specification rather than on any abstract. Note that the polyamideimide can only be searched in very general fashion in the RAPRA File, but the use is readily searchable.

CONCLUSIONS

It is clear from the results obtained in the different databases that the CAS Files are the only databases which lead directly to hit-RNs, i.e., hit-structures, and then via crossover to the CA File to hit-abstracts. All the other investigated databases produce only candidates from which false drops have to be discarded by subsequently reading the abstracts or documents. On the other hand CAS' restrictive indexing policies can lead to the failure to retrieve documents encompassed by generic patent claims or lacking hard data.

In regard to the CAS Registry File, I believe it is important that related SRU-based and monomer-based RNs should be indexed in the same record, and vice versa, so that all polymer structures can be obtained with one search. In addition, I feel that an indication of the class of the polymer would be extremely helpful in carrying out some searches. Indexing the sequences of block and graft polymers in a separate field of the Registry File would ensure better identification of the block and graft structure.

In Plasdoc, I would especially like to see the number of frequently used "other" code positions be reduced. I am also persuaded that the indexing of functional groups which result in the polymer backbone should be supplemented by the indexing of all functional groups, including nonreacting ones, so as to ensure greater selectivity when making chemical structure searches.

I might point out that these and other possible improvements have already been proposed to CAS and Derwent in working groups. I am convinced that the early introduction of these improvements is essential to obtain more selective search results in the future.

ACKNOWLEDGMENT

I thank Mrs. E. Jorda for her assistance in testing the RAPRA databases. I also thank M. Schellenbaum, B. Somers, and J. Fryar for valuable assistance in preparing this article.

REFERENCES AND NOTES

(1) *CAS ONLINE*, Available on STN International—Searching for Polymer Information in CAS ONLINE; Chemical Abstracts Service: Columbus, OH, 1990.
(2) Langstaff, E. M.; Ostrum, G. K. Access to Polymer Information in Chemical Abstracts: *J. Chem. Inf. Comput. Sci.* **1979**, *19* (2), 60–64.
(3) *CPI Plasdoc Retrieval Manual, Instruction Manual No. 4A, Combined Plasdoc Punch Code and Key Terms*; Derwent Publications Ltd.: London, England, 1988.
(4) Lambert, N. How to Search the IFI Comprehensive Database Online... Tips and Techniques: *Database* **1987**, *10* (6), 46–59.
(5) Kaback, S. M. The IFI/Plenum Chemical Indexing Systems. *In Computer Handling of Generic Chemical Structures*; Barnard, J. M., Ed.; Gower: Aldershot, 1984; pp 49–65.
(6) RAPRA (Rubber and Plastics Abstracts Database). *Online User's Manual*; RAPRA Technology Ltd.: Shrewsbury, England, 1988.
(7) DKI (Deutsches Kunststoff-Institut) DKI Thesaurus–Kunststoffe Kautschuk Fasern; 1985.
(8) Kaback, S. M. Polymer Patent Information Systems Could Be Even Better. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, (4), 371–379.