



White Paper

Polymer Searching: An Achilles Heel of Information Specialists

Authored by Dr. Yogesh Nagpal

This white paper is written with an aim to understand the complexities involved in the scientific information retrieval process in the domain of polymer technology and to briefly discuss the various tools/means used by the industry experts to perform polymer search projects.

Table of Contents

1. Introduction.....	3
2. What constitutes polymer information	3
3. What makes searching polymer information a difficult task.....	3
4. Means of searching Polymer Information and their Evolution.....	4
5. Platforms for searching polymer literature (especially Patents).....	7
6. Important databases for Polymer Searching:	7
7. Conclusion:	7
8. Sources & References:	8

1. Introduction

Technology is advancing at an enormous pace. The outcome of this increased research activity is that the volume of scientific information which needs to be organised, administered, stored and evaluated has increased significantly. Finding out relevant information or documents from this enormous amount of information sources (such as patent applications, scientific articles, news, blogs, whitepapers, internet citations, internal databases, external databases and many more) is an extremely challenging task. The job of information professionals though looks easy to some but believe me, it is extremely cumbersome, tiring and at times frustrating. Though this challenging scenario is pretty consistent in almost all the technology fields, but when it comes to retrieving polymer science and technology related information, the complexities increase exponentially.

Information retrieval in case of polymer searching projects, creates almost a nightmare scenario for information specialists. This can be ascribed to the fact that the use of routine search techniques (i.e. keyword or patent classification based searches) fail to deliver the expected results. By this, I mean that either you do not capture the relevant documents or you end up with huge list of non-relevant documents hiding the relevant documents of your interest. Optimizing search strategy in terms of good precision and recall is a daunting task in the technology domain of polymer science.

2. What constitutes polymer information

Majorly the scientific literature dealing specifically with polymers are concerned about starting monomers, type of polymer, number of repeating units, final polymeric structure, post-treatment, polymer chemistry and properties, polymer processing, applications of polymers, polymer compositions and formulations, physical parameters,

reaction catalyst, initiators, plastics manufacture and processing etc. (Fig 1).

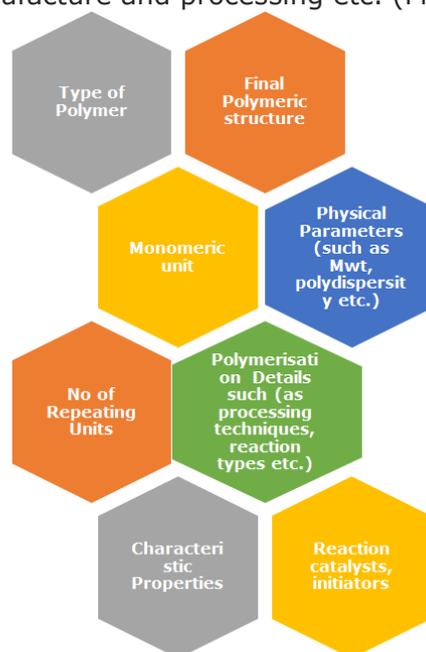


Figure 1 : What constitutes Polymer information

3. What makes searching polymer information a difficult task

Some of the prominent factors owing to which polymer searching is a challenging task are outlined below:

1. Nomenclature of Polymers: A single polymeric structure is named in multiple ways because of the inconsistency in naming methodologies of different systems. The four different systems i.e. IUPAC, CAS, MDL, SCION use different polymer nomenclature and styles. To add to this, very often polymers are represented by their industrial names/tradenames or as polymeric structures (such as structure I or structure II etc.).

2. Ambiguity of Polymeric Structures: Polymer structures stored in databases are frequently an ideal or theoretical representation, rather than a true representation of the actual polymer structure

3. Multiple Parameters: The number of parameters that can be varied during synthesis, formulation, and processing of polymer is extremely large. For example monomers, initiators, monomer/initiator

ratio, concentrations, temperatures, pressures, additives, stabilizers, etc.

4. Multiple Characterisation Techniques:

There is an extensive amount of characterization and screening data, originating from both classical polymer analysis (T_g , T_m , M_n , M_w , polydispersities) and other materials analytical techniques (conductivity, elasticity, hardness, blend formulations, etc.).

4. Means for searching Polymer Information and their Evolution

Information searchers perform various types of searches on multiple sources/search platforms/databases to capture patent/scientific documents of interests. Some of the most common types of searches are based on keywords, author, assignee, inventor, patent classification codes (viz. IPC, CPC, USCLA, ECLA, FI/F terms etc.), citations etc. But while dealing with polymer literature, all these types of searches become ineffective in terms of precision and recall.

In order to enable information specialists to conduct efficient polymer searches, Plasdac introduced polymer coding/indexing systems in the year 1966. This was the first attempt towards systematic indexing of polymeric information present in patent literature.

Ever since then, polymer coding system has evolved over its lifespan. In this evolution process, some indexing systems fell out of use as they get replaced by better indexing systems. (Figure 2)

Some of the most notable polymer coding/indexing systems are briefly discussed below

AM Punch codes: The polymer multipunch codes were introduced in the year 1966.

The punch card contains 80 columns and 12 rows which represented 960 possible punch positions for encoding data (i.e. chemical fragment). A single concept is represented by group of punch positions, as there is limit to number of positions for encoding data (i.e. only 960). Fig 3 represents the punch code representation of the concept acrylic acid monomer.

Both generic and specific searching can be done with the help of AM codes, but because of the limitation of multiple usage the search precision gets compromised.

This coding system has been discontinued from Derwent Update 199501. However, these codes can be searched from 1966 through the end of 1994. The multipunch searches can be done by using the

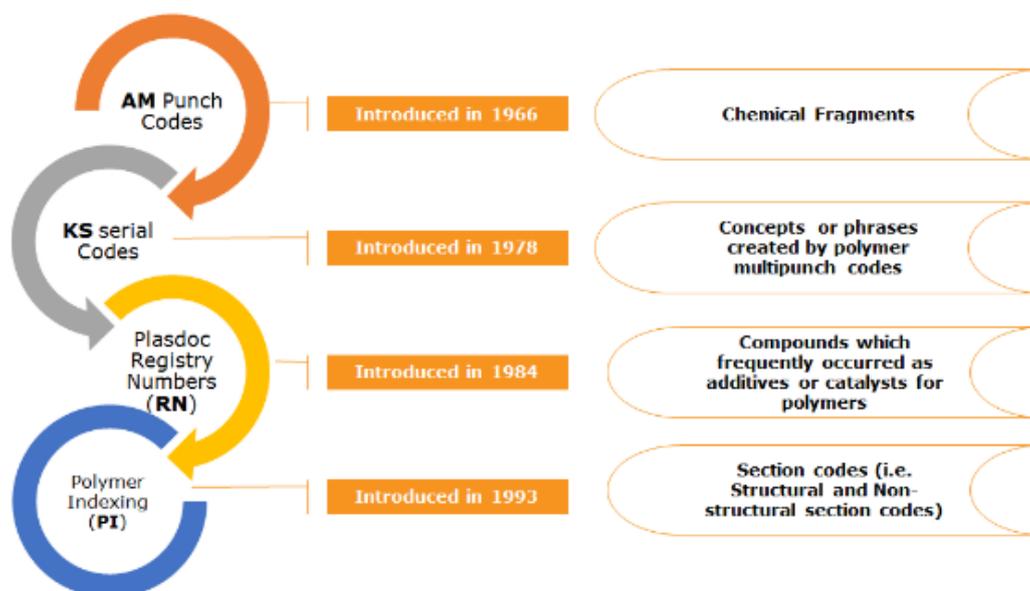


Figure 2 : Evolution of Polymer indexing system

qualifier PF= on Dialog, /AM on Questel Orbit and /AM or /FG on STN.

Punch Code & its description	
074	Monoolefinic Acrylic Compounds
075	Acid or metal salts
076	Acrylic Acid or Derivatives
343	Monomer or condensate

Concept : Acrylic Acid Monomer
(AM Punch Code Representation
074.075.076.343)

Figure 3 : AM Punch code representation of Acrylic acid monomer

KS (Key Serial) Codes: Key serial codes were introduced in the year 1978 with an aim to improve precision in searches. The primary advantage of KS codes was that they were unique for each concept. They represent single concepts or phrases created by combinations of Polymer Multipunch codes, which means key serial codes were assigned to pre coordinated groups of punch codes. For ex. Concept Acrylic acid monomer is represented as KS code 0414.

The distinguishing factor between the KS and AM codes are that KS codes are non-hierarchical whereas AM codes are hierarchical. The KS codes are available for searching DWPI from 1978 through to end of 1994. The key term serial number can be searched by using the qualifier /KS on Questel Orbit and STN; KS= on Dialog).

Plasdoc Registry Numbers (RN): In 1984, Plasdoc Registry Numbers were introduced to further improve search precision. These numbers were assigned to those compounds which frequently occurred as additives or catalysts for polymers. Polymer formers or monomers are not assigned Plasdoc registry numbers. Searching for specific compounds as additives and catalysts for polymers can be carried out from 1984 to 1994 by use of Plasdoc Registry numbers. The Registry numbers can be searched using the qualifier DR= on Dialog; /DR on Questel Orbit and/DRN on STN.

Polymer Indexing (PI): Enhanced polymer indexing system introduced in the year 1993 completely replaced the previously used AM and KS indexing systems. This indexing system adopts

hierarchical structure and inherent the benefits of both the previously used AM and KS concepts. These codes are applied to all claimed and claim-related example polymers found in the patent specification. Enhanced Polymer Indexing comprises two types of section codes i.e. Structural and Non-structural section codes, each of which is further subdivided into facets. The structural section contains all the chemical information, whereas the Non-structural section contains all the non-chemical information (Fig 4).

Structural Section	Non Structural Section
Polymer Formers	Novelty Descriptors
Chemicals	Universal Terms
Natural Polymers	Polymer Descriptors
Polymer types	Shape & Form
Modified Polymers	Additives
Chemical Aspects	Catalysts
	Chemical Processes
	Physical Operations
	Equipment
	Properties
	Applications

Figure 4 : Sections of Enhanced Polymer Indexing

Some of the salient features which makes this indexing system better over its earlier counterparts is outlined below

1. This indexing system has been designed to perform both specific as well as generic searching of compounds encountered in polymer domain. Specific compounds are represented by SCN (i.e. Specific Compound Numbers) whereas generic codes are also available for conducting broad searches by using generic aspects.
2. The functional aspects are also provided along with the chemical attributes
3. Provision of Linking functionality for improving searching precision and flexibility required in such a complex technology. This helps in associating

different concepts and thus improving precision in searching.

The Polymer Indexing system had undergone revision in the year 1996 and 2004 to introduce codes for new concepts.

Apart from these indexing systems, some other advanced options for searching polymer or chemical literature are briefly discussed below

Derwent Manual Code: It is a hierarchical classification and patent indexing system which is applied by the indexing team of Thomson Derwent. One of the major differentiating factor of Derwent Manual codes with other patent classification systems (such as IPC's) is that they are applied only to the inventive features described by the basic patent document. These codes cover both the novel technical details disclosed within the patent, and the invention's application. Thus, noise in the result set gets reduced significantly and these codes are extremely useful for conducting broad subject searching.

In order to use this indexing system for searching chemical patent literature (i.e. Derwent sections A-M) one needs to opt for special subscription option of Chemical Patent Index (CPI Manual Codes).

Polymer Class Terms: The polymeric compounds present in the Registry database are broadly classified into 42 different categories known as Polymer Class Terms. The categories are based upon the structural attributes of the polymeric compounds. Each category or Polymer Class term reflect

- 1) The types of linkages formed in the polymerization process
- 2) Linkages already present within the monomer backbone
- 3) Linkages present in structural repeating unit (SRU) backbones

Some of the representative examples of Polymer Class Terms and the kind of

results captured from using it are outlined in Table 1

Class Term	Code	Type of Polymer Retrieved
Amino Resin	AR	Condensation polymers of amines with aldehydes
Epoxy Resin	ER	Epichlorohydrin polymers with a diol. Polymers of monomers containing two or more epoxy group
Polyacetylene	PACT	Monomers containing a carbon carbon acyclic triple bond
Polyether	PETH	-O- ether linkages in backbone, polyoxymethylenes, polyoxyalkylenes, polyoxyarylenes, and polyoxyphenylenes

Table 1: Representative examples of Polymer class terms

Please note Polymer Class terms are assigned only to polymers, which means that oligomers such as dimer, trimer, tetramers etc. are not classified under Polymer Class Term Indexing scheme. Coordination compound polymers gets classified under the Polymer class term POLYOTHER or one of the MANUAL class terms. If the classification for a polymer is uncertain, it also receives the class POLYOTHER.

Control Term: Controlled Term vocabularies or database thesaurus is also pretty useful tool for performing polymer searching and overcome the challenges faced by information searchers. To perform control term based searches users need to refer database thesaurus and identify the relevant control terms (or indexed terms) first. Typically, these terms are indexed manually by the experts of database providers.

Chemical Registry Number: Registry has more than 1 million polymer substance records and is updated on daily basis. Each polymeric compound which is included in the Registry database is assigned a unique CAS Registry number.

Registry contains mainly two types of polymeric records (1) Monomer based Registry numbers and (2) Structural Repeating Unit (SRU) based CAS Registry numbers. The primary registration of polymer is done on the basis of monomeric units present. This means a polymeric compound containing component monomers will have CAS registry number and each monomeric compound will also have its own unique CAS Registry numbers. Further, monomer

based records for condensation polymers can be supplemented with a structural repeating unit based entry or registration number, provided SRU is well documented or it is the only chemical structure possible.

The same target polymer can be formed via different routes or different set of monomers. Further, a polymer can be registered via monomer based as well as SRU based registration. This creates problems in performing CAS Registry number based searches. This difficulty can be overcome by using POLYLINK command, with which you can retrieve CAS registry numbers for the monomer based and SRU representation for a given condensation polymer or set of polymers. Thus, Polylink command identifies all related condensation polymers.

Apart from CA/Caplus databases, CAS registry number is also included in some other important polymer databases such as CBNB, CIN, PROMT, IFIPAT, IFICDB, IFIUDB, USPATFULL, and USPAT2. Therefore, using Chemical registry number feature for strategizing polymer search projects is very effective option.

Chemical structure based search: Chemical structure searching is an invaluable tool in the arsenal of competent searcher, especially for accessing the information contained in patent literature. Polymer structure searching is similar to searching any routine organic molecule. If your target compound for searching is polymer, then you can perform the chemical structure based search on STN platform by drawing the monomeric unit or SRU like any other organic molecule. However in order to restrict your search results to polymers only you need to apply polymer screens. Provided below are

Polymer Screen	Definition
2043	Any polymer with a structure or a manual CAS RN
2067	Homopolymer or copolymer (non-SRU)
2068	Any SRU
2069	SRU with end groups
2070	SRU without end groups

Table 2: Polymer Screen/filters on STN

some of the examples of the polymer screens

The above mentioned polymer screens can be applied by using SCREEN command online (Ex. SCREEN 2043). Typically, structure based polymer search can be performed by drawing the monomer fragments or by drawing the SRU. The structures can be drawn offline using STN drawing tool.

5. Platforms for searching polymer literature (especially Patents)

The three most notable hosts/online vendors which offer access of various polymer searchable databases are STN, Dialog and Questel. The users need to have subscription of at least one of these search platforms, in order to perform specialized polymer searches on one or more databases. Apart from these Scifinder is also an excellent resource for chemistry focused searches. Though it is more user friendly, but it lags behind in its coverage.

6. Important databases for Polymer Searching:

Registry, Caplus, WPIX and IFICDB are amongst the most frequently used polymer information databases. All of these can be accessed via STN platform (Table 3).

Database	Content		Search Type		Coverage
	Patent	Literature	Structure	Bibliographic	
REGISTRY	Yes	Yes	Yes	Yes	1907 -
CAPLUS	Yes	Yes	Yes	Yes	1840 -
WPIX	World Patents	-	-	Yes	1966-
IFICDB	US-Patents only	-	-	Yes	1950 -

Table 3: Important Polymer databases on STN platform

7. Conclusion:

These days there are a number of means available at the disposal of searchers to strategize polymer searches viz. polymer indexing, Derwent manual codes, structure based searching etc. However, none of them will find all the relevant documents alone. The final search strategy will always be devised by combining these means. Performing high

quality and cost effective polymer searches, requires very good technical knowledge coupled with detailed know how of information tools and databases. There is enormous amount of information available, but to search and identify the ones that address your business needs is a challenging task. It is always advisable to seek professional help in strategizing and executing polymer focussed search and analytics projects.

At Winsome IP, we have a dedicated team of experienced information specialists comprising chemistry doctorates and postgraduates who perform these specialized searches. Our team possesses thorough knowledge of database functionalities and is competent to strategize such projects. We work closely and iteratively with you in order to develop efficient search strategies which are tailored to your needs and budget.

8. Sources & References:

1. NB Dahibhate, Chemical Patents: Search Strategy, Chapter 7
Link: http://shodhganga.inflibnet.ac.in/bitstream/10603/3782/17/17_chapter%207.pdf
2. Polymer indexing user guides
Link: <http://ip-science.thomsonreuters.com/support/patents/userguides/polymerguides/>
3. Plasdac coding systems
Link: http://ip-science.thomsonreuters.com/m/pdfs/mgr/Plasdac_Coding_Systems.pdf
4. Polymer Indexing System Description Introduction
Link: <http://ip-science.thomsonreuters.com/m/pdfs/mgr/polindsyst.pdf>
5. Edward S. Wilks; "Polymer Nomenclature and Structure: A Comparison of Systems Used by CAS, IUPAC, MDL, and DuPont. 1. Regular Single-Strand Organic Polymers"; *J. Chem. Inf. Comput. Sci.* 1997, 37, 171-192.
6. Nico Adams and Ulrich S. Schubert; "From Data to Knowledge: Chemical Data

Management, Data Mining, and Modeling in Polymer Science", *J. Comb. Chem.* 2004, 6, 12-23.

About the Author



Dr. Yogesh Nagpal

PhD, Organic Chemistry, Panjab University, Chandigarh
Registered Indian Patent Agent
WON certified Patent Information Specialist

Dr. Yogesh Nagpal is currently managing the operations of patent research and analytics firm Winsome IP. Prior to joining Winsome IP in 2014, he had worked for five years in one of the leading KPO based in Gurgaon.

Email: yogesh.nagpal@winsomeip.com

Winsome IP strives to be a leader in providing value added services to a global clientele at every stage of patent lifecycle. We are a group of professionals having extensive techno-legal experience and expertise in IP creation, management and monetization. Our clientele includes law firms, technology corporations, start-ups, venture capitalists, patent brokerage firms and individual inventors. Through our services, we help our clients make smarter business decisions based on critical insights into IP opportunities.

For more information contact: info@winsomeip.com

Disclaimer: The content provided in this article is intended for informational purposes only. The author/ Winsome IP does not guarantee or warrant the accuracy, reliability, completeness of the information in this publication nor its usefulness in achieving any purpose.