

高分子電気材料データベースの研究

西 岡 篤 夫

A Database of Polymeric Materials for Electrical Engineering

Atsuo NISHIOKA

A bibliographic database was prepared for the researches in the field of polymeric materials in electrical and electronic engineering. CAS database was used as the primary source of necessary references, and each record consisting of 15 items was condensed in about 750 bytes. As a model system, about 240 records were stored in a floppy disk, and from the results of retrieving the system, some improvements were suggested.

1. 諸 言

わが国における学術情報のデータベース化に関する研究は比較的新らしく、10年前の文部省特定研究「広域大量情報の高次処理」¹⁾が重要な契機となっているが、欧米に比しデータベースの国産化はおくれて居り、この方面で今後より一層の努力が必要と考えられている。

著者は1976年に文部省科学研究費による特定研究「情報システムの形成と学術情報の組織化」に参加して以来、主に高分子化合物の物性数値情報データベースの構築を目的として研究を行ってきた。^{2,3)} 研究グループをつくって最初に着手した高分子のCarbon-13核磁気共鳴データベースは文献の書誌事項と主に化学シフトの数値情報の両者を含む複合型でレコード数1000件以上、ファイル容量5 MB以上に達し、1983年10月から京都大学大型計算機センターで公開サービスが行われている。⁴⁾

高分子という学問それ自体が新しくまた他の多くの分野と学際領域を成す複合的性格をもつのでその全体を包含するデータベースの構築は大事業であり、それには核となる有用なサブデータベース群が必要である。この意味で著者は第2の試みとして高分子電気材料データベースの研究を始めることにした。

2. 高分子電気材料の特長

電気機器や電線ケーブルの絶縁あるいは誘電体として、またコンデンサその他の電気回路素子用材料として以前から各種高分子材料が使用されて居るが、近年エレクトロニクスの進歩と共に

情報記録, 電子素子封塞, 印刷配線基板等の電子工学あるいはコンピュータ関係の用途も増え, 更には最近の傾向である「機能性高分子」を目指した導電性高分子, 圧電性材料から超導性, スイッチング素子等の研究が盛んで, 高分子電気材料は古くして且つ新しい分野であるといえよう。

しかしながら「高分子」と「電気」それ自体が広い学問分野であるため高分子電気材料は多重複合領域と考えられ, 必要な文献集合を作るのは決して容易ではない。その意味でこの方面のよく整理された文献集合が利用可能となれば研究を進める上で役立つところが大きいと考えられる。

このように高分子電気材料の文献集合の作成はそれ自体で有意義であるばかりでなく, その物性数値情報も含めてデータの利用が可能となれば研究を進める上で有力なtool となると考えられる。

例として誘電特性を考えてみよう。誘電特性は誘電率, 誘電損失という特徴的な物理量のほかに物性論の立場から緩和の挙動が重要であり, またデータを得る手段として測定法もあわせて考えなければならない。しかもこれらの諸因子は物質の種類, 組成, 製造法等によることは勿論外部要因として温度, 周波数による変化が最も重視されるから, これらの大量のデータを如何に効率よく表現するかが重要な課題である。

従来このようなデータの集合は便覧あるいはハンドブックの形式で編集されているが, 編集に要する労力, 時間, 経費が大きいため内容の更新を頻繁に行なうことができず, 通常10年位が改版の周期と思われる。しかも研究のスピードが速いため更新された信頼度の高いデータが益々必要になっていることを考えると従来の形式のデータの編集方法では現状に追いつき得ない状況にある。これに対しコンピュータ可読の形式でデータを編集すれば更新, 追加, 削除等の操作は容易であるからデータベース作成の意義は大である。しかも包含する分野が広汎で利用者が多方面になる程データベースは有利となる。

またデータの表現形式についても最近普及しつつあるリレーショナルデータベースの方法を用いることにより少数の基本となる表を作成すれば, あとは各種コマンドの操作により必要な結果を得る事が可能となりつつあるからデータ構造が高度になる場合にも対応が可能であると考えられる。

このように高分子電気材料の分野は書誌事項を得る対象としても, また物性数値情報を得る対象としてもデータベースの構築に向いた特長を有していると考えられる。

3. 文献集合の作成

前述の観点からこのたび高分子電気材料データベースの研究に着手することになったが, 大きなシステムの構築は早急には困難なので本研究においては先ずそのモデルとして小規模の文献集合の作成とその検索を試みた。以下その経過について報告する。

さて高分子電気材料の文献集合を得る方法であるが, 高分子化合物という物質の面から考えるとこれは化学の分野であり, 他方用途としての電気材料の面からは電気工学, 電子工学, 情報工学等に属すると考えられる。

前者の文献集合のソースとしては先ずCAS(Chemical Abstracts Service)をあげねばならぬし、後者としてはINSPEC(INformation Service for the Physics Engineering Communties)をとりあげる必要がある。前者は当初CACon (Chemical Abstracts Condensates) のMTの形式で提供されていたのが、これに詳細な項目索引 (GSI: General Subject Index) と化合物索引 (CSI: Compound Subject Index) を加えた CAsEarch の名で提供されて居り、例えば東京大学大型計算機センターのTŌŌL-IR(TŌKyo university Ōn Line Information Retrieval) により検索可能である。他方後者はA (物理学), B (電気・電子工学), C (計算機・制御工学) の3部門より成り、何れも京都大学大型計算機センターのFAIRS (Facom Advanced Information Retrieval System) により検索可能である。

3.1 TOOL-IRによる検索

TOOL-IRによりCASにつき種々のキーワードを入力して検索を試みた。隔週毎に更新されるCAS NEWを用い、誘電特性と半導体の二分野を先ず検索した後、高分子に関するキーワードとの論理積をとる作業を数回行なった。元素文献集合の機械検索には一意の解はないのであって、希望する集合を完全に得ようとすれば所要時間と経費の増大を招き、他方ノイズを除くことに重点をおくと逆に必要なものが一部脱落する等の問題を生ずるので一般には適当な妥協が必要である。

今回は小規模のモデルということから専門分野をある程度限定し整理の負担も考えて月間数十件のhitを得るようにした。東大ではSDI(Selective Dissemination of Information) のサービスを行なっているので数回の試行により得た質問語集をSDIに登録し郵送されたLP出力を調査して文献の傾向を知るようにした。

SDIに登録した質問語集による融週検索の一例を次に示す。

検索番号	コマンド	質問語	文献数
1	SEARCH(略語S)	DIELECTRIC	132
2	"	CONSTANT	326
3	"	LOSS	118
4	"	RELAXATION	102
5	"	MEASUREMENT	251
6	OR 2, 3, 4, 5	(論理和)	763
7	AND 1, 6	(論理積)	62
ここでは高分子の集合をとらずに出力した。			
1	S SEMICOND		189
2	S ORGANIC		298
3	S POLYMER		1178

4	S	PLASTICS	199
5	S	RUBBER	227
6	S	FIBER	372
7	OR	2, 3, 4, 5, 6	1864
8	AND	1, 7	10

高分子半導体の集合は母集合に比べて非常に少ないことが分る。

9	S	ELECTRIC	713
10	S	CŌNDUCT	243
11	AND	7, 9	116
12	AND	7, 9	33
13	AND	11, 12	24
14	S	PŌLYACET	8
15	OR	8, 13, 14	38

これは高分子半導体，高分子の電気伝導，ポリアセチレン等に関する出力を得るものである。

その後電気的性質に関する SECTION NO を用い，検索時間の短縮と効率を考えて以下の如くにした。

1	S	PŌLYMER*, CŌPŌLYMER, PLASTIC*, RUBBER
2	S	DIELECTRIC
3	S	CŌNSTANT, LOSS, RELAXATION, MEASUREMENT
4	AND	2, 3
5	S	CA176*, CA177*
6	S	ELECTRIC. AND. CŌNDUCT*
7	OR	4, 5, 6
8	AND	1, 7
9	S	ORGANIC. AND. SEMICŌND
10	S	POLYMER* AND. SEMICŌND
11	S	POLYACE* AND. *ETYLENE
12	OR	8, 9, 10, 11

これによる出力は隔週50～70であったが有機半導体関係では超電導体に関するものが最近増加しているのでこの質問語集では不充分であることが分り，更に改良を検討している。

これらのコマンドにつき若干補足説明をすると，1は簡単すぎるきらいがあり，CA135*, CA136*, CA137*等を用いた方が網羅的になるが経費の都合で小さくした。それでもhitは1600件程度はあるので80%は入っていると思われる。2と3は誘電体関係で，これに BREAKDŌWN や INSULATŌRも加えたいところであるが規模の点から今回は省いた。CA176 と CA177 は夫々 ELECTRIC PHENOMENA と MAGNETIC PHENOMENA の SECTCON NO. であって前者に

は電気伝導，誘電性，放電，電子放出，圧電性，螢光，超電導，光電導，熱電気，質量分析，半導体，デバイス等の諸項目が，また後者には磁性，磁気抵抗，磁気歪，強磁性共鳴，フェライト，超電導磁石等を含む。6は5と重複するが出力をみた上で追加した。9と10は有機半導体関係で11は近年注目されているポリアセチレンを抽出したものであるが，最近の更に新しい材料を検索するためのコマンドを若干追加する必要がある。

3.2 INSPECの検索

INSPECは前述の如くA，B，Cの3部から成り京大には全部あるが，東大にはCしかない。京大のFAIRSによりINSPECはABSTRACTを含む詳細な出力を与えるがSDIサービスがなく，センターのTSS出力のみによる。東大には上記C部門につきIEEのDB名でSDIサービスがある。両者を比較するとINSPECは半年毎のファイルで内容が詳しいが文献数は少ない。しかも東大ではC即ちコンピュータ関係なので今回の目的には利用価値が低い。京大のINSPECを検索した経験では材料よりも電気工学に重点があるため測定方法に関するものが多いことが分った。しかし高分子化合物のキーワードを入力すると出力が若るしく減少してCASと同列で利用することは不适当であり，むしろCASを主体とし，INSPECを補完的に利用の方がよいという結論に達した。

3.3 レコードの内容

CASの出力は各文献毎の一意番号であるChemical Abstracts NO.が各巻毎にV/I/Nの見出しに続いている（Vは巻，Iは号，Nは番号）。

これと下記15項目の内容から成る。

- J： 雑誌名
- T： 表題
- K： キーワード（KWIC形式：句の集合）
- D： 印刷物の種類（雑誌，著書，特許，会議録）
- V： 原著掲載誌の巻（J：に対応）
- P： 同ページ範囲
- Y： 同発行年月日（830305の如く6数字）
- C： 雑誌略号（CAS）
- N： 著者の国籍
- A： 著者名（姓を先に書く）
- L： 著者勤務先
- O： 使用国語
- S： CAのSection No. (CA176等)
- Z： 便郵番号
- U： 州名（略号）

高分子電気材料データベースの見地からはどのような材料のどの性質を扱っているかが問題なので化合物名と物性を別の項目として抽出した方がよいと考えられるが、今回は時間の都合で上記のままとした。なおK：についてはKWIC(Keyword in context)の性格上単語の重複が多く全体としてデータ容量の3割程度を占めるのでPersonal Database(PDB)としては個別単語にする方が簡単で入力も容易となるが、これは文献数とシステムの規模にも関係するので今後の課題である。

上記の如く項目が多いのはレコードのformatを決定する上で固定長のものを別にする方が設計し易い事とスペースの有効利用を図る事によると思われる、また出力のモードを数段階に分けて精粗何れも選択可能となり、2次的なファイルの設計も可能にする等よく配慮されているといえることができる。我々の目的には若干不要と考えられるものもあるが、この改良は次の課題としておく。

4. ファイル設計

個人データベースの規模という立場でパソコンを使用し、8時のフロッピーディスクに1メガバイト(MB)のデータを入力することにしてファイル構成を設計した。今回の実験に用いた機種はSÖRDのM243型Mark VでFDDは2台附属している。1レコード長は256Bであることから原データのレコード長をしらべチャンネル数を決定することになる。SDIの3回分につき各項目毎の長さをしらべた結果は表1の如くである。これらのデータをもとにして3個のチャンネルを用いることにした。各チャンネルに複数の項目があってもそれらを1括して256バイトの1レコードとして扱った。

表-1 レコードの項目長(バイト単位)

項 目 名	固定/可変	項目長	項目名	固定/可変	項目長	平均長
V/I/N	固定長	14	V	可変長	1~3	
C	固定長	6	P	可変長	3~7	
D	固定長	1	Z	可変長	3~7	
Y	固定長	6	N	可変長	3~7	
S	固定長	8	A	可変長	6~136	41
O	固定長	2	L	可変長	8~82	40
U	固定長	2	J	可変長	4~126	23
			T	可変長	9~299	70
			K	可変長	20~480	105

チャンネル1: IDNO(このデータベースでの一意番号で4ケタ), V/I/Nと11項目(J, T, Kを除く), 項目間の空白は3バイト。

チャンネル2: JとT(改行)

チャンネル3: Kのみ(256バイトまで)

以上のファイル設計に基づきSDI3回分の出カリストにより7件のレコードを1枚のfloppy

diskに入力した。

5. データの入力と検索

いわゆるデータベースシステムとしてはその運用に必要な基本プログラムと応用プログラムがあり、前者としては書き込み、読み出し、検索、追加、更新、削除等の諸機能が、また後者としては複数のデータファイルと複数のユーザーを考慮した種々の管理機能があるが、今回の研究はまだモデルの検討の段階である事とデータの構成を考察する事に主眼をおいたので最も簡単な基本機能すなわち、Read, Write, Searchのコマンドをもつプログラムを作成し、これに若干の機能を加えた。

既述の如く1レコードをパソコンの3チャンネルに分割したのでRead出力はこれらを全部出力するほか、6種の異なる組合せを選択可能とした。Writeの場合は番号順に書き込むほか更新を行うため特定番号についても可能とした。Searchは行検索を基本に、2語のAND, ORをとるほか、各項目の頭(A:など)を利用して指定項目を選択して見れるようにした。更にまた項目長の分布を知るため最大、最小、平均を出力したり、全レコードについて国名や著者名等のSortingも可能とした。

このような簡単なプログラムでは文献数が増大すると検索時間が長くなるので逆転ファイルや作業ファイルが当然必要になるが、これらは空きチャンネルの利用により、またレコード数の増加に対してはformatの再検討によるほか小型ハードディスクの利用によりある程度対処することができる。これらの点については次の機会に検討することとし、以下このシステムにより入力した237件のレコードについて検索を行なった結果と考察について述べる。

(1) D : (document の種類)

B(Book), C(Conference), J(Journal), P(Patent), T(Text)のhitは夫々1,7,156,53,11 で特許が論文の $\frac{1}{2}$ もある事はこの方面の研究が開発的である事を示唆する。ちなみにD:PとN:JAPANのAND即ち日本の特許をみると27件もあり上記53件の半分でわが国のactivityの高いことが分る。

(2) N : (国名)

237件の全文献について国名でSortすると表2の如くでソ連、日本、米国が何れも40~50件で全体の7割を占め西独がこれに次いでいる。材料別にもっと詳細にしらべる事が勿論必要であるが、大勢の判断には役立つ。

(3) O : (使用国語)

ソ連の論文54件中53件までがロシア語であるが日本は50件中33件が和文である。これは特許公報が和文で書かれているためと思われる。頻度は英語(125), ロシア語(53), 日本語(33), 仏, 独

表-2 国別のレコード数分布

国 名	件数	国 名	件数	国 名	件数
USSR	55	ISRAEL	3	ARGENT.	1
JAPAN	50	POL.	3	HUNG.	1
USA	44	CAN.	2	MALAY	1
FED. REP. GER.	17	EGYPT	2	PEOP. REP. CHINA	1
FR	13	LIBYA	2	TOL.	1
INDIA	8	NETH.	2	AUSTRALIA	1
UK	7	SWED.	2	BRAZIL	1
BELG	4	ITALY	2	IRE.	1
GER. DEM. REP.	3	SWITZ.	2		

各(8), その他である。

(4) Y : (発行年月日)

使用したSDIが3回分なので年によるソートに大して意味はないが、特許には月日が入っているので優先権の点で重要である。

(5) J : (雑誌名) と C : (雑誌コード)

全部で129種類あったが特許公報もこの中に含まれている。なお雑誌名による検索は省略形も用いられているため入力ミスを生じ易い。これを防ぐには予め対照表を見てからCÖDENを入力する方がよい。

(6) S : (部門番号)

CASでは通常各文献毎に1個のSection No.がついて居るので多くのキーワードを用いることなく特定分野の検索を行う事ができる利点があるが、境界領域の場合などはこの方法はSection No.でANDをとることができない不便さがある。表3の結果を見ると高分子の分類に属する135~138と電気物性に関する176~177の2群に分れている事が分る。これはCASの寄稿者の考えにもよるのであろうが我々の立場からするとSection No.の欠点が出ていると思われる。これを補うものがKey word群であるが、項目名として例えばR:(高分子のReference code)とE:(Electrical Properties)の2つを追加することにより一層の改善を図ることができよう。

次に高分子化合物名と電気的性質の各々につきいくつかの質問語を入力して検索した結果を表4に示した。レコード数が少ないので統計的情報としては甚だ不十分であるが、高分子電気材料の研究動向の一面をとらえていると考えられる。

表-3 Section No. による分類

	1	C A165007	1	C A122000	1	C A166004	4
C A109007	1	C A166002	2	C A122009	2	C A168005	2
C A122003	1	C A168003	1	C A135003	3	C A172003	4
C A135002	2	C A168006	4	C A135006	2	C A173001	2
C A135004	4	C A172006	1	C A136002	2	C A173003	2
C A135008	3	C A173002	5	C A136005	1	C A173012	1
C A136005	20	C A173004	2	C A137005	4	C A174005	1
C A137003	2	C A174003	1	C A138002	2	C A174013	1
C A137006	10	C A174009	1	C A139003	1	C A175002	1
C A138003	16	C A175001	1	C A139012	2	C A176000	4
C A139009	1	C A175011	1	C A140007	1	C A176002	10
C A140004	1	C A176001	13	C A142004	2	C A176004	1
C A142002	1	C A176003	4	C A142010	1	C A176006	2
C A142007	3	C A176005	2	C A153006	1	C A176009	31
C A152000	3	C A176008	5	C A157001	1	C A176014	1
C A156004	1	C A176010	6	C A165000	1	C A177007	2
C A161001	2	C A177003	1	C A165002	1	C A138003	0
C A165001	3	C A138003	0	C A165006	2		
C A165003	2	C A106006	2	C A166001	1		

表-4 高分子化合物および電気的性質による検索結果

高分子名	件数	電気物性	件数
POLYETHYLENE	8	DIELECTRIC CONSTANT	1 3
POLYSTYRENE	1	DIELECTRIC LOSS	3
POLYPROPYLENE	2	DIELECTRIC POLARIZATION	2
RUBBER	1 0	DIELECTRIC RELAXATION	7
EPOXY	1 1	CONDUCTING	1 3
FILM	2 7	BREAKDOWN	2
FIBER	2	MICROWAVE	4
TEXTILE	2	INTEGRATED CIRCUIT	1
POLYACETYLENE	2 0	BATTERY	3
POLYAMIDE	4	COATING	1 1
POLYIMIDE	3	CABLE	4

6. 結 言

高分子電気材料の研究開発の効率化に寄与する目的をもってデータベース構築の研究に着手した。まず文献集合の作成方法とレコード内容及びその形式についてCASのSDIにより検討した。次にパソコンにより小規模のモデルシステムを作成し検索を試みた。

入力レコード1件当たり約750バイトの設計で二百数十件を入力して検索した結果、統計的情報を得るにはレコード数が不足しているが、データベースの機能としてはCASを基礎としこれに若干の改良を加えれば書誌事項を内容とするいわゆるBIBデータベースとして使用可能であること

が分った。

次期の研究課題としてはソフトウェアを一層充実して機能を強化すると共にデータ量の増加に対応してハードディスクによる実験を行ない、あわせてMTとの間のデータ転送を試み、汎用コンピュータによるDBMSの利用へも途を開く計画である。

本研究は昭和57年度より特別研究として開始され現在も継続中であるが、この実験に関し深い御理解と多大の御援助をいただいている金井兼造理事長、正田 強学長をはじめ関係の各位に深く感謝致します。また昭和57年度の卒業研究として本研究の遂行に多大の努力をされた北川滋幸、大西浩和、鍋倉伸嘉の諸君に厚く御礼を申し上げます。

参 考 文 献

- 1) 北川敏男, 島内武彦編「巨大学術情報システム」, 昭和50年1月, 東京大学出版会; 同 編「広域大量情報の高次処理」, 1977年1月, 東京大学出版会。
- 2) 西岡篤夫, データベースPCMRDBの解説(分子・結晶データシステム解説No.2, p 67), 昭和53年9月(特定研究「情報システムの形成過程と学術情報の組織化」C-1班報告書。
- 3) Y. Fujiwara, K. Hatada, T. Hirano, T. Kawawura, S. Kondo, K. Matsuzaki, A. Nishioka, Y. Tanaka and B. Tomita, A. Carbon-13 NMR Database for Advanced Research in Polymers, Codata Bulletin, NO.40, 35-38 (1981)
- 4) 西岡篤夫, 高分子のCarbon-13核磁気共鳴データベース-CNMRP- 京都大学大型計算機センター広報, 16, NO. 5, 253-265 (1983)