

医薬品に関する文献情報処理

武田薬品工業株式会社 中央研究所調査部調査課

角田喜治

〔1〕緒論

医薬品は化学、生物学、物理学の基礎科学、および薬理学、臨床医学等の応用科学を動員し創製される *Product* であるため、対象となる文献の種類は多い。たとえば、当社中央研究所図書館で管理している文献だけでも蔵書数約8万冊、年間購読雑誌数約1千種に及ぶ。これ以外に医薬事業部、外国事業部、中央研究所各研究所で各々必要を文献として別途購入している。これら学術文献に加え、企業にとって重要なものは、特許、各種規制に関するもの、学会講演要旨集、速報誌などである。

医薬品の研究開発を効率良く推進し、また自社品についての的確な情報を医師、薬剤師あるいは厚生省に提供するためには、これら文献類を迅速に処理することが必要である。当社の場合、研究開発に関連するものは、中央研究所調査部が、医師、薬剤師、厚生省等への情報提供については、医薬事業部学術情報サー、オニ部が主担している。

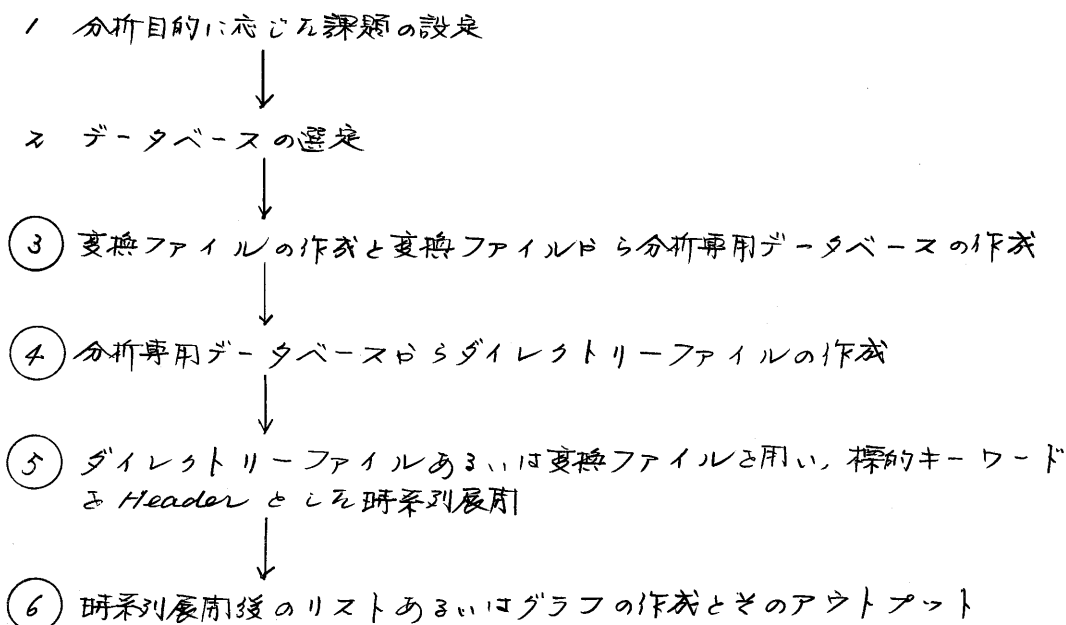
さらに、オンライン検索の普及に伴い、社外データベースの活用も、文献情報処理の一手段として重要と暮ってきた。当社では社外からのオンライン検索を利用する以外、*Ringdoc (RD)*, *Excerpta Medica (EM)*, *Central Patent Index (CPI)* を購入し、この中 *EM* については医師、薬剤師等への情報サービスにも利用している。しかしながら、これらデータベースの購入価格は高く、年々値上りしている。

そこで、著者は当課鈴木らの協力を得て、情報検索以外、たとえば医薬品の研究開発動向の把握などにも利用し、高価なデータベースを有効に活用できているの研究を行っている。今回は今までに「情報管理」、「武田研究所報」および昨年の「日本薬学会100年会」で発表しものを総括して報告するが、その応用範囲は広く、たとえば文献から抽出した創作情報解析などにも利用できると考えている。

〔2〕 動向把握のためのシステムデザイン

Computer readable database を用いて、医薬品の研究開発動向を把握するに先立ち、これらデータベースに格納されている書誌的事項、キーワードを利用し、分析目的に対応した研系列展開リストの作成、研究のライフサイクル分析のための各種キーワードの年別出現頻度数の算出およびクラスター分析に用いる変換ファイルの作成等に必要システムを開発した。

すなわち、



↓
 ア アウトプットされたリストあるいはグラフを用いての情報分析

(こゝで○印はコンピュータを利用する部分)

これら一連のシステムを演者らはALARMと呼称しており、現在使用しているデータベースは、前述の3種類である。このうち、CPIはDerwent社発行の特許に関するものであり、RDEMは下記の表に示すような特許性をもつデータベースである。なお、下表には医学部学領域で警用性のあつたMEDLARS(MD)について参考までに付記した。

Data base	E M	R D	M D
Covered journals	3882(1975年版 Journal List)	426(1976年版)	2398(1976年版 Index Medicus 用) + α (特定書誌用)
Contents of covered journals	臨床薬理や薬理に関する論文、あるいは自然科学や生化学などの基礎研究に関する論文を収載する雑誌が多い。医学関係では薬物による治験例をよく収載する雑誌が多く、薬物治験例以外の医学論文を収載する雑誌は多くない。	薬物による治験例関係、薬理、薬学および有機化学に関する雑誌が多い。	基礎および臨床医学、歯学、獣医学、環境・衛生医学、医薬工学、看護学から生物学、生化学、薬学等の関連分野もカバーしている。
Japanese journals	190(1974年版)	9(1976年版)	103(1976年版)
Covered reports/year	210981(1974年版)	45500(1976年版)	277592(1976年版)
Search coverage	1973年より	1965年より	1974年より
Thesaurus	MALIMET (Master List of Medical Indexing Term) Preferred Term 21万およびその同義語40万	主要索引語である Index Term (1355種類) とその関連用語である Free Term の関係を示すリスト	Annotated MeSH: 索引作成者、検索者用の MeSH でディスクリプタに必要な注記 (Annotation) がある。用語数は13312(1976年版)
Indexes	Preferred Term Code: 疾病名等の生物医学用語および医薬品名等、個々の Preferred Term に付されている索引コードで新タームが増える毎に新コードがつけられている。 Classification Code: 内容を体系的に表わす階層分類コードで4段階に階層分類されている索引コード (4,600種類あり) Item Index: 投与ルート、実験動物等主題に対する補助的概念に付されている索引コード (90種類あり) Secondary Indexing Term: 生物・医学用語、医薬品の商品名、治験記号、製薬会社名等が自由な表現で表わされている索引	Index Term: Ringdoc が設定している標準分類用語で化学、薬物、疾病等を表わすタームがあり、その数は新タームが増えてもその都度 Index Term になるとは限らない。 Free Term: Index Term の内容をより具体的に表現するために用いられるターム (Index Term の補足語) Ringcode: タームでは表わしにくい化学構造の細部を索引するために、タームのデータベース (Codeless Scanning Data Base) とは別に準備された索引方法でフラグメントコードで格納されている。	MeSH 付名 a. Major descriptors, Minor descriptors: MeSH に登録されている医学用語、薬品名、実験動物等についての Key word からの索引ならびにカテゴリ別に階層分類された索引コード b. Qualifier: 上記の descriptors に結合してその意味・内容を限定する機能を持つ索引
Retrieval system by computer	データベースから Preferred Term Code, Classification Code および Item Index を抜き出しそれぞれにコードあるいは索引コード (ユニークな番号) をつけた変換ファイル (Inverted File) をあらかじめ作成し、そのファイルから検索を行なう (このように変換ファイルを Directory File という)。	Index Term, Free Term 表、I' の言葉として抜き出し、それぞれにターム・インデックス番号をつけた変換ファイル (Inverted File) をあらかじめ作成し、そのファイルから検索を行なう。	JOIS I (JICST On-line Information System)

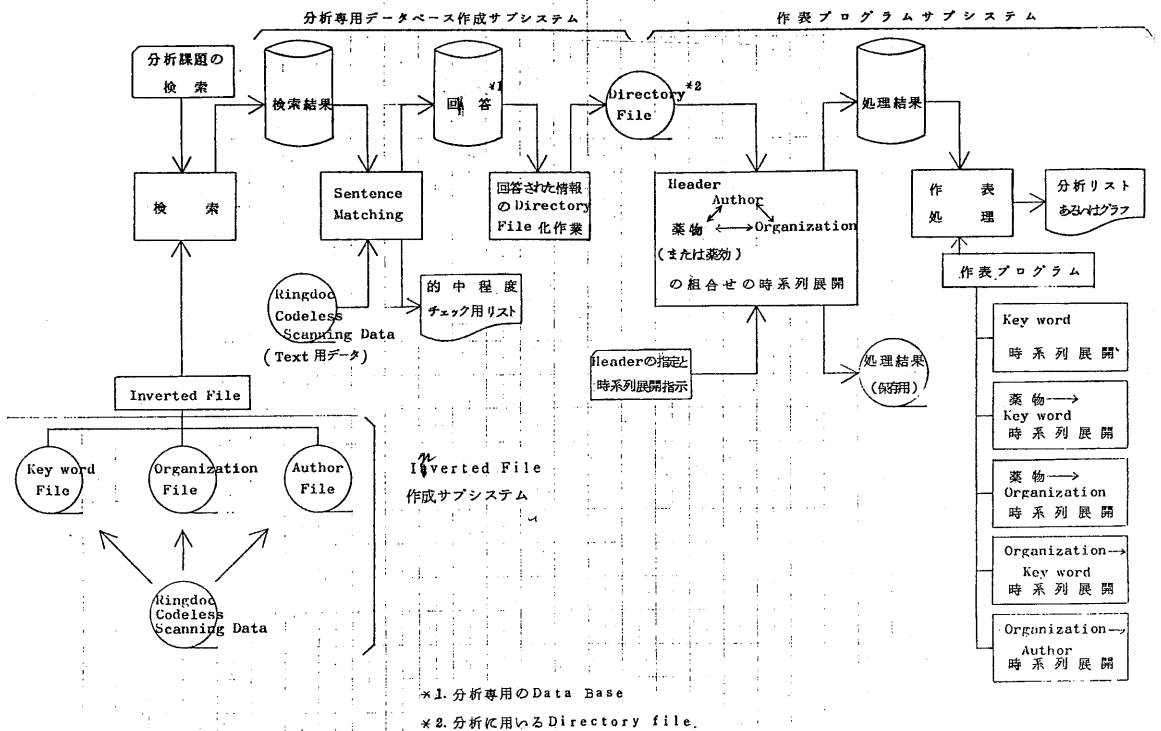
引用文献 「JOIS 利用のてびき」 日本科学技術情報センター 1976
 「EM-MD-RD の検索比較」 医学図書館 23(4)204-218(77)

RDを用いた場合のALARMのフローを代表例として次頁に示すが、CPI、EMの場合も基本的には同じであり、

- ・ 誰が、
- ・ どの研究機関で、
- ・ どのような研究で、

・いつの日か行っている
すなわち, Author, Organization, Theme (keyword) と分析の要素とし
それを迅速に処理するのが ALARM である。

Ringdoc を用いた場合の ALARM



RD の場合, 上記5種の時系列展開が可能である。

[3] 分析の事例

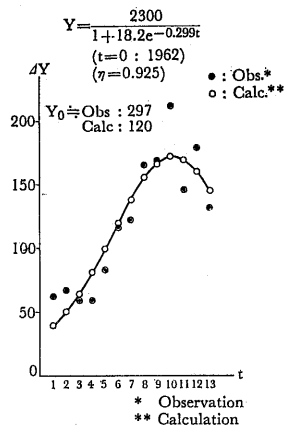
RD, CPI を用いた場合は, ペニシリン, セファロスポリン, β -ブロッカー, ベンジジアゼピンについて, EM を用いた場合はクロニジンについて分析し, 得られた知見のうちのいくつかを以下に示す。

✓ 新規ペニシリンの研究開発は下降している。但し, 臨床報告は多い。

00004 - PENICILLIN DER. : CASE REPORT

YEAR	COUNT	200	400	600	800	1000	1200	1400	1600	TOTAL
1965	220	*****								220
1966	219	*****								439
1967	267	*****								706
1968	473	*****								1179
1969	487	*****								1666
1970	641	*****								2307
1971	663	*****								2970
1972	690	*****								3660
1973	953	*****								4613
1974	1088	*****								5701
1975	1513	*****								7134
1976	1331	*****								8465
1977	1513	*****								9978

Time-series Frequency Subject Graph for Penicillins (RINGDOC)



	Obs.	Calc.
t = 1 (1963)	62	39
2 (1964)	67	50
3 (1965)	59	64
4 (1966)	59	81
5 (1967)	83	99
6 (1968)	116	119
7 (1969)	122	138
8 (1970)	165	155
9 (1971)	168	166
10 (1972)	212	172
11 (1973)	146	169
12 (1974)	179	160
13 (1975)	192	145
14 (1976)		127
15 (1977)		107
16 (1978)		88
17 (1979)		71
18 (1980)		56
19 (1981)		43
20 (1982)		33

Fig. 11. Logistic Curve of Patent Applications for Penicillins (I) CPI

2. 新規セファロスポリンの場合もペニシリンと約5年間のズレが同じ傾向が伺える。

3 β-ブロッカーについてはC社の研究開発力が最も大きい。

** BETA BLOCKER **		ORGANIZATION FREQUENCY LIST											PAGE 00001		
ORGANIZATION NAME		*65	*66	*67	*68	*69	*70	*71	*72	*73	*74	*75	*76	*77	TOTAL
ABBOTT	0	1	2	0	1	0	0	0	0	2	0	0	0	0	6
ALBERT	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
ALDRICH	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
ALFA	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ALLEN & HANBURY	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
ALLEN&HANBURY	0	0	0	0	0	0	0	0	1	0	0	1	1	0	3
ALLEN&HANBURY	0	1	1	1	1	3	3	0	1	0	0	3	8	7	28
ALCANTARA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

4 ベンゾジアゼピンの場合、有効性より安全性に関する報告が増加している。

Index term	Total	1967**	1972**	1977**
Side effect	2028	46	160	305
Depression	622	24	57	84
Anxiety	605	19	66	73
Epilepsy	618	15	44	95

上記研究開発リストを利用した研究開発分析、さらには研究開発リスト中のデータをロジスティック曲線に適用した、いわゆる研究のライフサイクル分析以外に、複数のキーワードを組み合わせとその類似性を利用したクラスター分析を行った。ここでは、EMを用い、クラスタリングについて分析した結果を報告する(スライド参照)。

なお、キーワードの組み合わせによる分析は、クラスター分析以外、数量化理論を用いる方法も考えられ、この理論を用いることによって、各企業の研究開発の

併効、あるいは特定の薬効とまつ化合物群の副作用の併効などを把握すること
可能と思われる。

〔4〕 総 括 — 文献情報処理の今後の一方向 —

新しい知見の迅速な把握は、有効性にしろ、副作用にしろ、あるいは基礎的
ものである、一次資料を活用するのが基本であるが、その際の重要ポイント
個々の一次資料の内容などのように評価し、重要度を選別するにせう。その
ためには、企業であれば技術情報管理部門、病院の場合はDI室など、いわゆ
るスタッフ部門による評価・選別のみでは不十分で、そこには、医師、薬理学者、
化学者等の専門家の参画が必要である。現にこのようなシステムを導入し、着
々と成果を上げていくところもあり、こゝでスリーミングされた文献情報は
その企業、研究機関に必要なデータベース作り(An house用)の基礎となる。

文献情報については、周知のとおり、MEDLARSが、CAB、さらにはJAP-
IC-DOCなども国家的レベルで利用性を意図した、いわゆるビッグデータベ
ースの開発・充実が進められているが、今後はこれと併行して、上記のようき
An house useを目的としてデータベースの開発も必要と思われる。その主な理由
は、ビッグデータベースで各々の企業・研究機関ごとの情報活動をすべて満足し
てくれるとは限らず、いからである。当然の事ながら An house 用 データベース
の開発に当っては、ビッグデータベースの特性を十分把握しておく必要がある。

たとえば

- ・ 研究開発段階の化合物: $RD > EM > CAN > MD$
- ・ 新 薬: $EM > RD > CAN > MD$
- ・ 既発売の医薬品: $EM > RD > MD > CAN$

- 送字領域：MD > EM > RD > CA

をどう把握して置く必要がある。

In house用のデータベースの開発には、人的資源を含め充分の投資と各々の専門家の協力が必要である。