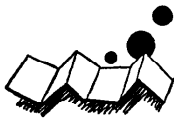


解説



医療におけるデータベース技術†

開原 成 允††

1. 医療データの特徴とデータベース技術の必要性

病院情報システムや地域医療情報システムの普及によって、医療の世界にも比較的大きなシステムが、作られるようになった。システムが大きくなればなるほど、データの構造は複雑となり、データベース技術の必要性は増大し、今では、大きな医療情報システムは、必ず中心に何らかのデータベースを管理するシステムをもつ。

しかし、医療と一言でいっても、その中には様々な業務が含まれ、必要とする技術も異なっているから、まず、「対象」と「必要とする技術」を整理するところからはじめなければならない。

医学上の必要性から、考えると、データベース技術は、表-1のごとく4つの場合に必要となる。

第1は病院情報システムにおけるデータベース技術であり、日常の病院の業務を行うために、患者のデータをデータベースとして管理するものである。

第2は臨床研究用のデータベースであり、ある目的をもって過去の病歴の詳細なデータをデータベースとして管理するものである。第1のデータベースと、混同されることがあるが、目的も、データベースとしての要件も、全く異なったものである。

第3の保健のためのデータベースとは、「ある地域のすべての人々」というように、大量の保健データをデータベースとしてもつものである。通常、項目数等も多くなく、レコード数が大きいことが特徴的である。

第4の情報サービスとは、医学文献や、薬剤のデータをデータベースとしてもち、オンラインで検索して、必要な情報を得るシステムである。これは、データベースではあるが、情報検索が主体のシステムであ

表-1 医療においてデータベース技術が必要な対象

	目 的	データベース技術からみた特徴
病院情報システム	病院において、病院の日常業務を行うために患者データを扱うシステム	「検索」を主にした通常の意味でのデータベース管理システム
臨床研究	病歴をある目的のために集積し、それを解析するためのシステム	データベース管理システムと統計解析システムが連結したシステム
公衆衛生・保健	ある地域の人々の保健に関するデータを集積し、それを能率よく扱うためのシステム	大容量で構造が簡単なデータを扱うシステム
情報サービス	薬剤のデータや、医学文献のデータを検索するためのシステム	情報検索システム

り、データベース管理とは、異なっているのです、ここでは、述べない。

以下に、最初の3者について、必要とされるデータベース技術、および、現状を解説する。

2. 医療におけるデータベース技術の現状¹⁾

2.1 病院情報システム²⁾

病院情報システムにおけるデータベースの特徴は、次の3点であると思われる。第1は、データが少量多種ということである。これは、他の業種のデータと比較した表-2³⁾によって、容易に理解されるであろう。第2は、データが時系列データであるということである。時系列であることは、「時間」という「項目」が増えただけのようにも考えられるが、実際は、時系列特有の処理が多くでてくるため、必ずしも、他項目と同様には扱えない。第3は、データベースの内容が、たえず流動していることである。入院患者をたとえれば、平均在院日数15日の病院では、15日たてば、内容はすべて入れ替わっていることを意味している。

このような、特徴はあるが、必要とされる処理は、データの集合から、必要なデータを検索し処理をする

表-2 各種システムにおいて取り扱う情報の相違

	最繁忙時処理量 (1時間当り)	入出力の形式	処理の形式
予約システム	125,000	2	5
銀行システム	40,000	10	5
病院情報システム	1,400	120	25

† Data Base in Health Care by Shigekoto KAIHARA (Director, Hospital Computer Center, University of Tokyo Hospital).

†† 東京大学医学部病院情報処理部
本研究の一部は文部省科研費試験研究 56890003 で行われた。

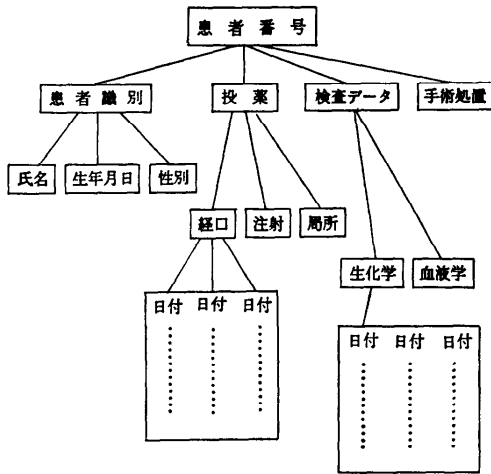


図-1 病院データベースのデータ構造の1例

というものであり、汎用のデータベース管理システムを利用できる可能性が十分ある。

事実、現在の病院情報システムは、この中核に、汎用データベース管理システムをもつものが多くなっている。しかし、日本の病院情報システムは、事務処理からはじまった関係で、病院のすべてのデータがデータベース管理システムによって管理されているというよりは、後の図-1にみるような病院の医事会計データの処理に、データベース管理システムが使われているというのがほとんどの病院の現状である。例をあげれば、富士通の病院情報システムでは、AIM、日本電気の病院情報システムでは ADBS、IBM の病院情報システムでは IMS、電電公社のシステムでは、DIMS-1 等であるが、データ構造は日本では、患者番号を中心とした比較的簡単なものである。しかし欧米ではより総合的な情報について、データベース技術が使われている⁹⁾。

病院情報システムにおけるデータベース管理のもう1つの型は、MUMPS を用いて、行う方法である。MUMPS は、データベース管理システムというよりは、プログラミング言語であるが、その機能の中に Disk 上の Tree 構造のデータを直接扱える機能があり、これが、最初に述べた病院のデータの特徴から考えると著しく、よく適合した機能である。それは、MUMPS が、これらの機能を満足させるために開発されたものであるから、当然のことであろう。

現在小型コンピュータを複数台用いた、いわゆる分散型のシステムにおいてはほとんどすべて MUMPS が用いられている(千葉大学、京都大学、大阪成人病セ

ンター、名古屋保健衛生大学病院等)。MUMPS をデータベース管理システムと考えることには異論があるが、MUMPS によるデータベースの扱いは将来のデータベース管理システムの方向を考える上で大きな示唆を与えるものと考えられる。図-1 に MUMPS によるデータ構造の1例を示す。

2.2 臨床研究用データベース

臨床研究用データベースとは、ある目的をもって限られた患者(たとえばある病気の患者)の詳細な病歴の内容を多数集めたものである。レコード数は数百から時に数万に及ぶ。

たとえば図-2は、その1例で、癌の患者のデータを集めたものの内容である。1患者のデータが時系列的に連なっていることが理解されるであろう。

INITIAL REGISTRATION

Item No.	data type
Identification	
1	Institution Code
2	Patient Identification Number and Check Digit
3	File Number
4	Birthdate
Demographic Information	
5	Birthplace
6	Race/Ethnicity
7	Sex
8	Residence at Time of Admission
Diagnosis	
9	Date of First Admission to Center for This Tumor
10	Sequence
11	Date of Initial Diagnosis
12	Primary Site
13	Laterality
14	Histology
15	Diagnostic Confirmation
16	Date of Best Diagnostic Confirmation
17	Stage of Disease at Time of First Therapy at Center
Therapy	
Cancer Therapy Prior to Admission to Center	
18	Surgery
19	Radiation Therapy
20	Chemotherapy
21	Endocrine Therapy
22	Immunotherapy
23	Other Cancer Therapy
24	Date of Initial Therapy at Center
Initial Therapy After Admission to Center	
25	Surgery
26	Radiation Therapy
27	Chemotherapy
28	Endocrine Therapy
29	Immunotherapy
30	Other Cancer Therapy
Patient Status	
32	Date of Last Contact/Death
33	Autopsy
34	Cancer/Treatment Related to Death
FOLLOW-UP ITEMS	
Identification as before	
Follow-Up Information	
35	Date Report Prepared
36	Method of Follow-Up
13	Laterality
32	Date Last Contact/Death
33	Autopsy
34	Cancer/Treatment Related to Death
PROTOCOL ITEMS	
Identification as before	
Protocol Information	
37	Date Entered on NCI Protocol
38	NCI Protocol Identification
37	Corrected Date Entered on NCI Protocol
38	Corrected NCI Protocol Identification

図-2 ある癌登録データベースの内容

このようなデータベースを処理する上で要求される機能は、第1に検索、第2に集計及び統計解析である。この検索は、しかし、あらかじめ Key の定められた検索ではなく、事前には、何が Key となるかは全く予想できない。検索は通常、試行錯誤で、くり返し Key を変更して行われる。検索されたデータに対しては、様々の集計や、統計解析が行われる。その解析方法は、平均値や標準偏差の計算から時には多変量解析にまで及び⁹⁾。

もし、データが時系列データである時は、処理はさらに複雑となり、最初に、時系列的に処理を行った上で、上記のような、検索や解析を行うことが必要になる。

このように、処理は、複雑なものが要求されるが、処理に対する時間的制約は、あまり厳しくはない。1回の検索や集計に、仮に数分かかったとしても許容し得る範囲である。

以上のような特徴をもった処理を、どのようなシステムで行うのが適当であるかについては、まだ定説はない。

1つの方向としては、これを満足するようなシステムを独自に開発しようとするものがある。たとえば米国スタンフォード大学の G. Wiederhold によって開発された TOD (Time oriented database system) はその典型であり、現在多くの臨床研究に使われている⁴⁾。

もう1つの方向は、検索に対する効率は無視し、汎用の、統計プログラム・パッケージを用いることとしそれに、簡単な検索システムを併用するものである(たとえば富士通の INTERAC-QUERY と SPSS を併用した例など)。最近の汎用統計パッケージには、ある程度のデータベース管理機能がある(たとえば SAS)ので、これで、不満足ながら一応の目的が達せられることもある。その他、データ構造が変化なく長期に亘ってデータを集積する時には、汎用の DBMS を使った例もある(たとえば、長崎大の原爆被爆のデータベースなど)。

汎用の DBMS を使う場合は検索は容易であるが統計解析は行いにくく、統計パッケージを使うと検索が行い難くなる。

臨床研究用データベースに対し、もし、relational database システムが、適応できるとすれば、ビューの柔軟性の問題は解決されると思われるが、現在までのところ、これらのデータベースに本格的に適用された

例はない。

2.3 保健データベース

これは、一定の地域の人々の健康に関するデータを収集・蓄積しておき、これらの人々が医療を受けるときにデータを提供するとともに、地域の医療計画の基礎としても用いようとするデータベースである。

このような構想は、医療が社会化されている国において実現された。たとえば、スウェーデン、ストックホルム地方では、その地域の人々約 80 万人の医療の要約が、データベースとして貯えられている。ここで使われているのは、汎用の DBMS (現在は IBM の IMS) である。データベースの技術的観点からすると、基本的なデータと詳細なデータをどのように分けてもつと効率がよいかという点が最も問題となる。また、このような大きなデータベースになると、データの秘密保持の問題も大きな問題となる。このためスウェーデンでは、秘密保持のための方法を独自に IBM とともに開発し、これを用いている。

日本においては、このように地域全体の住民のデータをデータベースとしてもつことは、小規模の実験的事例を除いては行われていない。これは、データベースの技術的問題というよりは、医療の社会的条件が異なっているためと思われる。

しかし、日本では、企業等がその構成員を対象として、データを収集して保持している例は多い。しかし、これらのデータは、医療データとしての特徴はあまりなく、したがって、データベース技術的に特に記すべき点はないように思われる。

3. 医療におけるデータベースの問題点と将来

3.1 医学的問題

医学領域にデータベース技術を応用する際に、まず考慮しなければならないことは、医療データをデータベースとしてもつことが、医学的に意味があるか否かという点である。医療データは、本来、公開されるべき性質のものではなく、個人と医師の間の信頼関係の中で保有されるべき性質のものである。したがって、多数の利用者を想定したデータベースの考え方とはなじまない面もある、したがって、データベースとして、保有することの必然性が、まず、厳格に問われなければならない。

これまで、あげた3つの領域は、その必然性が比較的確はっきり認められる領域であるが、それにしても、

その運用体制等をよく考えた上でこれを作っていくかなければならないことはいうまでもない。

第2の問題点は、医学データは、データベースとして扱うためには、未整理の部分が著しく多いという点である。同じ、データが、使う人によって別の内容をもっている時には、これをデータベース化して、処理を効率化しても、意味はない。現在、医学用語のソーラス等を作る努力はあるが、まだその完成には時間がかかるであろう。医療データのデータベースを作っても、内容の不備のために、それが利用できないということがないように、データの内容について細心の注意が必要である。

このような問題があるにしても、医療へのコンピュータの普及とともに、データベース技術は医療にとって不可欠のものとなることは確かであり、医療の中の環境の整備が整えば将来は急速に利用が進むであろう。

3.2 技術的問題

医療からのデータベース技術への要請は多くある。これらを、マンマシンインタフェースの問題、時系列データの問題、統計処理とのつなぎの問題、自然言語処理の問題等に分けて、以下に述べる。

3.2.1 マンマシンインタフェース

医療においては、データベースの利用者はすべて、計算機のみならず専門家であり、いわゆる「end user」である。したがって、データベース管理システムと利用者との間のインタフェースが著しく重要となる。

これは、単にデータベースの利用者のみではない。データベースを作っていく人々も、多くの場合、医療関係者であり、このためには、データベースの作成者と、DBMS の間のインタフェースもまた、著しく重要である。

こうした観点から、今の DBMS を眺めた時には、利用者との間のインタフェースは改善されてきているが、データベースの作成は、まだ著しく複雑で、高度のプログラミング技術を要する。今後の方向として、データベースを作成するための end user tool の開発が望まれる。

この意味で大変興味があるのは、米国の Duke 大学病院で開発された Patient Care System (略称 PCS) とよばれるプログラミング tool である。これは、ディスプレイの画面設計を著しく容易にするとともに、それが、DBMS とも連結していて、高度のプログラミング技術を要せずに、データベースを作成し、運用す

ることができる。これは医療の中で開発されたものであるが、今後は、他分野でも多く使われるようになると思われる。日本では高知医科大学で採用されている⁶⁾。同様な意味で、MUMPS が医療の中でなぜ広く使われているかについても注目すべきと思う。MUMPS の Tree 構造の file を作成する過程は、実に自然で理解しやすい。汎用 DBMS のデータベースの作成が、これと同じ程度に容易に行い得るとすれば、データベースはより利用者にとって身近なものとなる。

3.2.2 時系列データの処理

すでに述べたごとく、時系列データは、時系列上で、他と異なった処理を必要とする場合が多い。したがって、これを容易にするデータベース技術があることが望まれる。これは汎用の DBMS で不可能ではないにしても、たとえば TOD にみるような時系列に対する考慮は必要である。これは、単に医療のみでなく、他分野にも広く適用できる問題と思われる。

3.2.3 統計処理

今のデータベース管理システムは検索に重点がおかれ、検索されたデータの処理に対してはあまり配慮が払われていない。しかし、医療データベースで必要なのは、むしろ、検索されたデータの処理である場合も多い。

また、Table 型のデータによる検索すなわち、いわゆる relational データベースの技術の発展への期待も大きい。第2に述べた臨床研究用データベースは relational データベースの最もよい適用分野の1つと考えられる。

3.2.4 自然言語処理

医学的問題の項でも述べたように、医療データは、まだ定義のはっきりしないものも多い。このようなデータを扱うためには、Text を検索したり、また、検索のための別の知識ベースをもつといった技術が必要になる。語の検索は、情報検索の中で多く研究され、「前方一致」、「後方一致」、「語の切り出し」等は情報検索の中では、繁用されている。しかし、これらの技術は現在の汎用 DBMS と結びついているものは少ない。

医療の世界ではこれらと合わせたようなデータベース技術が非常に必要とされている。

3.2.5 画像データベース

やや異質な問題であるが、医療の世界では、画像データの重要性から、文字のデータと画像データへ統合化されたデータベースが著しく望まれている。

これは、本特集とは別の問題であると思われるので、一言触れるに留める。

3.3 社会的問題

医療データベースに関する社会的問題は、秘密保持の問題と Cost/effectiveness の問題である。

いずれも、著しく重要な問題であるが、紙面の関係で末尾の文献を参照されたい。

なお、最後に文献について、一言触れるならば、文献1)は、現在のところ、最もよくまとまった単行本である。日本の病院の現状については、文献2)の資料を基礎にして本年中に単行本が発刊される予定である。

参 考 文 献

1) Wiederhold, G.: Database for Health Care,

Lecture Notes in Medical Informatics Series, Vol. 12, Springer-Verlag, Berlin (1981).

- 2) 病院情報システム, 第12回医療情報学研究会資料, 情報処理学会, 東京 (1982).
- 3) 關原成允, 稲田 紘: 医療情報学, 情報処理学会, オーム社 (1980).
- 4) Wiederhold, G., Fries, J.F. and Weyl, S.: Structured Organization of Clinical Databases, Proc. of the 1975 NCC, AFIPS, Vol. 44, p.479 (1975).
- 5) 北添康弘他: 総合医療情報システム—IMIS—の開発研究, 病院管理, 19: 137 (1982).
- 6) 徹底したベンチマーク・テストを実施した米公衆衛生局のケース, 日経コンピュータ, 特別縮小版, 東京 (1981).

(昭和57年7月2日受付)