

医学側から見たデータベース言語の検討

郡 司 篤 晃

(厚生省医務局総務課

医療システム開発調査室)

I. はじめに

これまで医療専用データベース言語として作られたものはMUMPSぐらいである。MUMPSは、ボストン市のMGJにおいて、医療用言語として1960年後半に開発された。医療用であることの条件として、① man-machine interface が柔軟であること、② 可変長の string data があつかえること、③ 樹状構造のデータをファイルで束ねること、④ データベースの管理が容易であること等を満たすものとされた。MUMPS はその後10年の歴史の中で多くの改善が行なわれ、近年はその標準化の動きもある。しかしMUMPSとて、HISやMEDに適用するには多くの欠点がある。また、Barnett自身もMUMPSは必ずしも医療用語ではないという。近年多くのデータベース言語が作られたが、医療の中における普及は必ずしもほかほかしくはない。この原因は一体何か。また、その普及を医療自身の特殊性が妨げているとしたら、それはいつかいつかあるのか。以下このような問題意識のもと、医療情報システムとは何か、また医療情報システムの特徴について述べ、医療用データベース言語の満たすべき条件について考察する。

II. 医療情報システムにおけるデータベースシステム

医療分野におけるコンピュータの応用が進むにつれて、医療情報システムの概念自身も発展しつつある。しかしデータベース技術の応用される範囲となると、(1)病歴情報システム(HIS) (2)文献検索システム、その他のデータバンク、(3)NOMED等の特殊応用が、主たる領域である。しかし、以下においては、問題を把握するため、HISについてのみ検討することとする。

南原はHISについてはこれをさらに表1の8に、Hospital Data Baseと、Medical Record Data Baseに分類する必要を強調している。前者は病院内の日常業務のためのシステムで、処理はおおむね単純な検索で、その内容もあらかじめ決められている。一方後者は臨床研究用であつて、処理は色々の処理の組み合わせになり、処理内容もあらかじめ決めておくことはできない。そして必ずしもリアルタイムである必要はないのだからという。

表I HOSPITAL DATABASE AND MEDICAL RECORD DATABASE

	OBJECTIVES	TYPE OF PROCESSING	CONTENTS OF PROCESSING	TIME FOR PROCESSING	NUMBER OF DATA
HOSPITAL DATA BASE	DAILY OPERATION OF THE HOSPITAL	SIMPLE PROCESSING MAINLY RETRIEVAL	PREDETERMINED	REAL TIME	LIMITED BY THE NUMBER OF THE PATIENTS OF THE HOSPITAL
MEDICAL RECORD DATA BASE	CLINICAL RESEARCH	COMBINATION OF VARIOUS PROCESSING	REQUIRED PROCESSING CANNOT BE PREDETERMINED	MAY WAIT FOR DAYS	UNLIMITED

しかし現在で多く行なわれつつあるようなHISであれば、病歴情報システムであるから、本質的には他の領域

の事務的システムと大差はないであろう。従つてHospital Data Baseなるものをこのような事務システムと解すれば、医療システムの特殊性というものは、データや用法の意味内容を別にするれば、ほとんどないといえる。しかしながら、病歴管理をその業務の一つに加える、医師に直接入力させ、Weed, L.L. のシステムのようにPOSを設け、さらにWarner, H.R. のHELP システム^{*2}のようなもの

まで考えると、必ずしも研究用ではないが、主わめて医療システムの特徴がで
くるのではなからうか。これ等のシステムは必ずしも普及してこのシステムでは
ない。また純粋に研究用にしても同様で、現在の段階では、費用対効果の差から
いつて普及するものではない。従つて現状からはこの二つの分類になり得る。
しかしいわゆる Medical Record Data Base と Hospital Data Base の発展形態を考
えるに、その方向に何つて研究開発が行なわれており、かつまた、二つのシステム
が必ずしも不連続な概念でないと考えられるので、以下この二つを峻別せずに論
議を進めることにする。

Ⅱ. 医療情報システムの特長

(1) システム環境の特長

病院における主たるマンパワーは医師、おまひ看護婦であるが、彼等はコンピ
ュータを操作することを目的とし従業員ではない。彼等の関心は常に一義的に患
者のケアにあり、この目的のために直接的に役立つ仕事でなければ、その仕事
に対するモチベーションはさわめて少ない。従つて、これ等の人々への入力と要請
する場合には、システム側がそのロードを大きくかぶらなければならぬ。

従つてシステムはすべて user oriented でなければならぬ。これとは言語を
使用する場合にはその end use language は、高級で簡単なものであることが必要と
なる。この点 MUMPS は DDL が無いというところが利点になっている。またファイル
の管理も、各ユーザーに分けて持つようなことも考えなければならぬ。

次に医療の業務はこれを行なつていく本人であるため、見直しをしない場合が多い。
従つてあらかじめ作られたシステムを実行に移す後になると主わめて強い拘束感
を与えることになる。この点 MUMPS のようにユーザーが簡単にシステムを作
れるような、いうならば自己増殖型のシステムであることは有利である。しかし
このような場合には、全体のシステムの効果は犠牲にせざるを得ないという欠点
がある。

(2) 用語、記述形式が定式化しにくい。

用語が特殊であるのはやむを得ない。医療の場合にはこの用語自身データ
となることが多い。たとえば病名、処置名、薬品名などは、主わめて重要なデータ
となるものである。

用語の定義、概念との対応が必ずしも明確でない場合が多く、また表現の方法
がいく通りもある。たとえば、病名についても Gordon の調や^{※3}とこ^{※3}と、
雑誌の論文のタイトルに出てきた実質 2,700 ~ 2,800 ほどの病名が、実に 24,000 通
りに表現されていくという。従つて病名について自由な形式での入力をゆるし
た場合には、それを単一のコードと対応づけるためには、ソースを参照するとい
う一連の処理が必要となる。

診療記録の記載は原則として文章である。従つてこれを入力することは主わ
めて困難である。これを標準化することは現在では不可能である。医学は常に医療
の標準化の努力はしてきて、最近の POS の中でも新しい努力がくりかえされ
ているが、医療は個人のケアを目的として以上、文章生産とは根本的に異
なり、例外的、個別的な事項が常に重要になる。

(3) データの種類が多い

アナログデータ、しかも次元のものが多く、その媒体も多様である。文字、

句、文章のデータが多い。

(4) データの構造

医療データは原則として、個人について発生する。従ってデータの論理構造上個人IDをrootとする複雑な樹状構造である。またデータに多くのqualifierがつく。データの発生した日時、データ取得の方法(たとえば、臨床検査の検査方法)などである。また記憶の形式をシリアルに展開したような場合には、ID、データ名、発生場所、発生源等の情報も付け加えなければならないので、たとえば有効教字子樹ぐらゐのデータをファイルするのに40バイトものフィールドを必要とすることになる。従ってこれをどう扱うにしたらよいか、どう処理するかは医療データベースの一つの課題である。たとえばある臨床検査データを検索する場合、ある時点で多くの関連ある項目のデータを見る場合と、ある項目を時系列的に見たいという場合がしばしば起る。

(5) データの検索が多様である。

医療情報システムが個人のケアを目的としたものである以上、当然個人単位のファイルの検索、updateが多いであろう。しかし現実にはhospital data baseの場合には事務管理システムがあるもので、それぞれが部所ごとの目的別のファイルが必要となる。たとえば外来予約においては、日時、科、診療室、医師をkeyとする検索が多く使われる。また検査結果の精度管理のような場合には、検査項目、日時等がkeyになる。その他入院予約、厨房、薬局、等々でそれぞれが業務目的のためのファイルの検索が行われる。

日報、月報、年報の集計、保険の請求業務のような集計業務が多い。

謂ゆる研究目的というものは、検索キーは一般に前もって決めることは不可能である。一応全てのデータがkeyになり得ると考えなければならぬ。しかし研究目的の処理は一般にバッチ処理であるので、特別を例を除けば必ずしもkey directoryを持つ必要があることにはつながらない。しかし、症例検索システムのような場合には、この点で特別な工夫が必要となる。このような場合でもinverted fileはデータが数値であったり、データ項目、データ量が多い場合には、現実には利用が困難であろう。

Hospital data baseとmedical record data baseは、丁字型の関係にある^{*4}。これ等は論理的にlinkageが取られていゝべきである。研究はこのようなシステムを前提にして、プロシヤクタイプにすることによって、そのデータ処理はまわめて楽になるわけである。実際には、このようなシステムを実現してみれば実際に行われる処理というものは、単純な統計処理が主であるという。(表2)

このような統計処理を行なうためには、マスターファイルからの対象の抽出が必要である。

検索にはある程度の論理的式による検索が容易でなければならぬ。先にも述べたように医療データはqualifierのついた同種のデータが多い。

たとえば、同じ検査がくりかえし行なわれることが多い。従って、この場合は「ある検査値が少く以上の人」という表現では充分であつて、「ある検査値が少なくとも一欠以上の人」というような表現でなければならぬ。このような検索式が比較的容易に書ける

表2 要求のあったデータ処理の割合

単純な作表	5.8%	独立性の検定	0.4%
SORTING	1.2%	ノンパラメトリック検定	1.2%
平均値・標準偏差の計算	57.0%	その他の検定	0.2%
度数分布表の作成	9.4%	回帰分析(重回帰を含む)	0.4%
クロス集計表の作成	5.3%	判別分析	0.1%
項目間の簡単な計算	1.1%	分散分析	0
2つ以下の条件による検索	0.8%	共分散分析	0.1%
2つ以上の複雑な条件による検索	0.1%	因子分析・主成分分析	0.1%
層別の集計及び平均値等の計算	3.5%	数量化(林)	1.5%
相関係数(相関行列)の計算	3.5%	特殊な計算	1.8%
平均値の差の検定	5.8%	文字の処理	1.0%

これは、医療データの検索処理には有利であろう。

(6) システムリファインメント上の特徴

① ファイルの種類、規模がきわめて大きい。

患者の基本ファイル、病名、その他の医学用語、薬品名、入出力ファイル等々をまとめた多種類のファイルがある。1日の外来数1,000人以上の病院が多い。M国では毎年50,000件程度の病歴がふたつある。カイザーのシステムの用語集には70,000語がファイルされている。バーモントのWeedのPOSでは60,000の画面が用意されているという。

② ファイル処理機能

バーモントのシステムの60,000画面のうち12,000画面が、単なるプログラムのための画面であり、データの入力ファイルの平均アクセスは4回でありという。

③ response time

ファイルの出し入れが多いので、レスポンスタイムが長いと、作業に長時間を必要とすることになる。エルカミソの病院の場合にはピーク時のresponse timeを3秒以内におさえている。

(7) 特殊な処理

自動問診のような例では、string dataの処理、syntax処理等が必要になる。日本のシステムでは、漢字の処理が可能なことは、man machine interfaceを著しく改善されるであろう。

IV. まとめ

医療の場合には、経済的な理由から、できるだけ小型のCDUを前提としない。また、大きなファイル容量と処理能力を前提とし、データベース言語には次のようなrequirementsが考えられる。

① できるだけend user languageに近いこと。

② 基本的な樹構造のデータ構造があつたこと。

③ いくつかの目的別のindex fileをもつて、早いfile accessが可能なこと。

④ 検査、あるいはデータ保護機能を与えていること。

⑤ バッチ的のファイル形式の変換が行なえること。

reference

1. 南原成允, 医療情報処理研究会資料, 1976. 2. 18.
2. Homer R. Warner, et al. HELP-A program for medical decision-making. Computer and Biomedical Research 5:65-74, 1972.
3. Gordon, B.L., Terminology and content of medical record, Computer and Biomedical Research 3:436-444, 1970.
4. 郡司 篤見 ほか, 循環器疾患における医療データ管理のコンピュータシステム手続, 26:174-180, 1972.
5. Schultz, J.R., et al, An initial operational problem oriented medical record system. AFIPS Conference proceedings, 38-, 1971.