

イメージ文書を基盤とし時間軸管理によるデータベースシステムの構築に関して

平 井 博 久

〒229-0034 相模原市共和1-8-22

hnhhh@mail3.alpha-net.or.jp

近年、広く「構造化されたデータベースシステム」が業務に用いられている。しかしこのデータベースによるシステム開発には、多くの知識と経験が必要とされている。さらにシステム開発後に環境の変化、例えば、利用目的の変化、利用者の視点の変化などによりシステム改定が行われる場合が多くこの作業には多大な労力とコストがかけられている。本データベースシステムは電子化文書を器とし、この電子化文書に含まれている情報及び背景データを要素とし、これを時間軸で管理を行なっている。そのデータ利用に関しては、利用者の指定した時間軸を中心とし、実行時に動的にデータベースが構築される。文書を器としたデータベースである為に修正、変更が容易であり、その結果非常に自由度が高く、耐久性の強いシステムとなろう。

Development of document database system under time-base control

H i r o h i s a H i r a i

hnhhh@mail3.alpha-net.or.jp

"Structured database systems" are widely used in the field of business. However, system development based on databases needs a lot of knowledge and experience. Furthermore, substantial labor and cost are necessary for revision of systems, which frequently occurs as a result of the change in the environment after system development such as change in the purpose of usage and change in users' points of views. This database system is based on electronic documents, and the information and background data contained in such documents are subjected, as elements, to time-base control. According to the time-base a user specifies, a database is established dynamically using relevant data. Since documents are used as bases, correction and modification are easy, and as a result, the system will have a high degree of freedom and high durability.

1. はじめに

近年、構造化されたデータベースシステムが各分野で業務に広く利用されている。しかしこのデータベースシステムを利用してシステム開発を行う場合、多くの知識と経験が必要とされている。また開発後に利用目的の変化、利用者の視点の変化などによりシステム改定が行われる場合が多くこの作業には多大な労力とコストがかけられている。本稿は、この原因が「構造化されたデータベースシステム」の持つ構造上の問題によると考え、この問題解決の為に新しい考えに基づいた「時間軸管理による電子化文書データベース」（以下本システムと呼ぶ）を提案するものである。

しかしながら本提案は、電子化文書を器としてデータベースを構成している為、一方文書管理上の問題に対処しなければならない。この点に関し最初に問題提起する。

さて、昨今インターネットの普及により、ストック情報を利用する従来型の電子化文書の利用形態と異なり、ネットワークを通じて流入するフロー情報を収集し時間軸で管理し、後で活用する形態が一般的になってきている。この情報としては文字データ、画像データがあり、これを基にした情報が多く流されている。これらの情報管理を鑑みた時、本データベースシステムの問題は今後多く利用できるのではないかと考える。

本稿の構成は以下のとおりとなっている。

第2章でシステム開発上、文書管理上の問題を挙げ、その解決として第3章で、新しい考えに基づくデータベースを提案し、第4章でそのシステム概要を述べる。

2. 問題提起

2.1節で構造化データベースシステムによるシステム開発およびその後のシステムメンテナンスの問題、さらに2.2節で文書管理上の問題を述べる。

2.1 システム開発上の問題

以下現在広く運用されている構造化データベースシステムの主な問題点を列記する。

(1) 管理項目の選定

データベースに登録するデータ項目とそのデータ内容の決定は、ある視点から見て行われる。しかしこの視点は、時間的推移により変化し、また各利用者の各人各様の視点がある。この為の開発時に全目的に合致するシステムを設計する事は出来ない。従ってシステム開発後に利用者からたびたびシステムの部分的または全面的な改定の要求が出される事となる。

(2) データ管理

現在データベースシステムで管理されているデータは本来管理すべき価値のあるデータであろうか。利用頻度の低いデータまで登録していないか。この結果システムを複雑にし開発コストを押し上げていないか、等が問題として挙げられる。実際、野口悠紀雄氏（注）は「まず9割は使わない」との指摘もある。

（注）野口悠紀雄著「超整理法」中公新書

もし、データベースで管理すべきデータとして挙げるとすれば、データの再利用、例えば同じデータを複数の帳票に表示するようなデータが対象となるのではないか。

(3) キー項目設定の必要性

現在使われているデータベースはリレーショナルデータベースが多い。このデータベースは、キーデータを与え複数のファイルを検索し、当該データを各ファイルから拾い集めている。このリレーショナルデータベースの特徴は、データの重複性を排除する事、正規化ファイル構造により、物理的にデータ保存領域を節約している点にある。当稿は、この構造化がシステムを複雑にすると主張する。前者は、データ登録時の時間管理を行わない結果であり、後者はハードコストに意識を巡らした結果ではないかと考える。

(4) 画像処理機能

画像データは、現在インターネットの画面に多くの取り扱われているがしかし、業務システムとしては、コンピュータ上にデータベース化されていないものはまだまだ非常に多く、その為に紙による管理を余儀なくされている場合が多い。紙による管理とコンピュータによる管理を併用しなければならない点、以前に比して煩雑性を増している。

(5) データ入力上の問題点

システムへの入力作業の容易さはシステムの生命線である。特に多量のデータ入力、専門職に依存させる必要があり、これが大きな負荷となっている。

(6) その他

現在インターネットは多く使われているが、その目的は情報検索である。この場合に「情報検索」とは言わず「リンクを張る」と言い用語の定義が変わっている。このインターネットの画面は電子化文書形式のものが多いが次のような問題がある。

- ・ 1対1のリンク構造である。
- ・ 情報はディレクトリー構造のもとに管理されているので文書を移動する時にはシステムの作り替えが必要になってくる。

2. 2 文書管理に関する問題点

本システムは、電子化文書をデータベースの器としている。本システムは、紙を電子化文書で置き換えたものであるために通常の紙による文書管理から発生する問題は避けられない。まずこの基本的な問題について言及する。

○問題点

現在、紙による多種多様の文書が多く創られているが、これらの文書管理には通常、まず「分類体系」を定め、それにしたがって分類保管されてきた。しかしながら分類は一意的に決まるのではなく様々な見方があり多くの分類体系が存在する。また、その決定は予め決められているのではなく、文書の量が溜まった段階で決められるものが多い。もし、分類を予め決めてしまうとするやり方を採れば、野口悠紀雄氏(注)によれば

- ・分類時の複数属性の問題、境界領域の問題、分類後の誤入の問題、どこへ入れたか忘れてしまう、等の問題が発生すると述べられている。このような考察の結果、分類する事自体に問題があり、さらに情報処理の原理上の問題があるのではないかと指摘している。

(注) 野口悠紀雄著「超整理法」中公新書

3. 時間軸電子化文書 データベース

そこで本稿では、このような問題点解決の為に時間軸管理に基づく電子化文書データベースシステムを提案する。

設計概念

時間軸管理の電子化文書管理データベースとは、以下の7個の設計概念に基づき構成される、電子化文書をデータベースの基盤(器)とし、この中に含まれる情報を時間軸により管理し、利用時に動的にデータベースを構築するシステムである。

(1) データベースの基盤

従来「データベース」と言う言葉は実体に関する「格納データ」とその「データを定義する記述体」の意味で使われている。しかしながら「データの基地」とする「ベース(容器)」に関する明確に定義づけし

たデータベースシステムは未だない。当システムは、電子化文書を「データの基地」に持つ。

(2) データベース情報空間

上記(1)の電子化文書群で構成された空間を言う。空間を構成する要素は、大きく分けて管理データと背景データからなる。通常、文書には文書作成の「目的」があり、この「目的」に従って文書中には、必要とする「データ項目」とその「データ内容」がちりばめられている。管理データとしては、文書中にある①データ項目名、②データ内容、③さらに当文書作成の目的(つまり文書形式)からなる。①と②を要素という。背景データとは、管理されないがこの文書構成の一部を成すものであり、例えば、文書中の注意書き、手書き文字、表、画像などを言う。この「背景データ」は、電子化文書の文書構成の一部として同時にシステムに保存される。

「背景データ」の重要性

背景データは、管理すべきデータの発生事由(イベント)を示す重要なデータである。この背景データの中から将来利用目的の変更により管理データとして取り上げられるものもある。

(3) 時間的管理

電子化文書はすべて登録時に登録日、登録時間がシステムとして付与され時間的管理がなされる。本システムではこの時間情報が全システムを通じて唯一のキー項目であり、このキーを通じてシステムが運営さ

れる。

(4) キーの概念の排除

現在多く使われているデータベースはリレーショナルデータベースであり、このデータベースは、キーデータを基にして複数のファイルからデータをかき集めている。しかしながら本システムでのキーはシステムが付属的に付ける時間データ以外はキーの概念はない。従って空間内では、同一項目に関して複数のデータが存在する。どのデータを採用するかはアプリケーションに委ねられる。空間内に1個しかない場合が従来のキーである。もし同一項目で異なるデータが複数ある場合、直近の電子化文書に記述されているデータを採る場合には、これが従来のキーデータの概念と同等となる。従ってリレーショナルデータベースシステムの理論は本データベースシステム理論に包含される。

(5) 情報項目とデータの選定登録

電子化文書の中に書かれている情報は利用者が後日情報検索または情報参照する目的の為に登録管理を行う。この作業は、文書が発生の都度行なう事を原則とするが、利用者の利用タイミングを考え利用者に委ねられている。定型文書のように形式が同一であり、さらに指定領域が定まっている文書は、最初1回領域設定の作業を行いこれをテンプレートとして登録する。次回以降は、このテンプレートと呼び出しそれを利用してデータ領域を指定していけば良い。通常データ領域は2個以上あるので個別に領域を指定し、各々項目名を付与していく事となる。文書フォームを変えた場合、項

目の追加削除がありこのテンプレートは変遷させていく事ができる。システムは、このテンプレートの変遷をサポートする。従って、本システムは、データ項目の自由な変遷に耐えるデータ項目増殖型データベースである。

また、イメージにて文書が処理される為に現在不必要とされた文書中の情報も将来活用する事が出来る。

(6) 情報の検索および利用に関して

時間軸を指定して利用者の利用目的に合ったテンプレート番号を端末画面に入力する。本システムは同種または異種のテンプレートを時間軸に沿って合成し、合成されたテンプレートが画面に表示される。この機能は、従来の「情報検索」の用語より「リンクを張る」の用語の方が適当であると考えるがここでは通常どおり「情報検索」の用語を用いる。

(7) 文書保管場所

電子化文書の総ては同一の保管場所に格納される。従ってホルダー、拡張子などの概念は払拭される。従来発生していたディレクトリ分類による文書管理上の弊害は回避される。

4. システム概要

電子化文書データベースの構成と操作の流れを図1で示す

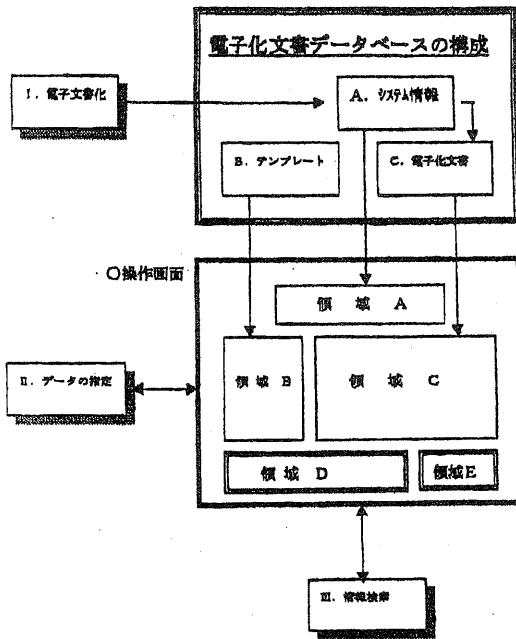


図 1

4. 1 電子化文書データベースの構成

電子化文書データベースは、一登録文書に関し通常（1）システム情報、（2）電子化文書、（3）テンプレートから構成される。但し上記（3）テンプレートは文書の形態により生成されない事もある。なお本システムにおけるキー（識別子）は全システムを通じて時間軸である。これが本システム中唯一のキーである。

（1）システム情報

紙形式の文書をイメージスキャナー機器、ファックス機器を通じて電子化文書として登録する。この時システムは次の情報をシステム情報とし附与し、電子化文書と共に管理する。

- ・登録日、登録時刻
- ・文書サイズ
- ・OCR情報（文書の先頭行又は全行）

（2）電子化文書

紙形式文書を電子文書化したものである。この文書中に含まれる管理すべき情報は罫い枠を持って指定される。この位置はテンプレートにて管理される。

（3）テンプレート

電子化文書中に含まれるデータ内容の領域を指定し、それに係わるデータ項目名とデータ内容をテーブル形式にて構成する。登録情報としては、

- ・テンプレート番号
- ・データ項目名（手入力、又はOCR化）
- ・データ内容（同上）
- ・データアドレス

なお本システムでは、デフォルト値が入力されているテンプレートと空白値のテンプレートがある。

4. 2 操作画面の構成

図1にて操作画面の構成を示す。その内容は以下の通り。

（1）電子化文書に関する情報の表示

- ・領域A：システム情報表示
- ・領域B：テンプレート
- ・領域C：電子化文書

（2）利用者操作領域

- ・領域D：検索情報の入力
入力内容は、検索時間軸、テンプレート番号
- ・領域E：操作ボタン群（前頁、次頁へ移動等）他

4.3 操作手順

本システムの運用は、以下の3つのフェーズにより行われる。

I. 電子文書化フェーズ

紙を媒体とする文書をイメージスキャナーやファックス機器を利用し本システムへ電子化文書とし変換登録する。

II. データの指定フェーズ

当該電子化文書中のデータ領域を指定し登録する。この作業はテンプレート機能を用いるが、この場合次の2ケースがある。

(1) テンプレートの生成

新しい形式の文書に対し、システムの操作機能を用いデータ領域を指定しながらテンプレートを生成する。

(2) テンプレートの利用

同一形式の文書に対しては、作業軽減の為に前に生成したテンプレートを利用してデータ内容を入力していく。

○なお、当テンプレート機能の応用として出力帳票を容易に設計する事ができる。

帳票レイアウトをイメージスキャナーで背景データとして読込み、テンプレートにより当該データをこの帳票に埋め込むものである。

III. 情報検索フェーズ

検索を行いたい電子化文書に対しテンプレートを用い、検索データを入力する。

4.4 テンプレート

本システムで中心的な役割を果たす機能であり、その主とするところは、利用者の時間軸指定に基づき、関連する電子化文書群をデータベースの器とし、データ項目名とデータ内容の一覧表を作成し、関連電子化文書を表示する。

○テンプレートの合成

本データベースは、テンプレートの合成により行われる。まず利用者は、対象とする電子化文書の作成された期間と、対象とする電子化文書の形式を示したテンプレート番号を指定する。

本システムはその電子化文書が持つテンプレートを集約し新しいテンプレートを合成する。この機能は、システムが、実行時に電子化文書を基盤とし動的にデータベースを構築する事を意味する。

テンプレート合成は下記2ケースがある。

(1) 同一形式の文書を対象として

検索の対象は同一形式の電子化文書である。他形式の電子化文書を参照しない場合である。通常コンピュータ用語として使われているファイル1個の意味である。ただし本システムでは、データ項目名の変遷(追加、削除、変更)は許している。

その動きを下記図2に基づき説明する。

この図は3枚の電子化文書が時間を異なえて登録され、またデータ項目名も変化している事を示している。

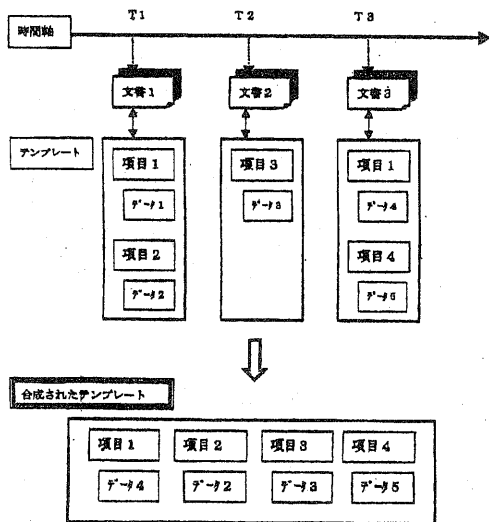


図 2

電子化文書1では、データ項目名1に対しデータ1、データ項目名2に対しては、データ2が登録されている事を示している。

以下同様であるが、電子化文書3では、データ項目名1に対してデータ4が登録されている事を示している。

これを時間軸 T1 から T3 を指定して1枚の検索用のテンプレートとして合成している。その合成されたテンプレートのデータ項目名とデータは図2のようになる。なお、データ項目名1に対してデータ4となる。これは、本システムが直前のデータ値を採用している帰結である。

データ値が2個以上ある場合には操作ボタンで順次画面を捲っていく。但し、画面上では現在使用されているものがアクティブであり、他電子化文書の内容を見なければフォーカスの移動が必要である。

(2) 異種の文書形式が混在する場合

これは、例えば本文書から他文書の標準データを参照したい場合である。

この場合、本文書および他文書に関するテンプレート番号を並列に指定する。システムは、関連する項目名間にリンクを張り一覧表として形成する。なお参照の電子化文書は画面の決められた位置に表示される。この場合、参照文書間のテンプレートの成ロジックは前(1)の場合と同様である。

なお、テンプレート作成時に、システムとして色々なデフォルト値を前もって設定しておけば後の運営に便利である。

5. おわりに

本稿では、管理データと背景データからなる電子化文書を器とし時間軸管理に基づく新しい形のデータベースを提案した。これらの情報は、総て時間軸で管理され利用時に時間軸の指定のもと動的にデータベースが生成できる。この為に本データベースは、いつでも容易にデータ項目の追加削除が出来き、その結果種々のシステム開発およびその後のシステムメンテナンスが容易になる事が期待できる。

謝 辞

本研究は新医療施設開発振興財団からの助成を受けています。現在、本研究をベースとした医療関係のアプリケーションシステム開発に取り組んでいます。システム開発支援を戴いた事に深謝致します。