
第8章 データベースの基礎

高田 良宏

1. はじめに

文化資料の調査・整理・保存・再利用は、長い間アナログ的手法で進められてきたが、情報技術の進歩により、さまざまな文化資料が電子化され、より有効に活用可能な状況となった。さらに、近年の情報機器の低価格化と社会への浸透により、膨大な量のデータ（電子化された資料とその関連データ）が生産されるようになってきた。それら膨大な量のデータを共有データとして再利用するには、データベース化が不可欠といってもよい。

2. データベースとは

データベース (Database) とは、広義の解釈では、「特定のテーマや目的のために収集されたデータの集まり」であり、紙上のアドレス帳や名刺をファイリングしたものもデータベースである。さらには、図書館もデータベースと言える。一方、狭義の解釈では、「特定のテーマや目的のために収集され、コンピュータから容易に再利用できるように整理整頓されたデータの集まり」であり、コンピュータから利用することを前提としている。

データベースの語源は、第二次大戦後の米軍が、そこにアクセスすればすべての情報が得られるように、点在していた膨大な量の資料をひとつの基地に集約して効率化を図った際に、その場所を **Data** (情報) の **Base** (基地) と呼んだことにあると言われている。

データベース化することの利点を次に示す。なお、これ以降の説明は、狭義、すなわちコンピュータ上のデータベースについて行う。

- ①データを共同利用できる
- ②データを複数の人と共有でき、データ管理コストを削減できる
 - 膨大なデータを最小限にまとめることができる
 - 各所に散在している同種類のデータをまとめることができる

③データ間の矛盾をなくすことができる

個人、部門ごとに管理すると、データ相互間に矛盾が生じる。データベース化によりこれらの矛盾を排除し、データの信頼性向上をはかることができる

④いろいろなデータアクセスが可能

わかりやすいデータベース操作言語を用いて、いろいろなデータアクセスが可能になる（例えば後述のリレーショナル・データベースでは、ISO 等で規格化されており、製品間の相違が少ない）。

⑤データの機密保護と安全確保が容易（セキュリティ）

データの暗号化やアクセスのためのパスワードを付けることで、データの機密保護と安全対策が可能

⑥データアクセス管理の強化

複数のユーザから1つのデータへのアクセスを調整し、正確な更新を可能にする

⑦業務の標準化を図る

データベース化はコード化、用語統一、作業の標準化を伴って推進される。したがって、データベース化は業務の標準化、合理化に大きく貢献することになる

データベース自体はコンピュータ上のデータの集まりであって、利用者やアプリケーションソフトウェアから直接データベースを操作することはしない。データベースに対する操作は、データベース管理システム（DBMS : DataBase Management System）を介して行う（図 1）。DBMS はデータベース上のデータを参照したり、データを追加、変更、削除するなどの操作やデータベースの管理のための仕組みを提供するためのソフトウェアである。データベースソフトとして提供されるものはこれにあたる。Oracle, MS SQL, MS Access, PostgreSQL, MySQL などがよく知られている。データベースと DBMS を合わせてデータベースシステムというが、我々がデータベースという場合は、このデータベースシステムをさすことが多い。

DBMS の役割の概要を右に示す。

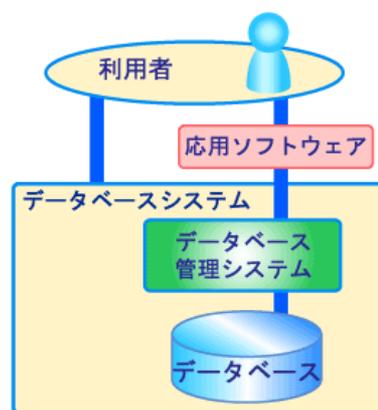


図1 データシステムの概要

①データベース管理

問合せ処理、操作履歴管理

データの一貫性チェック（有り得ないデータや矛盾の排除）

物理的なデータ管理（インデックス作成）

②同時実行制御

排他制御（ロック・アンロック）

③機密保護機能

④トランザクション管理（トランザクション：関連する複数の処理を一つの処理単位としてまとめたもの）

⑤障害回復機能

利用者は、直接データベースシステムを利用することもできるが、一般的には応用ソフトウェア（図1）を介して行う（応用ソフトウェアがデータベース管理システムで提供されている機能呼び出す）。応用ソフトウェアとは特定の目的のために作られたアプリケーションソフトウェアのことで、普段我々が利用している情報システムはこれにあたる。

データベースは世の中に存在する多くの情報システムで利用されている。その用途の例を以下に挙げる。

- ✓ 住所簿等（出席簿，株主名簿，選挙人名簿，電子電話帳，…）
- ✓ 検索エンジン（Google，Yahoo，…）
- ✓ 業務系システム（銀行の預金処理，クレジットカードの決済処理，株取引，各種切符購入・予約，生産管理，在庫管理，人事管理，電子カルテ，…）
- ✓ 各種情報DB（研究者DB，企業DB，国勢調査，電子百科事典，IPDL（特許電子図書館），OPAC，化学DB，機関リポジトリ，…）
- ✓ 金沢大学内を見てみると…
アカンサスポータル，教務（学籍，シラバス，履修，成績），落とし物，就職支援，人事，給与，財務，実績，図書，リポジトリ，…

3. データベースの仕組み ～リレーショナル・データベース～

データベースに格納されるデータはある特定の規則によって管理される。この特定の規則をデータモデルといい、その管理法の違いによって幾つかの種類に分けられる。

- ①階層型データベース：データを階層型に格納/整理する仕組みをもったデータベース。もっともイメージしやすい例は、会社の組織図である。階層型データベースでは、データはツリー構造で表し、ある1つのデータが他の複数のデータに対して、親子の関係をもつ。したがって、データにアクセスするためのルートは一通りしかない。データの重複が問題となる。
- ②ネットワーク型データベース：データは網の目の形で表現される。それぞれのデータ単位（ノード）が繋がっていることから、ネットワーク型と呼ばれている。ネットワーク型データベースでは、複数の親データへのアクセスが可能になり、階層型

で問題となっていた冗長性を排除する仕組みになっている。

③オブジェクト指向データベース：オブジェクト指向システム開発および、オブジェクト指向プログラミングの考え方がデータベースシステムに取り入れたものである。オブジェクト指向の基本的な考え方には、抽象データ型（abstract data type）、継承（inheritance）およびオブジェクト識別性がある。抽象データ型は、現実世界の実体をデータとそれに対する操作または、手続きとでカプセル化して表現するという考え方である。

④リレーショナル型データベース：縦横2次元の表を基本にデータを管理する。表の縦一列が項目（列）に、横一列がレコード（行）にあたる。この表に対して、射影（プロジェクション）、選択（セクション）、結合（ジョイン）といった操作を行うことにより、自分の望むデータを取り出すことができる。構造が簡単なために理解しやすく、応用ソフトウェアを構築しやすい。

（カード型データベース：）ソフトウェアの便宜的な分類。データベースのデータモデルとしては存在しない。

一般に、データベースというと、このリレーショナル型データベースをさす。通常は、リレーショナル型の型も省略し、単にリレーショナル・データベース（RDB：Relational DataBase）と呼ばれることが多い。リレーショナル・データベースは、商用DBの80%以上を占めると言われている。リレーショナル・データベースを管理するDBMSを特にRDBMS（Relational DBMS）という。リレーショナル・データベースには次のような特徴がある。

①各データは2次元の表によって表現される（図2）。

	列名		列		
	fname	fname_kana	sex	code	phone
行	高田	タカタ	男	3-7777	076-222-3333
	笠原	カサハラ	男	2-1234	
	GOTO		NULL		

表，列，行単位で処理が行われる

- ・表：テーブル（Table）
- ・列：カラム（Column），項目，フィールド（Field）
- ・行：ロー（Row），レコード（Record）
- ・列名：カラム名（ColumnName），項目名，フィールド名（FieldName）

一行が一件のデータ

- ・行：0～n件，列は固定

図2 テーブルと各部の名称

②複数の表のデータを関連付けることで、すべてのデータをひとつの巨大なデータベースとして活用する（図3）。

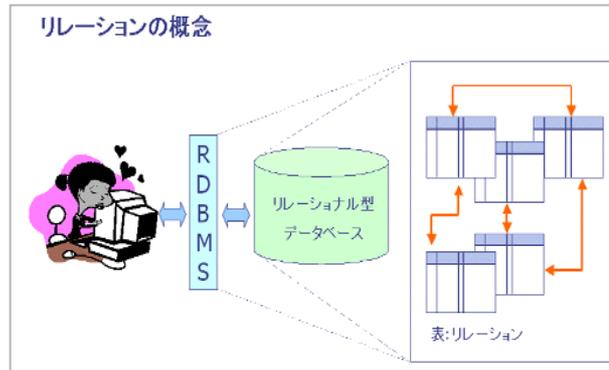


図3 リレーシヨンの概念

③リレーシヨナル・データベースに対応するデータベース言語としてSQL（Structured Query Language：構造化問合せ言語）がISOとJISを取得している。SQLは、以下の3言語で構成されている。

- DDL（Data Definition Language）（データ定義言語）
例：CREATE TABLE（表の作成）、DROP TABLE（表の削除）、ALTER TABLE（表の属性変更）
- DML（Data Manipulate Language）（データ操作言語）
例：SELECT（行の表示）、INSERT（行の挿入）、UPDATE（行の更新）、DELETE（行の削除）
- DCL（Data Control Language）（データ制御言語）
例：GRANT（権限委譲）、REVOKE（権限削除）、COMMIT（変更確定）、ROLLBACK（ロールバック）

さらに階層型データベースやネットワーク型データベースと比べると、

④プログラムとデータの独立性が高いため、データベース構造に変更が入ってもプログラムへの影響は極めて小さい。

表計算ソフト（エクセル等）とリレーシヨナル・データベースの違いを簡単な例で説明する。図4は文房具の出荷状況をエクセルで管理している例である。図5は、図4と同じ出荷状況をリレーシヨナル・データベースソフト（MS Access）で管理している例である。図4では、得意先名の情報や商品の情報が繰り返されて入力されており、①誤入力する可能性が高い、②あらかじめ得意先情報や商品情報の管理を行うことができない、③得意先や商品情報に変更があった場合は対象となる全レコードを修

出荷明細

	A	B	C	D	E	F	G	H	I	J	K
1	ID	日付	得意先コード	得意先名	得意先の住所	得意先のTEL	得意先のFAX	商品コード	商品名	単価	風量
2	1	2000/01/0	101	オレンジ商店	和歌山県	11-1111	12-1111	A101	えんぴつ	10	10
3	2	2000/01/0	102	アップル屋	青森県	22-2222	23-2222	A102	消しゴム	50	10
4	3	2000/01/0	103	すいか商店	熊本県	33-3333	34-3333	A101	えんぴつ	10	10
5	4	2000/02/0	101	オレンジ商店	和歌山県	11-1111	12-1111	A101	えんぴつ	10	10
6	5	2000/02/0	102	アップル屋	青森県	22-2222	23-2222	A102	消しゴム	50	10
7	6	2000/02/0	103	すいか商店	熊本県	33-3333	34-3333	A101	えんぴつ	10	100
R											

図 4 出荷状況（エクセルで管理した例）

正する必要があるなどの問題が起きる。図 5 では、関連する項目ごとに分けて管理しており、上記①～③のような問題は発生しない。つまり、リレーショナル・データベース化することで、データ間の整合性を保つことが可能であり、さらに、業務量の減少にもつながる。

The screenshot shows three tables in MS Access:

- 出荷数: テーブル** (Sales Table):

ID	日付	得意先コード	商品コード	数量
1	2000/01/01	101	A101	10
2	2000/01/01	102	A102	10
3	2000/01/01	103	A101	10
4	2000/02/01	101	A101	10
5	2000/02/01	102	A102	10
6	2000/02/01	103	A101	10
*	0			0
- 商品: テーブル** (Product Table):

商品コード	商品名	単価
+ A101	えんぴつ	10
+ A102	消しゴム	50
*		0
- 得意先: テーブル** (Customer Table):

得意先コード	得意先名	得意先の住所	得意先のTEL	得意先のFAX
+ 101	オレンジ商店	和歌山県	11-1111	12-1111
+ 102	アップル屋	青森県	22-2222	23-2222
+ 103	すいか商店	熊本県	33-3333	34-3333

図 5 出荷状況（MS Access で管理した例）

4. モデル化：ER 図とレコード

モデル化（モデリング）とはデータ（管理対象となる物や事象）をルールに基づき整理し、構造図として表わす手法である。データベースの基本設計段階では、ER 図（Entity Relationship Diagram）という構造図を作成する。ER 図とは、データ間の関連によってデータ構造をモデル化する手法「ER モデル」を図で表したものである。現実の世界の管理対象、たとえば、顧客、商品、部品、サービス、人などの実体（Entity）と関連（Relationship）の 2 つの概念を用いて表す。

ER 図にはいくつかの記述法がある。図 6 に ER 図の記述方法の一例を挙げる。管理対象（実態）は四角で囲む。管理対象間は、管理対象間の関係（1 対 1、1 対多、または、多対多）に応じて、直線または矢印で結ぶ。ER 図的には、多対多の関係でも問題ないが、多対多の関係は、管理対象間の親子関係を特定できず、リレーショナル・データベースとして設計できないので、1 対多の関係に分解する。

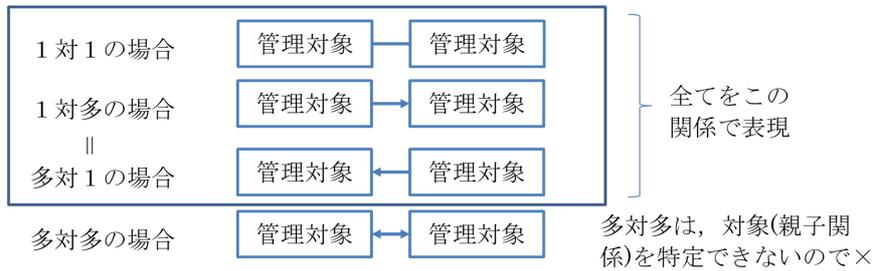


図6 ER図の記述方法の例

図7は、小学校での管理対象間の関係を表したER図である。生徒とクラス、クラスと学校、教員と学校の関係は1対多の関係、校長と学校、教員とクラスの関係は1対1となっている。図7のように1対1および1対多等の関係を、明記することができる。

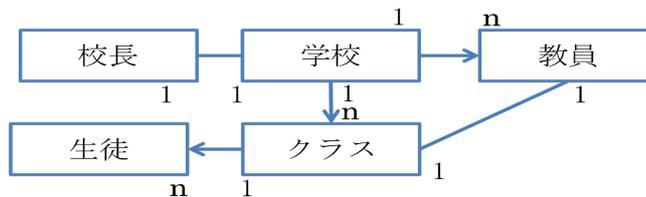


図7 学校での管理対象の関係

次に、図書の貸し出しを例に、ER図を作成してみる。ある図書館において、従来はエクセルで管理していた貸出簿をデータベース化するものとする。

図8はエクセルでの管理簿である。この場合の問題点は、本を借りるまで本の存在を特定できず、同様に、本を借りるまで利用者の特定ができないことである。つまり、本と利用者を事前に管理できていないことである。

図書ID	書名	氏名	貸出日	返却日	備考
M1090	DB入門	高田 良宏	2012.4.1	2012.5.1	
M1091	DB応用	森 雅秀	2012.4.4		
M1090	BD入門	島三郎	2012.4.20		
M1092	UML入門	高田 良宏	2012.5.1		
M1512	SQL入門	嶋三郎	2012.6.1		

図8 図書の貸し出し簿（エクセル）

そこで、図書と利用者を管理対象としてER図を作成した(図9)。ところが、ある本は複数の利用者から借りられる。ある利用者は複数の本を借りるので、両者の関係は多対多となる。つまり、親子関係を特定できないことになる。

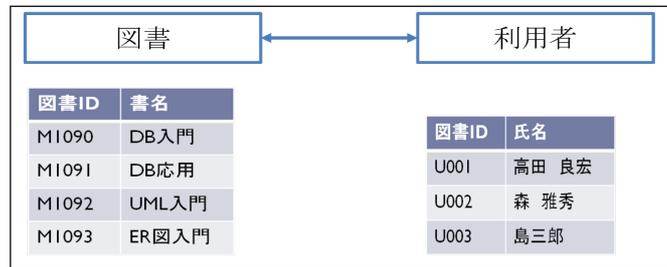


図9 図書と利用者を管理対象とした場合のER図（上段）
検証のために作成したレコード（下段）

図9で、貸出情報を記録できなかったことに着目し、新たに図書貸し出しという管理対象を加え関係を整理したものを図10の上段に示す。管理対象に図書貸し出しを加えたことによって、図9で問題であった多対多の関係が解消され、全て1対多の関係で表すことができる。

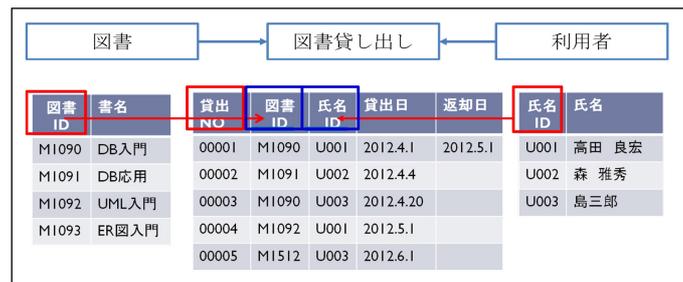


図10 図書の貸し出しを管理するER図（上段）
検証のために作成したレコード（下段）

テーブルには、行を一意に特定するための項目を設定する。その項目を主キーという。図10の3つの管理対象においては、左からそれぞれ図書ID、貸出NO、氏名IDが主キーである（図10の下段）。列と列に依存関係がある場合、参照する側の項目を外部キーとして設定する。外部キーを設定することにより、親子関係を特定できる。図書貸し出しの図書IDと氏名IDはそれぞれ、図書の図書ID、利用者の氏名IDを参照するので、外部キーに設定する。

5. 簡単な例および文化資源学での応用

データベースの基本設計の簡単な例および文化資源学での応用例を付録の5、6に示すので参照されたい。