

4M-6

大規模DB向け正規化手法

川下 満 関根 純 池田 哲夫
NTT情報通信処理研究所

1. はじめに

特定のDBMSを前提としたDB設計の前に、DBで管理すべきデータを整理することの重要性が唱えられており、その際に中心となる手法が正規化である。

正規化は、データ項目間の関係(従属関係)を整理することから始まる。企業の業務に用いるDBでは正規化の対象となるデータ項目は一般に多量であり(数千~数万項目)、これらの間の関係も莫大となる。従って、正規化を効率よく行うにはデータ項目間の関係を簡潔に表現することが重要である。

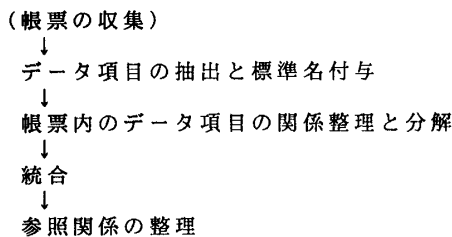
データ項目間の関係を表現する方法として、パブルチャート^[1]、Backus-Naur式表記^[2]、等が提案されているが、前者は図形表現でありコンパクトな表記は難しく、また後者はデータ項目名を演算子で繋いだ表現であり直感的にわかりにくい。また、データ項目のグループを正規化により分解する際の転記が煩わしく、ともに数千~数万のデータ項目を持つ大規模DBの正規化に用いるには効率が悪い。

筆者等は、設計者がデータ項目の関係を理解しやすい単位(例えば業務で用いる帳票)毎にデータ項目の関係を整理して正規化を行う手法を採用し、帳票内のデータ項目の関係と正規化の経緯が容易にわかり、しかも、分解の際の転記をなくした表記法を考案した。

本稿ではこの表記法とそれを用いた正規化の手順を報告する。

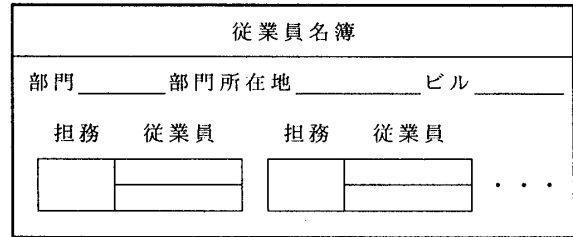
2. 正規化手法

採用した正規化の手順を示す。なお、正規化は第3正規形までとした。



3. 正規化と表記法

図1の帳票を例として、正規化の手順と表記法を説明する。



<仮定> 部門: 部門所在地 = 1: 1
担務: 従業員 = 1: N

図1 正規化する帳票の例

3. 1 データ項目の抽出と標準名付与

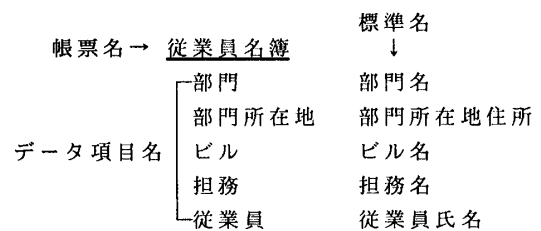
集めた帳票よりデータ項目を抜き出し、帳票名と共に記述する。

また、帳票のデータ項目名を標準名に変換し、帳票より得た名称の横に記入。これにより類似のデータ項目は類似の名称となる。標準名は統制された単語を用い、次の形式をした名称である^[3]。

標準名 = 修飾語 + 主要語 + 区分語

【言葉の説明】

- 区分語: データ項目が表す値の範疇、単位。
(例) 名、住所、コード
- 主要語: 区分語を修飾する用語の内特に重要なもの。関与する実体が候補。
(例) 部門、ビル、担務
- 修飾語: 区分語、主要語を修飾する用語。
(例) 新、定期、最大、以上



3. 2 帳票内のデータ項目の関係整理と分解

(1) 繰返し項目の排除(第1正規形の作成)

帳票内のデータ項目に、階層(繰返し項目のネスト)の深さに従ってレベルの番号を付ける。

①レベル1のデータ項目を1つのグループとし、このグループ内でキーとするデータ項目には「K」（K: Key）を、他のデータ項目には「N」（N: Non key）を付ける。グループに名前を付ける。

【キーについての補足】

- ・キーとなるデータ項目が無ければふさわしいデータ項目を（）を付けて追加する
- ・キーが複数のデータ項目より構成される（連結キー）場合はそれらのデータ項目の全てに「K」を付ける

②レベル2のデータ項目とレベル1で「K」を付けたデータ項目を1つのグループとして、グループ内のデータ項目に「K」と「N」を付ける。レベル1で「K」を付けたデータ項目はレベル1のグループのキーに対する参照キーとなる。

③更に下のレベルのデータ項目についてこれを繰り返す。

従業員名簿	レベル	グループ名		
	↓	組織	担務	担当
部門名	1	K	N	
部門所在地住所	1	N		
ビル名	1	N		
担務名	2		K	N
従業員氏名	3			N
(従業員番号)	3			K

(注) 従業員氏名には将来同姓同名があるりうが従業員番号なら従業員を一意に識別できる。

(2) 従属関係の抽出 (第2、第3正規形の作成)

帳票内の全データ項目を見渡して既にグループとして分解した以外の従属関係を探し、新たなグループとする。また、このグループ内の従属される側の(キーとなる)データ項目に「K」を、従属する側のデータ項目に「N」を付ける。このグループに名前を付ける。

従業員名簿	レベル	組織	担務	担当	ビル
部門名	1	K	N		
部門所在地住所	1	N			N
ビル名	1	N			K
担務名	2		K	N	
従業員氏名	3			N	
(従業員番号)	3			K	

更に次の処理を行う。

①あるグループ(G2)で「K」の付いたデータ項目の全てに他のグループ(G1)で「K」が付いているなら、G1とG2で共に「N」の付いたデータ項目をG1より除く。

②あるグループ(G3)で「K」の付いたデータ項目の全てに他のグループ(G1)で「N」が付いているなら、G1とG3で共に「N」の付いたデータ項目をG1より除く。

帳票xx	G1	G2	G3
データ項目1	K		
データ項目2	K	K	
データ項目3	≠	N	
データ項目4	N		K
データ項目5	≠		N

従業員名簿の例ではグループ 組織 の部門所在地住所の「N」に取消線をする。

以上の処理により、各グループは第3正規形となる。

3. 3 統合

各帳票より作成したグループのうち、同じキーを持つ、または、候補キーをキーとするグループを次の手順で統合する。

①各グループの「N」のデータ項目でキーとなり得るデータ項目には全て「C」を書き加える(C: Candidate key、候補キーの意)。

各グループの「K」と「NC」の付いたデータ項目名を集め、主要語と区分語の対で分類する。同じ組に分類されたデータ項目名について同一のデータ項目であることを確認した後、それらのデータ項目を持つグループを統合する。

②統合先のグループのデータ項目名に統合するデータ項目を付け加える。統合されたグループはグループ名に取消線をする。後で参照関係を知るために統合されたグループには統合先を、統合先のグループにはどのグループを統合したかを記録する。

③グループを統合したら、新たに生じたデータ項目間の従属関係を探し、別グループとする。

3. 4 参照関係の整理

キーと参照キーの関係は、3. 3 ②の記録により知ることができる。また、グループの分解で生じた以外の隠れた参照関係は3. 3 ①と同じ手法で探せばよい。

4. おわりに

大規模DBの正規化を効率よく行うための表記法と、それを用いた正規化の手順を示した。今後は業務用DBの設計に適用して問題点を探り、改善する予定である。

文献: [1] ジェームズ・マーチン「データベース環境の実現と管理」日経マグロウヒル社(1987)
 [2] Tom DeMarco「構造化分析とシステム仕様」日経マグロウヒル社(1986)
 [3] 関根、川下、鈴木「ネーミング手順と支援ツール」信学技報、Vol.89、No.63(1989)