

## 統合的データモデリング支援環境“Darwin”(5)

4 E - 8

### ーデータ管理支援ツール“Darwin/DA”の紹介ー

丹羽隆 加藤 雅樹



NTTデータ通信株式会社 技術開発本部

#### 1 はじめに

近年になって、DOA(Data Oriented Approach: データ中心アプローチ)によるシステム開発の普及が進み、データ資源の管理や再利用を重視するようになってきた。それに伴って、データ管理者(Data Administrator)の作業が注目されるようになり、その重要度が増してきた。ただ、当然その作業は複雑化・多様化していくことになり、CASEツール等による効率化が望まれるようになってきた。本稿では、統合的データモデリング支援環境“Darwin”のうち、データ管理のフェーズを支援するためのツールである“Darwin/DA”のコンセプトと特徴的な機能について紹介する。

#### 2 コンセプト

データ管理者は、データ分析・データベース設計を進めていくに従って、リポジトリに蓄積されていく様々な情報(データ)を管理する必要がある。“Darwin/DA”はそれを支援することを目的としている。

本ツールでは、データ資源の一元管理や再利用を図るために、すべての設計情報をリポジトリ(サーバ上のデータベース)に格納する。リポジトリ上には、開発プロジェクトに特化した情報と、プロジェクトに独立に再利用可能な情報とを分割して置き、高度なデータ資源管理を実現するための環境を提供している。再利用可能な情報としては、以下のものを格納している。

##### (1)用語

本ツールは、異音同義語の発生によるデータ項目重複を排除するために、拡張Durell命名規則<sup>[1]</sup>による名称標準化機能<sup>[2]</sup>をサポートしている。この機能の実現のために、モデルにおいて用いられる用語を辞書に格納する(用語辞書)。

##### (2)ドメイン

本ツールでは、同音異義語の発生によるデータ項目重複を排除するために、アトリビュートの値の範囲やデータ型は、ドメインによって規定している。ドメインの情報はデータモデル設計情報の一部であるが、プロジェクトに独立なものとして再利用可能なので、ドメイン辞書として別に切り出している。

これらデータ辞書に関しては、基本的に追加・変更できるのはデータ管理者だけである。したがって、これらの構築はデータ管理者に大きな負担がかかる作業となっている。そこで、この作業に注目し支援機能を考案・実

装した。

#### 3 用語辞書構築支援

名称標準化機能の実現のために、データモデル内で使われる言葉(単語)は、すべて用語辞書に登録されている必要がある。2章に述べたように、用語辞書の構築はデータ管理者の大きな作業である。

用語辞書構築の作業は次のような手順で行われる。

(1)文書・帳票等から、業務に用いられる用語を抽出する。

(2)用語を意味によって分類する。同義語・類義語同士を1つのグループにする。

(3)各同義語グループについて、標準用語を決定する。(図1)

(4)各標準用語の文法的な役割を分析し、それによって用語類別<sup>[3]</sup>を決定する。

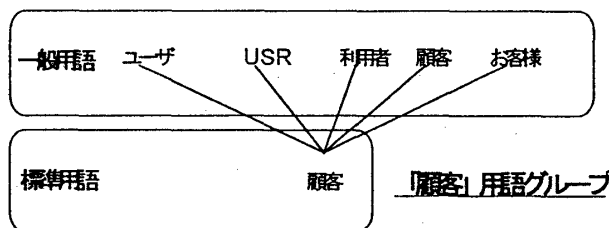


図1 用語類別

これらの作業は、人間の手作業に負う部分が多い。特に(1)の作業は、紙上での書き移し・手入力でのコンピュータへのデータ投入を多く必要とし、効率化が望まれる部分である。

本ツールでは、最も手間がかかる手入力作業を極力避けるために、次のものを提供する。

##### (1)基本用語辞書

一般的な用語が蓄積された基本用語辞書を提供する。この辞書は、通常よく用いられる用語を、同義語・類義語によってグループ化して蓄積したものである。

用語辞書を構築する場合を考えたとき、この基本用語辞書を用いることで、業務に特化した用語のみを新たに登録すればよいことになる。

##### (2)文書ファイルからの用語抽出機能

電子化された文書から単語を抽出する機能を提供する。これは、入力文書ファイルに対して簡単な解析を行い、用語の候補となるものを抽出し、出力するものである。データ管理者は、抽出された用語の中から、基本用語

#### Integrated Data-Modeling Environment “Darwin”(5)

##### ~Introduction of Logical Data-Modeling Tool “Darwin/DA”

Takashi Niwa, Masaki Kato

Research and Development Headquarters,

NTT DATA COMMUNICATIONS SYSTEMS Corp.

辞書にない（業務に特化した）単語を辞書に追加登録してやればよい。

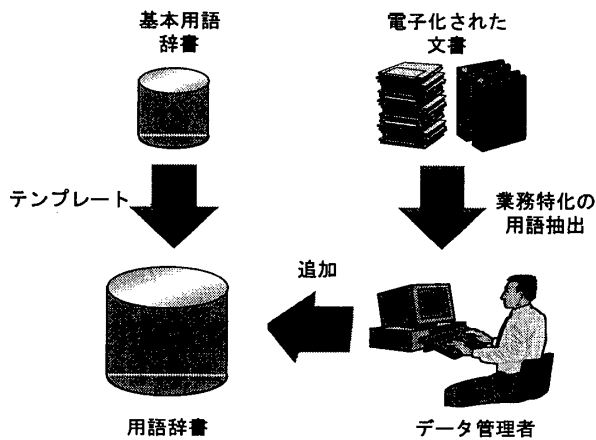


図2 用語辞書の構築

(1)と(2)の機能を用いることにより、用語辞書の構築をほとんど手入力作業なしで行うことができ、作業効率の改善が期待できる。

#### 4 ドメイン辞書管理支援

アトリビュートの詳細を決定する段階では、必要なドメインはすべて辞書に登録されているのが理想である。しかしながら、一般にドメインとなる項目の抽出をトップダウンに行うことは難しい。現実には、データ分析者がボトムアップにドメインとなる項目を挙げてくることが多い。しかし、分析者が自由にドメイン辞書を変更できるとすると、各分析者の意識の違いにより、同音異義語の発生の怖れが出てくる。

このことを解決するために、ドメインの仮登録/認証という方法を考案した。この方法では、ドメインの追加は次のような手順で行われる。

##### (1)仮登録（データ分析者）

データ分析者がドメインとして挙げたい項目をドメイン辞書に登録する。（この段階では、そのドメインは登録した設計者のみに有効である。）

##### (2)審査（データ管理者）

仮登録されたドメインをデータ管理者が辞書に登録するものとして適当かどうかを判断する。

##### (3)承認（データ管理者）

適当と判断されたドメインは、正式に辞書に登録される。（この時点で初めて、そのドメインは全員に対して有効になる。）

この手順に従ってドメインの追加を行うことで、ボトムアップにドメイン項目の抽出を行いたいという要求と、データ管理者が一元的に管理したいという要求を両方満たすことができる。

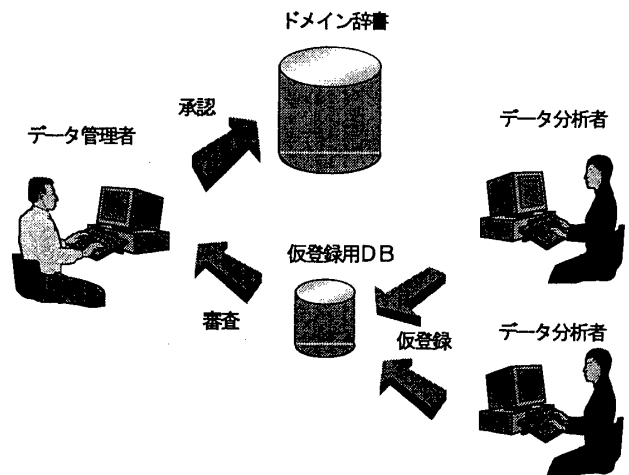


図3 ドメイン辞書の構築

本ツールはこの方式に則った、データ分析者からのドメイン仮登録機能・データ管理者による承認機能をサポートしている。これにより、ドメイン辞書の構築・管理のための効率的な作業環境を提供している。

#### 5 おわりに

本稿では、データ管理ツール“Darwin/DA”の、特徴的な機能について報告した。

今後の課題として、リポジトリ上のメタモデル情報の管理作業の整理と支援機能の考案、ドメイン辞書のテンプレートの用意などを考えていく予定である。

#### 参考文献

- [1] 加藤 雅樹：Durell命名規則の拡張によるデータ項目名称付与の自動化，情報処理学会第49回全国大会，3V-07，1994
- [2] 加藤、他：総合的データモデリング支援環境“Darwin”(1)，情報処理学会第51回大会，4E-04，1995
- [3] Durell.R.W：データ資源管理，日経マグローヒル，1987