

データベースビューについて

東北大学電気通信研究所 増永良文

1. はじめに

データベースは社会の貴重な共有資源である。しかしながらデータベースを利用しようとするとき、一般に蓄積されているデータやその操作言語が利用者の使用したい形で提供されているという保証はない。互いの利用者がいれば百様の利用形態を望むといっても過言ではない。このようなときデータベースシステムがビュー (view) をサポートし、データベースと元の操作言語を仮想化して利用者に提示し、データベースに格納されている情報を有効に利用させることができれば大変都合が良い。それが実世界の利用者の視野を現実的にサポートすることになるからである。データベース管理システムのビューサポート機能は現在のところ十分に機能している。例えば関係データベースシステムとして知られている IBM 社の SYSTEM R, QBE, ある...

のカリフォルニア大学バークレー分校の INGRES ではビュー定義機能も提供しているが、検索を対象にしているだけで断片的な更新操作はサポートと述べていない。CODASYL方式のデータベースシステムではいわゆるサグスキーマの定義が利用者の視野を定義する概念に相当するが、ビュー定義機能は極めて小さい。尚、良く知られているようにANSI/X3/SPARCはデータベースシステムアーキテクチャとして内部、概念、外部のスキーマからなる構造を提案しているが、上述ビューサポートの概念はデータベース管理システムが概念スキーマと外部スキーマ間の検索、更新用のインターフェイスを具備することに相当している。

ビューサポートのメカニズムについては1975年頃から研究が開始されてきている。しかしながら数学的平明さを有する関係データベースシステムの枠内でも現在十分なメカニズムの解明がなされているとはいえない。それは本質的に問題解決がシンタクティックな枠組だけで行なえず、必ずセマンティックな枠組を必要としてしるからである。しかしながら、著者は問題解決に次の三ステップの段階を踏むことを提案し、ビューサポートの問題解決を目指している。

(ステップ1) データベースビュー及びそのサポートと17問

かを与えられたデータモデルの枠組の中で明確に定義する。
 (ステップ2) この枠組の中で、どのような機能が十分働き、何がそうでないか、従って何が問題となるのかをシンタクティックかつセマンティックの両側面から明らかにする。

(ステップ3) 問題解決のために何をなすべきかを明確にする。具体的には次の二つになる。

(i) 与えられたデータモデルの枠組でのサポート限界を明示し、この範囲内で完全に機能するメカニズムを構築する。

(ii) (明らかとなった問題点は満足できないものであるとして) 新しいデータモデルを考察し、問題解決をはかる。

本稿ではステップ1及びステップ2の一側面を明らかにすべく関係データモデルの枠組の中で以下若干の議論を展開する。

2. ビューサポート概念

一つの組織体がデータベース構築の対象とした実世界をデータベース実世界(DBRW)と呼ぶ。DBRWのデータ構造をなるべく忠実にデータベースシステム内にとり込みべく通常のデータモデルが使用されてデータベース概念世界(DBCW)が構築される。通常組織体には多数のデータベース利用者がいる。それら各利用者は実世界に対してそれぞれ個

有の世界観に於てビューを持ち、データベース外部世界(DB
EW)を形成する。図-1にビューサポートの概念を示す。

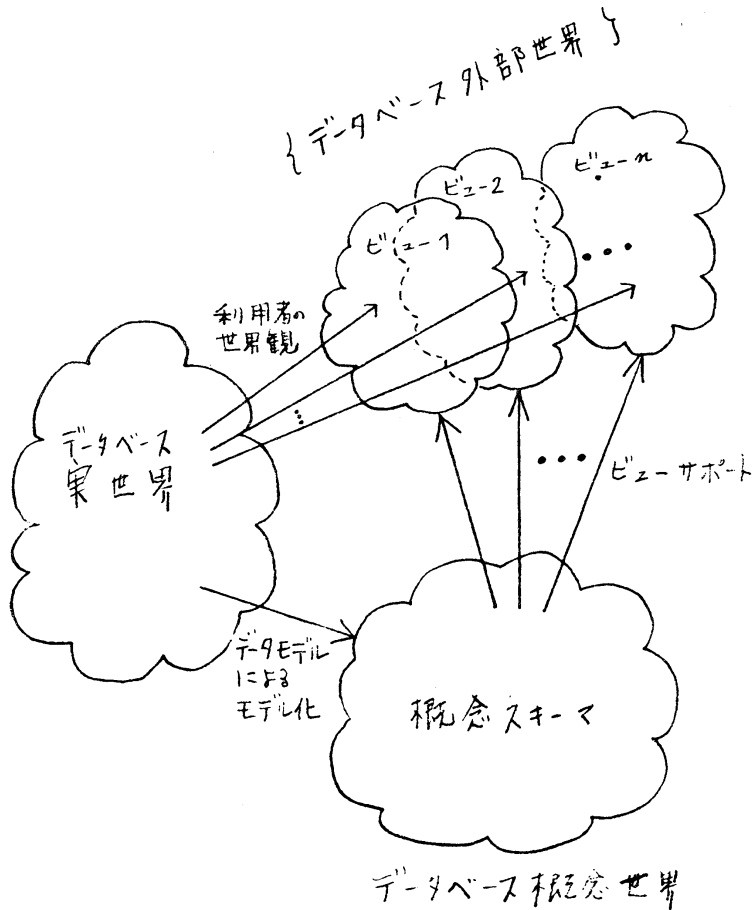


図-1 ビューサポート概念図

3. ビューとビューサポートの基礎的定式化

関係データベースモデルの枠組の中で行なう。 R を関係名,
 A_i ($i = 1, \dots, n$) を属性名, $D_i (= \text{dom}(A_i))$ ($i =$
 $1, \dots, n$) をドメイン名とする。 $R(A_1/D_1, \dots, A_n/D_n)$ を
 フリー関係スキーマという。 二つの定義は次のとおり

である。 $R \triangleq \{ r \mid r \subseteq D_1 \times D_2 \times \cdots \times D_n \}$ 。 $c_j \in 2^{D_1 \times \cdots \times D_n}$ 上の述語, $C_R \triangleq \bigwedge_j c_j$ はリ-関係スキーマ R 上に課せられた保全制約条件とする。 C_R に因り R の関係スキーマ R_C (以後単に関係 R_C という場合も可) は次のように定義する。 $R_C \triangleq \{ r \in R \mid r \text{ は } C_R \text{ を満たす} \}$ 。写像 $u_R : R \rightarrow R$ が関係 r ($\in R_C$) の更新 (挿入, 削除, 書換) 演算であるとは $u_R(r) \in R_C$ が成立するといえる。 $U_R \triangleq \{ u_R : R \rightarrow R \}$ とする。

次に $R_1, \dots, R_m$ はリ-関係スキーマとし, $D = (R_1, \dots, R_m)$ はリ-データベーススキーマということにする。 $c_R \in 2^{\text{dom}(R_1)} \times \cdots \times 2^{\text{dom}(R_m)}$ 上の述語, $C_D \triangleq \bigwedge_R c_R$ とする。 C_D は D 上の保全制約条件である。 $D \triangleq \{ (r_1, \dots, r_m) \mid r_i \in R_i, i=1, \dots, m \}$ であり, $D_C \triangleq \{ (r_1, \dots, r_m) \in D \mid (r_1, \dots, r_m) \text{ は } C_D \text{ を満たす} \}$ とする。写像 $u_D : D \rightarrow D$; $\equiv \equiv u_D \triangleq (u_{R_1}, \dots, u_{R_m})$ で $u_{R_i} : R_i \rightarrow R_i$ (つまり $u_D(r_1, \dots, r_m) \triangleq (u_{R_1}(r_1), \dots, u_{R_m}(r_m))$ である) と定義する, がデータベース d ($\in D_C$) の更新演算であるとは $u_D(d) \in D_C$ が成立するといえる。 $U_D \triangleq \{ u_D : D \rightarrow D \}$, $U_D^* \triangleq \{ 1 \} \cup U_D \cup U_D^2 \cup \cdots$ とする。 $\equiv \equiv \forall d \in D, \Lambda(d) = d$ である。 $U_D^2 = U_D \times U_D$ である。

そこで V を フリー関係スキーム とする。 D_C が V の 部分集合としての ビュー スキーム $\mathcal{A}(D_C)$ を サポート するという概念を明らかに する ためには、ビュー - 定義写像 を $\mathcal{A}: D_C \rightarrow V$ と 定義 する。 $\mathcal{A}$ は D_C から V への 構造的写像 である。 $\mathcal{A}(D_C)$ は D_C 上の $\mathcal{A}$ の もとでの ビュー という こと になる。 写像 $u_v: V \rightarrow V$ が ビュー v ($\in \mathcal{A}(D_C)$) の 更新演算 である といふ $u_v(\alpha) \in \mathcal{A}(D_C)$ が 存在 する こと を いう。 $U_V \equiv \{u_v: \mathcal{A}(D_C) \rightarrow \mathcal{A}(D_C)\}$ と する。

定義 1. $D_C' \subseteq D_C$ と する。 D_C' が $\mathcal{A}(D_C')$ を u_v ($\in \{u_v: \mathcal{A}(D_C') \rightarrow \mathcal{A}(D_C')\}$) について サポート する といふ 条件 が 成立 する こと。

$(\exists! \theta: \{u_v\} \rightarrow U_D^*) ((\forall \alpha \in D_C') (u_v(\mathcal{A}(\alpha)) = \mathcal{A}(\theta(u_v)(\alpha)))$ かつ $(\theta(u_v)(\alpha) \in D_C')$)。

ここには θ を 操作的写像 といふ。 一般に D_C の 部分集合 D_C' の 規定 の 仕方 には 大別 して 二通り の 方法 がある。 一つ は ビュー の 更新操作 に 依存 する 場合 であり、 他 の 一つ は そう ではなく 専ら ビュー の 構造的写像 に 依存 する 場合 がある。

ここ では 二つ 以上 言及 しない。 図-2 に 上の 条件式 の 前半 分 を 図示 している。 後半 分の 条件 は $\theta(u_v)$ が C_D の 更新操作 と なる こと を 要求 している。 $\exists!$ は 唯一 の 存在 を 表わ しているが 二 の 唯一性 は 変換 の 意味論的曖昧さ を 禁止 する こと

と表わしていい。

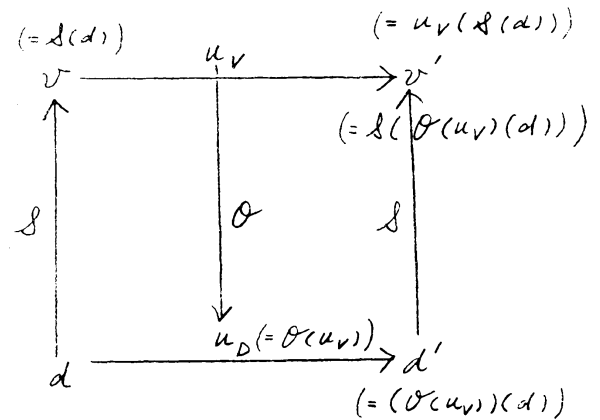


図-2 ビューポートの可換図

$U_{s(D'_c)} \triangleq \{ u_v : s(D'_c) \rightarrow s(D'_c) \}$ と分類する。

$U_{s(D'_c)}^D \triangleq \{ u_v \in U_{s(D'_c)} \mid \forall v \in s(D'_c), |u_v(v)| \leq |v| \},$

$U_{s(D'_c)}^I \triangleq \{ u_v \in U_{s(D'_c)} \mid \forall v \in s(D'_c), |u_v(v)| \geq |v| \},$

$U_{s(D'_c)}^M \triangleq \{ u_v \in U_{s(D'_c)} \mid \forall v \in s(D'_c), |u_v(v)| = |v| \}.$

ここで $|w|$ は関係 w のタプル数を表わす。

定義2 D'_c が $s(D'_c)$ にタプル削除操作によってサポートされることは $U_{s(D'_c)}^D$ の任意の元 u_v について定義1の条件式が成立すること。

タプル挿入、書換操作のサポートについても同様定義する。

定義3 D'_c が $s(D'_c)$ にサポートされることはタプル削除、挿入、書換操作についてサポートされること。

尚詳細な定義関係については機会を更けて述べてみたいと思ふ。ここではこれ以上行わない。

4. ビューサポートの問題点

もっとも基本的なビューの定義概念は INGRES の SYSTEM R がサポートしようとしたものがある。例えば INGRES⁽¹⁾ では QUEL の構文を使って次のようにビューを定義する。

RANGE OF X IS ED

RANGE OF Y IS DM

DEFINE EDM (EMP = X.EMP, DEPT =
X.DEPT, MGR = Y.MGR)

WHERE X.DEPT = Y.DEPT

ここで $ED(EMP, DEPT)$, $DM(DEPT, MGR)$ は基本関係または既に定義されているビューであり、この二つの関係の DEPT 上の自然結合としてビュー EDM が定義された。通常このタイプのビューではビューを定義するのに次の二種類の演算が使用される。

(1) 関係演算 (関係代数の用語では直積, 和, 差, 交り, 射影, 制限, (自然)結合, 割り算等)

(2) 非関係演算 (関係名, 属性名の変換, ドルと円に変換するようなドメインの変換, 更に複雑な関数 (平均, 最大, 最小等のアグリゲート関数とドメイン間の複雑な計算関数など))。

このタイプのビズーについては関係データモデルが世に出
てしばらくしてから、いち早くそのサポート可能性が議論
されることになった。当初のアプロ-4の17の題意があると
直ちにビズー更新と禁止するという考えであったが、最近
データ意味論とともからみ深い考察がなされるようになったと
共に、このようなシンプルな概念のもとに定義されるビズー
についてより問題は極めて複雑であることがわかってきてい
る。二れり議論の歴史的カーベイは本稿では行なわれないが
、意味論とともからみビズーサポートとはこのようなものだとい
うことを示す格好の例を挙げてこの章の使命とした。

関係 $CP (CHILD, PARENT)$ を親子関係を表わ
している関係とする。親あるいは子供が死亡すれば両者間
に関係 CP は消滅するとする。二のと主子供と祖父母の関
係を表わすビズー CG を次の様に定義する。

RANGE OF X IS CP

RANGE OF Y IS CP

DEFINE $CG (CHILD = X.CHILD,$
 $GRANDPARENT = Y.PARENT)$

WHERE $X.PARENT = Y.CHILD$

いま $(a, b), (b, c) \in CP$ であつたとしなう。

ビズー CG の定義から $(a, c) \in CG$ である。さて、

b が死亡したとすると (a, b) 及び (b, c) は CP から削除される。それら二つの対の消滅は自動的にビュースキーム CG からの (a, c) の消滅を誘引する。これは CG が CP と CP の自然結合の射影として定義されているのだから当然である。ビュースキームはそのようなものと理解している限りは問題ない。しかし CG に本来子役と祖父母の関係を表わしている関係だとユーザの見方を押しつくと問題となる。何故ならもし CG が本来の子役と祖父母の関係を表わしている子役 a と c が CG の関係にあることは a の親 b (b は c の子役でもある) の生存、死亡には無関係に維持されねばならぬからである。繰り返すが CG では子役にとって親が生存しているという制限付の子役と祖父母の関係が維持されているということなのである。

以上自然結合の射影として定義されたビュースキームの持つ意味と、それが本来の子役-祖父母の関係を表現していると捉えようとするとは問題点のあることを述べた。しかしながら問題を更に複雑にするのは場合によって自然結合の射影として定義されるビュースキームが逆に本来の姿を捉えていると考えられる世界が存在するからである。次いで定義される従業員と上司の関係を表わすビュースキームを考察する。

RANGE OF X IS ED

RANGE OF Y IS DM

DEFINE EM (EMP = X. EMP, MGR = Y.
MGR)

WHERE X. DEPT = Y. DEPT

このとき $(e, d) \in ED$ か $(d, m) \in DM$ で、結果として $(e, m) \in EM$ であるとしよう。このとき従業員 e とマネージャー m とを結びつけた部局 d が廃部になれば、 e と m との関係 EM が消滅すると考えるのは極めて自然である。

上に示した二つのビュースキーマのシンタクティックな構造は同一である。しかし現実世界の意味論と照らし合せて、ビューとしての制限が前の例では妥当性に欠け、後の例 EM では妥当のようと思われる。このようにビューポートの問題はセマンティクスに深くかかわっている点がある。これらの問題の解決が促されるビューポート機構を実現するうえで望まれていくことになる。

5. おわりに.

ビューポートをみがかしくしているセマンティカルな側面と側面を使ってオ4章で示した。本来ビューはデータ

モデル及びその操作言語の仮想化技法として捉えることが出来、著者は次に示す幾つかのビズーをタリで分類している。⁽⁴⁾

- A. スキーム変換型ビズー
- B. DML変換型ビズー
- C. データモデル変換型ビズー
- D. 抽象化型ビズー
- E. 知識型ビズー
- F. 時変型ビズー

これらのビズーのサポート構造や目的については別の機会に議論を展開した。

[謝辞]

本研究に関連しこれ迄御討論下さった諸氏に深謝する。

尚、本研究の一部文部省昭和55年度科学研究費補助金、課題番号568006の援助のもとで行なわれたことを付記する。

[参考文献]

- (1) M. R. Stonebraker 他, "The Design and Implementation of INGRES," ACM TODS 1, 3 (1976)
- (2) U. Dayal 他, "On the Updatability of Relational Views," Proc. 4th VLDB Conf., p. 368-377 (1978)
- (3) F. Bancilhon, "Supporting View Updates in Relational Data Base," Proc. LFIP TC-2 Working Conf. on Data Base Architecture, p. 198-219 (1979)
- (4) 増永, "仮想化技法としてのデータベースビュース", "情報処理学会データベース管理システム研究会資料 No. 24-2, (1981.3 A19A) .