CHARACTERIZATION OF WELL-BEHAVED DATABASE SCHEMATA AND THEIR UPDATE SEMANTICS

ANNE VERROUST

LRI Université Paris-Sud, Orsay
INRIA, Rocquencourt.

Abstract

A simple interface to deal with a set of facts is presented. The notion of a closed set of facts is defined. The problem of inserting or deleting facts in such a model is investigated. An alternative representation in terms of a unique relation leads to a new definition of decomposition which allows the presence of incomplete information.

I. INTRODUCTION

Consider a set of attributes, for instance

COURSE, STUDENT, GRADE, TEACHER

There is a number of groupings of these attributes that correspond to predicates from real life or to sentences while others do not. We call them objects (cf. Sciore [Sci]).
For instance :

(A) COURSE STUDENT
(B) COURSE STUDENT GRADE
(C) COURSE TEACHER
(D) COURSE STUDENT TEACHER
(E) COURSE STUDENT GRADE TEACHER

corresponding to the following sentenses

(a) student s takes course c
(b) student s got grad g in course c
(c) teacher t teaches course c
(d) student s takes course c from teacher t
(e) student s got grad g in course c taught by teacher t.

Each one of these sentences makes sense by itself i.e. represents facts that do not relate to any other side information. Another way to look at it is to say that each object corresponds to update units in an information system : each fact from (a) to (e) could be "entered" as a new fact in a database.

On the contrary other groupings of attributes do not make sense by themselves.

For instance :

(F) STUDENT GRADE
(G) COURSE GRADE
(H) TEACHER GRADE
(I) TEACHER STUDENT
(J) COURSE GRADE STUDENT

(F) corresponds to :

(f) there exists a course c such that student s got grade g in course c

(G) corresponds to :

(g) there exists a student s that got grade g in course c

(H) corresponds to :

(h) there exists a student s and a course c such that student s got grade g in course c taught by teacher t

(I) corresponds to :

(i) there exists a course c such that student s takes course c from teacher t

(J) corresponds to :

(j) there exists a student s taking course c from teacher t and who got grade g in that course.

The sentences (f) to (j) all derive from facts through the introduction of unknown values : (f) and (g) derive from (b), (h) and (j) from (e), (i) from (d). Thus they cannot represent facts. (We suppose that there is no unknown, or existential value in the facts).

Some of the objects correspond to "atomic" or undecomposable sentences. We call them atoms.

For instance :

(A) COURSE STUDENT
(B) COURSE STUDENT GRADE
(C) COURSE TEACHER

are the atoms

. (D) is not an atom because (d) can be understood as the conjunction of (a) and (c)

(d) is : student s takes course c taught by teacher t
(a) (c)

. It is the same for (E) :
(e) is : student s got grade g in course c
(b)

taught by teacher t
(c)

The facts corresponding to the atoms are called atomic facts.

Given a set of atoms and a set of atomic facts, new facts can be deduced from them through essentially two processes : projection and join.

Deriving facts using projection

The first obvious remark we can make from the examples is that if X is an atom then $X' \subseteq X$ is not necessarily an atom.

For instance COURSE GRADE STUDENT is an atom while COURSE GRADE is not.

It is however possible for both X and X' to be atoms. Consider : COURSE STUDENT and COURSE STUDENT GRADE. In this case, we shall be able to infer new facts using projection.

For instance if
"John got C in Math "

then we can deduce that "John takes Math courses"

Deriving facts using join

One important property of the set of objects is that it is closed under union of non disjoint objects : i.e. if X ans Y are objects and $X \cap Y \neq \emptyset$ then $X \cup Y$ is also an object.

For instance the union of any two among the set of objects (A),(B),(C),(D),(E) produces objects.

Furthermore we can infer new facts using join as follows. If "John takes Maths" and "Martin teaches Maths" then we can deduce that
"John takes Maths from Martin".

Intuitively, we suppose that there exists a sort of "join dependency" between the atoms. Although this point of view is quite restrictive, it certainly corresponds to many real life situations. It also derives from the definition of the notion of object as an information that makes sense by itself.

Given a set of atoms S, we shall say that a set of atomic facts is closed under S if no new atomic fact can be derived from it using the projection and join.

For instance, if we have the set of attributes :

STUDENT,COURSE,TEACHER,DATE

the atoms are

STUDENT,COURSE
STUDENT,COURSE,DATE
TEACHER,COURSE
TEACHER,COURSE,DATE

The following set of atomic facts containing the three facts

(1) "John takes math on Monday"
(2) "Martin teaches maths"
(3) "Smith teaches English on Thursday",

is not closed under S.

By joining (1) and (2) together and projecting the result on TEACHER,COURSE,DATE, we obtain :

(4) "Martin teaches maths on Monday"

which is not in the set.

Now the set of atomic facts consisting of (1), (2), (3) and (4) is closed under S.

Intuitively, given a set of facts, the information it represents is the closure of this set under the projection and join derivation rules. It is therefore important to define properly this closure and to be able to compute it efficiently.

Then one can worry about the updating problem : what happens when we update a set of facts ? If we consider the problem of inserting or deleting a fact in or from a closed set of facts, the result is not closed anymore in general.

For instance, if we insert "Max got C in Algebra" we might have (if it is not already there) to insert the fact "Max takes Algebra".

If we want to delete "Max takes Algebra" we might have to delete also the facts "Max got x in Algebra" that could exist in the set of facts. The problem is then to define and compute the closed set of facts that represent the result of an update.

In this paper we first present our formal framework by defining atomic facts, derivation rules and studying the closure of a set of facts.

Then we study the insertion problem. It is shown that defining the result of an insertion is an easy task, while computing it can be a complex one. We give characterizations, in terms of set of objects, where this computation is easy. These characterizations are presented using the hypergraph formalism.

In section IV we study the deletion problem. It is shown that defining the result is a difficult problem because, in general, deletions lead to ambiguity. We give a complete characterization, in terms of set of atoms, of the cases where the result is uniquely defined and show that, in these cases, the computation is easy.

We also give in APPENDIX an alternative representation in terms of a single relation having nulls. This representation leads us to a new definition of decomposition including relations with nulls.

II. SETS OF FACTS CLOSED UNDER S

Let U be a set of attributes. For each A in U the domain associated with A is denoted D(A).

Let $X \subset U$, a fact on X is a mapping

$x : X \to \bigcup_{A \in X} D(A)$ s.t. $x(A) \in D(A)$ (i.e. a tuple on X)

X is called the definition set of x and denoted $\delta(x)$

We first define the notion of set of objects in this context.

Definition : The set of objects is a covering of U such that if X and Y are objects and $X \cap Y \neq \emptyset$ then $X \cup Y$ is also an object.

The objects correspond to all possible sentences or real life assertions.

Example 1. If U = {COURSE,STUDENT,GRADE,TEACHER} the corresponding set of objects is

COURSE STUDENT
COURSE STUDENT GRADE
COURSE TEACHER
COURSE STUDENT TEACHER
COURSE STUDENT GRADE TEACHER

For a given set of objects $\mathcal{O}$, the database consists of a set of facts defined over the set of objects $\mathcal{O}$.

From two facts x and y, we can deduce a new fact using the "join" operation.

Definition : Let X, Y be in $\mathcal{O}$ s.t. $X \cap Y \neq \emptyset$. Let x, y be facts over X and Y such that

$x_{|X \cap Y} = y_{|X \cap Y}$

Then x and y are compatible and $x * y$ is the fact over $X \cup Y$ defined by $x * y_{|X} = x$ and $x * y_{|Y} = y$.

We extend the join operation to sets of facts. Let $\mathcal{A}$ and $\mathcal{B}$ be two sets of facts.

$$\mathcal{A} * \mathcal{B} : \{x * y \;/\; (x \in \mathcal{A}) \wedge (y \in \mathcal{B}) \wedge (x \text{ and } y \text{ compatible})\}$$

In order to avoid redundancy, we consider a particular subset of the set of objects.

Definition : Let $\mathcal{O}$ be a set of objects. A set of atoms S for $\mathcal{O}$ is the minimal subset of $\mathcal{O}$ such that for each X in $\mathcal{O}$ there is a connected (*) $\mathcal{C} \subseteq S$ verifying $X = \bigcup_{Y \in \mathcal{C}} Y$.

(*) $\mathcal{C}$ is connected if there exists an ordering $X_1 \ldots X_n$ of the elements of $\mathcal{C}$ such that $\forall i \in \{2,..,n\}$ $\exists j < i$ such that $X_i \cap X_j \neq \emptyset$.

It is clear from the definition that for each set of objects $\mathcal{O}$, a set of atoms can be constructed. As a matter of fact, this set is unique.

Example 2. The set of atoms for the set of objects of example 1 is

COURSE STUDENT
COURSE STUDENT GRADE
and
COURSE TEACHER

Given a set of atoms S, an atomic fact x is a fact such that $\delta(x) \in S$.

From now on, the set of atoms S will be fixed unless otherwise mentioned. $\mathcal{A}, \mathcal{B}$ will denote finite sets of atomic facts.

Definition : Let $\mathcal{A}$ be a finite set of atomic facts, $\mathcal{A}^*$ is the limit of the sequence $\mathcal{A}^n$ defined by $\mathcal{A}^1 = \mathcal{A}$ and $\mathcal{A}^{n+1} = \mathcal{A}^n * \mathcal{A}$
(This sequence converges after a finite number of steps since it is non decreasing and has an upper bound).

We can note that :

- $\mathcal{A}^*$ is closed under join operation
- All the facts of $\mathcal{A}^*$ are defined on $\mathcal{O}$.

Definition : Let x and y be facts on X and Y, $x \geq y$ iff $X \supseteq Y$ and $x_{|Y} = y$
(This corresponds to Zaniolo's [Za 2] "≥" or more informative).
Denote $\bar{\mathcal{A}}$ the set of maximal elements for $\leq$ of $\mathcal{A}^*$

$$\bar{\mathcal{A}} = \{x \in \mathcal{A}^* \mid \forall y \in \mathcal{A}^*,\ (y \geq x) \Rightarrow x = y\}$$

Note that :

- $\mathcal{A} \subseteq \mathcal{A}^*$ and $\bar{\mathcal{A}} \subseteq \mathcal{A}^*$
- $\forall x \in \mathcal{A}$, $\exists y \in \bar{\mathcal{A}}$ s.t. $y \geq x$

But in general $\mathcal{A} \not\subseteq \bar{\mathcal{A}}$

Example 3. Let U be {COURSE,STUDENT,GRADE,TEACHER} that we shall abbreviate {C,S,G,T}

S = {(C,S),(C,S,G),(C,T)}

$\mathcal{A}$:

C	S	G	T
Latin	Suzan		
Latin	Paul		
Math	Albert	15	
Math			Martin
Latin			Dupont
English			Smith

$\mathcal{A}^* = \mathcal{A} \cup$ {Latin Suzan Dupont, Latin Paul Albert, Math Albert 15 Martin}

$\bar{\mathcal{A}}$ = {Latin Suzan Dupont, Latin Paul Dupont, Math Albert 15 Martin, English Smith}

In this example $\mathcal{A} \not\subseteq \bar{\mathcal{A}}$ □

Let $\mathcal{A}$ be a set of facts and S the set of atoms. Define the projection of $\mathcal{A}$ on S by :

$$\pi_S(\mathcal{A}) = \{x \mid \delta(x) \in S \wedge (\exists y \in \mathcal{A}, \ y \geq x)\}$$

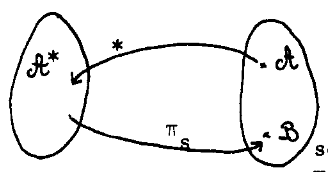
Note that :

1) $\pi_S(\bar{\mathcal{A}}) = \pi_S(\mathcal{A}^*)$ for any set of atomic facts

2) $\pi_S(\mathcal{A}^*) \supseteq \mathcal{A}$ for any set of atomic facts

Thus $\mathcal{A}^*$ is the set of facts implied through join while $\pi_S(\mathcal{A})$ is the set of facts implied through projection. This allows us to define the notion of closure under S.

Definition : Let $\mathcal{A}$ be a finite set of atomic facts. $\mathcal{A}$ is closed under S iff $\pi_S(\mathcal{A}^*) = \mathcal{A}$.
i.e.



$\mathscr{O}$: set of objects

Thus $\mathcal{A}$ is closed under S iff $\mathcal{B} = \mathcal{A}$

III. INSERTING A NEW FACT

Let $\mathcal{A}$ be a set of facts closed under S. Let x be an atomic fact. In general $\mathcal{A} \cup \{x\}$ is not closed under S. Thus we must be able to determine the side effects induced by this insertion. A reasonable definition for the result of this insertion is the smallest set of facts $\mathcal{B}$ closed under S such that $\mathcal{A} \cup \{x\} \subseteq \mathcal{B}$.

Example 4. Let U be {Employee,Department,Manager} that we shall abbreviate {E,D,M} and S be {(E,D), (D,M,(E,M)}
(We suppose that the manager of each department is the manager of all of the employees of that department).

Let $\mathcal{A}$ be :

E	D	M
Albert	Shoes	
Paul	Shoes	
Jane	Books	
Jane		Smith
	Books	Smith

$\mathcal{A}$ is closed under S.

Suppose we want to insert the atomic fact

"Brown is Paul's manager"

then $\mathcal{B} = \mathcal{A} \cup \{\langle \text{Paul,Brown} \rangle\}$ is not closed under S since $\pi_S(\mathcal{B}^*) = \mathcal{B} \cup \{\langle \text{Shoes,Brown} \rangle\} \neq \mathcal{B}$

Theorem : Given a finite set of atomic facts $\mathcal{A}$, there exists a unique minimal $\hat{\mathcal{A}}^S$ s.t.

1) $\mathcal{A} \subseteq \hat{\mathcal{A}}^S$

2) $\hat{\mathcal{A}}^S$ is closed under S

The proof follows from the next theorem.

Theorem : Let $\mathcal{A}$ be a set of facts over S. $\hat{\mathcal{A}}^S$ is the limit of the sequence $\mathcal{A}_n$ defined by :

$$\mathcal{A}_0 = \mathcal{A} \quad \text{and} \quad \mathcal{A}_{n+1} = \pi_S(\mathcal{A}_n^*)$$

Proof. (a) We first prove that $(\mathcal{A}_n)\, n \in \mathbb{N}$ is stationary

- for a given $\mathcal{B}$, $\mathcal{B} \subseteq \mathcal{B}^*$, and π_S is non decreasing thus $(\mathcal{A}_n)\, n \in \mathbb{N}$ is an increasing sequence

- for each attribute A in U and for each integer n, the values of A appearing in the facts of $\mathcal{A}_n$ are in

$$\mathcal{V}(A) = \{x(A) / (x \in \mathcal{A}) \wedge (A \in \delta(x))\}$$

$\mathcal{A}$ is finite, thus $\mathcal{V}(A)$ is finite.
Let $\mathcal{B}$ be the set of facts defined as follows :
For each X in S, $X = (A_1 \ldots A_k)$ such that $\mathcal{V}(A_i) \neq \emptyset$, $\forall$ i.
For each $(a_1 \ldots a_k)$ in $\mathcal{V}(A_1) \times \ldots \times \mathcal{V}(A_k)$

$$x : X \to \bigcup_{A \in X} D(A), \quad A_i \to a_i \qquad \text{is in } \mathcal{B}$$

and that is all.

$\mathcal{B}$ is a finite set of facts such that $\mathcal{A}_n \subseteq \mathcal{B}$ for any integer n.
Thus there exists N such that $\mathcal{A}_N = \mathcal{A}_{N+k}$ $\forall k \in \mathbb{N}$
More specifically $\mathcal{A}_N = \mathcal{A}_{N+1} = \pi_S(\mathcal{A}_N^*)$
Thus $\mathcal{A}_N$ is closed under S.

(b) Let $\mathcal{B}$ be a set of facts closed under S such that $\mathcal{B} \supseteq \mathcal{A}$
We have $\mathcal{B} = \pi_S(\mathcal{B}^*)$ ($\mathcal{B}$ **closed under S**)

$$\supseteq \pi_S(\mathcal{A}^*) \quad (\mathcal{B} \supseteq \mathcal{A})$$

Then by induction on n we have $\mathcal{B} \supseteq \mathcal{A}_n$ $\forall n \in \mathbb{N}$
Thus $\mathcal{B} \supseteq \mathcal{A}_N$ and $\mathcal{A}_N = \hat{\mathcal{A}}^S$ Q.E.D.

Going back to example 4, we have :

$\mathcal{B} = \mathcal{A} \cup \{\langle \text{Paul,Brown} \rangle\}$

$\mathcal{B}_1 = \pi_S(\mathcal{B}^*) = \mathcal{B} \cup \{\langle \text{Shoes,Brown} \rangle\}$

$\mathcal{B}_2 = \pi_S(\mathcal{B}_1^*) = \mathcal{B}_1 \cup \{\langle \text{Albert,Brown} \rangle\}$

$\mathcal{B}_3 = \pi_S(\mathcal{B}_2^*) = \mathcal{B}_2$

Thus $\hat{\mathcal{B}}^S = \mathcal{B}_2$:

E	D	M
Albert	Shoes	
Paul	Shoes	
Albert		Brown
Paul		Brown
	Shoes	Brown
Jane	Books	
Jane		Smith
	Books	Smith

Thus in this example we obtain the result of this insertion after two steps.

In general the sequence $\mathcal{A}_n$ reaches its limit in a finite but unbounded number of steps.

Theorem : If U is infinite or if there exist three attributes in U such that the cardinality of their domain is infinite then for each integer n there exists a set of atoms S_n and a set of atomic facts $\mathcal{A}$ associated such that

$$\mathcal{A}^{S_n} \neq \mathcal{A}_n$$

The proof is given in Appendix 1. From a computational point of view it is important to characterize the cases where the sequence converges in few steps. (Note that the number of steps conditions the number of joins and projections necessary to compute the result).

We shall say that the set of atoms S converges in n steps iff for each set of atomic facts $\mathcal{A}$, $\hat{\mathcal{A}}^S$ is computed at most in n steps (i.e. $\hat{\mathcal{A}}^S = \mathcal{A}_n$ where $\mathcal{A}_0 = \mathcal{A}$ and $\mathcal{A}_{k+1} = \pi_S(\mathcal{A}_k^*)$).

We will give a characterization of this property for n=0 or 1. We identify S with a hypergraph (U,S) where U is the set of nodes and S is the set of edges.

Definition : If $H = (\mathcal{X}, \mathcal{E})$ is a hypergraph and X an element of $\mathcal{E}$, we call minimum precycle description of X a subset C of $\mathcal{E} \setminus \{X\}$ such that :

(i) there exists a numbering $X_1 \ldots X_n$ of the elements of C such that $X_1 \cap (X_i \cup \ldots \cup X_{i-1}) \neq \emptyset$, $\forall i \in \{2,\ldots,n\}$ (C is connected)

(ii) $X \subseteq \bigcup_{Y \in C} Y$

(iii) C is minimum in the following way : every proper subset of C fails to satisfy (i) and (ii).

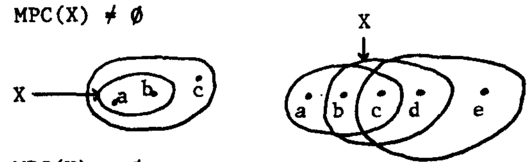
MPC(X) will be the (possibly empty), set of the minimum precycle description of X.

Proposition : S converges in 0 steps iff

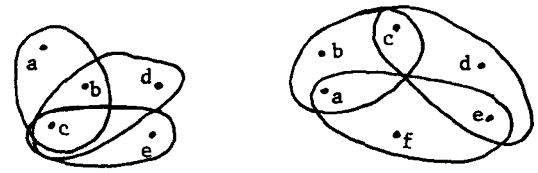
$$\forall x \in S,\ MPC(X) = \emptyset$$

Examples :

. $MPC(X) \neq \emptyset$

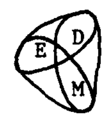


. $MPC(X) = \emptyset,\ \forall X \in S$



In the example 4 : U = {E,D,M}

MPC((E,D)) = {{(M,D),(E,M)}}
MPC((M,D)) = {{(E,D),(E,M)}}
MPC((M,E)) = {{(E,D),(M,D)}}



Proof.

1) Suppose

$$\exists X \in S,\ \exists C \subset S,\ C \in MPC(X)$$

We define $\mathcal{B}$ as follows :

$$x : X \to \bigcup_{A \in X} D(A) \qquad A \to 0 \qquad \text{X is in } \mathcal{B}$$

For Y in C $\quad y_Y : Y \to \bigcup_{A \in Y} D(A) \qquad A \to 1 \qquad y_Y$ is in $\mathcal{B}$

and that is all.

(We have assumed, for the sake of simplicity that each attribute A has {0,1} in its domain) then

$$z : \bigcup_{Y \in C} Y \to \bigcup_{A \in \bigcup_{Y \in C} Y} D(A) \qquad A \to 1 \qquad \text{is in } \mathcal{B}^*$$

by projection on X we obtain :

$$z' : X \to \bigcup_{A \in X} D(A) \qquad A \to 1$$

$$z' \in \pi_S(\mathcal{B}^*) \setminus \mathcal{B}$$

Thus S does not converge in 0 step.

2) Conversely, suppose that $\mathcal{A} \neq \pi_S(\mathcal{A}^*)$ for a set of facts $\mathcal{A}$ over S.
There exists a sequence $t_1 \ldots t_m$ of elements of $\mathcal{A}$ such that

$t \leq (\ldots(t_1 * t_2)\ldots * t_m)$ (1) $(t \in \pi_S(\mathcal{A}^*) \setminus \mathcal{A})$

We take a minimal sequence $t_1 \ldots t_k$ such that (1) is true.
We have

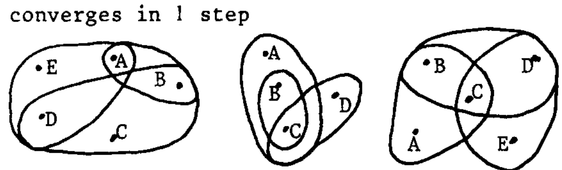
- . $\{X_1 \ldots X_k\}$ satisfy (i) and (ii) since (1) is true
- . $\{X_1 \ldots X_k\}$ satisfy (iii) since $t_1 \ldots t_k$ is minimal.

Thus $\{X_1 \ldots X_k\} \in MPC(X)$ Q.E.D.

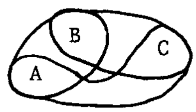
Finally there exists a complete characterization in terms of hypergraph for the property "converges in 1 step" (See [Ver] for full description).

Examples :

. S converges in 1 step



. S does not converge in 1 step



Nota :
When S is tree-structured [Ban]
i.e. S is such that :

- . $\bigcap_{X \in S} X \neq \emptyset$
- . $\forall X \in S$, $\{Y \cap X | Y \in S\}$ forms a chain for the set inclusion

then S converges in 1 step.
Furthermore, given S tree-structured and $\mathcal{A}$ a set of facts over S then

$$\mathcal{A} \text{ is consistent w.r.t. S iff } \pi_S(\mathcal{A}) \subseteq \mathcal{A}$$

In this section we have studied the problem of an insertion in a set of facts. We have shown that the result of an insertion is uniquely defined. We have shown that, in the general case, the number of joins and projections necessary to compute the result is unbounded and we have characterized the sets of atoms where this number is limited to 0 and 1.

IV. THE DELETION PROBLEM

Similarly, given a set of facts $\mathcal{A}$ closed under S and $x \in \mathcal{A}$, in general $\mathcal{A}\setminus\{x\}$ is not closed under S. We want to define the result of the deletion of x by the largest $\mathcal{B}$ closed under S such that $\mathcal{B} \subseteq \mathcal{A}\setminus\{x\}$. Unfortunately :

Theorem : In general, given a set of atomic facts $\mathcal{A}$ there is no unique $\mathcal{B}$ closed under S s.t.

(i) $\mathcal{B} \subseteq \mathcal{A}$

(ii) no proper super set of $\mathcal{B}$ satisfies (i)

Example : Let us take the previous example
S = {ED,DM,ME}

$\mathcal{A}$:

E	D	M
Albert	Shoes	
Paul	Shoes	
Jane	Books	
Jane		Smith
	Books	Smith

If we want to delete "Smith is Jane's manager"

$\mathcal{A}\setminus\{$< Smith,Jane >$\}$ is not closed under S

We have two different solutions for the deletion :

$\mathcal{B}$:

E	D	M
Albert	Shoes	
Paul	Shoes	
Jane	Books	

The effect of the deletion is : Smith is not Jane's manager because he is not the manager of Books department

or $\mathcal{B}'$:

E	D	M
Albert	Shoes	
Paul	Shoes	
	Books	Smith

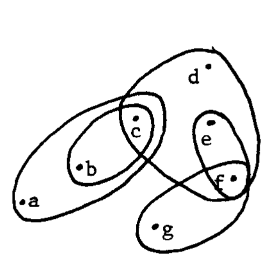
The meaning of the deletion is : Smith is not Jane's manager because Jane is not an employee of the Books department.

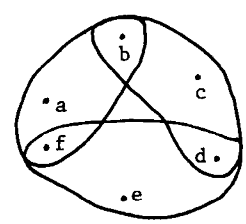
We can characterize the sets of atoms which permit to uniquely define a deletion as follows :

Theorem : Given a set of atoms S, for any set $\mathcal{A}$ closed under S and for any atomic fact x in $\mathcal{A}$, there exists an unique set $\mathcal{B}$ such that $\mathcal{B}$ closed under S, $\mathcal{B} \subseteq \mathcal{A}\setminus\{x\}$, and $\mathcal{B}$ maximum iff :

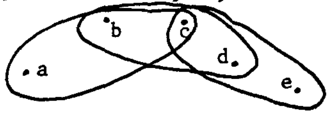
(for any X in S if $MPC(X) \neq \emptyset$ then for any C in MPC(X) $|C| = 1$)   (a)

Examples : S permits to uniquely define a deletion :





S does not permit to uniquely define a deletion :



<u>Proof</u>. 1) Let S be a set of atoms such that (a) is true, $\mathcal{A}$ a set of atomic facts and t an element of $\mathcal{A}$. Let us note that if $\mathfrak{B}$ is closed under S and included in $\mathcal{A}\setminus\{t\}$ then

$\forall\, r \in \mathfrak{B}$ we have $r \not\geq t$.

Let $\mathfrak{B} = \mathcal{A}\setminus\{r \mid r \in \mathcal{A} \text{ and } r \geq t\}$
Then for any $\mathcal{C}$ closed under S and included in $\mathcal{A}\setminus\{t\}$, $\mathcal{C} \subseteq \mathfrak{B}$
Suppose that $\mathfrak{B}$ is not closed under S.
Then there exists r in $\pi_S(\mathfrak{B}^*)\setminus\mathfrak{B}$

So there exists a minimal sequence $t_1 \ldots t_k$ of elements of $\mathfrak{B}$ such that

$r \leq (\ldots(t_1 * t_2) * t_k)$ so $\{\delta(t_1 \ldots \delta(t_k)\} \in MPC(\delta(r))$

S satisfies (a) so k=1.
Then $t_1 \geq t$ and $t_1 \in \mathfrak{B}$ which is impossible.
Thus $\mathfrak{B}$ is the solution for the deletion of t in $\mathcal{A}$.

2) Suppose that S contains X such that $\exists\, C \in MPC(X)$, $|C| > 1$. Consider the set of atomic facts

$$\mathcal{A} = \{t \mid (\delta(t) \subseteq \bigcup_{Y \in C} Y) \wedge \delta(t) \in S \wedge (t(A)=0, \forall\, A \in \delta(t))\}$$

$\mathcal{A}$ is closed under S.
Let s be the element of $\mathcal{A}$ defined on X.
We will show that there exist at most two solutions for the deletion of s in $\mathcal{A}$.

By the definition of MPC(X) there exists a numbering $Y_1 \ldots Y_k$ of the elements of C such that

$Y_i \cap \bigcup_{j<i} Y_j \neq \emptyset$
and, by the minimality of C, $Y_k \cap X\setminus(\bigcup_{j<k} Y_j \cap X) \neq \emptyset$.
Consider the set :

$$I = \{\mathfrak{B} \mid \mathfrak{B} \text{ closed under S}, \{t \mid t \in \mathcal{A} \wedge \delta(t) \in C\setminus\{Y_k\}\} \subseteq \mathfrak{B}, \mathfrak{B} \subseteq \mathcal{A}\setminus\{s,s'\}\}$$

where s' is the element of $\mathcal{A}$ defined on Y_k
I is not empty by the choice of Y_k.
I has at most a maximal element $\mathcal{C}$ for the set inclusion.
$\mathcal{C}$ is also a maximal element of :

$$J = \{\mathfrak{B} \mid \mathfrak{B} \text{ closed under S}, \mathfrak{B} \subseteq \mathcal{A}\setminus\{s\}\}$$

Let $\mathfrak{B} \in J$, $\mathfrak{B} \supseteq \mathcal{C}$.
If $\mathfrak{B} \supset \mathcal{C}$, $\mathfrak{B} \notin I$ thus $s' \in \mathfrak{B}$
Then $\{t \mid t \in \mathcal{A} \wedge \delta(t) \in C\} \subseteq \mathfrak{B}$
Thus $\mathfrak{B} \ni s$ which contradicts the definition of $\mathfrak{B}$
J has another maximal element since

$$\mathfrak{B} = \{t \mid t \in \mathcal{A} \wedge \delta(t) \subseteq Y_k \wedge \delta(t) \in S\}$$

is an element of J, and $\mathfrak{B}$ and $\mathcal{C}$ incomparables.

Q.E.D.

We can now address the problem of computing the result of a deletion when it is uniquely defined.

<u>Theorem</u> : When S permits to define a deletion then, given $\mathcal{A}$ closed under S and x in $\mathcal{A}$, the result of the deletion of t in $\mathcal{A}$ is $\mathfrak{B}$:

$$\mathfrak{B} = \mathcal{A}\setminus\{y \mid y \in \mathcal{A} \text{ and } y \geq x\}$$

<u>Proof.</u> The proof follows directly from the first point of the proof of the previous theorem.

We can note that when S permits to define a deletion, then S converges in 1 step. Thus it allows easy insertions (see [Ver]).

In this section we have studied the problem of a deletion from a set of facts. We have shown that, in the general case, the result of a deletion is not uniquely defined and we have characterized the sets of atoms where the result is unique and shown that in this case the computation is simple.

V. CONCLUSION

We have described a model that represents facts over a set of attributes. We have introduced the notion of set of facts closed under a set of atoms.

In this framework we have investigated the problem of inserting or deleting facts. This led us to the characterization of well behaved sets of atoms in terms of hypergraphs.

It was shown that, in these cases, the insertion and deletion could be :

1) uniquely defined.
2) easily computed.

We have given also an alternative approach of this model in terms of a single relation with null values.

APPENDIX A.

<u>Proof of the theorem</u>

If U is finite or if there exist three attributes in U such that the cardinality of their domain is infinite, then for each integer n there exists a set of atoms S_n and a set of atomic facts associated such that

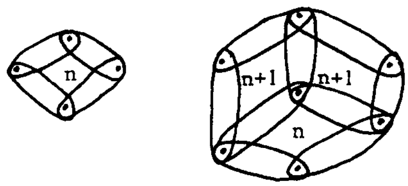
$\mathcal{A}^{S_n} \neq \mathcal{A}_n$

(A) Suppose U infinite

$\forall n \in \mathbb{N}$, we construct S_n recursively as follows :

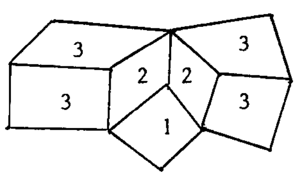
S_1 is 

S_{n+1} is obtained from S_n by transforming each



in adding 3 new vertices and 5 new edges to each 

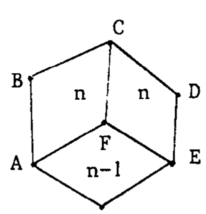
for example



we represent it by a graph since all the edges have the same cardinality = 2.

Given S_n, $\mathcal{A}$ set of atomic facts such that $\mathcal{A}_n$ is not closed under S_n is defined as follows :

. For each couple n-n (when n > 1) :



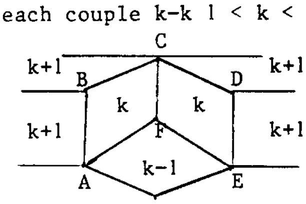
$t_n : \left| \begin{array}{l} B \to 1 \\ A \to 1 \end{array} \right.$ $t'_n : \left| \begin{array}{l} C \to 2 \\ F \to 1 \end{array} \right.$

$r_n : \left| \begin{array}{l} B \to 1 \\ C \to 1 \end{array} \right.$ $r'_n : \left| \begin{array}{l} C \to 2 \\ D \to 2 \end{array} \right.$

$s_n : \left| \begin{array}{l} C \to 1 \\ F \to 1 \end{array} \right.$ $s'_n : \left| \begin{array}{l} D \to 2 \\ E \to 2 \end{array} \right.$

are in $\mathcal{A}$.

. For each couple k-k 1 < k < n :

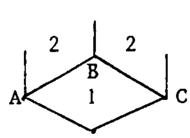


$t_k : \left| \begin{array}{l} C \to 2 \\ F \to 1 \end{array} \right.$

$r_k : \left| \begin{array}{l} C \to 1 \\ F \to 1 \end{array} \right.$

are in $\mathcal{A}$

. For 1 :



$t_1 : \left| \begin{array}{l} C \to 2 \\ D \to 1 \end{array} \right.$

$t_1 \in \mathcal{A}$

Then $t \left| \begin{array}{ll} A \to 1 & t \in \mathcal{A}_{n+1} \\ D \to 1 & t \notin \mathcal{A}_n \end{array} \right.$ □

(B) Suppose that there exist three attributes A, B, C in U such that D(A), D(B) and D(C) are infinite.
We will show that S = {AB,BC,AC} does not converge in n steps for any integer n.

(We assume, for simplicity that D(A), D(B) and D(C) contain $\mathbb{N}$).

Let n be an integer, n > 1. Then the set of atomic facts $\mathcal{A}$ defined as follows is such that $\mathcal{A}_n$ is not closed under S :

on AB, BC, AC | We have $\{(i,j) \mid |i-j| \le 1$ and $\{i,j\} \subset \{1,\ldots 2^{n+1}+1\}\}$

Then $\pi_S(\mathcal{A}^*) = \mathcal{A}_1$ is composed of the facts

on AB, BC and AC | $\{(i,j) \mid |i-j| \le 2$ and $\{i,j\} \subset \{1\ldots 2^{n+1}+1\}\}$

and $\mathcal{A}_k$ is composed of the facts, for any $k \le n+1$

on AB, BC, AC | $\{(i,j) \mid |i-j| \le 2^k$ and $\{i,j\} \subset \{1\ldots 2^{n+1}+1\}\}$

Thus $\mathcal{A}_n \neq \mathcal{A}_{n+1}$ and $\mathcal{A}_n$ is not closed under S.

APPENDIX B

Alternative approach : representing facts by a unique relation.

Given a set of facts $\mathcal{A}$, the information contained in $\mathcal{A}$ is all the possible facts that can be deduced from $\mathcal{A}$ i.e. $\mathcal{A}^*$. It can be represented by $\bar{\mathcal{A}}$, the minimal (in terms of facts) set of facts "information equivalent" (as Zaniolo's [Za 2]) to $\mathcal{A}^*$.

In general, it is not possible to represent $\bar{\mathcal{A}}$ by a simple relation unless we use nulls : a course might have students and no teacher, a student might not yet have a grade. The solutions proposed for the formal treatment of incomplete relations in database operations ([Cod], [Lip], [ImL] [Bis], [Sag], [Vas], [KoU], [Za1], [Za2], ...) show up that the interpretation of null values adds computational complexity and semantic problems. Thus we propose a simple model, following Zaniolo in his representation of nulls.

Given a set of facts $\mathcal{A}$, we associate with it a relation over U as follows :

1) Extend each domain by a special null value :

$\bar{D}(A) = D(A) \cup \{-\}$

'-' means "no information" as Zaniolo [Za2]

2) For each fact x in $\bar{\mathcal{A}}$, extend x to U as follows

$$\bar{x}|_{\delta(x)} = x$$

$$\bar{x}(A) = - \quad \forall A \in U \setminus \delta(x)$$

3) We call this set of tuples Rel($\mathcal{A}$)

Example :

Let $\mathcal{A}$ be :

COURSE	STUDENT	GRADE	TEACHER
French	Suzan		
French	Paul	15	
Math			Martin
French			Dupont

then Rel($\bar{\mathcal{A}}$) is :

COURSE	STUDENT	GRADE	TEACHER
French	Suzan	-	Dupont
French	Paul	15	Dupont
Math	-	-	Martin

□

Conversely, given a relation R with nulls and a set of atoms S we define an associated set of atomic facts $Fact_S(R)$ as follows :

$$Fact_S(R) = \{x \mid \delta(x) \in S \wedge (\forall A \in \delta(x), x(A) \neq -) \wedge (\exists y \in R,\ y \geq x)\}$$

$\geq$ is extended to tuples :

$$x \geq y \text{ iff } x[A] = y[A] \text{ or } y[A] = \text{'-'} \quad \forall A \in U$$

$Fact_S(R)$ is the total projection of R over the set of atoms S. Let R be a relation eventually with null values. We say that R has a lossless decomposition (LLD) with respect to S iff

$$R = Rel(Fact_S(R))$$

Theorem : If R has LLD with respect to S then $Fact_S(R)$ is closed under S.

Proof:

a) Let us consider the composition of operations :

$$Fact_S \circ Rel : \{\text{set of atomic facts}\} \rightarrow \{\text{set of atomic facts}\}$$

To apply Rel to a set of atomic facts

(1) compute $\bar{\mathcal{A}}$

(2) complete eventually the elements of $\mathcal{A}$ by '-' to obtain an inocmplete relation over U

To apply $Fact_S$ to an incomplete relation we :

(3) eliminate the '-' in the tuples

(4) project upon S the set of facts resulting of (3).

Thus we can eliminate (2) and (3) in the process. We have : for any set of atomic facts

$$Fact_S(Rel(\mathcal{A})) = \pi_S(\bar{\mathcal{A}})$$

b) Suppose R has LLD with respect to S

then $Fact_S(R) = Fact_S(Rel(Fact_S(R))$

thus $Fact_S(R) = \pi_S(\overline{Fact_S(R)})$ by (a) Q.E.D.

Nota : The converse is not true.

But : if $\mathcal{A}$ is closed under S then Rel($\mathcal{A}$) has LLD w.r.t. S.

Theorem : If R is total (i.e. without nulls) and has LLD with respect to S then R satisfies the join dependency $\bowtie$ S. The converse is not true.

Example : U = {Name,Child,Car,Registration number, Address}

that we shall abbreviate {N,CH,C,RN,A}

R :

N	CH	C	RN	A
Dupont	Zoe	Renault	XYZ2375	Paris
Dupont	Zoe	Citroen	UU1245	Orléans
Dupont	Zazie	Renault	XYZ2375	Paris
Dupont	Zazie	Citroen	UU1245	Orléans

S = {{N,CH}, {N,C,RN}, {N,A}, {N,C,A}}

i.e. each car has a registration number and is associated with an address.

But : $Fact_S(R)$ = {Dupont Zoe, Dupont Zazie,
Dupont Renault XYZ2375,
Dupont Citroen UU1245,
Dupont Paris, Dupont Orléans,
Dupont Renault Paris,
Dupont Citroen Orléans}

By joining the three facts Dupont Renault XYZ2375, Dupont Orléans and Dupont Zazie, we obtain Dupont Zazie Renault XYZ2375 Orléans which is not in R.

ACKNOWLEDGEMENTS

The author would like to thank Serge Abiteboul, François Bancilhon, Witold Lipski and Nicolas Spyratos for their comments and helpful suggestions.

REFERENCES

[Ban] Bancilhon & al. : "VERSO : A relational Back End Data Base Machine", Proceedings of the International Workshop on Database Machines, San Diego, 1982.

[Bis] Biskup J. : "A Formal Approach to Null Values in Database Relations", formal Bases for Data Bases, Toulouse, Dec. 1979.

[Cod] Codd E.F. : "Extending the Database Relational Model to Capture More Meaning", ACM Transaction on Database System, 4, 1979.

[ImL] Imielinski T., Lipski W. Jr. : "On representing Incomplete Information in a Relational Database", 7th International Conference on Very Large Data Base, Cannes, 1981.

[KoU] Korth H.F. and Ullman J.D. : "System/U : A Database System Based on the Universal Relation Assumption", XP1 Workshop on Relational Database Theory, Stony Brook, June/July 1980.

[Lip] Lipski W. Jr. : "On Database with Incomplete Informations", JACM 28, 1981.

[Sag] Sagiv Y. : "Can we use the universal instance assumption without using nulls ?", 7th ACM-SIGMOD Internationl Conference on Management of Data, 1981.

[Sci] Sciore E. : "The Universal Instance and Database Design", Technical Report 271, Princeton University, 1980.

[Vas] Vassiliou Y. : "A Formal Treatment of Imperfect Information in Database Management" Technical Report CSRG - 123, Toronto, 1980

[Ver] Verroust A. : "Relations incomplètes et décomposition", Séminaire Bases de Données Toulouse, Nov. 1982.

[Za1] Zaniolo C. : "Relational Views in a Database System Support for Queries", IEEE, Computer Society Computer Software and Applications Conference, Chicago, 1977.

[Za2] Zaniolo C. : "Database Relation with Null Values", extended abstract, ACM SIGACT SIGMOD Conference on Principles of Database Systems, 1982.