

# ADQUISICION DE NUEVO CONOCIMIENTO POR COMPARACION Y CONSTRUCCION DE LA JERARQUIA DE CLASES FORMALES

**Henri Briand**

IRESTE-MIAGE, Université de Nantes (Francia)

**José A. Brugos**

Departamento de Matemáticas, Universidad de Oviedo (España)

**Covadonga Fernández**

Departamento de Lenguajes y Sistemas Informáticos e Ingeniería del software,  
Universidad Politécnica de Madrid (España)

**Concepción Pérez-Llera**

Departamento de Matemáticas, Universidad de Oviedo (España)

**Rafael Portaencasa Baeza**

Departamento de Lenguajes y Sistemas Informáticos e Ingeniería del software,  
Universidad Politécnica de Madrid (España)

## Resumen

En este trabajo se construye a través de un proceso basado en comparaciones sobre tripletas de tipos de objetos obtenidos a partir de los hechos elementales (objetos y características) que establece el experto, el primer nivel de la jerarquía de Generalización /Especialización, que denominamos Jerarquía de Clases Formales. Este nivel clasifica los tipos de objeto del Universo del Discurso y genera una partición de las características obligatorias de los tipos de objeto en clases de equivalencia por niveles.

**Palabras Clave:** Jerarquía de Generalización/Especialización, Característica Obligatoria.

## 1.- Introducción

La utilización de S.G.B.D. orientadas a objeto supone un importante avance pero no puede ser un fin en sí misma. La etapa siguiente de la evolución de los S.G.B.D. consiste en unificar las Bases de Datos y las Bases de Conocimiento en un mismo sistema. Tales S.G.B.C. han de ser adaptados tanto para gestionar datos clásicos como los conocimientos u objetos complejos [9].

Se han realizado numerosos esfuerzos en el campo de las bases de datos en la línea de mejorarlas semánticamente para que ofrezcan mayor flexibilidad y sobre todo en la de poder incorporarlas en formalismos del campo de la I.A. como red semántica [4], [5], [2], 'frame' [1] modelo particional [6].

El esquema de representación del conocimiento que utilizamos como base de nuestro formalismo es el modelo O.A.S. (Objeto-Asociación Semántica) [11], muy cercano al Modelo Entidad-Relación Binario [10] pero donde se aporta una nueva visión de la definición convencional de asociación. La asociación la planteamos como una fusión de esquemas de características que independiza el hecho de que un objeto posea una característica del hecho de que ésta tenga asociado un valor, con lo cual también podemos independizar las distintas evaluaciones que pueda tener una característica. Tal planteamiento constituye a nuestro entender un punto de aproximación muy importante de cara a la integración de los modelos conceptuales de B.D. en el campo de la representación del conocimiento.

En la sección 2 se plantean dos conceptos básicos del modelo O.A.S. que son el esquema de características de un tipo de objeto y tipo de asociación de esquemas de características.

En la sección 3 se esboza la forma de obtener, lo que llamamos, tipos de objeto primitivos y sus asociaciones irreducibles, a partir del estudio de los hechos elementales.

En la sección 4 se ofrece el método para adquirir nuevo conocimiento, a partir de la lista de tipos de objeto primitivos, por comparaciones sobre tripletas de tipos de objeto, a la vez que se construye la jerarquía de Clases Formales.

## 2.- Esquema de Característica y Tipo de Asociación entre Esquemas de Característica.

Un hecho elemental es una aseercción simple sobre el Universo del Discurso (U.D.). El U.D. es visto como un conjunto de objetos que cumplen unos predicados. Los hechos elementales son aseercciones sobre qué objetos particulares cumplen qué predicados particulares. La clase más simple de hecho elemental afirma que un objeto cumple un predicado unario (juega un papel [I0]. Definimos esquema de característica,  $C(O)$ , por un nombre de predicado unario  $C$  (por ej. Tiene\_olor) y un tipo de objeto  $O$  (por ej. flor), cuya representación gráfica será,

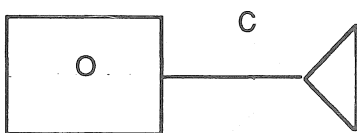


Fig. 1

Una característica  $c$  sobre el esquema de característica  $C(O)$  es un conjunto finito de objetos  $c = \{O_j / O_j \in c, O_j \in O\}$  que tienen un cierto significado definido por el cumplimiento del predicado  $C$ , por ejemplo  $c = \{\text{rosa, clavel, jacinto, ...}\}$ .

Una tupla de característica es una instanciación de la característica sobre un objeto  $O_j$  particular, por ejemplo Tiene\_olor (rosa).

Según la representación gráfica al predicado  $C$  le damos la dimensión del papel del modelo E-R (papel E-R), aunque este modelo asocia el concepto de papel a la función que los objetos juegan dentro de una asociación, con lo que aquí se le da un sentido más general cercano a la línea del modelo NIAM (papel NIAM); y esto, lo diferenciamos manejando, en nuestro modelo, los dos conceptos: esquema de característica y papel E-R.

Un esquema de característica  $C(O)$  puede ser obligatorio u opcional, dependiendo de que a todos los objetos, registrados en la B.C., del tipo de objeto  $O$  les exijamos su cumplimiento o no.

Una tupla de característica puede tener una evaluación, a través de una asociación con otra tupla de característica correspondiente a otro esquema de característica.

La diferencia entre esquema de característica y papel E-R o NIAM, radica en que el esquema de característica es interpretado sobre la población del tipo de objeto, y el papel E-R y el papel NIAM, en el caso de una asociación, son interpretados sobre la población de la asociación donde participa ese tipo de objeto a través de un esquema de

característica. Con respecto a la obligatoriedad del esquema de característica/papel, la relación entre ellos es:

"Si un esquema de característica es opcional, entonces el papel es opcional pero no a la inversa".

"Si un papel es obligatorio, entonces el esquema de característica es obligatorio, pero no a la inversa"

Por ejemplo, puede ser que el tipo de objeto Flor posea obligatoriamente la característica de Tener\_olor, sin embargo eso no significa que el papel Tener\_olor sea obligatorio, es decir que se registre obligatoriamente en la B.C. un olor determinado para cada flor, porque puede haber razones para hacerlo opcional como, por ejemplo, no conocer los olores de algunas flores.

La asociación entre tipos de objeto la planteamos como una fusión de esquemas de características, así, por ejemplo, un tipo de asociación binaria entre esquemas de características se define como un conjunto de pares de objetos (i,j) de dos características  $c_i, c_j$  sobre los esquemas de características  $C_i, C_j$  respectivamente. Su representación gráfica será:

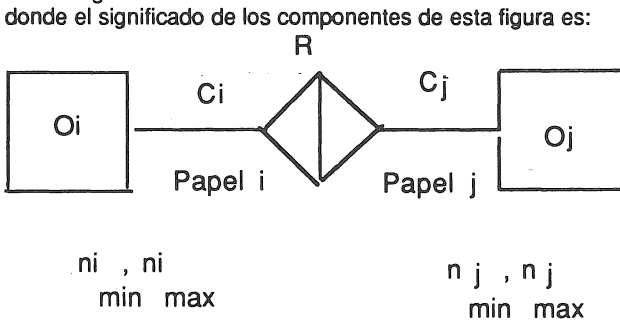


Fig. 2

$O_i$  : Conjunto de objetos sobre el que se define el predicado  $C_i$

$O_j$  : Conjunto de objetos sobre el que se define el predicado  $C_j$

$R$  :Nombre del predicado que cumplen las parejas de objetos asociados (i,j) que pertenecen a las características  $c_i, c_j$  respectivamente; o simplemente un identificador de la asociación.

$n_i$  : Cardinalidad del segmento i (número de elementos de  $R$  asociados a un dado de  $C_i$  ( $O_i$ )) puede ser máxima o mínima.

$C_i, \text{ papel } i$  : Característica y papel (E.-R) que poseen los objeto de  $O_i$  que se asocian.

Cuando la cardinalidad mínima del segmento  $i$  es 1 se dice que el esquema de característica es obligatorio, en el caso de que sea 0 se dice que es opcional.

La diferencia entre  $C_i$  y el papel  $i$  (E-R o NIAM) radica en que restricción de papel obligatorio es interpretada en términos de la población del tipo de objeto y no en términos de la población de la asociación.

Decimos que  $R$  tiene papeles  $y$ , que  $O$  tiene características. Las características funcionan como papeles de la asociación, cuando se evalúan.

$C_j$ , papel  $j$ : Característica y papel (E-R) que poseen los objetos de  $O_j$  que se asocian.

Lo interesante de esta definición de asociación, es que aquí, la asociación la planteamos como una evaluación de una característica, con lo que podemos independizar el hecho de que un objeto tenga una característica, como por ejemplo tener sabor, del hecho de que esa característica pueda tener un valor, como por ejemplo sabor dulce, y además podemos independizar las distintas evaluaciones que puede tener una característica. Esto es sumamente interesante de cara a la integración de los modelos conceptuales de B.D. en representación del conocimiento. Lo anterior puede ser resumido en la siguiente figura, donde se exponen las distintas tablas de instancias de los hechos que queremos diferenciar siendo éstos, por una parte, las tuplas de la característica y, por otra las tuplas de la asociación (evaluación):

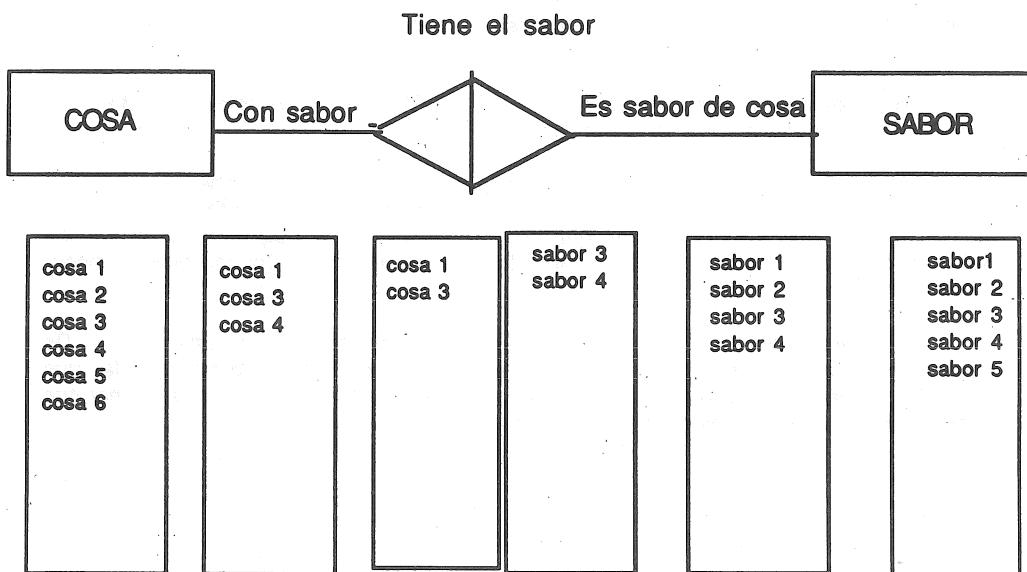


Fig. 3

### **3. Obtención de los Tipos de Objeto Primitivos y sus asociaciones Irreducibles**

La fase inicial de adquisición de conocimiento consiste en la recolección de los tipos de objeto y sus características y/o asociaciones, definidas por el experto irreducibles (aquellas que no se pueden descomponer en asociaciones más simples sin pérdidas de semántica), es decir de hechos elementales ligados a sentencias elementales en lenguaje natural, informes, documentos, ejemplos familiares,...

Nuestro interés básico es dividir el U.D., inicialmente de tal forma que cada objeto sea clasificado en exactamente un tipo de objeto. Por eso al seleccionar los tipos de objeto debemos asegurarnos que sean mutuamente exclusivos, esto es, que no tengan instancias en común, combinando tipos de objeto que se solapan en un único tipo de objeto.

Para asegurarnos de que los tipos de asociaciones sean irreducibles nos apoyaremos en métodos formales basados en las restricciones o condiciones de integridad estáticas sobre esos hechos como son las cardinalidades máxima y mínima de un tipo de asociación (de donde se obtienen las Dependencias Funcionales), las Dependencias Multivaluadas y las Dependencias de Unión, y en el proceso de normalización del Modelo Relacional hasta la 5 FN.

### **4. Adquisición de nuevo conocimiento y construcción de la Jerarquía de Clases Formales.**

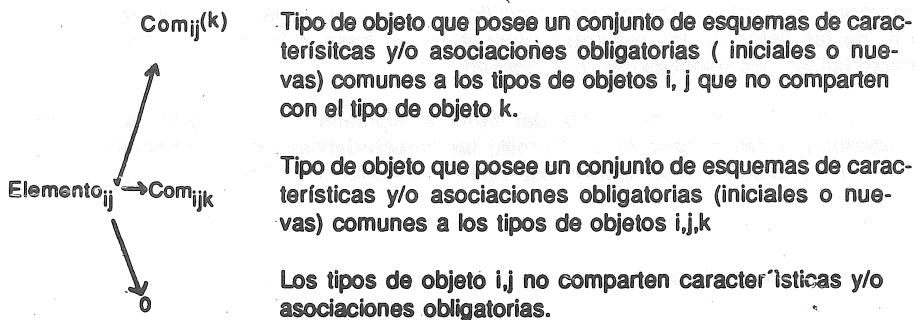
Los tipos de objetos primitivos obtenidos a partir del estudio de los hechos elementales se colocan en la lista LISTA\_OBJETOS\_PRIMITIVOS, junto con sus características (hechos elementales unarios) y asociaciones (hechos elementales n-arios) iniciales.

Los tipos de objetos de LISTA\_OBJETOS\_PRIMITIVOS se colocan en tripletas (Obj,Obj,Obj) y el experto ha de establecer:

a) características y/o asociaciones que compartan dos de los objetos (i,j) y que los diferencien del tercero (k). De entre estas características y/o asociaciones (que pueden ser iniciales o nuevas), las que sean obligatorias, se asignarán a un tipo de objeto al que se le suministrará un nombre formal:  $comij(k)$ , que además puede poseer un nombre natural.

b) características y/o asociaciones que compartan los tipos de objetos i,j,k. De entre estas características y/o asociaciones (que pueden ser iniciales o nuevas), las que sean obligatorias, se asignarán a un tipo de objeto al que se le suministrará un nombre formal:  $Comijk$ , que además puede poseer un nombre natural

	Ob <sub>k</sub>	Ob <sub>1</sub>	Ob <sub>2</sub>	.....	Ob <sub>i</sub> .....	Ob <sub>k-1</sub>	Ob <sub>k+1</sub>
Ob <sub>1</sub>		0					
Ob <sub>2</sub>			0				
Ob <sub>i</sub>					[Com <sub>ij</sub> (k), Com <sub>ijk</sub> ]		
Ob <sub>k-1</sub>						0	
Ob <sub>k+1</sub>							0



**Fig. 4**

Si Com<sub>ij</sub>(k) en elem<sub>ij</sub> es distinto de ∅ entonces las características y/o asociaciones que comparten los tipos de objetos i,j (que les diferencien del tipo de objeto k) se eliminan de éstos y se crean relaciones hijo-padre entre ellos y Com<sub>ij</sub>(k).

Si Com<sub>ijk</sub> en elem<sub>ij</sub> es distinto de 0 entonces las características y/o asociaciones que comparten los tipos de objetos i,j,k se eliminan de éstos y se crean relaciones hijo-padre entre ellos y com<sub>ijk</sub>.

Si se dispone de solamente dos tipos de objeto primitivos es obvia la generación de niveles, pues sólo se solicitarían al experto las características y/o asociaciones que comparten, las cuales se almacenarían en Com<sub>ij</sub>.

En vez de manejar k tablas (una por cada k), donde cada elemento de cada una de las tablas posea dos valores, lo que hacemos es almacenar en elem<sub>ij</sub> una lista de (k-2)\*2 elementos.

El propósito de esta técnica es utilizar las características y/o asociaciones que utiliza el experto para discutir los conceptos del dominio del problema y diferenciar objetos, por tanto clasificarlos. El proceso de construcción de la jerarquía clases formales parte de LISTA\_OBJETOS\_PRIMITIVOS y está basado en la aplicación del siguiente procedimiento recursivo que genera los niveles de la jerarquía.

**GENERA\_NIVEL** (L: lista\_objetos)

**PARA CADA** pareja de tipos de objetos  $i, j$  distintos ( $j > i$ ) **HACER**

**PARA CADA**  $k \neq i, j$  **HACER**

**SI** los objetos  $i, j$  poseen nombres naturales

**ENTONCES** Generar  $Comij(k)$  y  $Comijk$  de la forma expresada más arriba.

**SI NO** Generar  $Comij(k)$  y  $Comijk$  por simple comparación conjuntista entre los conjuntos de características y/o asociaciones de los objetos  $i, j, k$ ;

**FIN\_DE\_HACER**

**PARA CADA** elem  $ij < >$  lista-vacía **HACER**

Renombrar a nombres únicos los  $Comij(k)$  y  $Comijk$  en la lista del elemento  $elemij$  a partir de las listas de los restantes elementos, en base a la igualdad del conjunto de características y/o asociaciones que posean;

Construir a partir de la lista del  $elemij$  el siguiente nivel del grafo eliminando en los objetos  $i, j, k$  (en el caso de un  $Comijk$ ) las características y/o asociaciones que estén en los  $Com$  y creando una relación hijo-padre entre ellos y el  $Com$ ;

**FIN\_DE\_HACER**

lista  $\leftarrow$  unión de las listas de los elementos  $elemij$ ;

**SI** lista no está vacía **ENTONCES** ejecutar **GENERA\_NIVEL** (lista)

**SI NO** acabar;

El resultado es una red, que llamaremos *grafo de clases formales* que se corresponde con una jerarquía de generalización, de la forma:

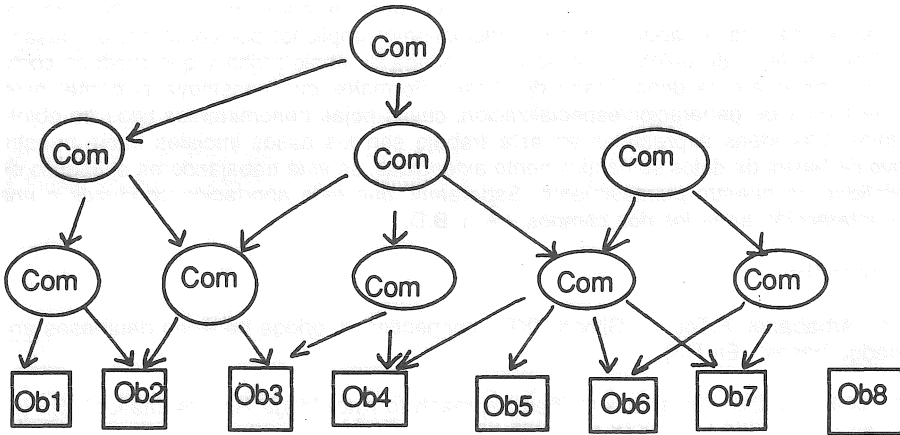


Fig. 5

Las hojas son los tipos de objeto primitivos, y se derivan utilizando el principio de predicabilidad [12], según el cual clases de cosas diferentes tienen clases de predicados distintos, y, de este modo, las cosas pueden ser clasificadas de acuerdo con los predicados que se les aplican. Los predicados pueden estar ligados con las categorías a varios niveles dentro del grafo, y una vez que el predicado está ligado a una categoría se aplica a todas las categorías dominadas por ella.

Para esta red se cumple que los nodos y las hojas son disjuntos dos a dos, en lo que respecta sus conjuntos de asociaciones y/o características obligatorias, es decir, se ha realizado una partición del conjunto global de características y/o asociaciones obligatorias. Este conjunto estará formado por el conjunto de características y/o asociaciones iniciales más las características y/o asociaciones nuevas que han podido ir apareciendo en el proceso anterior basado en tripletas.

Las flechas que llegan a un objeto/clase representan las herencias de características y/o asociaciones que le corresponden.

### Conclusiones y futuros trabajos

En este trabajo se plantea la asociación como una fusión de esquemas de características, aportando una nueva visión de la definición convencional de asociación del modelo Entidad-Relación. Tal planteamiento, el cual es básico en el modelo (O.A.S.) sobre el que subyace este concepto, constituye un punto de aproximación importante en lo que respecta a la integración de los modelos conceptuales de B.D. en el campo de la representación del conocimiento.

Desde la concepción anterior, se ofrece una metodología que proporciona al diseñador la manera de adquirir nuevo conocimiento (implícito) por comparación basada en triplas de tipos de objeto, proceso que procede de abajo/arriba y que produce como resultado una jerarquía denominada de Clases Formales que constituye el primer nivel de la Jerarquía de generación/especialización, cuyas hojas denominamos tipos de objeto primitivos. Las ideas expresadas en este trabajo son los pasos iniciales hacia nuestro objetivo de bases de datos semánticamente extendidas, se está trabajando en el estudio de la viabilidad de nuestro planteamiento. Esperamos que esta aportación conduzca a una mayor interacción entre los dos campos, I.A. y B.D.

## **Bibliografía**

- [1] R.M. Arbabanel, F.Tou, Y. Gilbert. "KEE connection: A bridge between databases and knowledge bases". En [13].
- [2] Addanki "KL-DB: Towards a Unified Approach to Knowledge Representation". En [8].
- [3] M.L. Brodie, J. Mylopoulos (Eds.). "On Knowledge Base Management Systems: Integrating Artificial Intelligence and Database Technologies". Springer-Verlag, 1986.
- [4] H. Briand. "Un modèle étendu entité-association et sa représentation dans un base de faits d'un système expert". En [7].
- [5] H. Briand. "Translation from ad Entity-Relationship diagram into an object oriented database". TOOLS, Paris 1990.
- [6] C. Fernández Baizán. "Expresión de Dependencias en el Modelo Particional: Análisis de sus propiedades". Actas del IV Congreso de Lenguajes Naturales y Lenguajes Formales, pp. 579-594. Barcelona 1989.
- [7] Actas Congreso INFORSID, Toulouse 1985.
- [8] R.A.Meersman, A.C. Sernadas (eds.). "Data and Knowledge (DS-2)". North-Holland, 1988.
- [9] J. Mylopoulos. "On Knowledge Base Management Systems" en [3]
- [10] G.M. Nijssen, T.A. Halpin. "Conceptual Schema and Relational Database Design". Prentice-Hall, 1989.
- [11] M.C. Pérez Llera. "Metodología para la obtención del modelo O.A.S.". Tesis Universidad de Oviedo, 1991.
- [12] L. Shastri. "Semantic Networks: A formalization". Artificial Intelligence Vol. 39 pp. 283-355, 1989.
- [13] M. Yazdani, M. Richer (Eds.) "AI Tolls and Techniques". Norwood 1988