

SOBRE LA POTENCIA DE DERIVACIÓN DE ESQUEMAS EXTERNOS A PARTIR DEL ESQUEMA DE BASE DE DATOS

Saltor, Félix

Facultat d'Informàtica de la U.P.C. 08034 BARCELONA

Resumen. Esta comunicación investiga los esquemas externos que podrían en general derivarse de un esquema de base de datos, y, utilizando de un modo sistemático como metamodelo un modelo ERA (Entity Relationship Attribute) ampliado con subtipos, establece una clasificación de ellos en cuatro clases, con las correspondientes subclases. Esta clasificación puede ser utilizada para analizar la potencia de derivación de esquemas externos de un determinado modelo de datos, a base de evaluar cuáles de las clases y subclases son obtenibles en el modelo en cuestión. En esta comunicación se aplica este método a los modelos más representativos de 3 familias de modelos: en red, jerárquicos, y relacionales.

INTRODUCCIÓN

Entre los aspectos a estudiar de todo modelo de (base de) datos es importante su capacidad de derivar esquemas externos a partir del esquema de base de datos ("esquema de base de datos" según la terminología ISO, por desdoblamiento del "esquema conceptual" de la terminología ANSI/SPARC). Aunque la información que se obtenga utilizando un esquema externo deba necesariamente derivarse de la información en la base de datos (a nivel "extensión"), esto no implica que el esquema externo deba ser un subconjunto del esquema de base de datos (a nivel "intensión").

Para analizar sistemáticamente los modelos externos, utilizaremos como meta modelo el siguiente modelo de la familia ERA (Entity Relationship Attribute, es decir, Entidad **Interrelación** Atributo): las interrelaciones pueden ser n-arias y tener atributos; los valores de los atributos son simples y no compuestos, únicos y no múltiples, y pueden ser nulos. El modelo se amplía con la generalización: un tipo de entidad puede ser un subtipo de otro (supertipo). Este modelo corresponde, en la clasificación de modelos E-R de Chen (81), a la categoría GERM, subcategoría 1.

Este metamodelo nos permitirá clasificar las transformaciones del esquema de base de datos (EBD en lo sucesivo) en esquemas externos (EE) en cuatro clases, cada una de ellas con sus subclases. Esta clasificación podrá ser utilizada para analizar la potencia de un determinado modelo de datos, a base de evaluar cuáles de las clases y subclases son obtenibles en el modelo en cuestión. Podrán también compararse mo-

delos de datos, analizando si uno de ellos permite derivar clases o subclases de EEs que no son derivables en el otro modelo.

Aplicaremos este método a los siguientes tres modelos:

- un miembro de la familia de modelos en red, las especificaciones CODASYL (Olle,78);
- un miembro de la familia de modelos jerárquicos, el modelo de IMS y DL/I (IBM.84), al que llamaremos "modelo de doble arborescencia" de acuerdo con (Saltor,76); y
- un miembro de la familia de modelos relacionales, el modelo relacional básico (Codd,70), con nulos y "vistas" (esquemas externos).

Nótese que consideramos que en red, jerárquico y relacional no son modelos sino familias de modelos, y que hemos escogido el modelo más conocido y usado de cada familia. Para más detalles ver (Saltor,84). El método se ha aplicado a un modelo E-R equipado con el álgebra de Borgoña en (Saltor,85).

CLASIFICACIÓN

Esquemas externos subconjuntos del esquema de base de datos.

La primera clase está formada por los EEs que son subconjuntos estrictos del EBD, es decir:

1. cada tipo de entidad ha de ser un tipo de entidad del EBD, con el mismo significado;
2. cada tipo de interrelación ha de ser un tipo de interrelación del

"trampa de las conexiones".

Interrelaciones reflexivas: Llamaremos reflexivo a un tipo de interrelación binario en que los dos tipos de entidad interrelacionadas son el mismo, por ejemplo ES HIJO DE entre PERSONA y PERSONA. Mediante la combinación seguida de la proyección obtendremos ES NIETO DE, ES BIZNIETO DE, etc. La operación de unión de todos ellos dará el cierre transitivo de la interrelación reflexiva, en el ejemplo ES DESCENDIENTE DE.

Interrelaciones agregadas: Se obtienen análogamente a la agregación de atributos.

Interrelaciones restringidas: Un EE podrá tener un tipo de interrelación R2 derivado de un tipo R1 del EBD, con las ocurrencias de R2 restringidas a aquellas ocurrencias de R1 cuyos atributos satisfagan una condición.

Aplicación a los tres modelos. Las especificaciones CODASYL no soportan nuevas interrelaciones. El modelo de doble arborescencia soporta sólo dos casos: la agregación de una interrelación 1 a n (padre a hijo) en una 1 a 1 mediante la inversión de estructuras, y la combinación (en particular de interrelaciones reflexivas, tan importantes en empresas manufactureras).

El modelo relacional básico con nulos y vistas soporta todas las subclases de nuevas interrelaciones, excepto el cierre transitivo de una interrelación reflexiva.

Esquemas externos con nuevos tipos de entidad

Levantamos también la primera de las restricciones y obtenemos la cuarta clase de EEs, en que aparecen tipos de entidad que no figuran en el EBD.

Unión de tipos de entidad: De dos (o más) tipos de entidad A y B del EBD puede derivarse un tipo de entidad externo AOB por unión. Podrá participar en tipos de interrelación derivados de los B, y de sus uniones.

Restricción de tipos de entidad: De un tipo de entidad del EBD tal como EMPLEADO podemos derivar un tipo de entidad externo EMPLEADO EN LA DIVISION 200 por restricción. Es un tipo nuevo, ya que no existe con este significado en el EBD, aunque sus ocurrencias sean subconjunto de las ocurrencias de EMPLEADO.

Agregación de tipos de entidad: De un tipo de entidad del EBD tal como FECHA se obtiene por agregación en un EE el

tipo de entidad MES. Si VENTAS DEL DIA interrelaciona FECHA y PRODUCTO, su agregación producirá VENTAS DEL MES entre MES y PRODUCTO.

Paso de interrelaciones a entidades: Un tipo de interrelación del EBD puede aparecer en un EE como tipo de entidad. Es el único modo de disponer de sus ocurrencias sin tener que incluir los tipos de entidad que interrelaciona.

Aplicación a los tres modelos. Las especificaciones CODASYL y el modelo de doble arborescencia no soportan nuevos tipos de entidad. El modelo relacional básico con nulos y vistas soporta todas las subclases de EEs de esta cuarta clase.

CONCLUSIONES

Es posible clasificar los esquemas externos derivables del esquema de base de datos en cuatro clases:

- . subconjuntos estrictos del esquema de base de datos;
- . con nuevos atributos, que no existen en el esquema de base de datos;
- . con nuevos tipos de interrelaciones; y
- . con nuevos tipos de entidad.

Cada clase consta de varias subclases (obtenidas por herencia, por agregación, por cálculo, etc.).

Esta clasificación permite evaluar la potencia de derivación de esquemas externos de un determinado modelo, analizando cuáles de las clases y subclases permite derivar.

Hemos aplicado este método al modelo más representativo de la familia de modelos en red, las especificaciones CODASYL, y al modelo más representativo de la familia de modelos jerárquicos, el modelo de doble arborescencia. Se observa que el segundo es algo superior al primero en este aspecto de potencia de derivación de esquemas externos.

El modelo más representativo de la familia de modelos relacionales, el modelo relacional básico con nulos y vistas, permite derivar todas las clases y subclases de esta clasificación (excepto el cierre transitivo de una interrelación reflexiva), y es pues considerablemente más potente que los dos anteriores en este aspecto.

Por otra parte, el hecho de que sea posible derivar esquemas externos que no son subconjuntos del esquema de base de datos hace totalmente inadecuado el uso del término "subesquema" para refe

- EBD, con el mismo significado, y los tipos de entidad que interrelacionan han de aparecer en el EE; y
3. cada atributo ha de ser el mismo atributo del correspondiente tipo (de entidad o de interrelación) del EBD.

Aplicación a los tres modelos. Esta clase de EEs está soportada por los tres modelos. Sin embargo, el concepto de subtipo no está directamente soportado por ninguno de ellos.

Esquemas externos con nuevos atributos

En esta segunda clase levantamos la tercera restricción de la clase anterior, y permitimos que un tipo externo tenga atributos que no tiene el correspondiente tipo en el EBD. Tiene varias subclases.

Atributos heredados: Un tipo de entidad del EE puede tener atributos que en el EBD son atributos de un supertipo del tipo correspondiente. También puede heredar atributos de un subtipo, con valores nulos en su caso. Un tipo de interrelación puede heredar atributos de los tipos de entidad que interrelaciona. La recíproca no está permitida directamente, pero es posible la agregación.

Atributos agregados: Varias ocurrencias de un mismo atributo en el EBD pueden agregarse para derivar una ocurrencia de un nuevo atributo de un EE. Un tipo de entidad de un EE puede tener un atributo obtenido por agregación de un atributo de un tipo de interrelación del EBD en que participa el correspondiente tipo de entidad. También puede derivarse un atributo por agregación de un atributo de un tipo de entidad interrelacionado.

Atributos calculados: A partir de dos o más atributos del EBD puede calcularse un nuevo atributo de un EE. También pueden intervenir en el cálculo atributos heredados o agregados.

Atributos restringidos: Restringiendo los valores de un atributo que pueden aparecer en un EE, y asignando el valor nulo a los otros, obtenemos los atributos restringidos (por ejemplo, ver todos los salarios inferiores a un millón, obteniendo el valor nulo para todos los otros empleados).

Aplicación a los tres modelos: Las especificaciones CODASYL no soportan nuevos atributos. El modelo de doble arborescencia no soporta, en general, nuevos atributos. Sólo si un subtipo se representa como hijo de su supertipo, y dicho hijo tiene otro padre, un EE en que el hijo aparezca como hijo del segundo padre puede heredar atributos

del supertipo.

El modelo relacional básico con nulos y vistas soporta la herencia de atributos mediante la combinación ("join") y la restricción mediante la selección o restricción, y muchos sistemas soportan los atributos agregados y calculados. La herencia desde un subtipo requiere, además de valores nulos, el poder obtener el resultado de la combinación exterior ("outer join").

EE con nuevas interrelaciones.

Si levantamos además la segunda restricción de la primera clase, permitiremos tipos de interrelación en un EE que no existen en el EBD. Encontramos diversas subclases.

Interrelaciones heredadas: Un tipo de entidad en un EE puede participar en tipos de interrelación en los que en el EBD no participa él, sino un subtipo o supertipo.

Unión de interrelaciones: Si en el EBD dos tipos de entidad, A y B, están interrelacionadas por dos tipos de interrelación, R1 y R2, un EE puede contener un nuevo tipo de interrelación entre A y B, obtenido por unión de R1 y R2. Ello se generaliza a más de dos.

Interrelaciones proyectadas: De un tipo de interrelación ABC entre tres tipos de entidad A, B y C del EBD, pueden derivarse tres tipos de interrelación que pueden aparecer en un EE: AB entre A y B, BC entre B y C, y CA entre C y A. Ello se generaliza a derivar tipos de interrelación n-arios a partir de tipos m-arios ($m > n$), lo que llamaremos proyección.

Interrelaciones combinadas: Si el EBD contiene un tipo de interrelación AB entre los tipos de entidad A y B, y otro AC entre A y C, un EE puede tener un nuevo tipo de interrelación, digamos ABC, entre A, B, y C. Esta operación, que llamaremos combinación, se generaliza a derivar, de un tipo de interrelación m-ario y otro n-ario, un nuevo tipo $(m+n-c)$ -ario, donde c es el número de tipos de entidad tomados como pivotes.

Combinación con proyección: Tanto combinación como proyección son transitivas y pueden aplicarse en sucesión: de AB y AC del ejemplo anterior puede derivarse BC por combinación seguida de proyección. Esto puede ser importante para la seguridad: permitir a ciertos usuarios ver B,C y BC, sin que puedan ver, no sólo los atributos de A, sino el hecho de la existencia de A. Por otra parte, hay que tener cuidado para no caer en la famosa

rirse a los esquemas externos.

REFERENCIAS

Chen(81): "A preliminary framework for Entity Relationship models", en P. Chen (ed.) Entity Relationship Approach to Information Modeling and Analysis, ER Institute.

Codd(70): "A Relational Model of Data for Large, Shared Data Banks". Comm. A.C.M., 13, 6.

IBM(84): IMS/VS 1 General Information Manual, Forms GH20-1260.

OLle(78): The CODASYL Approach to Data Base Management, Wiley Interscience.

Saltor(76): "Dels fitxers clàssics a les bases de dades: Dades i models estructurals", Novàtica, nº 11.

Saltor(84): "On external schemata not subsets of the Database schema", Facultat d'Informàtica de Barcelona.

Saltor(85): "On the power to derive external schemata from the Database schema", Facultat d'Informàtica de Barcelona.

SUMMARY

External schemata derivable from a Database schema are classified into four classes and corresponding subclasses. The power of a data model to derive external schemata may be evaluated by applying this classification. It is shown that the bihierarchical model is somehow superior to the CODASYL specifications, and that the relational model is far more powerful than the other two.