

# ESTEREOTIPOS DE CLASES ENTIDAD Y CLUSTERING DE OBJETOS EN SISTEMAS DE GESTION

Servetto, A.<sup>1</sup>; López, G.<sup>2</sup>; Jeder, I.<sup>2</sup>; Echeverría, A.<sup>2</sup>; Grossi, M.D.<sup>1</sup>

1. Laboratorio de Sistemas Operativos y Bases de Datos - Facultad de Ingeniería. UBA.  
Tel: 54-11-4343-0891. Extensión 142  
e-mail: aserve@gmail.com
2. Laboratorio de Informática de Gestión - Facultad de Ingeniería. UBA.  
Tel: 54-11-4343-0891. Extensión 141  
e-mail: glopez@fi.uba.ar

## CONTEXTO

Esta línea de investigación forma parte de un proyecto del Laboratorio de Sistemas Operativos y Bases de Datos cuyo objetivo es el desarrollo de un Manejador de Bases de Objetos para Sistemas de Gestión, respetando los principios del software libre [1] y siguiendo los lineamientos del OMG [2].

## RESUMEN

En las metodologías de desarrollo de software de gestión con objetos, los modelos de dominio se realizan en la etapa de elaboración o de análisis, y a medida que se avanza en el desarrollo las clases entidad de estos modelos se insertan en distintos diagramas de diseño correspondientes a la realización de casos de uso que responden a modelos de ejecución y no de persistencia, y el modelo integral de dominio se va refinando según esquemas de navegación que surgen de los diagramas de diseño [3]. El rendimiento de los manejadores de bases de objetos se puede optimizar si se consideran estereotipos de clases entidad propios de los sistemas de gestión que permitan clusterizar objetos en función de los patrones de acceso característicos.

Los propósitos de la línea de investigación y desarrollo que se presenta en este trabajo son establecer estereotipos de clases entidad para sistemas de gestión que faciliten el diseño de bases de objetos y estudiar y ensayar técnicas de clustering para manejadores de bases de objetos que permitan optimizar la recuperación.

## PALABRAS CLAVE

bases, objetos, clusterización, estereotipos, clases, gestión

## INTRODUCCION

Las clases de objetos en un sistema de gestión se pueden clasificar o estereotipar según la clase de datos que clasifiquen. En general pueden ser

- *De Datos Maestros*: objetos de un sistema de información que representan entidades de existencia real o ideal, por ejemplo productos o servicios, o valores de referencia para determinar características o atributos de otros datos (dominios de atributos definidos por extensión).
- *De Datos Transaccionales*: registros de hechos o eventos relacionados con datos maestros, por ejemplo de ventas de productos o de prestaciones de servicios.

Los datos transaccionales a su vez pueden diferenciarse según la posibilidad de que las transacciones que representan sean actualizables o no:

- Las *transacciones actualizables* son aquellas de las cuales se registra su previsión o programación en el tiempo, antes de que la transacción se produzca, y una vez producida se completan sus características (por ejemplo turnos para servicios que devienen en registros de

servicios).

- Las *transacciones no actualizables* son aquellas que se registran luego de producidas, como si fuera en una bitácora, y luego no se actualizan más (por ejemplo registros de facturación o de operaciones bancarias).

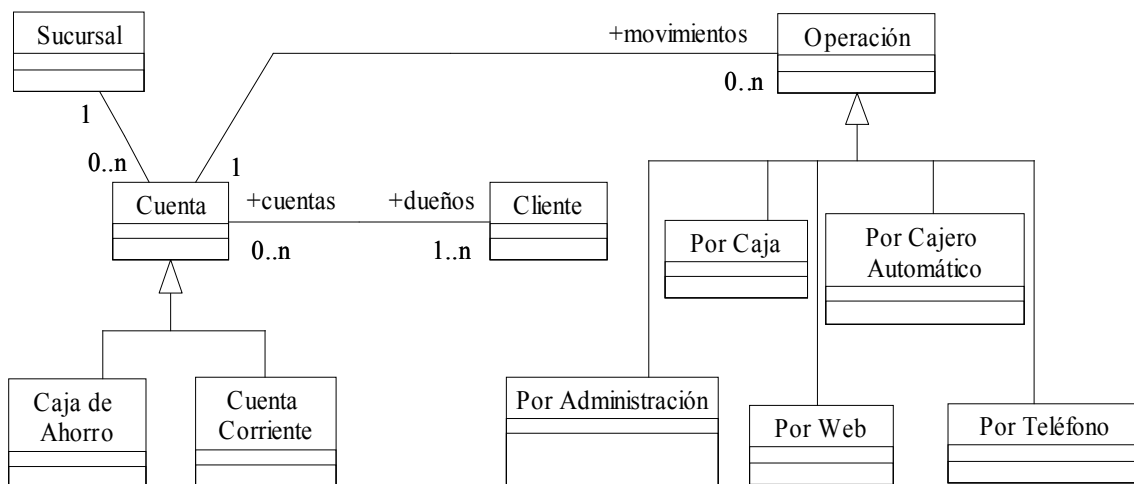
Las **operaciones con objetos** pueden ser de navegación o de consulta. Las operaciones de navegación implican accesos por identificador automático o del manejador, y las de consulta, expresadas en OQL (*Object Query Language*), por el índice de identificación de los objetos en el sistema de gestión o por otros criterios, que pueden sustentarse en otros índices.

En las bases de objetos los **identificadores de objetos** son automáticos y transparentes al usuario, por lo que no garantizan la unicidad de los objetos (es posible instanciar más de una vez un mismo objeto para el sistema de información pero con distintos identificadores): para garantizar la unicidad de los objetos se impone definir un criterio de identificación que sea propio del sistema de información. En el manejador que se está desarrollando se exige que para toda clase se precise qué atributo o conjunto de atributos identifica unívocamente a cada objeto, como es característico en el diseño conceptual de bases de datos.

## LINEAS DE INVESTIGACION Y DESARROLLO

En aras de la flexibilidad y en atención a que en los sistemas de información los objetos transaccionales se suelen identificar naturalmente a partir del identificador del objeto maestro con el que se relaciona la transacción más un indicador de tiempo, en el manejador de bases de objetos de este proyecto se admiten identificadores mixtos y externos [4], es decir, compuestos por identificadores de objetos maestros más un discriminante propio, en el caso de los mixtos, o compuestos exclusivamente por identificadores de objetos maestros, en el caso de los externos.

Para la composición de los identificadores mixtos y externos, en el ODL (*Object Definition Language*) se usa el nombre de las clases de objetos maestros, que en los objetos se traduce al identificador automático o del sistema. Por ejemplo, para definir objetos transaccionales que representen operaciones de una cuenta bancaria se define una clase Operación, dependiente de una clase Cuenta:



La definición en XML (como estándar de representación) quedaría:

```
<clase nombre="Operación" tipo="TNA" instanciable="no">
  <atr nombre="momento" tipo="tiempo"/>
  <atr nombre="movimiento" tipo="(DE|CR)"/>
  <atr nombre="tipo" tipo="Tipo Operación"/>
  <atr nombre="monto" tipo="fracc"/>
</clase>
```

```
<id tipo="mixto">
  <comp tipo="ext" pos="1" clase="Cuenta"/>
  <comp tipo="int" pos="2" atr="momento" orden="desc"/>
</id>
</clase>
```

En el fragmento de XML precedente se define la clase Operación, de tipo Transaccional No Actualizable y no instanciable (abstracta). Los atributos de esta clase son momento, que registra fecha y hora de la operación; movimiento, que registra si es un DE(bit) o un CR(édito); el tipo (definido por el administrador por extensión en la clase Tipo de Operación) y el monto. El identificador de la clase en el sistema de información es mixto, y su primer componente, externo, es el identificador automático de la cuenta a la cual afecta la operación, y el segundo componente, interno, es la fecha y hora de la operación en orden descendente. Cada operación se asocia con una (multiplicidad mínima) y sólo una (multiplicidad máxima) cuenta, que sólo debe definirse en caso de que se pretenda navegar desde una operación hacia una cuenta.

La posibilidad de definir identificadores mixtos o externos representa una innovación en las bases de objetos, ya que al no requerirse explícitamente navegación desde una operación hacia la cuenta a la que pertenece, en una base de objetos tradicional no se caracterizaría a los objetos de Operación con un identificador de la cuenta y en el manejador de este trabajo sí. La introducción de este identificador posibilita una clusterización de las operaciones por la cuenta a la que afectan y en orden cronológico inverso, optimizándose así las recuperaciones de navegación.

La implementación de las asociaciones entre una clase de objetos maestros y una de objetos transaccionales con identificador mixto no se realiza mediante una colección de identificadores en los objetos maestros, que sería muy voluminosa, sino que, a la manera relacional, se maneja accediendo a los objetos transaccionales asociados por el identificador definido.

Como la clase Operación es transaccional no actualizable, los patrones de acceso característicos son para inserción o recuperación (no hay bajas ni actualización de objetos). Si la identificación de objetos en el sistema de información fuera por un número de operación independiente de la cuenta de pertenencia, una organización adecuada para la persistencia sería la secuencial indexada sin agrupar registros en bloques: todo nuevo registro se agrega al final del archivo, y el criterio de indexación de secuencias podría ser por umbrales horarios (por ejemplo se indexarían secuencias de operaciones por cada hora del día, por cada x cantidad de horas de cada día, o por cada día a partir de una hora determinada -por ejemplo la hora de clearing).

Pero en el caso de ejemplo, como la identificación depende de la cuenta a la que afecta, la agrupación o clusterización de operaciones se realiza según la cuenta, y las operaciones en cada cuenta se ordenan por cronología inversa; entonces la inserción de registros no puede realizarse siempre al final del archivo, sino al inicio de cada secuencia, siendo que cada secuencia es de operaciones en una cuenta en orden cronológico inverso. Por lo tanto la organización que se escoge es la secuencial indexada con registros organizados en bloques. Particularmente para esta combinación de clase o estereotipo de clase y definición de identificador se emplea árboles B# (árbol B+ con restricciones de contenido mínimo en nodos propias de los árboles B\*) con los objetos organizados por el identificador del sistema de información. El fundamento de esta elección es que es común que las aplicaciones pidan iteradores sobre objetos por la clave de identificación del sistema de información, ya sea para listados o para búsquedas aproximadas, cuando se trata de objetos maestros, y para objetos transaccionales también, para realizar operaciones de totalización o síntesis (cortes de control); entonces se piensa que tener los objetos almacenados según el criterio de recuperación más probable importa una mejora del rendimiento general.

En cuanto a los identificadores automáticos o del manejador, los define automáticamente el manejador, y para los identificadores mixtos y externos los componentes externos son los

identificadores automáticos de las clases principales.

Para toda clase se crea por defecto un índice para el identificador automático con organización directa extensible y función de dispersión basada en los bits menos significativos del identificador de los objetos (resto de la división del identificador entre el tamaño de la tabla de dispersión).

En cuanto a la organización de los objetos secuencial indexada sin bloques, el índice se organiza con estructura B# y como referencias se usa el desplazamiento en bytes a cada objeto a partir del inicio del archivo.

Para otros índices también se usa árboles B#. Se admite la definición de índices que se clasifican según la multiplicidad de objetos referidos por cada clave, como **índices de identificación** (un objeto por clave) o **índices de clasificación** (muchos objetos por clave). Como referencias a objetos correspondientes a cada clave de recuperación (sea de identificación o de clasificación), se usa la clave de organización: el identificador del sistema de información.

En general, las **organizaciones de archivo** que se emplean para la persistencia de los objetos se detallan en el siguiente cuadro:

	<b>Datos Maestros Actualizables</b>	<b>Datos Maestros No Actualizables (Dominios)</b>	<b>Transacciones Actualizables</b>	<b>Transacciones No Actualizables</b>
<b>Identificador Interno</b>	B#	Sec. Ind. s/bloques	B#	Sec. Ind. s/bloques
<b>Identificador Mixto o Externo</b>	B#	B#	B#	B#

Para almacenar **jerarquías de objetos** se usa un único archivo, correspondiente a la clase raíz, sin importar que ésta sea abstracta o concreta. No se admite herencia múltiple, y si se desea identificar objetos de alguna subclase con un identificador propio se debe definir un índice de identificación, pero el identificador formal siempre es el de la clase raíz (no se admite identificadores en subclases). Los índices de subclases, tanto para identificadores automáticos como los definidos para recuperación se implementan como **índices selectivos**: indexan sólo los objetos de la subclase sobre la que están definidos.

## **RESULTADOS OBTENIDOS/ESPERADOS**

En la implementación del manejador se usan técnicas relacionales, dada su probada eficacia, pero la mayor eficiencia del manejador en los casos planteados queda supeditada a la definición de la base, que si se efectúa sin usar identificadores mixtos o externos no se logra. En el caso de ejemplo, si las operaciones de identificar con un número de operación propio del sistema de información, la forma de implementar la asociación con la clase cuenta sería tener en cada objeto de cuenta una colección de identificadores automáticos de operación.

Todavía no se cuenta con resultados experimentales porque el manejador del proyecto se encuentra en una etapa de pruebas. Cuando el producto se encuentre estabilizado se estará en condiciones de comparar parámetros de rendimiento con productos similares.

Se está implementando paralelamente una versión del módulo de almacenamiento para el manejador de bases de objetos que implementa las asociaciones al estilo relacional, para comparar rendimientos con la propuesta de este trabajo.

## **FORMACION DE RECURSOS HUMANOS**

En el proyecto de investigación y desarrollo de referencia participan docentes auxiliares y alumnos de grado de la carrera de Ingeniería en Informática en el marco de las materias Taller de Programación II (desarrollo de módulos experimentales del manejador de bases de objetos), Organización de Datos (implementación de técnicas de clusterización y almacenamiento de objetos), Trabajo Profesional, y Tesis; también participan alumnos de la carrera de Licenciatura en Análisis de Sistemas en el marco de la materia Aplicaciones Informáticas.

En el proyecto de referencia también participan investigadores del Laboratorio de Informática de Gestión.

## **BIBLIOGRAFIA**

- [1] Malcolm Atkinson (University of Glasgow), François Bancilhon (Altaïr), David DeWitt (University of Wisconsin), Klaus Dittrich (University of Zurich), David Maier (Oregon Graduate Center) y Stanley Zdonik (Brown University), "The Object-Oriented Database System Manifiesto", Proc. of the First International Conference on Deductive and Object-Oriented Databases, Kyoto, Japón. 1990.
- [2] [http://www.omg.org/technology/documents/corba\\_spec\\_catalog.htm#CCM](http://www.omg.org/technology/documents/corba_spec_catalog.htm#CCM)
- [3] Jim Arlow, Ila Neustadt, "UML 2", Anaya Multimedia, 2006.
- [4] Carlo Batini, Stefano Ceri, Shamkant B. Navathe, "Diseño Conceptual de Bases de Datos: Un enfoque de entidades-interrelaciones", Addison-Wesley/Díaz de Santos, 1994.