

S.O.O.M. Un Modelo Semántico para Bases de Datos

Ing. Augusto Giancola, Ing. Francisco Marcano, Ing. Leonid Tineo
Ing. Ma. Esther Vidal, M.Sc. Enzo Chiariotti

Universidad Simón Bolívar, Caracas, Venezuela

Es creciente el interés en la incorporación a los manejadores de bases de datos de herramientas de modelación que expresen no sólo la estructura de los datos, sino también su interpretación y su papel en el sistema en que se encuentran. Como consecuencia de este interés encontramos han venido siendo explorados múltiples caminos para la modelación de datos.

Se presenta una revisión de los principales conceptos asociados a la modelación de datos y de diferentes paradigmas y proposiciones en el área, para, sobre la base de tales conceptos y proposiciones presentar la definición de un modelo de datos semántico, el modelo S.O.O.M., el cual constituido por un conjunto de abstracciones que favorecen la expresión de manera sencilla y natural de la estructura y significado de los datos los sistemas reales que se modelen.

INTRODUCCION

Los manejadores de bases de datos mayormente difundidos y conocidos, dejan de mano del programador el conocimiento y la preservación del significado de los datos que se manipulan; esto como consecuencia de la carencia en tales manejadores -o en los modelos que soportan- de herramientas de modelación suficientes para expresar no sólo la estructura de los datos, sino también su interpretación.

Se presenta en este trabajo la definición de un Modelo de Datos semántico, constituido por un conjunto de conceptos y abstracciones, dirigidas a constituir para el modelador de datos herramientas manejables y expresivas a los fines de recoger tanto estructura como significado de los datos de un sistema.

La definición del Modelo S.O.O.M. que aquí se presenta forma parte del proyecto de Manejador de Bases de Datos que se adelanta en la Universidad Simón Bolívar, y como parte del cual los autores han realizado también una definición formal del modelo y la definición e implementación de un prototipo de Lenguaje de Definición de Datos para el modelo.

La presentación del modelo S.O.O.M. se realiza en cuatro capítulos:

El primero expone un conjunto de conceptos, que constituyen el marco teórico-conceptual relativo a la modelación de datos, con especial atención sobre sus objetivos, tanto de carácter general, como aquellos asociados a la implementación de un modelo.

En el segundo capítulo se presentan los resultados de un análisis comparativo del universo de modelos de datos estudiados, realizado sobre la base de los aspectos que, en el marco conceptual del capítulo precedente, se consideran resaltantes en la evaluación de un modelo.

Partiendo la revisión de conceptos del segundo capítulo, el tercero, se hace una selección y justificación de aquellos que serán soportados por el modelo S.O.O.M, definido en el cuarto capítulo, que también incluye la representación gráfica de tales abstracciones y un ejemplo de modelación.

I. MARCO CONCEPTUAL

Se presenta una revisión de lo que se considera el conjunto básico de conceptos sobre los cuales, en los capítulos posteriores, se basará la presentación, análisis y evaluación de diferentes paradigmas de modelación de datos, y que orientarán la definición del modelo S.O.O.M.

1.1. Percepción y Expresión

La interacción del ser humano con su ambiente se verifica a través de dos procesos fundamentales, la **percepción** y la **expresión**. Por el primero, el hombre captura información de su ambiente, alimentando su visión particular del mismo; por el segundo, es capaz de transmitir a este mismo ambiente, la visión particular que ha elaborado del mundo real.

En el primer proceso, el hombre recibe de su entorno sensaciones, las cuales, enriquecidas con conceptos que ya ha elaborado, generan nuevos conceptos y, posiblemente, modifica los anteriores; a este proceso asociado a la percepción, lo denominamos **conceptualización**, y es a partir de él que cada hombre construye su visión del entorno.

Cuando el intercambio se verifica desde el ser humano hacia su entorno, es decir, cuando aquel desea expresar la visión del mundo real que ha construido, encontramos otro proceso de elaboración sobre el conjunto de conceptos que forma esa visión; a esta elaboración la denominaremos **modelación**; se trata de una reelaboración que represente de la manera más fiel y exacta posible ese conjunto de conceptos.

En el proceso de modelación, es necesaria la habilidad de ocultar los detalles y particularidades de los múltiples elementos que constituyen nuestro conocimiento, concentrándose en aquellos aspectos comunes de conjuntos de ellos, a los fines de reducir la complejidad y aumentar la significación de lo expresado. A esta habilidad la llamamos **abstracción**. Al conjunto de abstracciones y conceptos involucrados en la modelación,

constituye lo que denominaremos un **modelo**.

Para esa interacción del ser humano con su entorno –el cual incluye otros seres humanos–, son múltiples los modelos que ha elaborado, cada uno influenciado por diferentes y específicas necesidades de expresión: modelos más sofisticados permiten expresar fenómenos complejos. La necesidad de comunicación agrega la necesidad de encontrar modelos comunes para la expresión de ciertas ideas (lenguaje natural, formalismos matemáticos, etc).

1.2. Los modelos de datos

Como se mencionó, el hombre trata de describir los fenómenos que percibe de su entorno, y que constituyen la base del conocimiento de éste. Cuando hablamos de **datos** hablamos de estas descripciones que el ser humano hace de cada uno de los hechos (fenómenos) que percibe, y de las posibles relaciones que entre ellos haya percibido –todos estos hechos concretos–, o, inclusive, inferido.

El problema de la modelación de datos es el mismo que se ha descrito cuando hablamos de la expresión que hace el ser humano de su visión de mundo real –o de parte de éste–. Prácticamente cualquier forma de expresión conocida del ser humano constituye un caso de modelación de datos, cada una sobre su propio **modelo**, cada uno tratando de capturar la esencia del significado de los datos expresados.

Dos objetivos se reconocen claramente en la construcción y utilización de cualquier modelo de datos: expresividad y economía conceptual. El primero está relacionado con la necesidad de transmitir la esencia de lo representado, el segundo, apunta hacia la manejabilidad, del conjunto de abstracciones que constituyan al modelo. Las vías hacia cada una de estas metas no siempre coinciden, ni convergen, por lo que la modelación implicará siempre el compromiso de reducir el conjunto de conceptos (y abstracciones) utilizados, preservando el poder de representar la realidad de manera natural.

Consideraremos un **modelo de datos** como una herramienta intelectual que permite una interpretación del mundo real, suficientemente abstracta para favorecer su manejabilidad y, al mismo tiempo, suficientemente poderosa para ofrecer entendimiento sobre cómo están relacionados los datos sobre el mundo real. [Tsichritzis 82]

1.3. Modelos de Datos y Ambientes Automatizados de Manipulación de Datos

En la medida en que la representación de la realidad –o de una fracción de ella– se hace compleja, y

con esto, mayor el volumen de datos y relaciones que entre estos deben ser consideradas, crece la necesidad, y el interés, por la manipulación automatizada de estos datos y relaciones. Es este el caso del manejo de datos corporativos en una empresa, o del cálculo en un ambiente de investigación —para el que se desarrollan los primeros sistemas de procesamiento de datos—, pero también lo es el de la representación del conocimiento, los ensayos en inteligencia artificial —que, de manera particular, han enriquecido notablemente el área de la modelación de datos—.

De la necesidad de automatizar el manejo de los datos usados en la representación de la realidad, se desprende un tercer objetivo general a perseguir en la selección de las abstracciones y conceptos que constituyan el modelo: el de considerar la factibilidad de su implementación eficaz y eficiente en el ambiente automatizado que soportará esa interpretación de la realidad. Como en el caso de los enunciados anteriormente, este objetivo implica transitar un camino que no coincide con los de los otros, por lo que ha de establecerse un nuevo compromiso: cuáles abstracciones deben ser eliminadas del modelo por su dificultad de manejo automatizado, cuáles se siguen incluyendo, aunque no sean soportadas por una implementación —debiendo ser manejadas por el "usuario" del modelo—, cuáles deben ser preservadas —o agregadas— para facilitar la implementación.

Es sobre estos modelos de datos, orientados a la representación de la realidad en organizaciones complejas en vista del manejo automatizado de la información, que se desarrolla nuestro trabajo; y sobre los aspectos mencionados en el párrafo anterior y la satisfacción de los objetivos generales, que se conduce en el capítulo siguiente una revisión de los diferentes paradigmas de modelación de datos.

2. PARADIGMAS PARA LA MODELACION DE DATOS

Se presenta a continuación el resultado de la revisión de las principales orientaciones seguidas en la modelación de datos —como herramienta de diseño y en ambientes de manejo automatizado— hasta el día de hoy, señalando de qué manera y en qué medida, se han aproximado a la satisfacción de los objetivos generales enunciados en el capítulo anterior. Antecediendo a esta revisión, se indican los criterios para la agrupación de los modelos que se utiliza. Finalmente, se presentan dos cuadros comparativos, en los que se ubican los diferentes paradigmas, según elementos de evaluación que se describirán.

2.1. Criterios de Agrupación, Criterios para la Evaluación.

La revisión de paradigmas de modelación cuyos resultados aquí se refieren, se realizó agrupándolos

en dos categorías [Tsitchritzis 82]; el primero, corresponde a los más ampliamente utilizados y estudiados, que llamaremos **tradicionales**; el segundo incluye aquéllos orientados a la representación del significado de los entes y relaciones modelados, a los que, adoptando una denominación muy general, llamaremos **semánticos**. El análisis se realizó sobre la base de los objetivos de expresividad, economía conceptual -manejabilidad de conceptos- e implementabilidad, ya descritos, sin atender de manera particular a detalles de implementación, definición u operación de cada modelo.

2.2. Modelos Tradicionales

Entre los modelos tradicionales incluimos a **jerárquico, de redes y relacional**.

Estos modelos, en particular los jerárquicos y de redes, constituyen los primeros estadios en el reemplazo de los Sistemas de Archivos, utilizados hasta ese entonces en los ambientes de Procesamiento de Datos, por una representación de los datos que permitiera aprovechar las características particulares de estos, dándoles así significado, de manera independiente de las necesidades y aplicaciones a satisfacer -y de los cambios en éstas-.

Siendo este el marco en que surgen y se incorporan al mercado los modelos tradicionales, se explica que hayan seguido la vía de la implementabilidad, y que el conjunto de abstracciones soportadas fuera lo más reducido y manejable posible. La expresión del "significado" de los datos, aunque en cada modelo con diferente nivel, se encuentra orientada claramente a las operaciones (condiciones sobre ellas, especificación de las mismas), permitidas sobre los datos. (restricciones de inserción y permanencia en los de redes, control por aplicaciones en el relacional).

Característica común de los tres modelos es el uso de tablas para la representación de los "entes" del sistema a representar, estructura muy familiar para los usuarios en el ambiente de procesamiento de datos. Sin embargo, cada modelo, posee características que los diferencia en la satisfacción de los objetivos generales de la modelación.

Modelos de Redes:

Los modelos de redes -entre los que se puede considerar al jerárquico como un caso particular- utilizan al grafo (con algunas restricciones), como estructura para la modelación. De esta manera, estos modelos permiten **expresar** relaciones entre entes del sistema modelado, sobre las cuales se pueden definir las reglas de inserción y permanencia ya mencionadas. De manera que cierta expresividad es ofrecida por el

modelo. No obstante, la orientación a la implementación de estos modelos, induce limitaciones sobre la generalidad de las relaciones: en el caso de redes, debe preservarse "propiedad exclusiva" de un ente sobre sus miembros, en el jerárquico, las relaciones son de carácter estrictamente jerárquico. Otras relaciones más generales (muchos a muchos, p.ej.), pueden representarse, pero con la introducción de entes artificiales -no forman parte de la realidad modelada-.

Modelo Relacional

En este caso, la única estructura es la tabla, asociada a la única abstracción soportada: la agregación, vista como relación matemática. Esto reduce altamente la expresividad del modelo, pues, si bien permite representar de manera natural a cada ente del sistema, no hace lo mismo con las relaciones entre éstos, las cuales son representables también como tablas, pero comprometiendo al modelador a "recordar" (a través de la documentación) y preservar (por programas de aplicación) el significado de cada tabla. La menor expresividad va nuevamente en favor de la implementabilidad del modelo, pero de una manera distinta a la expuesta para los modelos de redes: las estructuras (red, jerarquía) permitidas inducen claramente una implementación (listas, listas doblemente enlazadas), en el caso relacional, las tablas constituyen un nivel de abstracción superior, el cual puede ser soportado a través de diferentes implementaciones.

Para la falta de expresividad que muestra el modelo relacional, el más utilizado actualmente, se han encontrado dos tipos de soluciones. La primera consiste en el uso de modelos más expresivos (E-R o E-R-E, [Chen 76], [Teorey 86]), en las tareas de diseño y documentación de la base de datos. La segunda, más formal, ha consistido en agregar al modelo algunos de los aspectos semánticos de los que carece (representación explícita de relaciones, preservación de integridad referencial); el intento más notable en este sentido lo constituye el modelo RM/T [Codd 79].

2.3. Modelos Semánticos.

Los modelos que se agrupan como semánticos se caracterizan por proveer conjuntos de abstracciones más completos, para la representación del significado de las diferentes relaciones que pueden encontrarse en el sistema que se trata de modelar, así como la posibilidad de expresar de manera explícita las diferentes restricciones que deban asociarse a cada ente o relación del sistema real.

Estos modelos se pueden clasificar en cuatro grupos [Tsichritzis 82]: Modelos Entidad-Relación, Modelos Binarios Semánticos, Modelos de Redes Semánticas y Modelos Infológicos. Todos estos modelos

persiguen la expresividad e involucran implementaciones más complejas. Los modelos presentan en orden creciente en cuanto a expresividad y, como es natural, de dificultades en su implementación.

2.3.1. Modelos Entidad-Relación

Estos modelos ([Chen 76], [Tsitchritzis 82], [Peckham 88]), basados en tablas y grafos, han sido ampliamente aceptados y utilizados en el ambiente de Bases de Datos, especialmente como herramienta para Diseño y Documentación de Sistemas, -en particular en ambientes que utilizan manejadores relacionales-.

Las características que permiten incluirlos entre los modelos semánticos son las siguientes:

- Diferenciación **entidades y relaciones**. Las relaciones son genéricas, y pueden agrupar a una o más entidades (n-arias), pudiendo establecerse para cada entidad participante en ella, su grado máximo de participación, soportando el carácter de relación de una determinada tabla.
- Permite establecer **restricciones de dominio** (conjuntos válidos de valores), para los diferentes parámetros que constituyen a una entidad o una relación, las cuales también deben ser preservadas por un manejador que soporte el modelo.
- Una adición semántica importante, la constituye el soporte específico de relaciones con características especiales, pero frecuentes en los sistemas que se trata de modelar [Peckham 88]. Así, en el caso del Modelo Entidad-Relación se tienen relaciones de Identidad y de Existencia, cuyas especiales características son preservadas por el modelo; en el E-R-Extendido [Teorey 86], estos casos especiales no se incluyen, mientras si se modelan las abstracciones de Clasificación y Tipificación que en el capítulo 4 serán expuestas, y que el modelo provee como jerarquías de Generalización y de Subconjunto.

3.3.2. Modelos Binarios Semánticos

Estos modelos, son los que muestran mayor economía conceptual entre los aquí agrupados; proveen como única estructura el **grafo**, sobre el cual se soportan los conceptos de **categoría y relación binaria**, conceptos sobre los cuales pueden construirse abstracciones más complejas (relaciones n-arias, agregación, restricciones de dominio, etc.). Esta limitación a categorías y relaciones binarias, aunque no impide, como se dijo, la representación de realidades más complejas, obliga al modelador a mantener el control de la semántica de estas construcciones -algo similar a la tarea del usuario del modelo relacional-. Sin embargo, el modelo ofrece dos herramientas valiosas en la reducción de esta tarea de control a nivel de programa de aplicación: a) la definición de restricciones de cardinalidad; b) la especificación -asociada a las relaciones y

categorías- de programas, con los que, a tiempo de modelación, se preservan las características especiales de cada relación o categoría, en el sistema modelado.

3.3.3. Modelos de Redes Semánticas

Estos modelos surgen como producto de la investigación en las áreas de Inteligencia Artificial y Representación del Conocimiento, y en consecuencia, se destacan por perseguir la representación de un subconjunto, suficientemente poderoso, de las abstracciones que utiliza el ser humano, en la expresión de su visión de la realidad. En particular, el área de Bases de Datos se ha enriquecido de este trabajo, tomando aquellas abstracciones que permitan aproximar el trabajo de modelación de un sistema de datos, a la visión que el modelador pueda hacerse de él. Como las características comunes más resaltantes de los modelos de redes semánticas se pueden destacar [Tsichritzis 82]:

- Las estructuras manejadas por el modelo son grafos (arcos y nodos), permitiéndose en un esquema la coexistencia de instancias (ocurrencias) y tipos (clases).
- Diferenciación entre tipo (clases) y *tokens* (instancias). Es el primero de los modelos presentados que introduce los términos de instanciación y generalización.
- Distancia Semántica: Nodos cercanos en el grafo se consideran semánticamente cercanos (los arcos de irrelevancia permiten "alejar" nodos, que en el grafo aparezcan cercanos).
- Partición: permite restringir el ámbito de una operación, a un subconjunto (de clases o de objetos) del esquema.
- Jerarquía de tipos: introduciendo las abstracciones de **Generalización (IS_A)** y **Agregación (PART_OF)**, estableciéndose explícitamente reglas de herencia de las propiedades (relaciones, atributos, restricciones) de una clase a otra.

Los nodos pueden representar: CONCEPTOS: instancias de tipos que se encuentran en la BD; EVENTOS: acciones entre conceptos; CARACTERISTICAS de un concepto; VALORES: de una característica. Por su parte, los arcos se distinguen como: ASERCIÓN (concepto-concepto); INSTANCIACIÓN (concepto-tipo); RELACIÓN BINARIA (tipo-tipo), la cual puede ser de jerarquía (IS_A o PART_OF), o genérica.

Se pueden especificar restricciones, como hechos (facts) acerca de *tokens* o clases, a través del cálculo de predicados; mientras otras más complejas, pueden especificarse a través de programas, asociados a los diferentes entes y operaciones, que forman parte del modelo como una metacalse predefinida.

Es importante mencionar que el crecimiento del conjunto de abstracciones que soporta el modelo, no afecta seriamente la economía conceptual, pues se trata de abstracciones que el ser humano maneja, y encuentra en su entorno, con frecuencia.

3.3.4. Modelos Infológicos

Estos modelos están orientados a usuarios terminales, tanto a tiempo de diseño como de operación ([Sudgren 74], [Tsichritzis 82]). A tiempo de diseño persiguen: a) facilitar la definición precisa de **qué** información debe contener la B.D., sin anticipar decisiones de **cómo** representarla en alguna estructura; y b) facilitar a expertos la transformación sistemática de la estructura de cierta información en una estructura eficiente de almacenamiento de datos. A tiempo de operación persiguen: a) definir una manera **no procedimental** de recuperar la información que requiere el usuario, sin tener que referirse a la estructura dataológica de la B.D.; y b) que el sistema manejador de Base de Datos **interprete** el requerimiento de información y seleccione una estrategia de procesamiento eficiente de manera dinámica.

El enfoque infológico muestra la orientación que siguen los modelos de Base de Datos, en cuanto a aumentar la expresividad del modelo, reduciendo la labor de traducción de la realidad que debe realizar el modelador, y la de interpretación que debe realizar el usuario. La entrada de un manejador que implemente el Modelo Infológico, debería aproximarse al lenguaje natural, la salida, debería aproximarse a lo que el usuario quiere saber. Esto representa el extremo del continuo de expresividad-manejabilidad sobre el que mencionamos se pueden ubicar los modelos de datos, pero también implica alcanzar el extremo contrario (el inferior) en el continuo de la implementabilidad –la otra medida en la evaluación de los modelos que se está haciendo–: implementar un modelo bajo la perspectiva infológica implica interpretación el lenguaje natural e interpretación y manejo del conocimiento, ambos, objetivos no alcanzados aun, y perseguidos por diferentes ramas de la computación.

2.4. El Paradigma Orientado por Objetos y las Bases de Datos

Desde Simula [Birtwistle 73], son muchos los aportes que se han dado en el campo de los lenguajes de programación orientados por objetos. [Nierstrusz 89] muestra un conjunto de conceptos relacionados con las construcciones y los mecanismos que deberían ser provistos por lenguajes que se desenvuelven sobre dicho paradigma.

Como consecuencia de este desarrollo, surge la idea de sistemas orientados por objetos, los cuales

proveen el soporte de tiempo de ejecución y el ambiente de desarrollo de aplicaciones orientadas por objetos. En relación al soporte de tiempo de ejecución, estos sistemas hacen referencia al manejo de objetos que engloba mecanismos tales como nombramiento de objetos, persistencia, concurrencia, distribución, seguridad y control sobre versiones de tipos y otros, como por ejemplo, recolección de basura. Por otra parte, proveen mecanismos poderosos para el desarrollo de aplicaciones, haciendo esta tarea más fácil, mediando el concepto de reusabilidad y herramientas de diseño y manejo de evolución de objetos.

Las manipulaciones anteriormente destacadas, pueden ser enfocadas desde el punto de vista de la tecnología de las Bases de Datos. Una Base de Datos Orientada por Objetos, es un sistema que provee soporte para objetos persistentes (transacciones, *queries*, etc.). Esto no es más que la extensión de los criterios tomados en el campo de lenguajes de programación orientados por objetos. Sin embargo, la incorporación de las ideas de este paradigma en el área de Bases de Datos, enfrenta algunos problemas por ser solucionados:

- inexistencia de un modelo de datos estándar;
- inexistencia de un lenguaje de manipulación de datos que sea equivalente al álgebra relacional.
- inexistencia de lineamientos para el diseño de Bases de Datos Orientadas por Objetos.

los cuales deberán ser resueltos a fin de preservar los principios fundamentales del paradigma.

2.5. Comparación

En la revisión se ha destacado la manera en que los diferentes paradigmas de modelación se han aproximado a los tres objetivos generales enunciados. Indicamos, el marco histórico de la aparición de los modelos tradicionales, para mostrar la influencia que éste ejerció; mencionamos la orientación semántica de los modelos que agrupamos como tales, los cuales se presentaron en orden creciente de expresividad.

Para complementar esta visión, y ofrecer una ubicación comparativa de los modelos expuestos, se muestran a continuación dos gráficos, en los que los modelos son colocados sobre sendos continuos, asociados con dos criterios que consideramos determinantes en la evaluación de los modelos. El Gráfico 1 ubica a los modelos, en función del esfuerzo de traducción que debe hacer el modelador para representar la situación que está modelando. Los extremos de este continuo, hombre y máquina, representan, la orientación del modelo, esto es, hacia la visión del modelador, o hacia la implementación del sistema modelado.

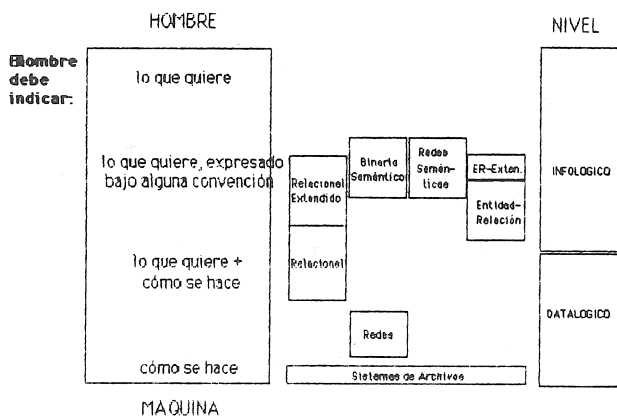


Gráfico 1

En el segundo (Gráfico 2) se ubican los modelos dentro de la perspectiva de la expresividad, esto es, de la manera en que permite representar la realidad, tal y como es.

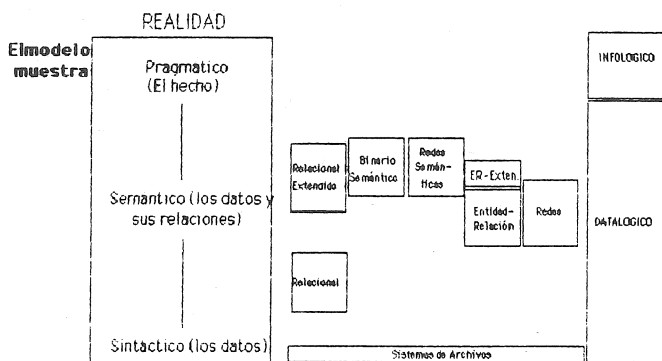


Gráfico 2

3. SELECCION Y DEFINICION DE UN MODELO

La formulación de un Modelo de Datos, implica la selección de características de uno o más paradigmas, que sirvan de base al conjunto de conceptos y abstracciones a ser soportadas por éste. La revisión de modelos cuyos resultados se reportaron en el capítulo anterior, constituyó un primer paso en esta tarea, en cuanto permitió conocer las orientaciones seguidas por cada modelo. El siguiente paso lo constituye la revisión de los conceptos y criterios de trabajo, que el estudio de la modelación de datos ha encontrado como fundamentales, a los fines de alcanzar expresividad en la modelación. Con estos conceptos se completa el

marco de análisis previo a la definición del modelo S.O.O.M.

3.1. Un Modelo Semántico

La orientación que sigue la modelación de datos es la de la expresividad, en cuanto a la representación del sistema real, en base a conceptos y abstracciones que lo reflejen fielmente. La definición de un modelo de datos, debe seguir esta orientación –precisamente la que sigue los modelos que llamamos **semánticos**- y es la que consideramos debía seguir el modelo cuya definición se reporta

Los aportes que los modelos semánticos ofrecen al campo de la modelación de datos (cada uno en diferente medida), y que justifican la selección de este paradigma para la definición de nuestro modelo, se pueden agrupar bajo cuatro criterios generales [Peckham 88]: Economía de Expresión, Preservación de la Integridad, Flexibilidad en la Modelación y Eficiencia en la Modelación.

La economía de expresión es aportada por los modelos semánticos al permitir extraer y representar un amplio rango de la información del sistema que modela, mediante un conjunto limitado de abstracciones, cuya naturaleza es similar a la de aquellas usadas por el ser humano en su comprensión del ambiente, facilitando el proceso de modelación. La introducción de tales abstracciones, favorece la preservación de la integridad de los datos, al proveer mecanismos, implícitos en cada abstracción o definibles por el modelador, para el soporte de las restricciones asociadas a cada ente de un esquema.

La flexibilidad en la modelación, se refiere a la posibilidad que ofrecen los modelos semánticos de representar las situaciones reales, utilizando los conceptos y abstracciones que mejor se ajusten al papel del sistema modelado, flexibilidad que no proveen los modelos "tradicionales", cuyas posibilidades de representación se circunscriben a las estructuras que proveen (tablas, tablas y redes o jerarquías).

Por último, la eficiencia en la modelación, se alcanza en la medida en que el modelador debe ocuparse menos de los detalles de implementación. Los modelos semánticos proveen, asociados a las abstracciones que soportan, mecanismos predefinidos para la preservación de su significado, evitando al modelador mantener este control cada vez que un mismo de tipo de relación o abstracción sea utilizada.

3.2. Revisión de Diferentes Modelos Semánticos:

Revisadas las ventajas generales del uso de la modelación semántica, la definición del modelo requiere la revisión de diferentes proposiciones que se encuentran dentro del marco de los Modelos de Datos

Semánticos, con el objetivo de conocer las facilidades que ofrecen, los aspectos en que coinciden y aquéllos en que difieren.

Los modelos revisados fueron : Entidad-Relación [Chen 76], RM/T [Codd 79], TAXIS [Nixon 87], SDM [Hammer 81], Funcional [Shipman 81], SAM* [Su 83], Event [King 84] y SHM+ [Brodie 84] -de los que se obtuvo información indirecta [Tsichritzis & Lochosky 82][Peckham & Maryanski 88]-; e IF0. De esta revisión, reportamos los aspectos en que las diferentes propuestas coinciden, y aquellos en los que se encuentran diferentes enfoques [Peckham 88]

Los diferentes modelos coinciden en:

- Proveer relaciones entre objetos de datos, que soporten la manera en que el usuario percibe el sistema real.
- Para tales relaciones deben ser provistos mecanismos para especificar estados, transiciones y respuestas válidas para la B.D.

mientras existen diferentes enfoques respecto a:

- La generalidad del modelo: ¿debe ser independiente de la aplicación u orientado hacia algún ámbito en particular?
- Debe proveerse en un alto nivel la preservación de la semántica de las relaciones soportadas, o dejar al modelador, la posibilidad de controlar las restricciones asociadas a ellas.
- Deben proveerse relaciones y abstracciones complejas (agregación, tipificación, colecciones, etc.) o, por el contrario, primitivas y generales (modelos binarios).

3.3. Conceptos y Abstracciones soportadas por los diferentes Modelos Semánticos.

La construcción de modelos semánticos implica la evaluación de diferentes conceptos y abstracciones, a los fines de establecer cuales deben ser soportadas por los modelos, cuáles no, y aquéllos que se pueden construir a partir de otras de menor nivel. Las abstracciones se presentan en la sección 3.3.1, mientras que en la sección 3.3.2, se harán algunas consideraciones sobre un concepto central en la modelación semántica: la definición por intensión -y no por extensión- como una manera de aumentar el poder de expresión y la flexibilidad de un modelo.

3.3.1. Abstracciones

Se presenta un conjunto representativo de las abstracciones utilizadas por el hombre en la expresión

de su visión de su entorno, las cuales son soportadas por diferentes modelos semánticos.

3.3.1.1. Agregación.

En la realidad, muchos objetos pueden ser vistos como constituidos por partes diferenciadas, cada una con significado propio. *La Agregación es la abstracción que permite definir un concepto compuesto a partir de objetos más simples (partes)* [Smith 77a,b].

Ejemplo 3.3.1.1: En un archivo, cada registro es la **agregación** de un grupo de campos:

Código	Nombre	Dpto.	Salario
--------	--------	-------	---------

3.3.1.2. Generalización.

En ciertas ocasiones, es necesario considerar un conjunto de objetos como similares, aún cuando puedan presentar diferencias estructurales. *Generalización es la abstracción que permite crear un **concepto genérico** a partir de conceptos con propiedades semánticas comunes.* [Smith 77b].

Ejemplo 4.3.1.2: Barcos, carros y aviones se pueden abstraer en el concepto genérico como vehículos de transporte, a pesar de sus diferencias estructurales.

4.3.1.3. Clasificación

Agrupar objetos de un mismo tipo con propiedades particulares comunes, en clases disjuntas (tales que su unión represente el conjunto inicial de objetos en cuestión) es el mecanismo provisto por la **clasificación**.

Ejemplo 4.3.1.3: En una fábrica de baldosas, estas se **clasifican** según el material que las constituyen, en baldosas de alta, media y baja calidad.

4.3.1.4. Tipificación

Un mecanismo similar a la clasificación, donde las clases resultantes no son necesariamente disjuntas y el criterio de pertenencia para cada clase se puede formar a partir de distintas propiedades, constituye la **tipificación**.

Ejemplo 4.3.1.4.: En una librería se quiere tener una Base de Datos sobre Temas de Libros. Podrían tipificarse los libros como Libros de Inteligencia Artificial, Libros referentes a Computación Gráfica y Libros con temas de B.D. Es claro que un libro puede hablar de todos los temas, o sobre ninguno. Por ejemplo, un libro que trate sobre el Modelo Semántico puede estar entre los libros sobre Inteligencia Artificial y entre aquéllos sobre Base de Datos, mientras uno de Ingeniería Civil, no corresponde a ninguna de las clases consideradas en el ejemplo.

4.3.1.5. Relaciones n-arias

Asociaciones generales pueden ser modeladas, mediante **relaciones**. Estas pueden ser de aridad múltiple. Estas asociaciones comprenden el concepto de interacción entre objetos con semántica particular, para cada sistema considerado.

Estas relaciones pueden también poseer características inherentes a ellas (cardinalidad, atributos), producto de la asociación entre los objetos.

Ejemplo 4.3.1.5: En un ambiente de procesamiento de datos, puede interesar conocer que usuarios accesa alguna aplicación, residente en algún equipo, a través de algún terminal. Esto se puede representar como una relación cuaternaria entre terminales, aplicaciones, usuarios y máquinas.

3.3.2. Definición de tipos por Intensión.

La mayoría de los conceptos que maneja el ser humano se defiende por intención, más que por extensión. Es decir, se tiene una descripción genérica de las condiciones que debe satisfacer un objeto para ser representante de un concepto. En modelos de datos tradicionales, se ha dado cierta preferencia a la definición de conceptos por extensión; esto posiblemente como producto de la influencia del contexto dentro del cual surgieron tales modelos, en el cual la orientación era estandarizar la representación y manipulación de datos corporativos.

Un caso diferente es el de los modelos semánticos, a cuyo surgimiento está asociado el deseo de representar de la manera más fiel y natural posible el conocimiento humano; este entorno resulta claramente propicio para preferir la definición de conceptos por intención. El modelo S.O.O.M. soporta y favorece definiciones por intención, mediante aserciones sobre los objetos que sobre él se definen.

4. EL MODELO S.O.O.M.

El producto de la selección de las características que debería tener un Modelo de Datos que se hizo en el capítulo anterior es el Modelo S.O.O.M., que se presenta a continuación, describiendo, ante todo, los conceptos y abstracciones que soporta, para luego definirlo más formalmente y mostrar la representación gráfica que se propone para su utilización en la modelación de casos reales.

4.1. Conceptos y Abstracciones Soportadas por el Modelo

El modelo soporta tres conceptos fundamentales: **clase, relación y jerarquía de clases**.

Clase: Un conjunto de entes con propiedades comunes, constituye un tipo o **clase**. Los entes que conforman una clase son llamados **objetos**. Una clase puede ser: un **tipo concreto**, esto es, predefinido por el ambiente (p.ej., enteros); o un **tipo abstracto**, constituido como abstracción de algún conjunto, compuesto o nominal de tipos. A cada clase se asocia un conjunto de **restricciones**, las cuales son explícitamente definidas a momento de modelación.

Relación: Una relación es, en términos matemáticos, un subconjunto del producto cartesiano de los dominios. Diremos que una relación es una **asociación** entre un conjunto de diversas clases (los dominios).

Una relación puede estar **parametrizada**, esto es, tiene atributos que no son propios de los objetos que asocia, sino de la asociación misma. Para cada dominio de una relación, se define su grado de participación en ella: **restricciones de cardinalidad**, en la cual se denotan las cotas inferior y superior para el número de objetos de esa clase (dominio) que pueden estar relacionados en una tupla de objetos del resto de los dominios. Otras restricciones: **aserciones** sobre los objetos participantes en la relación, pueden también asociarse a la relación, para denotar su **semántica**.

Jerarquía de Clases: Se puede establecer **jerarquías** entre clases, esto es, una clase puede tener propiedades (atributos o relaciones) que la identifican con otras clases. Esto se abstrae con los conceptos de **subclases** y **superclases**: diremos que una clase es una subclase de otra, la cual llamaremos superclase, si hay una jerarquía de la segunda a la primera.

Sobre la jerarquía de clases se establece un mecanismo que asocia a los objetos de la subclase, propiedades de la superclase, o viceversa: **herencia**. Se establecen tipos de herencia, según tipos de jerarquía. Cada tipo de jerarquía corresponde a una de las abstracciones presentadas en el Capítulo 3, y son:

Generalización: un conjunto de clases con características comunes (semánticas) puede abstraerse, mediante generalización, en una clase genérica, que puede ser atribuida por tales características. Esta abstracción establece un esquema de **herencia hacia arriba**, es decir, la clase resultante de la abstracción, hereda aquellas propiedades comunes que permiten la generalización.

Clasificación: Los objetos de una clase pueden ser agrupados en subclases **disjuntas**, mediante un conjunto de atributos que los discriminan. En una clasificación los objetos de la subclase heredan los propiedades de la superclase, esto es, **herencia hacia abajo**.

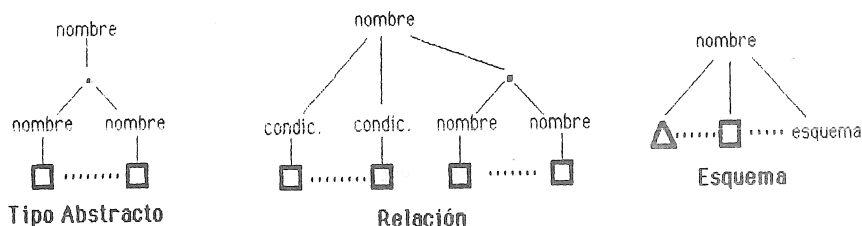
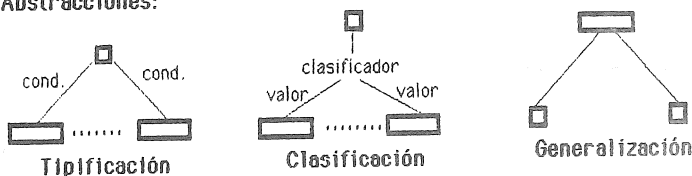
Tipificación. Dentro de una clase podemos observar objetos con características comunes que los agrupan, no necesariamente en forma disjunta. La abstracción que soporta tales agrupaciones es la **tipificación**. La tipificación induce un esquema de herencia hacia abajo.

2.4. Representación Gráfica de las Abstracciones y Conceptos soportados por el Modelo S.O.O.M..

Conceptos:



Abstracciones:



3.3. Ejemplo de modelación en S.O.O.M.

Para ejemplificar el uso de S.O.O.M. en la modelación de una Base de Datos, se presenta la Definición formal de un Diccionario de Datos para un Sistema de Bases de Datos que soporte el modelo. Esta definición forma parte de la implementación de un lenguaje de definición de datos para el modelo S.O.O.M. realizada en el proyecto de grado en el que fue definido del modelo que aquí se presenta.

3.3.1 Esquema en S.O.O.M. de un Diccionario de Datos.

Un diccionario de datos puede ser visto como una Base de Datos, que contiene datos sobre otras Bases de Datos, y su definición puede ser vista como la de un esquema.

Atendiendo a esta consideración, la definición del Diccionario de Datos, corresponde a la de un esquema en el modelo S.O.O.M. La definición se presenta utilizando la representación gráfica del modelo presentada en la sección anterior.

Esquema de la Visión del DDL del DD

Esquemas definidos como:

ESQUEMA

nombre

str

cuyos nodos son:

NODO_DEL_ESQUEMA

1:1

1:oo

NODO

ESQUEMA

siendo **nodo** una generalización de clases y relaciones:

NODO

CLASE

RELACION

CON:

RELACION

nombre

aridad

str

int

CLASE

nombre

discriminante

str

str

a cada **clase** se asocian:

CLASE_SE_ABSTRAE_EN

1:oo

1:1

CLASE

ABSTRACCION

donde:

ABSTRACCION

Arco_Generalización

Arco_Clasificación

Arco_Agregación

Arco_Tipificación

CON:

Arco_Generalización

ClaseGeneral

CLASE

Arco_Tipificación

ClaseTipific. predicado

CLASE

str

Arco_Agregación

identificador

tipo

offset

str

CLASE

int

Arco_Clasicación

ClaseClasific.

CLASE

asociada a:

CLASIFICADOR

1:00

1:1

Arco_Clasicación

Par_Atributo_Valor

Par_Atributo_Valor

con:

atributo

valor

Arco_Agregación

VALOR

donde:

VALOR

str

char

int

float

A cada relación se asocian:

PARAMETROS_DE_LA_RELACION

0:00

1:1

RELATION

Arco_Agregación

RESTRICCION SOBRE_RELACION

1:00

0:1

RELATION

- str

DOMINIOS_DE_LA_RELACION

1:00

1:1

RELATION

Arco_Relación

con:

Arco_Relación

CMin

CMax

dominio

offset

int

int

CLASE

int

CONCLUSIONES

Realizado el análisis de los principios básicos que debe cumplir un modelo (expresividad y economía conceptual), y una revisión de los paradigmas de modelación existentes en bases de datos, se definió un modelo de datos enmarcado dentro del paradigma semántico de modelación, el modelo SOOM.

El modelo S.O.O.M., aquí propuesto, maneja los conceptos de: clase (conjunto de objetos con características comunes) y relación (asociación entre clases); soporta las abstracciones: agregación (construcción de clases de objetos compuestos), generalización (clase que agrupa varias clases con semántica

común, clasificación (clases resultantes de la discriminación de objetos de una clase) y tipificación (clases resultantes de agrupar objetos pertenecientes a una clase con características particulares no excluyentes).

Presentamos el modelo S.O.O.M. como un modelo semántico porque: 1) distingue clases de relaciones; 2) provee abstracciones necesarias para la creación de nuevos conceptos a partir de los ya existentes; 3) favorece la definición de conceptos por intensión, mediante el uso de restricciones sobre clases y relaciones que denotan la semántica de éstas.

Además, el modelo S.O.O.M., ha sido enriquecido con la introducción de los conceptos de jerarquía y herencia entre clases (programación orientada por objetos), según las abstracciones de generalización, clasificación y tipificación.

REFERENCIAS

- Ballou, N, Chou, H-T., Garza, J.F., Woelk, D. "Features of the Orion Object Oriented Database System" en "Object Oriented Concepts, Databases and Applications", Part 3, Cap. 12, pp. 251-282. ACM Press, New York, 1.989.
- Beech, David. "Intensional Concepts in an Object Database Model". ACM OOPSLA Proceedings, Sept. 88.
- Bloom, T. y Zdonik, S. "Issues in the Design of Object-Oriented Database Programming Languages". ACM OOPSLA Proceedings, Oct. 87.
- Bretl, R, Maier, D., Otis, A., Penney, J., Schuchardt, B, Stein, J, Williams, E.H., Williams, M. en "Object Oriented Concepts, Databases and Applications", Part 3, Cap. 12, pp. 283-308. ACM Press, New York, 1.989.
- Brodie, M.L. "On the development of data models", en "On Conceptual Modelling Perspectives from Artificial Intelligence, Databases and Programming Languages, Brodie, Mylopoulos, Schmidt, Ed. Springer-Verlag, Nueva York, 1.984. pp. 19-48
- Caruso, M. y Zdonik, S. Report on "Object-Oriented Database Workshop: Semantic Aspects" ACM OOPSLA Addendum to the Proceedings, Oct. 1987.
- Chen, P. "Then entity-relationship model: Toward a unified view of data", ACM Trans. Database Syst. 1. pp. 9,36
- Codd, E.F., Extending the Database relational model to capture more meaning". ACM Trans. Database Syst., Vol. 4, Nº 4, Dic. 1.979, pp. 397-434.
- Date, C.J. An Introduction to Database Systems, Vol. II. Addison-Wesley Publishing Company, 1.984.
- Hammer, M. y McLeod, D. "Database Description with SDM: A semantic Database Model, ACM. Trans. Database Syst., Vol 6, Nº 3, Sept. 1981, pp. 351-386.
- Kim, W y Lochovsky, F. Edit. "Object Oriented Concepts, Databases and Applications" ACM Press, New York, 1.989
- Kim, W., Ballou, N., J., Chou, H., Garza, J. y Woelk, D. "Integrating an Object-Oriented Database Programming System with a Database System". ACM OOPSLA Proceedings, Sep. 88.
- Kim, W., Banerjee, J., Chou, H., Garza, J. y Woelk, D. "Composite Object Support in an Object-Oriented Database System". ACM OOPSLA Proceedings, Oct. 87.
- King, R. y MacLeod, D. "The Event database specification model". Proceedings of 2nd. International Conference on Database: Improving Usability and Responsiveness. (Jerusalem, Israel), 1.986, pp. 229-322.

- Maier, D., Stein, J., Otis, A., y Pardy, A. "Development of an Object-Oriented DBMS", ACM OOPSLA Proceedings, Sept. 86.
- Nierstrasz, O. "A survey on Object Oriented Concepts", en "Object Oriented Concepts, Databases and Applications", Part I, Cap. 1, pp. 3-32. ACM Press, New York, 1.989.
- Nixon, B, Chung L., Lauzen, I., Borgida A., Mylopoulos, J y Stanley, M. "Implementation of a compiler for a semantic data model: Experience with Taxis". Proceeding of the ACM SIGMOD Conference (San Francisco), ACM, 1.987.
- Peckham J. y Maryansky, F. "Semantic Data Models". ACM Computing Surveys, V. 20, #3, Sept. 88.
- Penney, D.J. y Stein, J.. "Class Modification in the GemStone Object-Oriented DBMS". ACM OOPSLA Proceedings, Oct. 87.
- Shipman, D.W., "The Functional data model and the data language DAPLEX", ACM Trans. Database Syst., Vol. 6, Nº 1, Marzo 1.981, pp. 140-173.
- Smith, J.M. y Smith, D.C.P., "Database Abstractions: Aggregation and generalization", ACM Trans. Database Syst. 2, 1.977 (b) pp. 105-133.
- Smith, J.M. y Smith, D.C.P., "Database Abstractions: Aggregation", Commun. ACM. Nº 20, 1.977 (a), pp. 405-413.
- Sundgren, B. "Conceptual Foundation of the infologic approach to data bases", Data Base Management, North Holland, Amsterdam, pp. 61-96, 1.974
- Teorey, T., Yang, D. y Fry, J.P.. "A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model". ACM Computing Surveys, Vol. 18, No. 2, Jun. 1.986.
- Thatte, S. Report on "Object-Oriented Database Workshop: Implementation Aspects". ACM OOPSLA Addendum to the Proceedings, Oct. 1987.
- Tsichritzis, D. y Lochovsky, F. Data Models, Prentice-Hall, 1.982.