

Organización del conocimiento y modelos de datos: la representación del universo del discurso

Jesús Tramullas Saz
Universidad de Zaragoza

0.1. Resumen

En este trabajo se analizan los principales mecanismos utilizados para la creación de modelos que representan objetos reales, con la finalidad de utilizarlos en la implementación de bases de datos. Se define el concepto de “modelo de datos”, insistiendo en la importancia de los mecanismos de abstracción utilizados por el diseñador. Por último, se revisa la importancia de estos mecanismos en la representación del conocimiento desde una perspectiva informativa.

Palabras clave: Bases de Datos. Modelos de Datos. Procesos de Abstracción.

0.2. Abstract

This paper analyzes the mechanisms applied to data models creation, which represents real objects, in order to use it in data base implementation. “Data model” concept is defined, Finally, abstraction tools in knowledge organization, from an informational perspective, are revised.

Keywords: Data Bases. Data Models. Abstraction. Processes and Tools.

1. Introducción

La Organización del Conocimiento (Currás, 1995), como disciplina, tiene la compleja misión de establecer mecanismos capaces de representar la realidad, enmarcando toda la posible riqueza de matices que ésta posee, al tiempo que debe proveer los mecanismos necesarios para que cualquier persona sea capaz de acceder, representar y comprender, casi con toda seguridad en diferente contexto, la realidad “original.” Esta realidad se compone de entidades, de características intrínsecas y extrínsecas pertenecientes a las entidades, y de las relaciones existentes entre las diferentes entidades (Dalhberg, 1995). Esta primera caracterización no puede, ni debe, hacer olvidar que todas las entidades se encuentran

inmersas en un contexto, el cual les otorga parte de su ser, pudiendo éste variar en función del propio contexto en el que se sitúe el objeto.

Lo delineado en el párrafo anterior sería perfecto, de no ser por un hecho: el ser humano, que por el momento cumple las funciones de describir, organizar e interpretar el conocimiento, es otro objeto situado en la realidad, sujeto a las mismas limitaciones que los objetos que describe, lo que añade un cierto nivel de subjetividad, y por ende de error. Para limitar en lo posible esa inseguridad, se ha creado un conjunto de métodos cuya finalidad es ofrecer al ser humano unos criterios rigurosos y permanentes para la representación de su objeto de conocimiento. Por supuesto, nos estamos refiriendo a los lenguajes documentales, en su más amplia acepción.

2. Representación del objeto documental

Los lenguajes documentales se utilizan para crear una representación característica e inequívoca de los objetos a los que teóricamente representan, con la finalidad, como ha señalado Maniez, de ordenar para recuperar (Maniez, 1992). Sin embargo, no presuponen ni suponen una representación de toda la complejidad del objeto: se refieren exclusivamente a su contenido, y limitan la representación de éste a un conjunto limitado de términos, probablemente significativos. Este mismo potencial de representación supone una cortapisa a la representación completa del objeto. En realidad, esta limitación hace que los lenguajes documentales se utilicen junto a una representación, más o menos formalizada, de otras características del objeto, generalmente relacionados con la apariencia física, o con procesos de producción y/o posesión legal del objeto.

De esta forma nos encontramos frente a una representación compleja, que se establece a través de un conjunto de atributos comunes a todos los elementos que pueden situarse dentro de una clase o tipo. Uno de esos atributos estará formado por la aplicación de un lenguaje documental, y el resto por otros atributos de diferente tipo, de forma que entre todos ellos ofrecen una imagen integrada. A este tipo de atributos, y a su estructura y organización, podemos definirla como “modelo de datos”.

3. Los modelos de datos

Tradicionalmente, se ha asociado el concepto “modelo de datos”, objeto de este trabajo, a la construcción de sistemas de bases de datos. Sin embargo, su importancia ha ido creciendo paulatinamente, de tal forma que la construcción de representaciones conceptuales de aspectos u objetos de la realidad, en cualquier campo, utiliza técnicas que se han originado y perfeccionado en el ámbito de las bases de datos, dado que el modelo de datos. La importancia del modelo de datos

en el proceso de construcción de un sistema de bases de datos se aprecia en la bibliografía sobre el particular (Elmasri y Navathe, 1989, p. 453-484).

Un modelo de datos es una representación conceptual de una parte del universo de discurso, seleccionada por su creador, que es utilizada para describir los elementos fundamentales de aquella, así como las relaciones que mantienen entre ellos. Como han señalado Miguel y Piattini, (1993, p. 162) llamamos modelo de datos

...al instrumento que se aplica a una parcela del mundo real (universo de discurso) para obtener una estructura de datos a la que denominamos esquema. Esta distinción entre el modelo (instrumento) y el esquema (resultado de aplicar el instrumento) es importante... Es importante también distinguir entre mundo real y universo de discurso, ya que este último es la visión que del mundo real tiene el diseñador... podemos definir un modelo de datos como un conjunto de conceptos, reglas y convenciones que nos permiten describir los datos del universo de discurso.

3.1. Modelo de datos y abstracción

El concepto fundamental que guía la creación y utilización de un modelo de datos es la abstracción. Si por abstracción entendemos la acción de: “1) considerar una cualidad, estado acción o fenómeno con independencia del objeto en que existe o por el que existe, o 2) formar la idea de un objeto separada de cualquier individuo en que se encuentra realizada”(1), resulta evidente que debe incorporar tanto la idea de totalidad del objeto, como la idea de las características individuales del objeto. Por supuesto, en el momento de crear un modelo de datos no se utilizan todas y cada una de las características del objeto a representar, sino aquellas que son pertinentes a la intención o intereses del creador.

Los procesos de abstracción utilizados en la construcción de modelos de datos pueden clasificarse dentro de uno de estos tres tipos (Batini, Ceri y Navathe, 1994, 18-22):

1. Abstracción de clasificación: se utiliza para definir un concepto como una clase de objetos de la realidad, que poseen propiedades comunes.
2. Abstracción de agregación: se usa para definir una nueva clase, que a su vez toma como punto de partida un conjunto de otras clases las cuales representan sus partes componentes.
3. Abstracción de generalización: establece relaciones de subconjunto entre los elementos de dos o varias clases.

Los modelos de datos definidos a través de los tipos de abstracción señalados deben ofrecer características de expresividad, simplicidad, minimalidad y formalidad.

3.2. Transformación del modelo de datos

El modelo de datos conceptual representa, a alto nivel, objetos y sus características objeto de tratamiento. Su representación no se encuentra limitada por ningún requerimiento externo. Simplemente, cumple los criterios definidos que facilitan la representación. Aspecto éste que cambia cuando se plantea, a un nivel inferior de tratamiento, la transformación del modelo conceptual representativo en una estructura manipulable por un sistema informático. En realidad, y volviendo al ámbito de las tecnologías de la información, concretamente de los sistemas de bases de datos, existe una notable diferencia entre un modelo conceptual, y su transformación (*mapping* es el término anglosajón) en un modelo utilizable por el sistema, al que se llama modelo lógico (Howe, 1983). Este modelo lógico puede corresponder en su totalidad, o no, al modelo conceptual, y resulta ser la representación del universo de discurso que se construye, como tal, en y para el sistema informático. Los modelos conceptuales son instrumentos para representar; los modelos lógicos son descripciones que simulan la realidad, nuevos objetos reales en sí mismos.

Esta transformación supone que la representación de la realidad original se ve constreñida por las limitaciones impuestas al seleccionar el tipo de modelo lógico a utilizar en la creación del sistema de bases de datos que representa a los objetos reales, sus relaciones y comportamiento. La tecnología utilizada determina las posibilidades de representación, en el nivel lógico. Los modelos de datos pueden dividirse en varios tipos (Korth y Silberschatz, 1993, p. 6-11):

1. Modelos lógicos basados en objetos: son los enfoques entidad/relación y orientación a objeto
2. Modelos lógicos basados en registros: son los enfoques relacional, jerárquico y de red
3. Modelo físicos de datos: unificador y de memoria de elementos

3.3. Objeto y acción

Las tendencias actuales se centran en los modelos lógicos basados en objetos. Los enfoques de entidad/relación y de orientación a objeto toman como punto de partida una representación, en el nivel conceptual, de los objetos presentes en el universo de discurso, y de sus características intrínsecas y extrínsecas. Dentro del mismo bloque se pueden colocar los enfoques (por ahora menos extendidos) infológico, semántico de datos y funcional.

Todos ellos se basan en una representación flexible y dinámica de la realidad. Esto supone que superan la concepción estática de las representaciones de objetos que ofrecen los modelos lógicos basados en registros. Los desarrollos más recientes en este campo tienden no sólo a representar los objetos, sino a incluir en

esta representación las acciones que ejecutan o reciben los objetos, o a dotar al sistema de bases de datos de cierto “conocimiento”, junto a reglas de inferencia, de tal forma que le sea posible generar nuevo “conocimiento” (por supuesto, entre comillas), y utilizarlo a su vez. De esta forma, la representación del universo de discurso se vuelve un sistema dinámico, no estático. Los nuevos, y ya cerca de su popularización como herramienta de usuario final, sistemas de gestión de bases de datos incorporarán la posibilidad de definir y utilizar objetos complejos, introducir jerarquías de objetos y aplicar reglas de inferencia (Chen y Dhar, 1990, p. 281-290).

4. Modelo de datos y calidad

La preocupación por la calidad se ha abierto paso en los últimos años con fuerza en el campo de las Ciencias de la Información, lo cual no es sorprendente si se analiza la creciente industria de la información electrónica. También se han desarrollado principios de calidad para la creación de modelos de datos, susceptibles de aplicación durante las diversas fases del proceso de creación. Los dos principales criterios son la corrección (conceptual y sintáctica) y la compleción (conceptual y semántica), entendiendo como tal la acción de completar todos los aspectos del modelo.

Como características deseables de un modelo de datos de alta calidad han sido definidas las siguientes (Reingruber y Gregory, 1994, p. 13-18):

1. Incluye los planes, políticas y estrategias de la organización
2. Utiliza un conjunto reconocido de reglas
3. Incluye a los expertos de los dominios de conocimiento
4. Puede transformarse en un diseño de alta calidad
5. Se crea en el contexto de otros elementos de la arquitectura de la organización
6. Se crea en el contexto de la organización completa
7. Se crea en el contexto de un ciclo de vida de calidad de los datos
8. Depende de la infraestructura de soporte
9. Implica a las personas adecuadas.

5. Conclusión

Todo lo anterior lleva a considerar la necesidad de contemplar tres panoramas diferentes, y a la vez complementarios, en el momento de establecer un modelo conceptual, tanto para sistemas de bases de datos, como para otros sistemas de representación:

1. En primer lugar, un modelo conceptual de información, entendiendo por tal un modelo de la estructura de información contenida en el universo de discurso representado, así como de la semántica que la acompaña y complementa.
2. En segundo lugar, un modelo conceptual de funciones, entendiendo por tal un modelo estructurado de las actividades, procesos o funciones que se tienen lugar en el universo de discurso representado.
3. Por último, un modelo conceptual de dinámica, entendiendo por tal un modelo que represente las dinámica, comportamiento y variaciones de los dos anteriores, desde una perspectiva espacial y temporal.

La aplicación de un modelo conceptual que responda a estos criterios fuerza a replantear la estructura de los lenguajes documentales que se utilizan en sistemas de representación basados en plataformas informáticas. Es evidente que los lenguajes documentales, en el momento actual, son un componente pasivo dentro de los sistemas de bases de datos. Se limitan a la utilización de un valor (o de varios) perteneciente a un dominio predeterminado. No toman en consideración la variación de los objetos que representan en el aspecto social, cultural, espacial o temporal. A este razonamiento cabría oponer que los objetos a cuyo contenido representan son estáticos. Sin embargo, la actual evolución de los documentos multimedia e hipertextuales, promovida en gran manera por la propia evolución de la publicación electrónica, y de las tecnologías que la sustentan, permiten variar el documento en forma y contenido tanto como se desee (Fong y Siu, 1996). Cuando varía el contenido, deben variar los elementos que lo describen y relacionan. Los modelos de datos ya han comenzado a incorporar este cambio en sus representaciones de los objetos contenidos en el universo de discurso al que hacen referencia. Al resultar la creación de modelos de datos una parte fundamental de la disciplina de organización y gestión del conocimiento, es forzoso plantear la posibilidad de cambios profundos en la estructura y dinámica tradicional de los lenguajes documentales. Al hablar de cambios estamos haciendo referencia a la creación y mantenimiento de lenguajes documentales que variarán a lo largo del tiempo, por lo que los mecanismos de control terminológico no sólo deberán apoyar en la introducción de representaciones de documentos en los sistemas: también deberán de ser capaces de cambiar, modificar o actualizar las descripciones automáticamente cuando sea necesario (Dalhberg, 1995, p. 20-22).

A su vez, estos lenguajes se utilizan como medio de representación y recuperación en el contexto de una organización determinada. El análisis de la organización, sus objetivos, estructura, funcionamiento y normas debe resultar obligado en la construcción de los modelos que representen a los documentos y a su contenido. Máxime cuando el concepto de “documento” está en permanente revi-

sión (Schamber, 1996: 669-671), especialmente en el momento en el que intervienen sistemas informáticos, bajo cuyo empuje ideas como “documento virtual”, “documento electrónico” o “documento dinámico” señalan de nuevo la necesidad de plantear modelos dinámicos de representación de los objetos que representan el conocimiento humano.

6. Notas

(1) Moliner, M. Diccionario de Uso del Español. Madrid : Gredos, 1990, t. 1, p.18.

7. Referencias

- Batini, C. ; Ceri, S. ; Navathe, S.B. (1994). Diseño conceptual de bases de datos. Un enfoque de entidades-interrelaciones. Wilmington : Addison-Wesley/Díaz de Santos, 1994.
- Chen, H.; Dhar, V. (1990). A Knowledge-Based Approach to the Design of Document-Based Retrieval Systems. // *Proceedings of Conference on Office Information Systems*, ACM SIGOIS Bulletin, 11 : 2-3 (1990), 281-290.
- Curras, E. (1995). Concierto y desconcierto en la organización del conocimiento actual y su intersección con el mundo de la información. // *SCIRE*, 1 : 1 (1995), 3-28.
- Dahlberg, I. (1995). Current trends in Knowledge Organization. // García Marco, F.J. (ed.) *Organización del Conocimiento en Sistemas de Información y Documentación*. Zaragoza: ISKO España y Vicerrectorado de Investigación de la Universidad de Zaragoza, 1995, p.10-14.
- Elmasri, R.; Navathe, S.B. (1989). *Fundamentals of Database Systems*. Redwood City : The Benjamin Cummings Pub.Co., 1989.
- Fong, J.; Siu, B. (Eds.) (1996). *Multimedia, Knowledge-Based and Object-Oriented Databases*. New York : Springer, 1996.
- Howe, D.R. (1983). *Data Analysis for Data Base Design*. Arnold, 1983.
- Korth, H.F.; Silberschatz, A. (1993). *Fundamentos de bases de datos*. Madrid : McGraw-Hill, 1993.
- Maniez, J. (1992). *Los lenguajes documentales y de clasificación. Concepción, construcción y utilización en los sistemas documentales*. Madrid : Fund. Germán Sánchez Ruipérez, 1992.
- Miguel, A. de; Piattini, M. (1993). *Concepción y diseño de bases de datos. Del modelo E/R al modelo relacional*. Madrid : RA-MA, 1993.
- Reingruber, M.C.; Gregory, W.W. (1994). *The Data Modelling Handbook*. New York : John Wiley & Sons, 1994.
- Schamber, L. (1996). What Is a Document? Rethinking the Concept in Uneasy Times. // *JASIS*, 47 : 9 (1996), 669-671