

La Fase de Abstracción Conceptual en Reingeniería de Bases de Datos mediante Análisis de Conceptos Formales¹

Carmen Hernández, Félix Prieto, Miguel A. Laguna, Yania Crespo

Dpto. de Informática, Universidad de Valladolid, España
Teléfono:+(34)983423670; Fax +(34)983423671
{chernan, felix, mlaguna, yania}@infor.uva.es

1 Contexto

La Reingeniería de Bases de Datos consiste, en primer lugar, en recuperar toda la información posible sobre la estructura de datos (conceptual, lógica y física) presente en una *Base de Datos Legada* (LDB) y en uno (o varios) *Sistema Software Legado* (LSS) para, posteriormente, rediseñar el esquema conceptual, transformar la LDB en la *Base de datos Nueva* (NDB) y el LSS en el *Sistema Software Nuevo* (NSS) y acometer la migración de los datos de la LDB a la NDB. La primera fase de este proceso se denomina *Ingeniería Inversa de Bases de Datos* y está formada por dos actividades principales que se pueden denominar *Análisis de Datos* y *Abstracción Conceptual* [6].

La actividad del *Análisis de Datos* intenta recuperar un esquema de datos lógico actualizado que esté completo estructuralmente y con su semántica bien identificada y documentada [5]. Este esquema lógico enriquecido es la entrada para la segunda actividad, la *Abstracción Conceptual*.

La *Abstracción Conceptual* intenta transformar este esquema lógico a un esquema conceptual equivalente. Las representaciones habituales de este esquema se hacen en términos del Modelo Entidad-Relación o del Modelo Orientado a Objetos. Ambos proporcionan el nivel de abstracción suficiente para acometer las posteriores actividades de *Reingeniería*.

El proceso de la *Reingeniería de Bases de Datos* no se ajusta a un modelo en cascada, en el que la actividad de *Abstracción Conceptual* no se habría de comenzar hasta que no se hubiera terminado la de *Análisis de Datos* [5]. Los ingenieros comienzan con las actividades propias de la Ingeniería cuando el esquema lógico aún está incompleto o es inconsistente. Según vayan acumulando más conocimiento sobre la estructura de la LDB volverán atrás en el proceso e irán completando el esquema lógico inicial.

Por esta razón se hace imprescindible definir un proceso para la obtención y transformación de estos esquemas que pueda ser automatizado, interactivo e iterativo, hasta conseguir obtener el esquema conceptual más conveniente y, además, un método que soporte la migración automática de los datos de la LDB a la base de datos que finalmente se construya.

¹ Este trabajo se desarrolla dentro del Grupo de Investigación en Reutilización y Orientación al Objeto (GIRO) de la Universidad de Valladolid, y ha sido parcialmente financiado por el proyecto CICYT TIC-2000-1673-C06-05.

2 Abstracción Conceptual mediante ACF

En este trabajo presentamos una propuesta que intenta responder a algunas de las tareas que se han de realizar en la fase de *Abstracción Conceptual*: la transformación del esquema lógico enriquecido a un esquema orientado al objeto y el soporte a las iteraciones. La propuesta consiste en abordar la consecución de dichas tareas aplicando técnicas de *Análisis de Conceptos Formales (ACF)*, de manera que va a ser posible construir una herramienta, basada en *ACF*, automática, gráfica e interactiva que soporte las iteraciones del proceso, ya que se van a poder marcar las zonas de impacto de las últimas modificaciones en el esquema lógico y comprobar hasta qué punto afectan al esquema conceptual obtenido. Con estas técnicas, se va a obtener un diagrama de clases que responda al modelo objeto definido por UML y que refleje el modelo que subyace en el esquema lógico de partida. En el diagrama de clases se obtendrán las clases significativas del problema, organizadas en una jerarquía especialización/generalización y relacionadas mediante asociaciones (como marco general para el resto de relaciones de UML, e.g. agregaciones), posiblemente caracterizadas por clases asociación.

El *ACF* [3] es una técnica matemática de organización de la información que permite poner de manifiesto las abstracciones subyacentes en una tabla de datos, formalmente un contexto, mediante la construcción de un retículo de conceptos, también conocido como retículo de Galois, asociado a ésta.

Partimos de un esquema lógico enriquecido y mediante aplicaciones de *ACF* obtenemos un diagrama de clases que responde al modelo objeto de UML.

El proceso que proponemos para la obtención del diagrama de clases parte de la siguiente información:

- Una lista de las relaciones y de sus atributos,
- las dependencias de inclusión [2] y
- las claves candidatas y las claves foráneas de las relaciones.

Como resultado, en el diagrama de clases obtenido se tendrán:

- las clases significativas del problema que hayan sido identificadas,
- la jerarquía de especialización/generalización que se da entre estas clases y
- las relaciones de asociación que se dan entre ellas. El establecimiento de las asociaciones puede dar lugar, si así lo requiere el esquema de partida, a la aparición de nuevas clases, como clases asociación, que caractericen la relación definida.

El proceso se divide en dos partes [4]. En la primera fase se tiene un doble objetivo: la identificación de las clases significativas y la clasificación de éstas en una jerarquía de especialización/generalización. La aplicación del *ACF* para cumplir este objetivo no es nueva. De hecho, la mayoría de los trabajos que se encuentran sobre aplicación del *ACF* van dirigidos hacia este objetivo. En [1] se afirma que la jerarquía que se puede obtener de la correcta aplicación del *ACF* en este sentido es la jerarquía que cumple con el criterio de mejor factorización de las características de las clases. La segunda fase pretende descubrir las relaciones de asociación y su caracterización mediante clases de asociación. Hasta donde nos consta, la forma de aplicación del *ACF* que proponemos para la segunda parte no se ha utilizado hasta el momento.

Este proceso se realiza de forma automática gracias a los algoritmos existentes para el trabajo con *ACF*. Una vez obtenido el nuevo esquema, el soporte a las iteraciones

vendrá dado por el análisis del impacto que tengan los cambios en el esquema lógico sobre el esquema conceptual. En este sentido las propiedades de los retículos de Galois permitirán aislar convenientemente estas modificaciones en el esquema conceptual, facilitando notablemente el proceso completo.

Referencias

- [1] M. Dao, M. Huchard, T. Libourel, and C. Roume. “*Evaluating and Optimizing Factorization in Inheritance Hierarchies*”. Proceedings of The Inheritance Workshop. ECOOP’2002 Workshop, June, 2002.
- [2] R. Elmasri and S. B. Navathe. “*Fundamentals of Database Systems*”. Benjamin/Cummings, Redwood City, 2nd edition, 1994.
- [3] B. Ganter and R. Wille. “*Formal concept analysis: mathematical foundations*”. Springer, 1999.
- [4] C. Hernández, F. Prieto, M.A. Laguna y Y. Crespo. “Formal Concept Analysis support for Conceptual Abstraction in Database Reengineering”. In *Proc. of the Database Maintenance and Reengineering Workshop (DBMR’2002)*, Montréal, Canada. October 2002. (ICSM 2002 Workshop).
- [5] J.-H. Jahnke and J. P. Wadsack. “*Varlet: Human-Centered Tool Support for Database Reengineering*”. Workshop on Software-Reengineering, Bad Honnef, Germany. Technical Report (Fachbericht INFORMATIK 7/99), University Koblenz-Landau. J. Ebert, B. Kullbach, F. Lehner (Editors). May 1999.
- [6] Hausi A. Müller, Jens H. Jahnke, Dennis B. Smith, Margaret-Anne D. Storey, Scott R. Tilley, Kenny Wong: "Reverse Engineering: A Roadmap," *The Future of Software Engineering Track at the 22nd International Conference on Software Engineering (ICSE 2000)*, Limerick, Ireland, June 4-10, 2000.