

UN SISTEMA EXPERTO PARA EL DISEÑO
DEL MODELO CONCEPTUAL
DE UNA BASE DE DATOS

Jorge L. del Mar Arzola
Lima - Peru

RESUMEN

El diseño de una Base de Datos (BD) ha sido considerado un arte, en vez de una ciencia, debido a los problemas existentes en esta área de la Informática.

La esencia principal de estos problemas se deben a la falta de una adecuada estructuración en las metodologías de diseño de BD y a la carencia de soporte automatizado para el desarrollo de este proceso.

En la actualidad, un enfoque para solucionar estos problemas es utilizar las herramientas que nos brinda la Inteligencia Artificial (IA) para facilitar el proceso de diseño de una BD.

En el presente trabajo se ofrece una descripción de las características y capacidades de un Sistema Experto (SE) que brinda soporte en el diseño del modelo conceptual de una BD.

Además, se presentan los criterios de diseño que se utilizaron en el desarrollo del SE.

1.0 INTRODUCCION

El desarrollo actual de la tecnología informática pone en evidencia una tendencia muy general en las Organizaciones por diseñar e implementar una BD.

Tradicionalmente, el diseño de una BD se ha realizado en base a los criterios y la experiencia de un "experto diseñador de BD" que obtiene los requerimientos informáticos de la Organización mediante entrevistas a los usuarios, revisando los documentos y los Sistemas de Información (SI) existentes. Luego, el diseñador de la BD integra los requerimientos de información, de forma tal, que se integran los requerimientos de los SI de la Organización (existentes en el presente y los que se se desarrollen en el futuro).

A continuación, el usuario y el diseñador validan el modelo conceptual obtenido después de varias interacciones.

Cuando se obtiene una version final del modelo conceptual el diseñador transforma el modelo conceptual de la BD en un modelo lógico de datos requerido para su implementación en una forma compatible con el Sistema de Gestión de Base/ de Datos (SGBD) que se utilizará en la Organización (por ejemplo, el modelo relacional).

Finalmente, el modelo lógico de datos es transformado en una forma compatible con la tecnología de Software y Hardware específico de la Organización.

La utilización del enfoque tradicional presenta los siguientes problemas:

1. El diseño de una BD se convierte en una tarea difícil y compleja que requiere la utilización de un recurso escaso: El experto diseñador de BD. Por otro lado, la calidad del diseño de la BD depende mucho de la experiencia y el conocimiento del experto diseñador de BD.
2. En la actualidad, el diseño de BD es un arte que tiene poco soporte de apoyo automatizado lo cual hace que su éxito dependa de la intuición y la experiencia del experto diseñador de BD. El diseño de una BD no es una tarea trivial que pueda realizarse fácilmente utilizando sólo lápiz y papel.
3. El éxito en el modelamiento de una BD, depende de la habilidad para percibir la configuración organizacional e identificar las entidades y relaciones en la Organización, este proceso generalmente es realizado con una mínima participación de los usuarios, y después que el experto diseñador de BD termina de elaborar el modelo conceptual de la BD, presenta los resultados a los usuarios para su refinamiento (generalmente el período de tiempo que transcurre entre la especificación de requerimientos de información y la presentación de resultados tangibles es muy

dilatado).

4. Generalmente, el experto diseñador de una BD no esta familiarizado en un nivel de dominio de los sistemas aplicativos de la Organizacin.

Muchos investigadores en el área de BD han propuesto que los SE son una poderosa solución que brinda soporte al diseño de una BD de forma que se superen las deficiencias del diseño basado en el enfoque tradicional.

Los SE son programas basados en técnicas de IA que permiten aplicar aspectos del conocimiento humano, y ejecutar tareas que generalmente son realizados en base a la experiencia de los seres humanos.

La realización de un SE para el diseño de una BD se sustenta tomando como base las siguientes ventajas:

1. Si la metodología de trabajo de un experto diseñador de BD puede codificarse en un programa, entonces podría ejecutarse varias veces y aplicarse cuando se necesite, de esa manera, se superaría el problema que origina la escasez de expertos humanos.
2. Al codificar la metodología de un experto diseñador de BD en un SE que interactue con el usuario final, entonces el usuario - que es el que mejor comprende la aplicación - puede convertirse en el diseñador de su propia BD.
3. Se reduciría las etapas de diseño, desarrollo e implementación de una BD.
4. Las Organizaciones son entidades dinámicas que

pueden sufrir cambios rápidamente, lo cual trae como consecuencia que los requerimientos de información pueden variar, con un SE dichos ajustes se realizarían en forma oportuna e implementando rápidamente dichos requerimientos, aumentando la productividad y eficiencia al mejorar las condiciones de trabajo como consecuencia de la automatización del proceso de diseño de BD.

5. Algunas de las metodologías de diseño de BD tienen la orientación hacia el desarrollo de una aplicación y no desarrollan una BD integrada para la Organización, de forma que se brinde apoyo a todos los SI.

Por otro lado, la creciente necesidad de aumentar la eficiencia y efectividad en las Organizaciones ha generado, a su vez, un interés por analizar y perfeccionar los aspectos que inciden en el desarrollo de una BD, es por esta razón que es necesario desarrollar Software que brinde apoyo en el diseño de una BD de forma tal que los problemas ocasionados por un enfoque tradicional sean superados para maximizar las ventajas y beneficios que se obtienen de la filosofía inherente de la tecnología de BD.

El principal objetivo del presente trabajo es llenar el vacío que en la actualidad existe en el área de BD en nuestro país, desarrollando un SE que brinde apoyo en el diseño del modelo conceptual de una BD orientado a un amplio rango de Organizaciones.

2.0 Sistemas Expertos y Diseño de BD.

El tema de la automatización del proceso de diseño de una BD ha sido enfocado por varios investigadores. NAVATHE y GADGIL argumentan que los principales problemas en el diseño de una BD son ocasionados por la carencia de una metodología de diseño estructurado y la falta de ayudas automáticas para desarrollar BD complejas NAVATHE (8), para dar solución a estos problemas, desarrollaron una metodología para automatizar el proceso de integración de "vistas" y, toma como base el modelo de vistas N - S NAVATHE (7).

En 1.984, J. MYLOPOULOS plantea que un SE puede recolectar información de las características de una BD que pueda utilizarse para generar el diseño de una BD en forma automática o semi - automática MYLOPOULOS (6).

Por otro lado, en 1.985, M. BOUZEGHOUB y un grupo de estudiosos de la Universidad de París aplican la tecnología de SE para producir un esquema interno a partir de las especificaciones de un esquema conceptual BOUZEGHOUB (2).

En 1.985, HAWRYSZKIEWYCZ desarrolla un SE que brinda apoyo al diseñador de una BD en el desarrollo del modelo E - R utilizando un dialogo en lenguaje natural HAWRYSZKIEWYCZ (3).

Luego, en 1.986, KAWAGUCHI desarrollo el

"INTELLIGENT INTERVIEW SYSTEM", que es un SE que brinda facilidades para que los usuarios construyan su BD en sus áreas de especialización KAWAGUCHI (4).

En 1988, VEDA STOREY desarrolló el "VIEW CREATION SYSTEM", que es un SE que recolecta los requerimientos informáticos de los usuarios y a partir de esa información genera "vistas" de la BD STOREY (9).

Por otro lado, en 1989 TAUZOVIK desarrolló "MODELER", un SE que recibe información (entidades, relaciones y atributos) en un lenguaje que es una extensión de "LESK" (un lenguaje de adquisición de conocimientos basado en lógica de primer orden muy similar al inglés y luego produce un modelo conceptual expresado en base al modelo E - R extendido TAUZOVIK (11)).

En la actualidad, muchos investigadores están desarrollando estudios para desarrollar SE que automatizen el proceso de diseño de BD, el objetivo final es tener un SE que realmente funcionen como expertos diseñadores del BD y que produzcan diseños de alta calidad, el camino para alcanzar este objetivo es aún muy largo, pero es un reto interesante.

3.0 DESCRIPCION DE LAS CARACTERISTICAS FUNCIONALES DEL SE PARA EL DISEÑO DEL MODELO CONCEPTUAL DE UNA BD

3.1 DESCRIPCION DEL SE

El principal objetivo en el diseño del SE fue formalizar el proceso de diseño del modelo conceptual de una BD de forma tal que se puede automatizar dicho proceso y además se incrementa la calidad y consistencia del diseño de la BD. Además, se debe reducir la dependencia de un recurso escaso y costoso: La persona experta en el diseño de una BD.

La meta específica en el diseño del SE fue formular un conjunto de reglas que sirvan como la base de conocimiento del SE y que permita la generación del modelo conceptual de la BD, es decir, formalizar el proceso de diseño del modelo conceptual.

El SE actúa como un diseñador de una BD en la determinación de los requerimientos informáticos de una Organización.

El SE está diseñado para trabajar interactivamente, de forma tal que se consulte a los usuarios que utilizarán la BD sobre sus requerimientos informáticos.

Con la ayuda del usuario el SE emplea procedimientos de normalización para transformar los requerimientos de información en un conjunto de relaciones en Tercera Forma Normal que representa el modelo conceptual de la BD. La selección de la representación relacional fue arbitraria

porque de igual forma se pudo utilizar cualesquier otra representación (Red, E-R, etc.)

Un aspecto importante del SE es que es independiente de la Organización o de la aplicación en que se le utilice debido a que no contiene ningún conocimiento específico para una Organización o aplicación en particular.

Por otro lado, la base de conocimiento contiene la información sobre el proceso de diseño de una BD.

3.2 CARACTERISTICAS DEL SE

En la actualidad, el SE para el diseño del modelo conceptual de una BD esta compuesta por 1.000 líneas de código fuente escrito con TURBO PROLOG y se ejecuta en un microcomputador IMB PC/XT o compatible.

El TURBO PROLOG fue seleccionado como el lenguaje para el desarrollo del SE por la facilidad que brinda para la aplicación de la técnica de prototipo y la flexibilidad que provee al mantenimiento y el desarrollo posterior del SE.

La base de conocimiento del SE está compuesto por 50 reglas, muchas de las cuales a su vez contienen Sub-Reglas.

El SE ha sido diseñado específicamente para recolectar los requerimientos de información de los

usuarios de una Organización, enseñándole al usuario los principios del modelo E-R y brindando el soporte necesario para que el usuario pueda especificar el modelo conceptual de la BD.

Las características principales del SE son las siguientes:

a) Base de Conocimientos del SE

El conocimiento del SE esta basado en la teoría de diseño de BD y en nuestra experiencia en el diseño de BD. Además, la base de conocimiento ha aumentado en función de las experiencias que se ha tenido en la utilización y refinamiento del SE. El conocimiento del SE esta representado como un conjunto de procedimientos y de reglas de producción.

b) Rendimiento del SE

El SE ha sido probado utilizando varios problemas de diseño para una variedad de Organizaciones.

El SE trabaja muy bien con usuarios que tienen algún conocimiento previo de la teoría de BD. Para el caso de usuarios que no han tenido ninguna experiencia previa con los conceptos de BD es recomendable que utilicen el Tutor de Conceptos de BD (incorporado en el SE), antes de realizar el diseño de una BD, de forma tal que se puede explotar en su totalidad al SE.

El software desarrollado para dar soporte al diseño del modelo conceptual de una BD esta clasificado como un SE porque incorpora la experiencia humana en el diseño de BD.

c) Soporte del SE al Diseño de una BD.

El SE inicialmente se orienta a la fase de análisis de los requerimientos informáticos de una Organización. Luego, en el diseño conceptual se representan los requerimientos de información en término del modelo E-R. Finalmente, brinda un soporte al diseño del modelo lógico de datos utilizando el modelo de datos relacional.

d) Usuarios del SE

El SE ha sido diseñado para usuarios de una Organización de quienes no se espera que tengan conocimientos previos de diseño de una BD. Inicialmente, es preferible que los usuarios del SE tengan algún conocimiento de los conceptos básicos de BD. En todo caso, el SE brinda soporte para que el usuario aprenda los conceptos del diseño utilizando el modelo E-R y para que el usuario pueda aplicar dichos conceptos para resolver sus problemas.

La interfase con el usuario es de texto orientador para lo que se utiliza facilidades de "ventanas" y de "menús".

e) Independencia del Dominio de Aplicación

El SE es independiente del dominio de aplicación, debido a que no contiene ningún conocimiento previo sobre alguna aplicación en particular. Por otro lado, el SE tiene una capacidad limitada para aprender que atributos son comunmente utilizados, en un dominio de aplicación particular.

3.4 Procedimientos del SE

La principal tarea en el diseño del SE fue la construcción de su Base de Conocimientos. La Base de conocimientos del SE representa la formalización del proceso de diseño del modelo conceptual de una BD.

La forma más simple de comprender la estructura interna del SE es revisar sus procedimientos. A continuación, describiremos los principales procedimientos del SE.

a) Procedimiento de Inicio

Una sesión de diseño con el SE se inicia con la identificación del usuario. Una vez que el usuario ingresa a la sesión, el SE le brinda al usuario instrucciones de como se desarrollará la sesión.

Luego, el SE le consulta al usuario si desea revisar conceptos y definiciones básicas sobre diseño de BD, para lo cual se emplea el tutor del SE.

b) Procedimiento de Estudio de la Organización

El SE le consulta al usuario si puede describir la Organización en función de 4 niveles:

- Area Funcional
- Función
- Proceso
- Actividad

En caso afirmativo, el SE le solicita al usuario que ingrese las Areas Funcionales de la Organización (en caso de que existan).

Por cada Area Funcional de la Organización, el SE le pide al usuario que ingrese las Funciones que desarrolla el Area Funcional.

Después, por cada Función que el usuario ha ingresado, el SE le solicita al usuario que ingrese los principales procesos que se requieren para dicha función.

Finalmente, por cada proceso ingresado el SE solicita al usuario que le indique las principales Actividades que se requieren para realizar dicho proceso.

c) Procedimiento de Definición de Entidades

En este procedimiento, el SE le solicita al usuario que le brinde un conjunto inicial de entidades que se requieren para realizar cada una

de las actividades definidas en el procedimiento anterior.

Además, el SE requiere que por cada entidad se le indique sus atributos y claves candidatas.

d) Procedimiento de Definición de Relaciones

El SE le consulta al usuario si desearia revisar las entidades que ha definido en el paso anterior antes de definir relaciones.

Si es afirmativa su respuesta, el SE le presenta una lista de las entidades definidas.

A continuación el SE le pide al usuario que le ingrese las relaciones entre las entidades y que se los dé en la forma:

(ENTIDAD 1) (VERBO) (ENTIDAD 2)

Además, por cada relación, el SE solicita al usuario que le indique la cardinalidad entre las entidades y los atributos de la relación.

e) Procedimiento de Validación

A partir de la información recolectada en los procedimientos anteriores, el SE realiza una validación de esta información para asegurarse que es consistente y trate de identificar las entidades y verbos que representan la misma información en diferentes formas.

f) Procedimiento de Identificación de Claves Primarias

El SE a partir de las claves candidatas determina en forma automática las claves primarias aplicando sus reglas internas.

g) Procedimiento de Normalización

Finalmente, el SE interactúa con el usuario para identificar y resolver las dependencias funcionales antes de presentar los resultados finales.

4.0 CONCLUSIONES Y RECOMENDACIONES

En el presente documento, hemos presentado un conjunto de criterios y experiencias en el desarrollo de un SE orientado al diseño del modelo conceptual de una BD; a éstos podemos añadir algunas conclusiones y recomendaciones.

4.1 Beneficios del SE

- Permite hacer operacionales las metodologías y técnicas de diseño de una BD.
- Implementa en forma práctica y controlada los conceptos de diseño de una BD e Ingeniería de la Información.
- Brinda facilidades para mejorar la calidad del diseño del modelo conceptual de una BD mediante su verificación en forma automatizada.

- Permite operacionalizar las técnicas de prototipo en el modelamiento de datos.
- Reduce la dependencia del experto diseñador de BD por parte de la Organización.
- Simplifica el mantenimiento del modelo conceptual de la BD.
- Permite reducir el tiempo de desarrollo de un Sistema de Base de Datos en una Organización.
- Brinda facilidades para que el diseñador de BD centre su trabajo en los aspectos creativos del diseño de la BD.
- Reduce los costos de diseño del modelo conceptual de una BD.
- Brinda facilidades para la participación de los usuarios en el diseño del modelo conceptual de una BD, de forma tal que aumente la calidad del diseño y una reducción de errores.

4.3 Desarrollo Futuro del SE

- a) El alcance del SE para el diseño del modelo conceptual de una BD, abarca la primera fase en el proceso de diseño de una BD, la orientación en el desarrollo del SE debe dirigirse a la conversión del modelo conceptual al modelo lógico para un SGBD específico y posteriormente la conversión del modelo lógico en un modelo físico.
- Además, se debe buscar la generación automática de programas de aplicación que utilicen la BD (con un

SGED específico) a partir del modelo conceptual de la BD y de las transacciones definidas por los usuarios.

- b) Por otro lado, se debe considerar que una mejora importante en el SE sería incrementar su conocimiento general y su habilidad para aplicar razonamientos llevados por el sentido común. Por ejemplo, un diseñador de BD generalmente es capaz de detectar que una entidad es una "persona" (es el caso de un "cliente", implícitamente sabemos que es una "persona").

El reconocimiento de este tipo de casos implica muchas cosas sobre los atributos y relaciones que el diseñador automáticamente puede tomar en cuenta. Además, incrementar la habilidad del SE para realizar este tipo de inferencias permite que la utilización del SE sea menos honerosa, de forma tal que esto permita mejorar considerablemente la calidad de los diseños producidos.

5.0 REFERENCIAS BIBLIOGRAFICAS

1. Borland International Inc.
"Turbo PROLOG"
Manual del Producto
1,986
2. Bouzeghoub M., G. Gardarin y E. Metais
"Database Design Tools: An expert System
Approach"
Project SABRE
Universite Paris
1,985
3. Hawryszkiewicz I. T.
"A Computer - Aid for E-R Modelling". Proc. 4 TH
International Entity - Relationship Conference
1,985
4. Kawaguchi A., Taoka N., Mizoguchi R., Yamguchi T.
y Kakusho O.
"An Intelligent Interview System for Conceptual
Design of Database"
Proc. ECAI
1,986

5. Mylopoulos J.
"Knowledge Representation and Databases"
A. Panel on AI and Databases Proc. Eighth
International Joint Conference on Artificial
Intelligence.
1,983
6. Mylopoulos J. y Borgida A., Greenspan S. y Wong H.
"Information System Design at the Conceptual Level -
the Taxis Project"
Database Engineering
Volumen 7, Número 4
1,984
7. Navathe S. B. y Elmasri R.
"Integrating User Views in Database Design"
IEEE Computer
1,986
8. Navathe S. B. y Gadgil S. G.
"A Methodology for View Integration in Logical
Database Design"
Proc. of the Eighth International Conference on Very
Large Data Bases
1,982

9. Storey V. C. y Goldstein R. C.
"A Methodology for the Creation of User Views
During Database Design"
ACM Transactions on Database Systems
Volumen 13, Número 3
1,988
10. Storey V. C.
"Knowledge - Based Approaches to Database Design"
University of Rochester
1,989
11. Tauzovich B.
"An Expert System for Conceptual Data Modelling"
Proc. 8 Th International Conference on Entity -
Relationship Approach
1,989