

Aplicación para normalizar bases de datos relacionales.

Ing. Sandor Escobar Ruiz.

Desarrollador en la Facultad 06, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

sescobar@uci.cu

Ing. Sisley Sosa Vazquez.

Desarrollador en la Facultad 06, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

ssosa@uci.cu

Msc. Yanet Espinal Martín.

Profesora de Programación, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

yespinol@uci.cu

Msc. Manuel Enrique Puebla Martínez.

Vicedecano de Producción e Investigaciones, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

mpuebla@uci.cu

RESUMEN.

El presente trabajo es el resumen de una investigación realizada con el objetivo de automatizar el proceso de normalización de bases de datos relacionales, a partir de un conjunto de relaciones con las que comienza el proceso y sus dependencias funcionales, además de permitir la generación del script asociado al modelo relacional para los principales sistemas gestores de bases de datos (SGBD).

Su elaboración estuvo soportada por el estudio de diferentes algoritmos de normalización, permitiendo la realización de análisis comparativos según algunos criterios de eficiencia; y para la implementación llegó a utilizarse, como entorno de desarrollo integrado (IDE), SharpDevelop 3.0.

La meta de esta aplicación es lograr tener un producto capaz de mejorar y automatizar todo el proceso de diseño de una base de datos relacional, así como que también pueda ser utilizarlo a modo de medio de enseñanza en cualquier centro de estudio.

Palabras claves: normalización de bases de datos; bases de datos; diseño de bases de datos.

INTRODUCCIÓN.

En el proceso de construcción de la mayoría de los sistemas informáticos, el diseño de las bases de datos (BD) ocupa un lugar importante, a tal punto que puede verse como un proceso relativamente independiente dentro del diseño de un sistema y compuesto por una serie de etapas para su desarrollo. Es por ello que resulta de interés el estudio de los problemas relacionados con el diseño de las bases de datos y la modelación de la información.

Cuando en 1970 el Dr. Codd propuso el Modelo Relacional, no podía pensar que, lo que se consideraba más bien una elegante teoría matemática sin posibilidad de implementación eficiente en productos comerciales, iba a convertirse, en los años ochenta, en la Segunda Generación de productos de base de datos, que actualmente aún domina el mercado.(1)

En el diseño de una base de datos relacional pueden encontrarse anomalías de inserción, eliminación, actualización de los datos, así como restricciones artificiales en la estructura de los mismos y dependencia entre ellos, implicando grandes problemas en la gestión y obtención de su información.

Para resolver esta situación se tienen que hacer reestructuraciones en el modelo de la base de datos, que afectan completamente la implementación de las diferentes funcionalidades del sistema y la estructura lógica de los datos almacenados, si los problemas no fueron detectados en la etapa de diseño, lo cual es muy común por no existir herramientas que comprueben y den solución a estas anomalías en la etapa del diseño. Por consiguiente, la no detección de este problema implicaría costo de tiempo a la hora de re-implementar la solución y llevaría a resultados incorrectos de la información que gestiona el sistema. Estos son problemas que dan lugar a la siguiente interrogante: ¿Cómo contribuir a eliminar la inconsistencia en los diseños de bases de datos relacionales?

Con vista a dar respuesta a esta pregunta se ha propuesto desarrollar un sistema para la normalización de bases de datos relacionales, que permita además la generación de script para los gestores más utilizados. Dicha herramienta constituirá un apoyo al diseño de la estructura lógica de los datos de cualquier sistema, y también se ha concebido como una herramienta educativa, que sirva para comprender y analizar cómo se llevan a cabo los procesos vinculados con la normalización.

DESARROLLO.

Definiciones básicas.

La normalización es una técnica para diseñar la estructura lógica de los datos de un sistema de información en el modelo relacional, desarrollada por E. F. Codd en 1972.(2) Es, además, la formalización de una serie de buenas prácticas o buenas intenciones a la hora de crear esquemas de bases de datos. Constituye una etapa posterior a la correspondencia entre el esquema conceptual y el esquema lógico, que elimina las dependencias entre atributos no deseadas.

Según el Grupo de Bases de Datos Avanzadas (LABDA), de la Universidad Carlos III de Madrid, consiste en la descomposición sin pérdida de información ni de semántica de la relación universal (o de una colección de relaciones equivalentes a la misma) en una colección de relaciones en el que las anomalías de actualización (inserción, borrado y modificación) no existan o sean mínimas.(1)

Se puede afirmar, basándose en los conceptos anteriores, que el proceso de normalización va reduciendo un conjunto de relaciones dado, a una forma más "deseable", o sea, más sencilla, más simple, "descomponiendo las relaciones" en otras relaciones mejores en términos de las operaciones que se ejecutan sobre ellas.

Para que un modelo relacional presente las condiciones mencionadas debe pasar por diferentes niveles de normalización o formas normales (FN), cada uno con sus características, que mejorarán las relaciones pertenecientes al modelo en cuestión. Se dice que una relación está en una determinada forma normal si satisface un cierto conjunto de restricciones, que se comentan a continuación.

Primera forma normal (1FN): En la primera forma normal todos los atributos son atómicos. Un atributo es atómico si los elementos del dominio son individuales. O sea, una columna no puede tener múltiples valores. Si a cada valor de X le corresponde un único valor de Y entonces a cada valor de Y le corresponde un único valor de X.

Segunda forma normal (2FN): Una relación está en segunda forma normal si está en primera forma normal y todos los atributos no clave (o que no formen parte de la clave) dependen por completo de la clave primaria.

Tercera forma normal (3FN): Una relación R está en tercera forma normal si está en segunda forma normal y todos los atributos no clave dependen de manera no transitiva de la clave primaria.

Forma normal Boyce-Codd (FNBC): Una relación está en forma normal Boyce-Codd si está en tercera forma normal y todo atributo no primario depende de la clave y no existe parte de la clave que dependa de un atributo no primario.

Solución propuesta.

El sistema Normalizador de bases de datos (NBD) cuenta con un diseño semejante al del editor ER/Studio (herramienta utilizada para el diseño de modelos conceptuales de bases de datos), pero su trabajo se inicia en el diseño del modelo relacional permitiéndole al usuario una interacción amigable y sencilla para su uso.

Además de las funcionalidades básicas como: crear un modelo con su nombre, atributos y dependencias funcionales, eliminar un modelo y/o editarlo, permite la normalización de dichos modelos hasta FNBC, en dependencia del criterio y la necesidad de cada usuario desarrollador de la BD, así como también brinda otros conjuntos de funcionalidades que a continuación se enumeran, propias de todo el proceso de estandarización y muy útiles para la comprobación y realización de estudios minucioso de las relaciones existente entre los datos que describen el modelo, dando la posibilidad de ver como suceden paso a paso para su estudio.

- Cierre de descriptor.
- Recubrimiento minimal.
- Cálculo de la clave.
- Implicación de una dependencia

- Equivalencia entre conjuntos de dependencias funcionales.

Los algoritmos de normalización que se desarrollaron están estrechamente relacionados con la teoría de la lógica matemática que describen dichos modelos o relaciones. En el caso particular para darle solución a la tercera forma normal, la aplicación permite seleccionar a los usuarios del sistema el algoritmo a usar, ya sea el algoritmo por descomposición planteado por Ullman o el algoritmo de síntesis desarrollado por Bernstein los cuales son los más conocidos y usados en esta temática, cada uno de estos con sus características específicas en dependencia de lo que se quiera. También el sistema brinda la posibilidad de llegar hasta FNBC, utilizando el algoritmo por descomposición extendido en caso de ser requerido por el usuario.

Es importante mencionar que existen otras formas normales como cuarta forma normal y quinta forma normal, pero para lograr un diseño adecuado y correcto, sin inconsistencia en la información que almacena y generada, basta con cumplir el tercer nivel, es por esto precisamente, que la aplicación fue desarrollada con vista a resolver sólo estas formas normales y no se profundizó en el estudio y desarrollo de las posteriores.

CONCLUSIONES.

Después de desarrollada la investigación descrita en este artículo, se ha llegado a las siguientes conclusiones:

- La utilización del sistema Normalizador de bases de datos, para el diseño de bases de datos relacionales, contribuye a mejorar y eliminar todos los problemas desencadenados provenientes de esta etapa.
- El sistema que se propone para elaborar bases de datos normalizadas, utiliza los algoritmos más eficientes para la automatización de todo el proceso de normalización o estandarización de bases de datos relacionales, pudiéndose convertir en una buena herramienta que potencia el trabajo relacionado con el diseño de bases de datos.
- La unión del proceso de normalización y la generación de script en el sistema Normalizador de base de datos, permitió la conformación de un sistema útil y de suma importancia para desarrollar cualquier sistema.

BIBLIOGRAFÍA.

- (1) Codd, E. F. "A relational model of data for a large shared data banks Com of ACM". *Communications of the ACM*. 13/6. 377-387. New York, ACM, 1970.
- (2) De Miguel, Adoración; Piattini, Marop; Marcos, Esperanza. *Diseño de Bases de Datos Relacionales*. Madrid, RA-MA, 1999.