

Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales

María A. Paz Menvielle, Juan C. Cuevas, Luis E. Damiano, Roberto M. Muñoz, Sergio Quinteros

Departamento de Ingeniería en Sistemas de Información

Facultad Regional Córdoba – Universidad Tecnológica Nacional

Maestro Marcelo López esq. Cruz Roja Argentina – Ciudad Universitaria - Córdoba

0351 - 4686385

pazmalejandra@gmail.com, juancarloscue@gmail.com, luis.damiano@gmail.com, rmunoz@sistemas.frc.utn.edu.ar, ser.quinteros@gmail.com

Resumen

El propósito de este proyecto de investigación es diseñar y construir un aplicativo que sea capaz de resolver situaciones reales en la obtención de diseños de modelos de datos en bases de datos relacionales, basado en el concepto informático de sistema experto y partiendo de un conjunto de datos desnormalizado.

El resultado de esta propuesta es lograr:

- Una herramienta a disposición del plantel docente y estudiantes de Gestión de Datos, asignatura de la Carrera de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Córdoba.

- Un aplicativo para la validación de conocimientos adquiridos por los estudiantes.

- Antecedentes para nuevos estudios relacionados y que amplíen lo logrado con ingeniería del conocimiento.

- La aplicación de conceptos y procedimientos de la inteligencia artificial en el proceso de normalización de una base de datos relacional.

- Una base de conocimiento en las dos especialidades: normalización de modelos de datos e ingeniería del conocimiento, que serán centrales en este proyecto y útiles como recursos de posteriores incursiones en investigaciones futuras.

Para todo esto es necesario desarrollar estudios y aprendizajes en el área de Ingeniería del

Conocimiento y del lenguaje CLIPS en particular, como lenguaje de programación a utilizar.

Palabras clave: normalización, modelo relacional, ingeniería del conocimiento, sistemas expertos, estructuras de datos, generador.

Contexto

En la asignatura curricular Gestión de Datos, del tercer nivel de la Carrera de Ingeniería en Sistemas de Información, se cumple con los contenidos mínimos fijados en la Ordenanza 1150/2007, donde la Universidad Tecnológica Nacional aprueba el Diseño Curricular de la Carrera y fija como contenido mínimo la técnica de “Normalización”, e implementa los descriptores de la Resolución Ministerial 786/2009, del Ministerio de Educación de la Nación, en la que explícitamente un descriptor es “Diseño, Administración y Gestión de Bases de Datos”, en la sub área Bases de Datos del área Tecnologías Aplicadas.

En la Ordenanza 1150/2007 se mencionan, para la asignatura, cinco objetivos generales, donde dos de ellos específicamente puntualizan:

- Desarrollar los conceptos de estructuración de los datos en dispositivos de almacenamiento

- Describir metodologías para el modelado de datos

El estudio de los modelos típicos y de sus variantes, cómo se comportan y el estudio de las

estructuras internas a los modelos y su caracterización, permitirá sistematizar este conocimiento.

El proyecto de investigación y desarrollo ha sido homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, reconocido con el código UTN1702 en el ámbito de la Universidad, por un período de dos años y a partir del 1° de enero de 2013. El proyecto se suma a los ya existentes en el Grupo de Investigación en Inteligencia Artificial, por su relación con la temática y porque está en el contexto de la línea de investigación del grupo.

Otros antecedentes, no menos importantes, son los proyectos de investigación desarrollados en el ámbito de la Facultad Regional Córdoba:

a) Administrador de Base de Datos Relacional TecnoDB. Finalizado en diciembre de 2007.

b) Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP. Finalizado en diciembre de 2010.

c) PROMETEO-Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales. Finalizado en diciembre de 2010.

Introducción

En este proyecto se pretende generar un aplicativo que, a partir de un conjunto de datos provistos por un usuario, a través de reglas y hechos expresados en lógica de predicados de primer orden, estructure un modelo de datos en el entorno del Modelo Relacional.

El alcance involucra la obtención de tablas normalizadas, hasta el cumplimiento de la tercera forma normal, y el señalamiento de las relaciones.

Este proyecto podría incluso ser ampliado, para alcanzar la implementación de mayor número de restricciones y reglas disponibles para situaciones especiales provenientes del mundo real.

La propuesta incluye el diseño y la construcción del aplicativo con un lenguaje de programación de sistemas expertos, siendo necesario desarrollar estudios y aprendizajes en el

área de Ingeniería del Conocimiento y del lenguaje CLIPS en particular.

En la implementación de la asignatura de referencia, Gestión de Datos, de la Carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional-Facultad Regional Córdoba, se desarrollan sólo temas relacionados a Base de Datos.

Los docentes de esta cátedra desarrollan funciones de administración, diseño y programación en bases de datos, aportando experiencia profesional en la solución de problemas. Uno de los temas que han generado debates en el seno de la cátedra, es el cómo abordar en forma ágil y clara el modelado de una base de datos. La discrepancia de cómo llegar al resultado, el cómo hacerlo, de qué forma, qué pasos realizar, los aspectos procedimentales o funcionales de las mecánicas que llevan al producto terminado, hace dificultoso establecer una metodología para la enseñanza. Se aclara que no está en dudas por ningún docente de la cátedra, que para lograr un modelado de una base de datos hasta tercera forma normal, se debe validar que el conjunto de las relaciones cumplan con satisfacer las reglas que establecen las formas normales, claramente definidas en [1], [2] y [3].

El grupo de investigación se planteó la posibilidad de concretar una investigación que brinde sostén y fundamento a una solución sistemática, viable para el tema de modelar una base de datos relacional.

La motivación de los docentes, por generar nuevos materiales y estrategias para la construcción de un espacio de investigación, permite que integrantes de la cátedra sean coautores en el libro Bases de Datos [4].

En el Grupo de Investigación en Bases de Datos (GIBD), del Departamento de Ingeniería en Sistemas de Información, de la Facultad Regional Concepción de Uruguay, generaron el proyecto 25-D040, referido a Métodos de acceso, consultas y aplicaciones en modelos de bases de datos no convencionales [5].

En cuanto a antecedentes de una herramienta que genere un modelo de datos relacional en forma automática, no se ha encontrado algo

disponible en el ámbito local académico ni en el empresarial. Sólo un producto comercial, generado por la empresa Axesor [6], de Granada (España).

Respecto de los conocimientos sobre la rama de Ingeniería del Conocimiento y Sistemas Expertos como productos, se hizo una investigación preliminar con acceso a bibliografía [7] [8].

El Sr. Sergio Antonio Becerra Zepeda realizó su Tesis para el Posgrado de Maestro en Ciencias Área Computación, Universidad de Colima (México), con el tema Bases de Datos Inteligentes, a utilizar como conocimientos válidos desde el ambiente académico [9].

Los Sistemas Expertos son una rama de la Inteligencia Artificial, que hace un amplio uso del conocimiento especializado para resolver problemas como un especialista humano. Enrique Castillo, José M. Gutiérrez y Ali Hadi definen: “Un sistema experto puede definirse como un sistema informático (hardware y software) que simula a los expertos humanos en un área de especialización dada” [10]. En su dominio de conocimiento, el sistema experto hace inferencias de la misma forma en que un especialista humano inferiría la solución de un problema. Al proceso de construir un sistema experto se lo llama ingeniería del conocimiento, y consiste en la adquisición de conocimiento a partir de un especialista humano o de otra fuente y su codificación en el sistema experto (García Martínez, R. - Britos, P.; 2004) [11]. Los sistemas expertos suelen estar diseñados de manera distinta a los programas convencionales porque los problemas no tienen generalmente una solución algorítmica y dependen de inferencias para obtener una solución razonable, considerando ésta como la mejor que se puede esperar si no hay un algoritmo disponible que ayude a obtener la solución (Giarratano, J. - Riley, G.; 2005) [12].

El lenguaje elegido inicialmente para el desarrollo del proyecto es CLIPS (C Language Integrated Production System), ya que permite la programación con paradigmas múltiples, proporcionando soporte para programación basada en reglas, orientado a objeto y por procedimientos, enfocándose principalmente en

los aspectos de la programación basada en reglas. Es una herramienta para el desarrollo de sistemas expertos, creada por la Software Technology Branch (STB), NASA/Lyndon B. Johnson Space Center (Alanis Barrera, Ricardo; 2005) [13].

Para la planificación, recopilación, generación de instrumentos y el desarrollo de la investigación se trabajará, entre otros, con el material bibliográfico “Metodología de la Investigación”, de César Augusto Bernal [14], y “Metodología de la Investigación”, de Hernández Sampieri, R. y otros [15].

Líneas de investigación y desarrollo

En el desarrollo de este proyecto se utilizará el método científico, en cuanto a la forma investigativa y el tratamiento de los datos, pero aplicando administración de proyecto con el método espiralado de seguimiento para la planificación del desarrollo, que estará basado en cuatro etapas a ser repetidas en forma cíclica hasta la culminación del proyecto: planeamiento, adquisición del conocimiento, codificación y evaluación.

La investigación abarca dos áreas de la informática, Base de Datos e Ingeniería del Conocimiento, y tomando en cuenta que hay inconmensurable conocimiento disponible en ambos campos, el tipo de investigación a aplicar es transversal, enfocada en conocimientos contemporáneos sobre ambos temas.

Este proyecto se enmarca en los recientes conceptos y métodos de Investigación, Desarrollo e Innovación (I+D+i), ligados a la mirada de desarrollo tecnológico y el ingreso del producto al mercado y su uso, basado en la secuencia: Síntesis y teoría; Explorar, hipotetizar y clarificar; Diseño, desarrollo y prueba; Implementación, estudio de eficacia y mejora; Aumento progresivo y estudio de eficacia.

Las acciones que se contemplan en este proyecto de investigación se describen en los próximos párrafos.

En la primera etapa de trabajo, se está haciendo la recopilación y detección de la existencia de diversas formas y criterios para obtener un modelo de datos normalizado. Según esto el dominio del problema es el área de ingeniería, específicamente en diseño de bases de datos relacionales. Con información recolectada y sistematizada, se hará un primer análisis de variables y se podrá comenzar con la selección de paradigma. Si bien el equipo entiende que el paradigma apropiado es el de programación en formato de Sistemas Expertos, el análisis de las variables, confirmará en su momento esta posición, o permitirá vislumbrar un paradigma diferente.

Determinado el paradigma de trabajo se realizará la selección de herramientas apropiadas para comenzar el desarrollo del producto, produciendo la adquisición del conocimiento y elegir el lenguaje de programación para un sistema experto. Como ya se indicó, se reconoce en CLIPS un lenguaje que puede brindar el potencial para el desarrollo.

Para adquirir conocimiento se utilizarán instrumentos de recolección de información y de medición de variables, dirigida a entrevistas con expertos en Bases de Datos, docentes de Gestión de Datos y diseñadores externos.

Teniendo el material recolectado, los análisis realizados y cerrando la etapa de adquisición de conocimiento, el equipo pasará al desarrollo del prototipo inicial. Se deberán definir las características de diseño y programación, la reingeniería del paradigma, si fuera necesario, la retroalimentación de ideas y conocimientos, perfilando el modelo de diseño y el producto.

Está previsto que durante el segundo año se comience el desarrollo del aplicativo, en donde se encaminará el diseño a una solución definitiva, pensando en el desarrollo de la herramienta de inferencia. Paralelamente se generará el plan de pruebas, sometiendo el producto a las pruebas para observar su comportamiento, produciendo una retroalimentación permanente. El mantenimiento y las mejoras de los detalles detectados, con la intención de que el producto logre un buen

grado de inserción y aceptación, generando innovación.

Un próximo paso será la obtención de un sistema mejorado y estable, que sea capaz de inferir o solucionar un conjunto de modelos y estructuras, incluyendo características deseables en una interfaz gráfica de usuario. Posteriormente está planificada la sociabilización del proyecto de investigación, a nivel académico y empresarial.

Resultados y Objetivos

El objetivo principal planteado es construir una herramienta capaz de resolver situaciones reales, obteniendo diseños de modelos en bases de datos relacionales, basada en el concepto de Sistema Experto y partiendo de un conjunto de datos desnormalizado.

El grupo se plantea los siguientes objetivos operacionales:

- *Permitir a los usuarios el cotejo y validación de los resultados obtenidos con la propuesta de normalización de la base de datos que brinde el producto, en relación a los datos suministrados.

- *Profundizar en el estudio y aprendizaje sobre Ingeniería del Conocimiento, con el fin de tener una visión acabada sobre el diseño del aplicativo.

- *Lograr que la herramienta emita un diagnóstico respecto de situaciones dudosas, realice sugerencia en relación a su base de conocimiento y efectúe preguntas adicionales sobre información de los datos suministrados.

- *Lograr incrementar el conocimiento avanzado sobre el modelado de bases de datos relacionales

- *Crear un método estandarizado para el abordaje y diseño de la normalización de bases de datos relacionales.

- *Brindar una herramienta didáctica al profesorado de la Cátedra de Gestión de Datos.

- *Introducir otra tecnología entre los recursos didácticos de que se dispone, pretendiendo que el aplicativo se encuadre como un producto de Tecnología de la Información y la Comunicación.

- *Difundir el conocimiento obtenido en el ámbito académico y público en general.

- *Transferir el conocimiento logrado en la temática de Base de Datos, Ingeniería del

Conocimiento, Diseño de Sistema Expertos, al ámbito académico y de investigación.

* Obtener un producto simple de utilizar y didáctico.

Formación de Recursos Humanos

El conocimiento obtenido y las conclusiones que se logren, en relación a temas de bases de datos y en sistemas expertos, aportarán seguramente al enriquecimiento en la formación de los integrantes del equipo de investigación. En dicho equipo trabajan una estudiante del último año de la Carrera de Ingeniería en Sistemas de Información, una ingeniera con título en trámite y tres estudiantes del cuarto nivel de la carrera, iniciando su formación en investigación científica y tecnológica. Los estudiantes podrán hacer la Práctica Supervisada de quinto año y acceder a Becas de Investigación de la Universidad.

Está planteado el desarrollo de la Tesis de la Maestría en Ingeniería en Sistemas de Información de un integrante docente, cursada en la Universidad Tecnológica Nacional.

Referencias

- [1] Elmasri, R. y Navathe, S. - Fundamentos de Sistemas de Bases de Datos-5ta Edición – Edit. Pearson-Impreso en España. 2007. ISBN: 978-84-7829-085-7.
- [2] Date, Christopher - Introducción a los Sistemas de Bases de Datos - Volumen 1 - Quinta Edición - Impreso en Estados Unidos. Edit. Addison Wesley Iberoamericana. 1993 – ISBN: 0-201-51859-7.
- [3] Silberschatz y otros- Fundamentos de Bases de Datos - Quinta Edición- Impreso en Estados Unidos. Edit. Mc Graw Hill- 2006. ISBN: 84-481-4644-1.
- [4] Reinoso, E.; Maldonado, C.; Muñoz, R.; Damiano, L.; Abrutsky, M. - Bases de Datos - Edit. AlfaOmega Editores- Argentina, 2012- ISBN: 978-987-1609-31-4.
- [5] Grupo de Investigación en Bases de Datos (GIBD) - Departamento Ingeniería en Sistemas de Información (DISI), Facultad Regional Concepción del Uruguay. Métodos de acceso, consultas y aplicaciones en modelos de bases de datos no convencionales.

Proyecto culminado: 25-D028 - Nuevas tendencias y aplicaciones en bases de datos.

<http://sistemas.frcu.utn.edu.ar/isi/index.php/gibd> - Accedido en mayo de 2012.

[6] Axesor - Sistemas Expertos en Normalización y Codificación de Bases de Datos- Producto Marketing Intelligence -

<http://marketing-intelligence.axesor.es/tratamiento-cualificacion/normalizacion> . Accedido en mayo 2012.

[7] Enrique Castillo, Ángel Cobo, José Manuel Gutiérrez y Rosa Eva Pruneda. Introducción a las Redes Funcionales con Aplicaciones – Un Nuevo Paradigma Neuronal. Editorial Paraninfo. Madrid, España. 1999. ISBN 84-283-2525-1.

[8] Ponce Cruz, Pedro - Inteligencia Artificial con Aplicaciones a la Ingeniería - Editorial ALFAOMEGA Grupo Editor, 2010 - ISBN 9786077854838.

[9] Becerra Zepeda, Sergio Antonio- Bases de Datos Inteligentes- Tesis para Maestro en Ciencias Área Computación - Universidad de Colima, México -Junio 1999.

http://digeset.uco.mx/tesis_posgrado/Pdf/Sergio%20Antonio%20Becerra%20Zepeda.pdf. Accedido en febrero 2012.

[10] Castillo, Enrique; Gutiérrez, José Manuel y Hadi, Ali S. Sistemas Expertos y Modelos de Redes Probabilísticas. Editorial Academia de Ingeniería. Madrid, España. 1998. ISBN 84-600-9395-6.

[11] García Martínez, Ramón - Britos, Paola Verónica - Ingeniería de Sistemas Expertos - Editorial Nueva Librería, 2004 - ISBN 9789871104154.

[12] Giarratano, Joseph - Riley, Gary - Sistemas Expertos Principios y Programación - Editorial Cengage Learning / Thomson Internacional, 2005 - ISBN 9789706860590

[13] Alanis Barrera, Ricardo- Sistemas Expertos e Inteligencia Artificial CLIPS - Burgos CyL., Febrero de 2005 - Universidad de Burgos: <http://es.scribd.com/doc/6732509/Clips>. Accedido en mayo de 2012.

[14] César Augusto Bernal - Metodología de la Investigación - Pearson Educación de México, 2006 - ISBN 970-26-0645-4.

[15] Hernández Sampieri, Roberto; Fernández Collado, Carlos; Baptista Lucio, Pilar - Metodología de la Investigación - MacGraw-Hill Interamericana Editores - México, 2006 - ISBN: 9701057538.