



INFERENCIA DE MODELOS DE BASE DE DATOS NORMALIZADOS

AGUSTIN CARRASCO
ASERMAX@GMAIL.COM

LUCAS GIACOMAZZI
LUCASGIACOMAZZI@GMAIL.COM

ANA BELÉN MERCADO PALAZZI
ABMERCADO19@GMAIL.COM

RESUMEN

La normalización de base de datos es una disciplina que se encuentra siempre presente en la tarea del Ingeniero en Sistemas e incluso de otros profesionales que trabajan con grandes volúmenes de datos.

Sin embargo, en la actualidad, esta tarea debe ser realizada por un experto en el área, o, en caso contrario, asumir el riesgo de generar un modelo subóptimo e incluso totalmente erróneo.

Se propone una posible solución a este problema, en base a la condensación del conocimiento en el área de la normalización de base de datos (a partir del estudio y elicitación de dicho conocimiento) en forma de reglas lógicas, que permitan a un aplicativo, en base a un conjunto de datos reales, la inferencia y generación de los modelos que puedan dar soporte a dichos datos y la forma en que los mismos se relacionan. De esta manera, el aplicativo se puede transformar en un asistente para todo aquel usuario que precise generar modelos y que no posea un experto al cual acudir.

ENTRADA

Para la entrada, se definió lo que denominamos un dataset; esta estructura define un dato particular perteneciente a un objeto del mundo real, como por ejemplo la fecha de una factura. Separando los datos en una porción atómica como ésta, se obtiene mucha flexibilidad en cuanto a los diferentes tipos de datos que se pueden insertar en el sistema.

En la estructura se define una entidad, un id único autogenerado y un atributo, que indican el tipo de objeto, la instancia en particular de esa entidad y el nombre del atributo al que el dato pertenece, respectivamente, y el dato en sí.

PROCESO

A pesar de que este tipo de aplicaciones no se ejecutan de manera lineal, a continuación se trata de explicar el flujo de datos para poder transformar los dataset de entrada en entidades con atributos y relaciones definidas, a partir del uso de reglas de inferencia definidas:

Captura de datos

Los datos ingresan al sistema de inferencia en forma de dataset, que se generan en base a la entrada del usuario o de otro sistema como forma de representación de los datos reales del sistema.

Transformación a modelos internos

A continuación se infieren las entidades básicas (que es muy probable que no estén normalizadas) que modelan dichos datos. Este procedimiento consiste en evaluar todos los dataset ingresados y aislar los diferentes atributos que conforman cada instancia de dicha entidad. A su vez se buscan patrones comunes que permitan encontrar propiedades de cada entidad, como potenciales claves primarias, posible subdivisión en otras entidades para normalizar la estructura, y así obtener un modelo inicial de las entidades que conforman el dominio a modelar.

Búsqueda de patrones relacionales y principios de normalización

Finalmente, se buscan patrones en las entidades creadas, y los datos utilizados para crearlos, que reflejen el carácter relacional de dichas entidades, es decir, las relaciones preexistentes entre los datos. También se evalúan otros patrones basados en los principios de normalización, que nos permitan dar una estructura al modelo con una mayor normalización.

Este procedimiento es el que más operaciones conocimiento-intensivo ejecuta, ya que se condensa en reglas de inferencia todo aquel conocimiento que permita diferenciar entidades que pueden (por la forma en que los datos ingresan al sistema) haber quedado juntas en una sola entidad, encontrar relaciones entre varias entidades cargadas individualmente, plantear definitivamente las claves primarias de cada entidad, así como también las claves foráneas que definen relaciones entre entidades.

RESULTADOS

La aplicación descrita previamente solo es capaz de procesar ciertos modelos predefinidos, por lo que no se puede decir que es un producto terminado. Sin embargo, los conceptos puestos en ella son sólidos y reflejan algunos de los conocimientos expertos en el área de normalización. Con ayuda del excelente motor de inferencia de CLIPS, es capaz de desempeñar su función correctamente para diferentes ejemplos de entrada, devolviendo como resultado modelos medianamente normalizados, y hasta cierto punto, usables para el diseño de una base de datos real.

Conceptos aquí expuestos son aptos para el desarrollo de una solución más completa para este problema.

A pesar de que solo se abarcó una pequeña parte del conocimiento referido al área de normalización, el análisis de dicho conocimiento desde el punto de vista de la Ingeniería de Conocimiento para el desarrollo de Sistemas Expertos resultó tremendamente provechosa, y, en nuestra opinión, da prueba de la potencia de esta disciplina y la viabilidad de solucionar este problema (y muchos otros) a partir de los conceptos y herramientas planteadas en este documento.

Contexto Académico

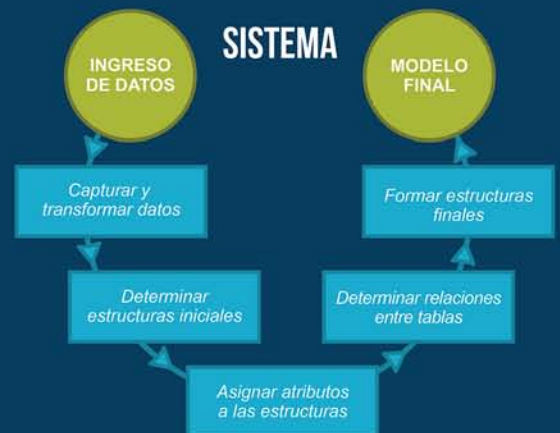
El seno de esta idea proviene del Proyecto de Investigación "Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales". Se trata de un proyecto homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, reconocido con el código UTN-1702 en el ámbito de la Universidad.

INGENIERIA DE CONOCIMIENTO

El problema analizado encaja en el área de la Ingeniería del Conocimiento, una disciplina de la Inteligencia Artificial que se concentra en el estudio y desarrollo de sistemas expertos.

Como explica Giarratano:

"...un programa de computación inteligente que usa el conocimiento y los procedimientos de inferencia para resolver problemas que son lo suficientemente difíciles como para requerir significativa experiencia humana para su solución... ¡Es decir, un sistema experto es un sistema de cómputo que emula la habilidad de tomar decisiones de un especialista humano!"



Arquitectura a futuro

La solución planteada es sin duda una porción de toda la extensión que puede adquirir los conceptos desarrollados en este trabajo. Es pertinente plantear una arquitectura a futuro que sea extensible en cuanto al tipo de entradas, los procesos de inferencia realizados y el tipo de respuestas de valor para el usuario; en base a esto, planteamos una organización básica del sistema (sin entrar en detalles), la cual consta de al menos tres capas bien diferenciadas:

Interfaces: esta capa contiene los módulos específicos que permiten la comunicación del sistema con el exterior. El desarrollo actual sólo permite el ingreso de datos por parte de usuarios, pero cabe la posibilidad de importar datos provenientes de otras fuentes.

Control: aquí se concentra la lógica común de control de la aplicación. En esta capa se procesan los datos de entrada y se les da el formato necesario para que puedan ser interpretados por el motor de inferencias. También servirá como nexo entre el usuario y la capa de conocimiento, facilitando el esclarecimiento de dudas, por parte del motor para la correcta generación del modelo.

Conocimiento: por último, la capa de conocimiento posee los diferentes módulos que concentran las estructuras de conocimiento (reglas y bases de conocimiento) a utilizar para la inferencia de los modelos y su validación.

INTERFACES	GUI	DBMS	...
CONTROL	RECOLECCION DE DATOS		
	LOGICA DE CONTROL		
CONOCIMIENTO	Inferencia de Modelos	Inferencia de relaciones	Validador de modelos de dominios específicos