

Tipo de artículo: Artículo original
Temática: Tecnologías de bases de datos
Recibido: 18/12/2013 | Aceptado: 7/05/2014

Un lenguaje natural controlado para especificar restricciones de integridad en el modelo ER

Controlled Natural Language for Integrity Constraints Specification in the ER Model

Alain Pereira Toledo ^{1*}, Alain Pérez Alonso ², Luisa M. González González ²

¹ Universidad de Sancti Spíritus “José Martí Pérez”. Avenida de Los Mártires, No 360, Sancti Spíritus, SS, Cuba. alain@uniss.edu.cu

² Universidad Central “Marta Abreu” de Las Villas. Carretera a Camajuaní km 3½. Santa Clara, Villa Clara, Cuba. apa@uclv.edu.cu; luisagon@uclv.edu.cu

* Autor para correspondencia: alain@uniss.edu.cu

Resumen

En la práctica los esquemas Entidad Interrelación generados durante la modelación conceptual no son capaces de especificar restricciones, necesarias para garantizar la integridad de un sistema de información. Aunque han existido intentos, la comunidad de diseñadores de bases de datos no ha adoptado una práctica para el diseño de la integridad de un sistema de información dado un esquema entidad interrelación del universo de discurso. Por tanto es necesario seleccionar un lenguaje con gran nivel de captación semántica para especificar restricciones. Para ello se eligió el lenguaje natural controlado RuleSpeak, unido al *Semantic of Business Vocabulary and Rules*. Finalmente, se muestra la especificación de aquellas restricciones de integridad definidas por el modelador e impuestas por un agente autorizado del negocio.

Palabras clave: lenguaje natural, restricciones de integridad, RuleSpeak, SBVR, sistemas de información.

Abstract

In practice, Entity Relationship schemes generated during conceptual modeling phase are not able to specify many of the constraints, to ensure the integrity of an information system. Although there have been attempts, the community of database designers has not adopted a practice for the design of integrity of an information system, from an ER

diagram of the universe of discourse. Therefore, it is necessary to select a semantically rich language, to specify constraints. In response, the controlled natural language RuleSpeak, bound to Semantic of Business Vocabulary and Rules was selected. Finally, the specification of integrity constraints defined by the modeler and imposed by an authorized business agent is shown. This approach helps to cover a lack of ER model in representing integrity, and helps in the understanding between the technical and business personnel.

Keywords: *information systems, integrity constraints, natural language, RuleSpeak, SBVR.*

Introducción

Desde que Peter Chen en su célebre artículo *The Entity Relationship Model: Towards a Unified View of Data* (Chen, 1976), lanzó al mundo su propuesta; el modelo Entidad Interrelación (ER) se ha convertido en una de las herramientas más populares dentro del diseño de bases de datos (Thalheim, 1999). No obstante, en la práctica los esquemas producidos mediante este modelo presentan algunas deficiencias (Fishman, 2003). Un buen modelo debe contener tres componentes fundamentales: la estructura, un conjunto de operadores que permitan la consulta y manipulación de los datos, y la integridad (Date, 2001). En el caso del modelo ER, solo existe soporte para las restricciones inherentes al modelo, pero prácticamente ninguna para aquellas impuestas por un agente autorizado del negocio. Si se toma en cuenta lo afirmado por (Date, 2001; Thalheim, 2011), más del 90% de los especificado durante el diseño de una base de datos, se circunscribe a las restricciones de integridad (RI).

Por su parte, el diseño de bases de datos comprende las etapas de diseño conceptual, lógico y físico. Por tanto, el diseño de la integridad debe transitar también todas las etapas. Sin embargo, aunque han existido intentos, la comunidad de diseñadores de bases de datos no ha adoptado una práctica para el diseño de la integridad de un sistema de información (SI), dado un esquema ER del universo de discurso (UoD). En (Thalheim, 2011) se muestra cómo aún permanecen abiertos algunos tópicos sobre áreas de las bases de datos, a pesar de ser relativamente antiguos. En especial, las RI no han recibido un tratamiento teórico unificado, y el modelo ER carece de un estándar, a diferencia de UML. No obstante, en lo tocante al diseño de la integridad en los SI, ya existen propuestas basadas en esquemas de transformación de lenguajes de modelado (Bajwa *et al.*, 2010).

Un enfoque muy utilizado, sobre todo por los practicantes del *Model Driven Architecture*, es el transformacional (Bowen and Hinchey, 1999). En especial, el enfoque *top-down* (Afreen *et al.*, 2011; Bajwa, Bordbar and Lee, 2010;

Bajwa *et al.*, 2011; Nemuraite *et al.*, 2010), establece la transformación de las restricciones desde el mayor nivel de abstracción (por lo general, a través de un lenguaje cercano al natural) hasta la tecnología de implementación. Por su parte, algunos han escogido un enfoque *bottom-up* para verbalizar las restricciones formales en un lenguaje natural (Pau and Cabot, 2008).

En respuesta a las deficiencias que muestra especialmente el diseño *conceptual* de bases de datos mediante el modelo ER, y en virtud de las tareas necesarias para contribuir a la solución; se propone en este trabajo:

- RuleSpeak como un lenguaje natural controlado, adecuado para la expresión de las RI durante la etapa conceptual.
- *Semantics of Business Vocabulary and Business Rules (SBVR)* como un lenguaje apropiado para, unido al RuleSpeak, especificar formalmente y con gran nivel de captación semántica, a las RI en la etapa conceptual.

No obstante, el alcance de tal propuesta no es definir un modelo definitivo para la especificación de RI. La intención es mostrar, más que demostrar, la posibilidad de representarlas, de manera que la comunidad relacionada con el UoD pueda entenderlas y que, a su vez, sean procesables por una máquina computadora.

Materiales y métodos

Aún son escasos los libros sobre bases de datos, que no dedican al menos un epígrafe para explicar los rudimentos de la construcción de esquemas ER. Quizás por ello, los diseños de bases de datos carecen de un correcto soporte para las RI que garanticen la integridad del sistema que se desea implementar. Y es que cuando Peter Chen ideó el modelo ER solo dio provisión para aquellas restricciones inherentes. Así que existen algunas restricciones que por su naturaleza o complejidad no pueden, o no deben, ser especificadas a través de diagramas (Fishman, 2003).

El modelo Entidad Interrelación

Según Olivé (2007), el compromiso ontológico es la forma o herramienta a través de la cual un modelador observa o representa el UoD. En especial, aquel modelo en que el mundo es descrito a través de entidades, atributos e interrelaciones se le denomina Modelo Entidad Interrelación (MER). Por tanto, todo esquema generado se representa a través de los tipos de entidades (conceptos que agrupan a entidades que poseen los mismos atributos), y de tipos de

interrelaciones que asocian a uno, dos o más tipos de entidades. En la Figura 1 se muestra un diagrama para un esquema ER de personas y sus cuentas bancarias, asociados a través del tipo de interrelación “tiene”.

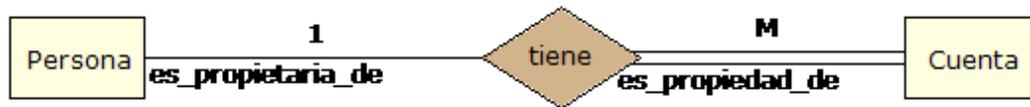


Figura 1. Esquema Entidad Interrelación de Personas y sus Cuentas

La amplia tradición de uso del MER en la comunidad de diseñadores de bases de datos, ha sido relacionada con su gran capacidad para representar significados de manera sencilla sin perder expresividad. Después de todo, el modelo ER es elogiado por haber dibujado la línea entre la expresividad y la sencillez en un buen punto (Badia, 2004). No obstante, algunos inconvenientes han surgido a partir de su popularidad, y es que ciertamente la utilización del enfoque ER no ayuda a los modeladores de bases de datos, a causa de la carencia de este modelo de un soporte adecuado para la integridad de los datos.

El MER solo da cabida a las restricciones inherentes, es decir, aquellas que de manera implícita son representadas a través de un esquema. Las restricciones de identidad de la entidad establecen que un conjunto de atributos identifican unívocamente cada una de las ocurrencias de un tipo de entidad. Por su parte, el modelo soporta algún tipo de integridad referencial a través de la restricción de participación obligatoria (llamada también de cardinalidad mínima uno), por la cual las instancias de un tipo de entidad tienen que estar interrelacionadas con al menos una entidad a través de determinada interrelación (Elmasri and Navathe, 2004). Ahora, restricciones tan sencillas como aquellas en que se limita el valor de un atributo de cierto tipo de entidad no pueden ser especificadas (Badia, 2004). En consecuencia, a través de un esquema ER solo es posible especificar la estructura del modelo de un SI.

El lenguaje SBVR

Ahora, ¿no sería el sueño de cualquier diseñador de bases de datos especificar las restricciones del sistema en un lenguaje lo más natural posible? En (Halpin and Morgan, 2010) se afirma que no solo es deseable, sino necesario expresar la descripción del UoD a través de un lenguaje natural.

El SBVR es un estándar adoptado por el OMG para servir de base a una descripción declarativa, formal y detallada de la lengua natural (OMG, 2013), es decir, es un lenguaje natural *controlado*. Para (Kuhn, 2013) un lenguaje natural controlado es un lenguaje construido, basado en algún lenguaje natural; y restrictivo en léxico, sintaxis y semántica a la vez que preserva sus propiedades naturales.

Como lo indica su nombre, SBVR se centra en la formalización del significado de las reglas y el vocabulario del negocio. Se afirma además que SBVR es una fórmula bien formada (OMG, 2013), por lo que una expresión del lenguaje siempre podrá ser validada sintácticamente. La ventaja de la formalización de un lenguaje natural, caso del SBVR, radica precisamente en la capacidad de ser analizado por sistemas informáticos.

Por otra parte, los conceptos en SBVR, están provistos de significado. Para ello se establece en (OMG, 2013) la manera de expresar la definición de cada término, notas aclaratorias y el esquema de referencia, que es el medio por el cual se establece a qué objeto del UoD se refiere. El resultado son conceptos con una expresión y significado asociado de gran nivel de precisión. Para los hechos y las reglas se proveen las formulaciones semánticas. A través de ellas es posible especificar y capturar la semántica de la expresión en el mismo registro natural. Para ello, SBVR establece un conjunto predeterminado de términos, es decir, un vocabulario para formular la parte semántica de la expresión.

Cómo expresar los términos, hechos y reglas del negocio no es normativo en SBVR. Esto significa que es posible elegir el medio de comunicación que desee el modelador, ya sea gráfico o textual. No obstante, en (OMG, 2013) se propone un lenguaje natural controlado como medio de expresión, el *Structured English*. Por su parte, SBVR se concentra en la intensión (significado), la representación (estructuras mediante las cuales son pensadas los significados), y en la extensión (a qué se refieren tales estructuras).

El lenguaje natural controlado RuleSpeak

Aunque *Structured English* es la propuesta de SBVR para el vocabulario y las reglas, otros posibles medios de expresión son mencionados en los apéndices de (OMG, 2013). Algunos son gráficos como el ya conocido ORM/NIAM de Terry Halpin (1998), y otros son textuales, como el RuleSpeak.

Ronald Ross es el creador de RuleSpeak (Ross, 1997), un lenguaje natural controlado, semejante al *Structured English*, y que ha sido adoptado por gran parte de la *Business Rules Community* (BRC), la cual agrupa a los

profesionales practicantes del enfoque de reglas de negocio. Originalmente fue concebido en inglés, pero ya cuenta con una versión al español (Catalá Cervera, 2009b), razón principal por la cual se eligió sobre su homólogo el *Structured English*. Además, cuenta con una mayor tradición en su uso, y con mayor número de recursos en cuanto a herramientas y entrenamiento patrocinados por la BRC.

La versión al español de RuleSpeak cuenta con una guía de plantillas para expresar las reglas según su tipo (Catalá Cervera, 2009b), y una guía sobre qué debe hacerse o no mientras se intenta escribir una regla del negocio (Catalá Cervera, 2009a).

Para escribir las RI en un CNL

Cuando se escribe una RI mediante una CNL, no se sigue ningún método formal. En realidad, es un proceso guiado por recomendaciones o principios que rigen la construcción “adecuada” de las restricciones (Witt, 2012).

Hay muchas maneras de escribir una regla, pero con un poco de entrenamiento, se puede lograr expresarlas con claridad y precisión, es decir, con un solo significado identificable. Téngase en cuenta que el entrenamiento necesario es mucho menor, comparado con técnicas tradicionales en las cuales el agente autorizado del negocio no conoce casi nada de la tecnología utilizada (Miguel Castaño and Piattini Velthuis, 1999).

El enfoque seguido parte del uso de plantillas. Las plantillas son estructuras sintácticas posibles en un lenguaje natural controlado, y se comportan como patrones que rigen la forma en que es posible expresar una sentencia, a la vez que crean un modo organizado y consistente de expresar las reglas.

Como se afirma en (Catalá Cervera, 2009), el propósito de las plantillas básicas es asegurar que las reglas escritas pueden comprenderse con facilidad. También ayuda a asegurar que los diferentes profesionales que trabajan sobre un gran conjunto de reglas expresen las mismas ideas, de la misma manera. Esta consistencia no se podría lograr si las reglas se expresaran de forma libre.

A continuación se muestra la restricción RI#1, para la cual se ha utilizado la plantilla que le sigue:

RI#1: La edad de una persona tiene que ser menor o igual a 30 años.

La|El <término1> {<cláusula de cualificación> | } tiene que <frase verbal> <cardinalidad> <término2>

Resultados y discusión

Con el MER, las plantillas RuleSpeak para SBVR, y la guía de principios y consejos sobre cómo escribir las restricciones, es que se tienen en la mano todas las herramientas y el método para iniciar el proceso de especificación.

Todos los términos utilizados durante el proceso deben estar contenidos en un esquema ER del UoD. Aunque el modelado en SBVR sigue el enfoque de modelo de hechos, es posible encontrar de manera muy práctica los términos y hechos fundamentales del negocio, implícitos en el esquema ER. Por tanto, la terminología utilizada en SBVR para referirse a las cosas que existen y sus relaciones en el UoD guarda relación, al menos parcial, con los componentes fundamentales del MER.

Tabla 1. Equivalencia ente la terminología utilizada por SBVR y ER para describir el UoD

SBVR	MER	Descripción
Noun concept	Tipo de entidad	Sustantivo común con un significado preciso en el UoD.
Characteristic	Atributo	Cualidad o rasgo de una persona o cosa.
Binary verb concept	Tipo de interrelación	Concepto de una frase verbal que involucra dos roles.
Role	Rol	Concepto que corresponde a las cosas basado en el papel que juegan.

La equivalencia no es necesariamente total

La equivalencia mostrada tiene un valor más práctico que formal, y solo tiene el propósito de encontrar la base para formular restricciones en SBVR que tomen sus constructos fundamentales de un esquema ER. En consecuencia, la Tabla 1 sirve de guía para identificar qué puede ser extraído del esquema ER con el objetivo de formular una RI. Las restricciones inherentes del modelo ER no serán analizadas ni expresadas en RuleSpeak con SBVR, pues ellas ya tienen cabida dentro del proceso tradicional de diseño de bases de datos.

En la Figura 1 se muestra una parte de un esquema ER de una entidad bancaria. *Persona* corresponde a un *noun concept*, mientras que *edad* es una *characteristic*, que en el diagrama de ejemplo no ha sido mostrado por razones de

espacio. Por tanto, es posible formular la restricción RI#1, pues los términos fundamentales exigidos por la plantilla están presentes en el esquema.

Aunque no es normativo, SBVR ha implementado un sistema de colores para identificar las partes de una regla. Ello se ha convertido en el sello de este lenguaje. RuleSpeak, que le antecedió por varios años, también ha acogido tal estilo cuando se utiliza en el contexto de SBVR. Así, la restricción RI#1 queda como sigue:

La **edad** de una **persona** tiene que **ser menor o igual a 30 años**.

En resumen, el diseñador/modelador elabora las RI en conjunto con un agente autorizado del negocio. Por su parte, las restricciones deberán estar basadas en los componentes de un esquema ER. La salida será un documento con las restricciones expresadas en RuleSpeak con SBVR. En la Figura 2 se condensa la práctica seguida en el proceso de especificación de las RI que no pueden ser representadas a través del esquema ER del UoD.

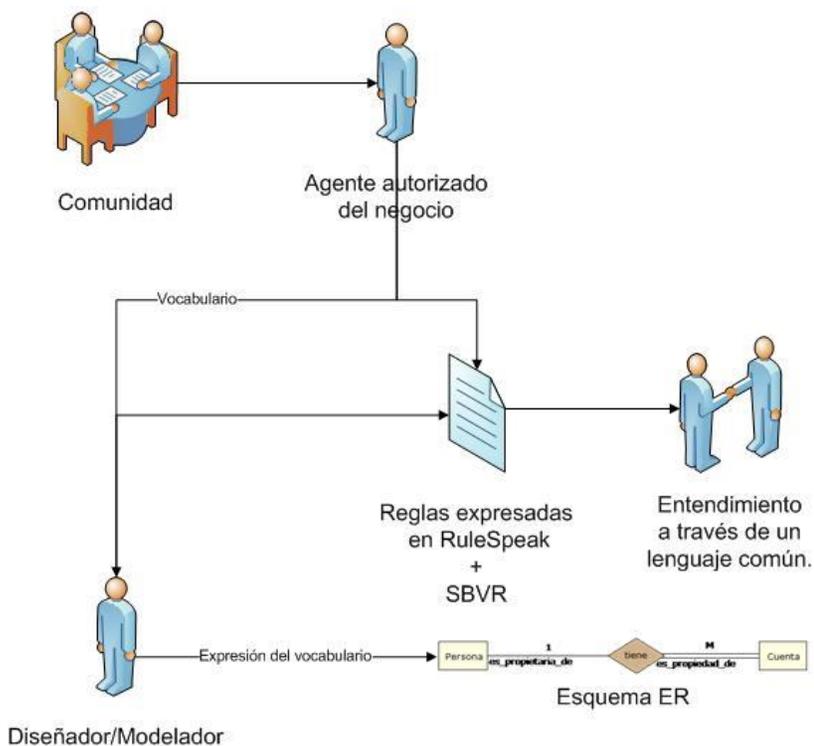


Figura 1. Esquema Entidad Interrelación de Personas y sus Cuentas

El entendimiento común entre el personal técnico y el del negocio se convierte en un resultado importante. No solo resuelve un problema muy común y antiguo en el diseño de sistemas de información, sino que ayuda en la búsqueda de una práctica en la comunidad de bases de datos, para tener en cuenta la integridad de los datos desde las primeras etapas hasta que finalmente es implementada. Esta aproximación ha sido suscrita por (Pereira Toledo *et al.* 2013a; Pereira Toledo *et al.* 2013b), donde se afirma que con el MER y las restricciones expresadas en RuleSpeak con SBVR, es posible encontrar una especificación destinada a la audiencia técnica, que pueda validarse a través de técnicas formales.

Conclusiones

Se encontró que RuleSpeak es no solo un lenguaje apropiado para expresar las RI asociadas a un esquema ER, sino que presenta recursos valiosos para el entrenamiento del personal técnico y del negocio involucrado en el diseño de la integridad de un SI.

SBVR, asociado a RuleSpeak para expresar (escribir) la RI, es un lenguaje adecuado que permite establecer el significado y la extensión, de los conceptos y sus restricciones. Su nivel de expresión natural, lo convierte en el lenguaje preparado para especificar las RI durante la etapa de modelación conceptual. Además SBVR goza de una correcta base formal que le hace computacionalmente analizable.

Finalmente se mostró la posibilidad de expresar aquellas RI que no pueden ser representadas por un esquema ER, el cual solo especifica los conceptos, características y relaciones que existen en el UoD, en conjunto con algunas restricciones muy comunes, tales como las restricciones de identidad de la entidad y las de cardinalidad.

Conocer cómo especificar adecuadamente las restricciones, dado un esquema ER del UoD, provee de un marco para lograr una generación automática hasta lograr una versión implementada en un futuro. Además, la expresión adecuada de las restricciones en un nivel conceptual, ayudará a cubrir una carencia del MER relativa a la representación de la integridad. No obstante, quedan preguntas abiertas sobre cómo expresar adecuadamente las restricciones y la estructura de manera tal que puedan ser validadas formalmente, qué reglas rigen la transformación hacia tales componentes formales y qué técnicas serán utilizadas en la validación formal.

Referencias

- AFREEN, H., I. S. BAJWA y B. BORDBAR. SBVR2UML: A Challenging Transformation. En *Frontiers of Information Technology (FIT), 2011*. 2011, p. 33-38.
- BADIA, A. Entity-Relationship modeling revisited. *ACM SIGMOD Record*, 2004, 33(1), 77-82.
- BAJWA, I. S., B. BORDBAR y M. G. LEE. OCL constraints generation from natural language specification. En *EDOC 2010 Conference*. 2010, p. 204-213.
- BAJWA, I. S., M. G. LEE y B. BORDBAR. SBVR Business Rules Generation from Natural Language Specification. En *AAAI Spring Symposium: AI for Business Agility*. 2011.
- BOWEN, J. P. y M. G. HINCHEY *High-integrity system specification and design*. Springer Heidelberg, 1999.
- CATALÁ CERVERA, A. Líneas guía básicas de RuleSpeak. Qué hacer y qué no hacer cuando se expresan reglas del negocio en lenguaje natural en castellano. v1.5, 2009a. Disponible en: <http://www.rulespeak.com/es/>
- CATALÁ CERVERA, A. Plantillas básicas de RuleSpeak. Especificando reglas del negocio en lenguaje natural en castellano. v1.5, 2009b. Disponible en: <http://www.rulespeak.com/es/>
- CHEN, P. P.-S. The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, 1976, 1(1), p. 9-36.
- DATE, C. J. *Introducción a los sistemas de bases de datos*. México: Prentice Hall, 2001.
- ELMASRI, R. y S. B. NAVATHE *Fundamentals of Database Systems*. Addison-Wesley, 2004.
- FISHMAN, N. *Server-Based Rules Enforcement*. Business Rules Solutions. LLC, 2003.
- HALPIN, T. Object-Role Modeling (ORM/NIAM). En P. BERNUS , K. MERTINS AND G. SCHMIDT eds. *Handbook on Architectures of Information Systems*. Berlin: Springer-Verlag, 1998.
- HALPIN, T. y T. MORGAN *Information modeling and relational databases*. Morgan Kaufmann, 2010.
- KUHN, T. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, 2013/06/26/ 2013, p. 1-50. Disponible en: http://dx.doi.org/10.1162/COLI_a_00168
- MIGUEL CASTAÑO, A. D. y M. G. PIATTINI VELTHUIS *Fundamentos y modelos de bases de datos*. RA-MA, 1999.

- NEMURAITÉ, L., T. SKERSYS, A. SUKYS, E. SINKEVICIUS, *et al.* VETIS tool for editing and transforming SBVR business vocabularies and business rules into UML&OCL models. En *16th International Conference on Information and Software Technologies*. 2010, p. 377-384.
- OLIVÉ, A. *Conceptual Modeling of Information Systems*. Edtion ed. Berlin: Springer, 2007.
- OMG. *Semantics of Business Vocabulary and Business Rules (SBVR) v 1.1*, 2013.
- PAU, R. y J. CABOT. Paraphrasing OCL Expressions with SBVR. En *NLDB*. 2008, p. 311-316.
- PEREIRA TOLEDO, A., A. PÉREZ ALONSO y L. M. GONZÁLEZ GONZÁLEZ. De SBVR hacia Alloy: una contribución a la transformación de restricciones de integridad. En *II Conferencia Internacional de Ciencias Computacionales e Informáticas*. La Habana, 2013a.
- PEREIRA TOLEDO, A., A. PÉREZ ALONSO y L. M. GONZÁLEZ GONZÁLEZ. Transformación de restricciones de integridad para el diseño de sistemas de información. En *Congreso Internacional COMPUMAT 2013*. La Habana, 2013b.
- ROSS, R. G. *The business rule book: classifying, defining and modeling rules*. Edtion ed.: Business Rule Solutions, Inc., 1997. 416 p.
- THALHEIM, B. The Strength of ER Modeling. En P.P. CHEN, J. AKOKA, H. KANGASSALO y B. THALHEIM. *Conceptual modeling: current issues and future directions*. Springer, 1999, vol. 1565, p. 227-242.
- THALHEIM, B. Integrity constraints in (conceptual) database models. En *The evolution of conceptual modeling*. Springer, 2011, p. 42-67.
- WITT, G. Writing Natural Language Rule Statements — a Systematic Approach: Part 1 — Basic Principles. *Business Rules Journal*, 2012, 13(7).