

MEDIDAS PARA LA EVALUACIÓN DE ESQUEMAS CONCEPTUALES¹

Piattini, M., Polo, M., Calero, C., Ruiz, F.

Grupo ALARCOS
Departamento de Informática
Escuela Superior de Informática
Universidad de Castilla-La Mancha
Ronda de Calatrava, 5
13071 Ciudad Real
{mpiattin, mpolo, ccalero, fruiz}@inf-cr.uclm.es

RESUMEN

Desde hace varios años se han propuesto diversas metodologías para modelado conceptual, pero en ninguna de ellas se proporcionan criterios cuantitativos que permitan evaluar los esquemas conceptuales. De hecho, la práctica totalidad de las métricas han sido desarrollada con el fin de controlar la mantenibilidad o complejidad de los programas, descuidando las bases de datos. En este trabajo se proponen diversas medidas para evaluar la calidad de los esquemas conceptuales, aplicándolas a un ejemplo.

Palabras clave:

Esquemas Conceptuales, Métricas, Calidad, Modelo E/R.

INTRODUCCIÓN

A la hora de representar estructuras conceptuales en la fase de modelado de bases de datos, suele utilizarse alguna variante del modelo E/R [1]. Desde hace varios años se han propuesto diversas metodologías para el modelado conceptual [2], pero ninguna de éstas proporcionan criterios cuantitativos que permitan evaluar los esquemas conceptuales, o elegir entre dos estructuras alternativas.

De hecho, Batini, Ceri y Navathe [3] proponen un conjunto de atributos de calidad pero no proporcionan para ellos ninguna medida numérica. Del mismo modo, Reingruber y Gregory [4] dan ciertas normas para construir esquemas de bases de datos de alta calidad, pero sin referencias cuantitativas. Kesch [5] constituye una excepción en este campo, ya que propone una serie de métricas para evaluar la calidad de esquemas conceptuales de bases de datos, pero a través de largos procesos manuales difícilmente automatizables.

Muy recientemente, Sneed y Foshag [6] han propuesto factores cuantitativos, cualitativos y de complejidad para bases de datos heredadas (*legacy*), restringiendo su aplicación, como es natural, a los esquemas lógicos y físicos, pero sin considerar los esquemas conceptuales.

Podemos afirmar que la elaboración de métricas para bases de datos (especialmente para esquemas conceptuales) es un área olvidada de investigación. La mayoría de las métricas propuestas desde el famoso número ciclomático de McCabe [7] hasta nuestros días se han centrado en la medida de la complejidad, calidad o mantenibilidad de programas [8-9].

Existen algunos proyectos en el área de las bases de datos que vienen desarrollándose desde mediados de los años ochenta, en los que diversos expertos han propuesto medidas con el fin de comparar “modelos” entendidos como formalismos para el modelado de datos. Diversos estudios [10-14] comparan diferentes técnicas (E/R, E/R extendido, NIAM, relacional, normalización, etc.) en cuanto a la facilidad para el modelado de datos. Pero hay que distinguir

¹ Este trabajo forma parte del proyecto MÁNTICA, parcialmente financiado por la CICYT y la Unión Europea (1FD97-0168).

claramente los modelos (como el E/R, relacional, etc.) considerados como técnicas, de los esquemas (resultados de aplicar los modelos a una parcela del mundo real).

Creemos más práctico la obtención de métricas para esquemas de bases de datos, que permitan a sus diseñadores evaluar de manera sistemática la calidad de los esquemas conceptuales a lo largo del proceso de modelado.

En esta comunicación presentamos un marco para la medición de atributos de calidad de estructuras conceptuales, siguiendo la filosofía de la norma ISO 9126 [15], que considera la calidad como la agregación de varios factores. El conjunto de factores de calidad que empleamos utiliza los propuestos en [3] complementados por algunos otros, todos ellos enfocados hacia la complejidad de las estructuras conceptuales.

Las medidas que proponemos utilizar para evaluar la calidad de los esquemas conceptuales se clasifican en medidas de Compleción, Corrección, Minimalidad, Expresividad, Legibilidad y autoexplicación, Extensibilidad y normalidad, Cohesión del esquema, Complejidad intra-entidad, Complejidad del esquema, Longitud interrelacional y Tamaño.

Este trabajo está organizado como sigue: en la sección 2 exponemos un conjunto de métricas para esquemas conceptuales. En la sección 3 aplicamos estas métricas a un ejemplo. Por último, exponemos nuestras conclusiones.

MÉTRICAS PARA EL ESQUEMA CONCEPTUAL

El objetivo del modelado conceptual es la obtención de una correcta representación de la organización de la información. Tal representación se realiza mediante un esquema conceptual, usando habitualmente el modelo entidad-interrelación.

Las medidas que proponemos utilizar para evaluar un esquema conceptual son:

Compleción

La completión es una métrica nominal que puede ser *alta*, *baja* o *muy baja*.

Corrección

Esta métrica puede ser también *alta*, *baja* o *muy baja*.

Minimalidad

La minimalidad se mide en función del número de atributos “sobrantes” (repetidos o calculables mediante operaciones matemáticas o inferencia lógica) presentes en el esquema. El ratio de minimalidad² se define como:

$$MR = 1 - \frac{NDA}{NAS} \quad \dots \text{donde:} \quad \begin{array}{l} \text{NDA es el número de atributos repetidos,} \\ \text{deducibles mediante operaciones o} \\ \text{derivables por inferencia lógica del esquema} \\ \text{NAS es el número de atributos del esquema} \end{array}$$

Expresividad

Proponemos los siguientes tres ratios:

$$SFAR = \frac{SFA}{NAS} \quad \dots \text{donde SFA es el número de atributos del esquema con} \\ \text{pleno contenido semántico y NAS es, como ya hemos} \\ \text{mencionado, el número de atributos del esquema} \\ \text{conceptual.}$$
$$SFER = \frac{SFE}{NES} \quad \dots \text{donde SFE es el número de entidades con pleno} \\ \text{contenido semántico y NES es el número de entidades} \\ \text{del esquema.}$$

² Las siglas utilizadas para nombrar las métricas corresponden a sus nombres en inglés.

$$SFRR = \frac{SFR}{NRS}$$

...donde SFR es el número de interrelaciones cuyos nombres tienen pleno contenido semántico y NRS es el número de interrelaciones del esquema.

De acuerdo con [4] un atributo, entidad o interrelación tiene pleno contenido semántico cuando su nombre es descriptivo y cumple las reglas de la organización.

Legibilidad y autoexplicación

Aunque esta métrica es la más subjetiva de las propuestas, no debemos olvidar a Fenton y Pfleeger [8] que afirman la importancia de “reconocer la utilidad de las métricas subjetivas, desde el momento en que entendemos su imprecisión”. Esta métrica es, como la compleción y la corrección, una métrica nominal que será *alta*, *baja* o *muy baja*.

Extensibilidad y normalidad

Ya que la tercera forma normal es la más natural para representar una base de datos, proponemos usar la siguiente métrica, debida a Gray [16]:

$$N = \frac{A_e \times E_e + R_e \times A_e + R_e \times E_e}{A_3 \times E_3 + R_3 \times A_3 + R_3 \times E_3}$$

...donde el numerador es calculado en función del número de atributos, entidades y relaciones del esquema real, y el denominador lo es en función del mismo en tercera forma normal.

“A” representa el número de atributos, “E” el de entidades y “R” el de interrelaciones.

Cohesión del esquema

Proponemos el uso del “Ratio de cohesión del esquema” para medir esta propiedad:

$$COSR = \frac{\sum_{i=1}^{[US]} NEUS_i^2}{NES^2}$$

[US] es el número de componentes conexas del grafo representado por el esquema conceptual.

NEUS_i es el número de entidades en la componente conexa i-ésima.

NES es el número de entidades del esquema.

Complejidad intra-entidad

La definimos como el número de interrelaciones de una entidad consigo misma. Ésta es una métrica para cada entidad; no obstante, siguiendo a Melton [9], podemos proporcionar una métrica válida para el esquema completo de esta manera:

$$IEC_{ESQUEMA} = (\text{media aritmética de IEC, desviación típica de IEC})$$

Complejidad del esquema

La complejidad del esquema se mide con el siguiente ratio:

$$SCR = \frac{SC}{MNR}$$

...donde: SC es el número de interrelaciones en el esquema y MNR es el número máximo teórico de interrelaciones en el esquema. Este “máximo teórico” viene dado por la siguiente expresión, en la que NES es el número de entidades del esquema:

$$MNR = \frac{NES \times (NES - 1)}{2}$$

Longitud interrelacional

Es la longitud del camino interrelacional más largo que podemos encontrar en el esquema, excluyendo las interrelaciones reflexivas.

Tamaño

Lo medimos como la siguiente tripleta, formada por el número entidades, el número de interrelaciones y el número de atributos del esquema:

$$S=(NES, NRS, NAS)$$

EJEMPLO

A continuación, aplicamos nuestras métricas a un ejemplo, tomado de [16], relativo a la base de datos de una biblioteca, cuyos requisitos se resumen en el siguiente cuadro, mostrándose el esquema conceptual correspondiente en la figura 1.

Se trata de diseñar una base de datos relacional que permita la gestión de préstamos de libros de una biblioteca. Si bien ya se han ido especificando a lo largo de estas páginas algunos aspectos de esta aplicación, vamos ahora a resumir y delimitar nuestro diseño.

Para ello consideramos que se parte de una serie de fichas que son las que se vienen utilizando hasta el momento en la biblioteca, y que son de dos tipos: fichas en donde se recogen las características de los libros y fichas relativas a los préstamos que se han efectuado, que recogen, por cada libro, la persona a la cual se le ha prestado, la fecha del préstamo y la de devolución. Además de estas fichas se ha recabado información sobre el sistema deseado mediante un conjunto de entrevistas con los usuarios que se puede resumir de la siguiente forma:

Para los libros interesa saber, además de lo que aparece actualmente en las fichas, el idioma en que están escritos.

Cada libro trata de uno o varios temas, lo que interesa reflejar para poder realizar consultas del tipo: "Libros o artículos que tenemos acerca de Bases de datos Multimedia", "Artículos que podemos consultar sobre el lenguaje QUEL",...

Los temas se pueden dividir en subtemas y así sucesivamente, por ejemplo, en el tema de DISEÑO podemos distinguir una serie de subtemas, como son DISEÑO FISICO, DISEÑO LOGICO y DISEÑO CONCEPTUAL.

De los autores, además del nombre, nos interesa conocer su nacionalidad y la institución donde trabajan, como Universidad de Stanford.

a) Compleción: podemos suponer, vistos los requisitos, que éstos se encuentran completamente recogidos en el esquema entidad-interrelación mostrado y que, por tanto, la completación es alta.

b) Corrección: es, evidentemente, *alta*.

c) Minimalidad: no encontramos atributos sobrantes (repetidos, deducibles o derivables) en el modelo entidad-interrelación, por lo cual:

$$MR = 1 - \frac{0}{9} = 1$$

d) Expresividad: supongamos que, según las reglas de la organización, todos los nombres utilizados en los diseños deben coincidir con los nombres reales utilizando, por ejemplo, notación húngara, excepto en el caso de los atributos clave, que deben ir precedidos del prefijo "Id". Por tanto:

$$SFAR = \frac{0}{9} = 0 \quad \dots \text{ningún atributo cumple las reglas de la organización}$$

$$SFER = \frac{7}{7} = 1$$

...todas las entidades e interrelaciones tienen nombres reales y significativos

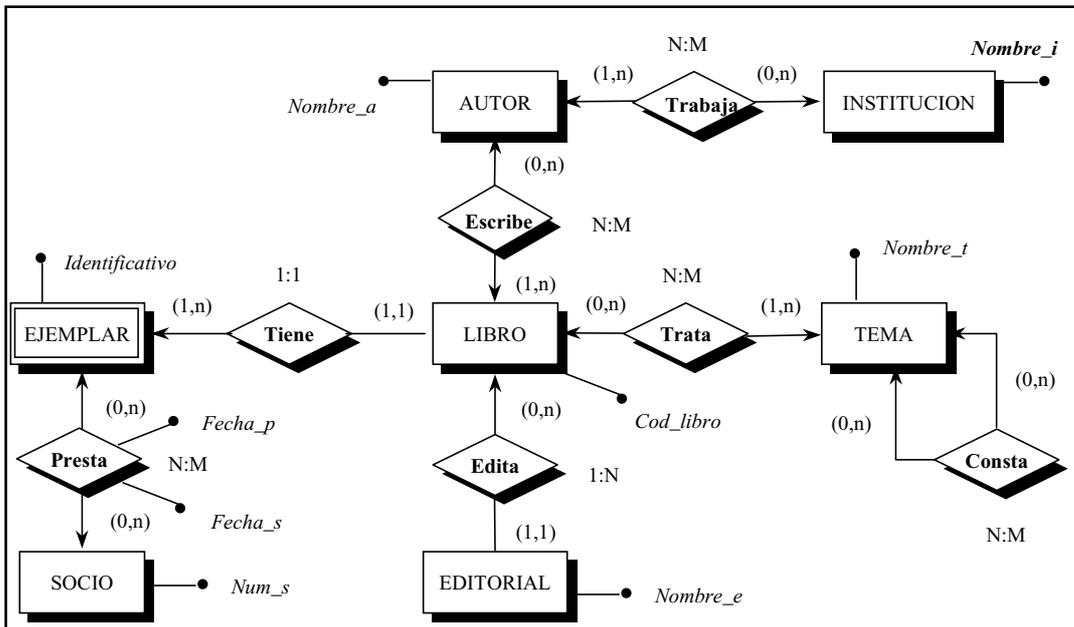


Figura 1. Esquema conceptual para el ejemplo.

$$SFRR = \frac{7}{7} = 1$$

e) Legibilidad y autoexplicación: es *alta*.

f) Extensibilidad y normalidad: como el esquema ER está en 3NF, su valor es 1.

g) Cohesión del esquema:

$$COSR = \frac{7^2}{7^2} = 1$$

...porque el grafo ER tiene sólo una componente conexa, formada por 7 entidades.

h) Complejidad intra-entidad: como sólo tenemos una entidad con interrelaciones reflexivas, el valor de esta métrica es:

$$IEC_{SCHEMA} = (0.14, 0.38)$$

i) Complejidad del esquema:

$$SCR = \frac{7}{\frac{7 \times 6}{2}} = \frac{7}{21} = 0.33$$

...porque hay 7 interrelaciones (numerador) y 7 entidades (denominador).

j) Longitud interrelacional: su valor es 4.

k) Tamaño: S=(7, 7, 9)

CONCLUSIONES Y TRABAJO FUTURO

Hemos propuesto algunas métricas para controlar la calidad de los esquemas conceptuales de bases de datos. La validez de estas métricas está siendo comprobada mediante una serie de experimentos, realizados tanto con alumnos como con la colaboración de diversas empresas y organismos. Además, intentamos obtener valores límite que nos permitan juzgar esquemas conceptuales.

Además, en la actualidad estamos trabajando en la obtención de métricas para bases de datos objeto-relacional.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Chen, P. P. The Entity/Relationship Model: Toward a Unified View of Data. *Communications of the ACM*, 1 (1), 1976.
- [2] De Miguel, A. y Piattini, M. *Concepción y Diseño de Bases de Datos*. Madrid, Ed. Ra-Ma. 1993.
- [3] Batini, C., Ceri, S. y Navathe, S. *Conceptual database design. An entity relationship approach*. Benjamin Cummings Publishing Company, 1992.
- [4] Reingruber, M. C. y Gregory, W. W. *The Data Modeling Handbook. A best-practice approach to building quality data models*. John Wiley & Sons, Inc. 1994.
- [5] Kesh, S. Evaluating the quality of entity relationship models. *Information and Software Technology*, 37 (12), 1991, 681-689.
- [6] Sneed, H.M. y Foshag, O. Measuring Legacy Database Structures. *Proc. of The European Software Measurement Conference FESMA'98*, Coombes, Hooft and Peeters (eds.), 1998, 199-210.
- [7] McCabe, T. A complexity measure. *IEEE Trans. Software Engineering*, 2 (5), 1976, 308-320.
- [8] Fenton, N. E. y Pfleeger, S. L. *Software metrics. A rigorous & practical approach*. International Thomson Computer Press, 1997.
- [9] Melton, A. (ed.). *Software Measurement*. International Thompson Computer Press, 1995.
- [10] Jarvenpaa, S. y Machesky, J. End user learning behavior in data analysis and data modeling tools. *Proc. of the 7th Int. Conf. on Information Systems*, San Diego, 1986, 152-167.
- [11] Juhn, S. y Naumann, J. The effectiveness of data representation characteristics on user validation. *Proc. of the 6th Int. Conf. on Information Systems*, Indianapolis, 1985.
- [12] Shoal, P. y Even-Chaime, M. Database schema design: An experimental comparison between normalization and information analysis. *Database*, 18 (3), 1987, 30-39.
- [13] Batra, D., Hoffer, J.A. y Bostrom, R.P. A comparison of user performance between the relational and the extended entity relationship models in the discovery phase of database design. *Communications of the ACM*, 33 (2), 1990, 126-139.
- [14] Kim, Y.-G. y March, S.T. Comparing Data Modeling Formalisms. *Communications of the ACM*, 38 (6), 1995, 103-115.

- [15] ISO. *Software Product Evaluation-Quality Characteristics and Guidelines for their Use*. ISO/IEC Standard 9126, Geneva, 1994.
- [16] Gray et al. Design metrics for database systems. *BT Technology*, 9, 1991.
- [17] De Miguel, A. y Piattini, M. *Fundamentos y modelos de Bases de Datos*. Madrid, Ed. Rama, 1997.
- [18] Piattini, M., Calero, C., Polo, M. y Ruiz, F. Maintainability in object-relational databases. *Proc. of The European Software Measurement Conference FESMA'98*, Coombes, Hooft and Peeters (eds.), 1998, 223-229.