

# Un abordaje informal, pragmático y eficaz para el aprendizaje de proyecto conceptual de banco de datos

**Vinícius Medina Kern**

UNIVALI, Cursos de Ciência e Engenharia de Computação

Rodovia SC 407, Km 4; 88122-00; São José-SC, Brasil

Teléfono +55 (48) 281-1500, Fax +55 (48) 281-1506

kern@eps.ufsc.br, <http://www.eps.ufsc.br/~kern>, <http://www.sj.univali.br/~kern>

## INDEXADORES

Proyecto de banco de datos, modelo Entidad-Relacionamiento, lenguaje natural, IDEF1X, aprendizaje

## RESUMEN

El proyecto conceptual de banco de datos trata de definir las estructuras de información importantes en el ambiente del banco de datos. El proyecto es generalmente diseñado en forma gráfica a través del modelo Entidad-Relacionamiento, que tiene base en la lógica de primera orden. Hay un lapso entre ese tipo de lenguaje y el lenguaje natural usada por proyectistas y expertos en el dominio del negocio, causa frecuente de errores de proyecto. Este artículo presenta un abordaje informal al aprendizaje de proyecto de banco de datos, basado en preguntas de carácter pragmático que hacen con que los estudiantes usen y testen su conocimiento de sintaxis y semántica. Los resultados son comentados.

## INTRODUCCIÓN

El proyecto de banco de datos tiene una etapa inicial, llamada conceptual, donde se identifican las estructuras de información importantes en el ambiente y restricciones estructurales sobre la información. Decisiones en esta etapa son críticas, pues establecen reglas para el modo como la información está organizada.

El proyecto conceptual de banco de datos es realizado usualmente en forma gráfica, a través del modelo Entidad-Relacionamiento (ER) [1] o uno de sus dialectos. En este trabajo usamos IDEF1X [2], un dialecto del modelo ER. El proyectista entrevista expertos en el dominio del ambiente para comprender cuales son sus necesidades de información, y entonces dibuja un modelo del banco de datos usando un lenguaje gráfico.

El modelo ER es un lenguaje semejante a la lógica de primera orden, mientras que los expertos hablan sobre sus necesidades de información en lenguaje natural, que es mucho más rica que la lógica citada, pues admite ambigüedad y variedad en la forma de enunciar reglas. Esta diferencia entre lenguajes es muchas veces la fuente de problemas de concepción y interpretación de modelos de banco de datos. La interpretación de modelos requiere sacar conclusiones pragmáticas a partir de frases representadas gráficamente según una sintaxis y semántica semejante a lógica.

Este trabajo presenta un abordaje informal y pragmático al aprendizaje de proyecto conceptual de banco de datos. Tras una breve introducción a la sintaxis y semántica de un lenguaje de proyecto, los estudiantes son confrontados con preguntas difíciles sobre el ambiente proyectado. Eso los exige sacar conclusiones pragmáticas a partir del gráfico.

En las próximas secciones se comparan el lenguaje de proyecto de banco de datos al lenguaje natural (español, por ejemplo) y se presenta un abordaje para facilitar el desenvolvimiento de conocimiento experto de análisis y diseño de modelos de banco de datos. Los resultados son comentados.

## LENGUAGE DE BANCO DE DATOS Y LENGUAGE NATURAL

Los elementos fundamentales de un modelo de banco de datos son: entidad, relacionamiento y atributo [1]. Chen [3] preconizó la forma de aproximar el lenguaje de modelos de banco de datos al lenguaje natural.

Un relacionamiento siempre puede ser leído como dos afirmaciones que siguen la forma “<Artículo indefinido> <sustantivo> <verbo transitivo> <cuantificador> <sustantivo>”. Los sustantivos son los nombres de las entidades (en función de sujeto y objeto); el verbo transitivo es el nombre del relacionamiento (o su inverso, si la lectura tiene sentido contrario a la migración de la llave); y el cuantificador trata del número de instancias de la otra entidad asociadas a través del relacionamiento. Por ejemplo, uno de los relacionamientos en la figura 1 puede ser leído como: “Un departamento ofrece cero, una, o varias disciplinas”, y “Una disciplina es ofrecida por exactamente uno departamento”.

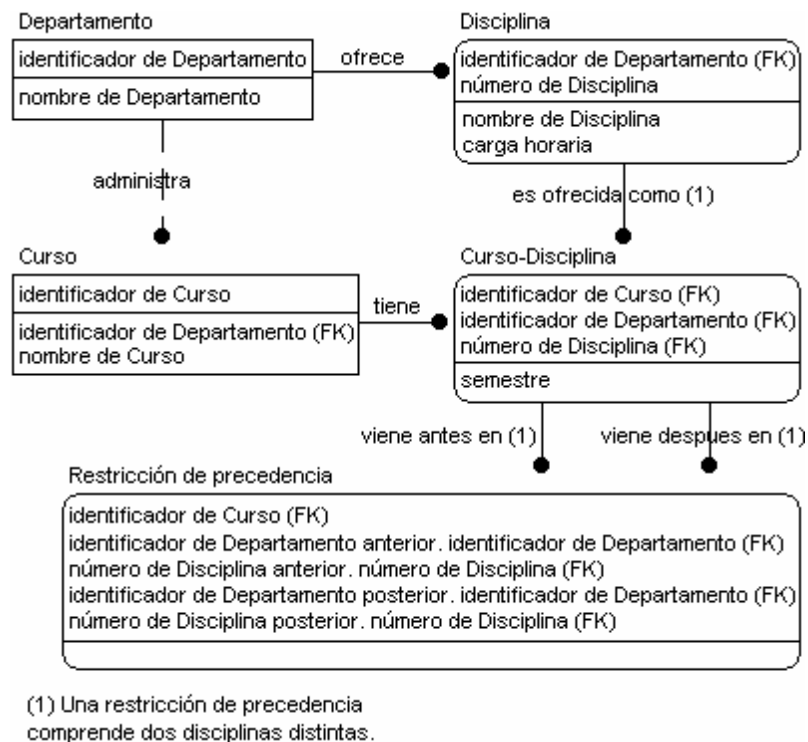


Figura 1. Modelo de banco de datos para una universidad con organización administrativa departamental

### SACANDO CONCLUSIONES PRAGMÁTICAS DE UN MODELO

Esa interpretación del modelo viene directamente de la sintaxis y semántica del lenguaje gráfico. Si pensamos en coordinar una secuencia de afirmaciones relativas a relacionamientos sucesivos, podemos sacar conclusiones pragmáticas sobre el negocio modelado, o verificar la veracidad de reglas que sabemos que deben ser verdaderas. Por ejemplo, para el proyecto ilustrado en la figura 1 en IDEF1X (un dialecto del modelo ER), se hacen las afirmativas:

1. Considerando una disciplina en el currículo de un curso (*Curso-Disciplina*), el departamento que ofrece la disciplina y el departamento que administra el curso no son necesariamente el mismo departamento.
2. Un curso puede tener varios currículos (lista y secuencia de disciplinas) al mismo tiempo.
3. Una restricción de precedencia no puede comprender dos disciplinas del mismo semestre.

Informalmente, uno puede esperar que las tres afirmativas sean verdaderas en un ambiente universitario departamental. Sin embargo, una verificación cuidadosa revela que solo la primera es verdadera. La segunda es falsa porque, si dos currículos pueden coexistir, entonces debe haber un modo de diferenciarlos, pero no hay – se puede diferenciar cursos a través de valores distintos de *idCurso*, pero no hay como diferenciar currículos, que son grupos de instancias o individuos de la entidad *Curso-Disciplina*. La tercera afirmación es falsa porque el gráfico dice (gráficamente) que

una restricción de precedencia es una asociación entre dos instancias de `Curso-Disciplina`, pero no hay restricciones sobre los valores de `semestre`.

## EL ABORDAJE PRAGMÁTICO AL APRENDIZAJE DE PROYECTO

En las primeras experiencias con este tipo de ejercicio, para sorpresa del profesor, los estudiantes tuvieron un promedio de cerca de 50% de respuestas correctas – la esperanza matemática para un teste de verdadero o falso. El profesor aplicó, entonces, el mismo tipo de teste usando un modelo abstracto, sin significado de negocios, como el ilustrado en la figura 2, con entidades A, B, C, etc. y afirmaciones sobre ese modelo. En este caso se verificó un índice de cerca de 70% de respuestas correctas.

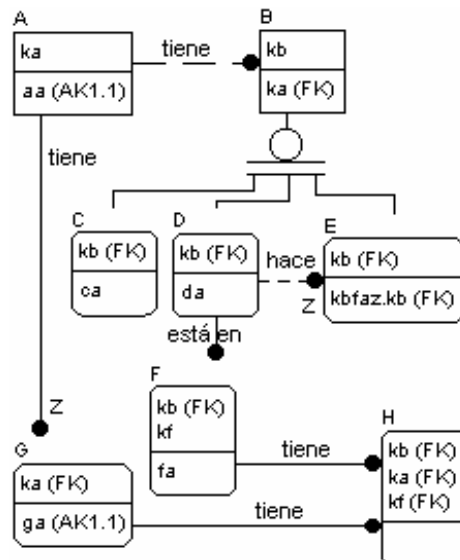


Figura 2. Modelo de banco de datos sin significado de negocios

Tras una discusión detallada sobre por que cada afirmación es verdadera o falsa, el teste fue repetido con otras afirmaciones para ambos modelos y los estudiantes tuvieron, en las dos series, un índice promedio de acierto de cerca de 70%. La interpretación del profesor para el fenómeno es que el experimento permitió a los estudiantes concentrarse en la observación del modelo según la sintaxis y semántica del lenguaje, y no según lo que creen que debe ser verdadero o falso en el ambiente de negocio.

Para cada modelo, en cada una de las dos series de respuestas, los estudiantes eligieran verdadero (V) o falso (F) para cerca de 10 afirmativas. La análisis de los resultados del experimento resumidas en la tabla 1 permite comparar el desempeño relativo a los dos modelos (universitario-real y abstracto) en las dos series. Es posible observar que los resultados son bastante homogéneos, a excepción de la primera serie de respuestas para el modelo universitario.

En la primera serie de respuestas para el modelo universitario los estudiantes se equivocaron más, quizás en consecuencia de su prejuicio sobre cual **debería** ser la regla en el banco de datos. Ninguna de las afirmativas en la primera serie del modelo universitario registró un suceso de 100%, mientras que, en la segunda serie, 2 de 10 tuvieron la respuesta correcta de todos los respondientes, de forma muy semejante a lo que pasó en las dos series para el modelo abstracto.

Es posible que los estudiantes busquen compensar su deficiencia en el dominio de la sintaxis y semántica del lenguaje usando opiniones propias sobre el ambiente de negocios (una universidad departamental, en nuestro ejemplo). Si eso es verdad para estudiantes, es posible que profesionales también interpreten modelos basados en sus prejuicios y no solamente en las reglas explícitas en el modelo. Eso representa un serio riesgo a la calidad de los sistemas de informaciones. Considerando las tres afirmativas de la sección anterior, que en un ambiente universitario se supone verdaderas pero en el modelo solo la primera es verdadera, se llegó al siguiente resultado, evidenciando la importancia del desenvolvimiento de la capacidad de análisis crítica de modelos de banco de datos:

- Afirmativa 2: suceso de 60% (60% de respuestas “F”).

- Afirmativa 3: suceso de 64% (64% de respuestas “F”).

Tabla 1 – Sumario de resultados del experimento

	Mayor índice de aciertos entre las afirmativas	Menor índice de aciertos entre las afirmativas	Mayor índice de aciertos de un estudiante	Menor índice de aciertos de un estudiante	Índice promedio de aciertos
Modelo universitario, 1ª serie	82 %	18 %	70 %	30 %	52 %
Modelo universitario, 2ª serie	100 %	36 %	90 %	40 %	70 %
Modelo abstracto, 1ª serie	100 %	36 %	92 %	50 %	72 %
Modelo abstracto, 2ª serie	100 %	27 %	92 %	46 %	70 %

Las dos afirmativas fueron respondidas en la primera serie. En los dos casos, casi la mitad de los respondientes se equivocaron. Ninguna de las afirmativas en la primera serie del modelo universitario registró un suceso de 100%, mientras que, en la segunda serie, 2 de 10 tuvieron la respuesta correcta de todos los respondientes.

Esos resultados revelan una mejora en el suceso de los respondientes. Como comparación, los mínimos y máximos de suceso en las dos series para el modelo sin significado fueron 36 y 100%, y 27 y 100% – poca variación, que puede ser considerada aleatoria.

## CONCLUSIÓN

Está claro que, de acuerdo con el índice promedio de 70% de suceso en ambas tentativas con el modelo abstracto (sin significado de negocios), no hubo un aumento en la capacidad de los estudiantes cuanto a su habilidad con la lógica. Sin embargo, considerando el progreso de respuestas correctas de 52 para 70% en el caso del modelo universitario (concreto, con significado de negocios), llama la atención que el índice promedio de aciertos es también 70% en el segundo caso. Aparentemente, la discusión detallada sobre los motivos para cada respuesta verdadera o falsa ayudó a los alumnos a trabajar mejor con los tres aspectos de un lenguaje: sintaxis, semántica y pragmática.

Es importante observar que los estudiantes que tuvieron más suceso en este tipo de ejercicio también lograron capturar con mejor suceso las reglas de negocios en sus ejercicios de concepción de modelos. Eso permitió al profesor ayudar a los estudiantes con desempeño menos efectivo a través de una orientación objetiva sobre cuáles son sus errores y como disiparlos.

Los resultados indican que los estudiantes mejoraran rápidamente sus habilidades de interpretación pragmática de modelos, únicamente por darse cuenta de que las conclusiones pragmáticas son estrictamente dependientes de la sintaxis y semántica de las afirmaciones elementares – las frases formadas para cada relacionamiento. El abordaje es, por eso, eficaz y eficiente para el aprendizaje de proyecto conceptual de banco de datos.

## REFERENCIAS

- [1] P.P. Chen, “The Entity-Relationship model - toward a unified view of data”. **ACM Transactions on Database Systems** 1 (1), p. 9-36, 1976.
- [2] NIST (National Institute of Standards and Technology). **Federal Information Processing Standards Publication 184: Integration Definition for Information Modeling (IDEF1X)**, Gaithersburg, MD, December 1993.
- [3] P. P. Chen. “English sentence structure and Entity-Relationship diagram”. **Information Science** 29 (2), Elsevier, p. 127-149, May 1983.