

Un modelo de datos para la construcción de tesauros

José Vicente Rodríguez Muñoz

Departamento de Información y Documentación. Campus de Espinardo. Universidad de Murcia (España).
E-mail: jovi@fcu.um.es

RESUMEN

En el marco de los sistemas de almacenamiento y recuperación de la información, los lenguajes documentales se presentan como medio para la mejora de la calidad del análisis y para garantizar una recuperación pertinente y precisa de los documentos. Asimismo los modelos de datos operan como herramientas conceptuales que permiten representar de una manera formal un determinado sistema.

Consideradas estas dos circunstancias, el propósito del presente trabajo consiste en representar la idea de tesoro mediante un esquema conceptual, con objeto de poder plasmar por medio del desarrollo de equipos lógicos un producto capaz de controlar el máximo número de incoherencias, proporcionando de este modo la deseada integridad de los datos.

ABSTRACT

In the framework of the information storage and retrieval systems, the documental languages are presented as middle for the improvement of the quality of the analysis and to guarantee a relevant retrieval process of the documents.

On the other hand, the data models operate as conceptual tools that permit to represent in a formal way a given system. Considering these two circumstances, the purpose of the present work is to represent the idea of a Thesaurus through a conceptual scheme, with the object of power to form by middle of the development of logical equipments a product capable of controlling the maximum number of incoherences, providing of this manner the necessary integrity of data

INTRODUCCIÓN

A principios de los años 60 aparecen los primeros tesauros,¹ que intentan dar solución a los problemas que empeñaban a plantear los lenguajes de clasificación de estructura más rígida, cuestión ésta que dio lugar a polémicas por la defensa a ultranza de los tradicionales sistemas de clasificación. En efecto, los tesauros son fáciles de utilizar, gracias a su presentación alfabética y sistemática; codifican el lenguaje natural a un lenguaje documental mediante relaciones semánticas de términos, de manera que proporcionan un conocimiento bastante completo del ámbito semántico en el que se encuadra cada concepto, todo lo cual constituye una gran ayuda para el análisis

documental y la posterior recuperación de la información.

Pero quizás la cualidad más importante de los tesauros es su gran flexibilidad, lo que les permite ampliar el número de términos, a medida que van apareciendo nuevos conceptos, e incluso suprimir los que han caído en desuso. Tratan en definitiva de articular un lenguaje documental muy flexible que sustituya la rigidez de otros para dar una respuesta eficaz al almacenamiento de un mayor volumen de información, el cual se incrementa cada vez más como consecuencia, entre otras cuestiones, de la rápida evolución de la ciencia.

Paralelamente, la aportación de la informática al mundo de la documentación generaría una nueva polémica entre el uso de lenguajes documentales controlados, tales como los tesauros, y lenguajes documentales libres, esto es, sin estructura semántica, pues se pensaba que, como consecuencia de las capacidades de almacenamiento y velocidad de proceso del instrumento, era posible la generación de un número ingente de términos sin mayor control. La polémica, entendemos, resultó estéril, como lo demuestra el hecho de que aún hoy plantea numerosas dificultades el tratamiento de grandes volúmenes de información.

1 N. Roberts. "The prehistory of the information retrieval thesaurus" *Journal of Documentation*. 1984, p. 271-285

Así, las exigencias de los usuarios respecto a sistemas de información más flexibles, adecuados y eficaces han obligado a dedicar una mayor atención a los datos y a su estructuración, se ha buscado entonces, una gestión más racional de éstos, pasando a ser considerados como un recurso fundamental de las organizaciones.

La creación de un sistema de información tiene dos aspectos claramente diferenciados, aunque haya entre ellos una fuerte interrelación: los datos y los tratamientos.² Si bien la concepción del sistema de datos y la del conjunto de tratamientos no puede realizarse de manera totalmente independiente, los problemas que han de resolverse son de naturaleza distinta, el objetivo de este trabajo se centra en los datos.

A pesar del esfuerzo realizado por numerosos investigadores y estudiosos del tema, e incluso de la comercialización de varias metodologías, no existe todavía, y posiblemente no existirá en un plazo de tiempo corto, una metodología consagrada de diseño. En nuestra opinión, el diseño de una base de datos, además de ser un difícil ejercicio intelectual, sigue teniendo algo de arte, lo que no significa que el contar con la ayuda de unos principios metodológicos, no sea imprescindible para conseguir un buen diseño.

En el proceso de diseño de bases de datos (BD) es necesario distinguir, a grandes rasgos, las siguientes etapas:

- 1) Requisitos de análisis y especificación.

- 2) Diseño conceptual.

- 3) Diseño lógico.

- 4) Implementación.

- 5) Diseño físico.

Se profundizará en la segunda etapa, la más importante en este desarrollo.³ Así, la característica principal de cualquier Sistema Gestor de Bases de Datos (SGBD) es el modelo de datos (MD) en que se basa. A nivel conceptual de la arquitectura de un SGBD,⁴ el MD cuenta con dos importantes funciones. Primero, suministra una metodología para representar objetos de una aplicación particular, y las relaciones entre esos objetos, a lo que se le llamaría la función "semántica" o conceptual. Segundo, el MD se estructura para permitir una traslación sencilla del esquema conceptual a la estructura física de los datos del nivel interno de los SGBD, lo que representa la función "representacional".

Las investigaciones en métodos de diseño de BD han desarrollado la idea de usar dos modelos distintos en las diferentes etapas del diseño,⁵ tal y como se ha señalado. Un MD conceptual proveerá de significados efectivos suficientes, para describir la BD, posteriormente se emplearía un modelo de datos representacional para conseguir la transformación de un esquema a estructuras físicas de datos. Así, el modelo relacional se podría emplear como MD representacional.

Ha aparecido un gran número de propuestas.⁶ Cada nuevo modelo pretende representar más sobre la semántica de los

datos, pero generalmente sólo son leves diferencias.

Uno de los más aceptados ha sido el modelo Entidad Relación (ER) de Chen.⁷ La contribución del modelo ER es la distinción entre entidades u objetos y relaciones entre éstos o conexiones entre los objetos.

En el modelo ER, se reconoce que las relaciones tienen ciertas propiedades semánticas que los distinguen de las entidades, como la relación de cardinalidad ($1:1$, $1:N$, $M:N$) y las dependencias de existencia.⁸

Investigaciones posteriores indicaron que las construcciones del modelo ER no eran suficientes para representar algunos datos semánticos importantes. Así, aparecieron los conceptos de subclase,⁹ también llamadas jerarquías ISA,¹⁰ y superclase, llamada generalización de categorías,¹¹ conceptos éstos no considerados directamente por el modelo ER. La importancia de estos conceptos se discute en "The semantic data model: A modeling mechanism for database applications." de Hammer y Mcleod; en "View representation in logical database design" de Navathex y Schkolnick; en "The functional Data Model and the Data Language DAPLEX" de Shipman; en "Extending the Database Relational Model to Capture More Meaning" de Cood, y en

2 R.W. Lanning. *An Entity-Relationship Approach to Model Management*. 1986, p. 65-72

3 K. Hogshead Davis. "Need for 'Flexibility' in a Conceptual Model." *Information & Management*. 1990, p. 231-241.

4 A. Klug y D. Tsichritz. *The ANSI/X3/SPARC Report of the study Group on Data Base Management Systems*. 1977.

5 V. Lum. *New Orleans Database Design Workshop Report*. 1979.

6 -N. Roussopoulos y J. Mylopoulos. "Using semantic networks for data base management." *Proceedings of ACM International Conference on Very Large Data Bases*. 1975.

-C.G. Davis et. al. *Entity-relationship approach to software engineering*. 1983.

7 P.P. Chen. "The Entity-Relationship Mode-Toward a Unified View of Data." *ACM Transactions on Databases Systems*. 1976, p. 9-36.
-C.G. Davis et. al. *Op. cit.*

8 -H. Schmid y J. Swenson. "On the semantic of the relational data model." *Proceedings of ACM SIGMOD Conference*. 1975.
- G. Wiederhold y R. Elmasri. "The structural Model for Database Design". *Proceedings of International Conference on the Entity-Relationship*. 1979.

9 -J.M. Smith y D.C.P. Smith. "Database Abstractions: Aggregation and generalization." *ACM...* 1977, p. 105-133.
-M. Hammer y D. Mcleod. "Database Description with SDM: A semantic Database Model." *ACM...* 1981, p. 351-386.

10 M. Minsky. "Computer science and representation of Knowledge" *Proceedings of National Computer Conference*. 1973.

11 J. M. Smithy D.C. P. Smith. *Op. cit.*

“The Structural Model for Database Design” de Wiederhold y Elmasri.¹²

Otro concepto semántico necesario no considerado inicialmente es el de la agrupación de entidades que no pertenecen necesariamente a un mismo conjunto de entidad, para participar en una relación con un rol dado. Este concepto de rol se introdujo en “The Role Concept in Database Models”¹³ y se discutió en “The functional data Model and the Data Language DAPLEX,” “Extending the Database Relational Model to Capture More Meaning” y en *Abstraction capabilities and invariant properties modelling within the entity-relationship approach*.¹⁴

En el modelo ER, sólo puede participar un conjunto de entidad simple en una relación con un rol dado. Como solución a esta restricción se introdujo¹⁵ una construcción adicional simple al modelo ER: el concepto de categoría. Así aparece el modelo denominado Entidad Categoría Relación ECR. El concepto de categoría maneja tanto la participación del conjunto de entidad múltiple en un rol dado de una relación, como los conceptos de subclase y superclase.

En el modelo ECR las entidades permanecen por sí mismas, y no necesitan ser identificadas por un atributo clave particular. Asimismo, los atributos en este modelo pueden ser multivalores, incluso ser definidos con el mismo esquema como atributos de valores simples. Estos se definen como funciones con dominio de la función conjunto de entidad, conjunto de relación, o categoría, teniendo

de rango el conjunto de posibilidades de un conjunto de valores. Finalmente el modelo ECR dispone también de una especificación más completa de la cardinalidad y propiedades dependientes de las relaciones. Se usa una especificación similar a ésta en “Data semantics,”¹⁶ pero extendida a las relaciones *n-áreas*, $n > 2$. Esta misma técnica también se usa para especificar restricciones en atributos.

Se acepta generalmente que un SGBD tenga diferentes tipos de interfases que sirvan a las necesidades de cada uno de los usuarios. Una de estas interfases la constituyen los lenguajes de consulta de alto nivel. Lenguajes formales no ambiguos, en esta clase, incluyen a “SEQUEL 2: a Unifies Approach to Data Definition, Manipulation and Control,”¹⁷ implementado en SYSTEM-R, y QUEL¹⁸ “THE DESIGN AND IMPLEMENTATION OF INGRES”, implementado en INGRES, así como el SQL An Introduction to Database Systems.¹⁹ Los usuarios para esta interfase incluyen personal técnico y administrativo que interactúa de un modo *ad hoc* con el SGBD. Parte de los lenguajes en esta clase se utiliza también como declaraciones incorporadas a un lenguaje de programación anfitrión para el acceso a la BD (por ejemplo el SQL inmerso en COBOL, C, etc.)²⁰ *An Introduction to Database Systems*.

Se utilizará como interfase el lenguaje de consulta de alto nivel GORDAS²¹ “A formal high-level query language for the entity-relationship model” (Graph-O Riented DATA Selection). GORDAS es único, en el sentido de que es un lenguaje de alto nivel formal, para BD orientadas a grafos de la clase de modelos de datos entidad-atributo-interrelación. Aunque se han propuesto varios de estos lenguajes para el modelo relacional de datos, muy pocos se han desarrollado para la clase anterior de modelos de datos.

Las construcciones de lenguaje de GORDAS se diseñaron para satisfacer dos requisitos:

Las construcciones de lenguaje de GORDAS se diseñaron para satisfacer dos requisitos:

1. Presentar al usuario un lenguaje en el que las demandas que hacen referencia a la información de la BD se expresen en términos cercanos a las percepciones del usuario, y
2. Conseguir un lenguaje formal y no ambiguo de modo que la traducción del lenguaje en estructuras de datos orientadas a la máquina sean relativamente directas.

Estos dos requisitos son algo contradictorios, por lo tanto GORDAS es un compromiso entre ambos. Precisamente por ello es el lenguaje elegido para representar los esquemas del modelo ECR.

Todo lo desarrollado hasta ahora está dirigido a modelar el tesoro, creando para ello un esquema conceptual, para llegar a construir un “esquema tesoro”. El proceso por el cual se ha llegado a la elaboración de éste, como ya se señaló, fue dado por las posibilidades que ofrecen, por un lado los MD y, por otro, los lenguajes documentales, para ser unificados en un desarrollo conjunto. Se siguió todo este ca-

12 – S.B. Navathe y M. Schkolnick. “View representation in logical database design.” *Proceedings of ACM...* 1978, p. 144-156.

– M. Hammer y D. Mcleod. “The semantic data model: A modeling mechanism for database applications.” *Proceedings of ACM...* 1978, p. 26-36.

– D.W. Shipman. “The functional Data Model and the Data Language DAPLEX” *ACM...* 1981, p. 140-173.

– E.F. Cood. “Extending the Database Relational Model to Capture More Meaning.” *ACM...* 1981, p. 351-386.

– G. Wiederhold y R. Elmasri. *Op. cit.*

13 C.W. Bachman y M. Daya. “The Role Concept in Database Models.” *Proceedings of 3rd IEEE International Conference on Very Large Databases*. 1977, p. 464-476.

14 – D.W. Shipman. *Op. cit.*

– E.F. Cood. *Op. cit.*

– P. Scheuermann et. al. *Abstraction capabilities and invariant properties modelling within the entity-relationship approach*. 1980, p. 121-141.

15 S.A. Weeldreyer. *Structural aspects of the entity-category-relationship model of data*. 1980.

16 J.R. Abrial. “Data semantics”. *Proceedings of IFIP Conference on Data Base Management*. 1974, p. 1-60.

17 D.D. Chamberlin. “SEQUEL 2: a Unifies Approach to Data Definition, Manipulation and Control.” *IBM Journal of Research and Development*. 1976, p. 560-575.

18 M. Stonebraker. “The Design and Implementation of INGRESS.” *ACM...* 1976, p. 189-222.

19 C.J. Date. *An Introduction to Database Systems*. 1986.

20 *Idem*.

21 R. Elmasri y G. Wiederhold. “GORDAS: A formal high-level query language for the entity-relationship model.” *Entity-Relationship Approach to Information Modeling and Analysis*. 1983, p. 49-72.

mino como paso previo para presentar la hipótesis de partida, y consecuentemente justificar por qué se eligió.

Se presenta el tesauro como un lenguaje controlado, al que podemos considerar como una herramienta de representación de conceptos, dado que cuenta con una estructura semántica importante. Para ello hemos seguido, bajo la norma ISO 2788, el trabajo presentado por Van Slype en su libro *Los lenguajes de Indización: Concepción, construcción y utilización en los sistemas documentales*²² ya que, para nuestros propósitos de ejemplificar éste nos presenta una visión más descriptiva del tesauro, incorporando, por ejemplo, los términos de Campo Semántico, pues representa un buen modo de agrupar a los términos cabecera de una forma temática.

Al final, lo que se pretende es dar una base teórica conceptual al tesauro, que permita en el mayor grado posible la eliminación de ambigüedades.

CONSTRUCCIÓN DEL ESQUEMA CONCEPTUAL

Para describir el tesauro en términos del MD elegido, se deben considerar los conceptos básicos del modelo, esto es, las entidades, categorías y atributos, para que a continuación se comience con la semántica del mismo, o lo que es lo mismo, tratar las relaciones. Se pasará después a definir las restricciones explícitas necesarias para acomodar el tesauro al esquema, y quedará así configurado lo que denominaremos esquema ECRT (entidadcategoría-relación para tesauros).

La presentación del esquema, además de emplear los diagramas ECR, se realizará mediante GORDAS. La utilización de este lenguaje se debe a que, como ya se ha dicho, permite recoger de una forma estructurada, ya no sólo el modelo ECR, sino otras cuestiones de interés no soportadas por el mismo.

Definición de los esquemas

Se considerará como definición de unidad léxica o término indizante del tesauro –a partir de ahora simplemente términos– una representación de un concepto, o lo que es lo mismo:



Se presenta el tesauro como un lenguaje controlado, al que podemos considerar como una herramienta de representación de conceptos, dado que cuenta con una estructura semántica importante



$$T = \Psi(c)$$

donde c es el universo del discurso y Ψ es una proyección del mismo, por tanto, se trata de una abstracción.

Definiciones de los conjuntos valor

Para describir un término se considerará el siguiente grupo de atributos, los cuales son suficientes para definir un término de forma unívoca, no obstante a la hora de describir un término, se podría ampliar, disminuir o simplemente modificar el conjunto aquí presentado. Estos atributos serían los siguientes:

- * **Nombre:** se trata de una secuencia de caracteres que identifican al término. Se identifica como atributo 'NOMBRE', y se representa en el dominio: $\text{dom}(\text{nombre}) = \text{string } 255$
- * **Código:** es un número entero de cinco dígitos que coincidirá con el número de registro interno del término. Se identifica como atributo 'CODIGO', y su dominio será: $\text{dom}(\text{código}) = \text{integer range } 0:99999$
- * **CDU (Clasificación Decimal Universal):**²³ con este atributo se trata de identificar los términos mediante una clasificación temática a través de una estructura arborescente (con una secuencia de caracteres alfanuméricos). El atributo será identificado como 'CDU', y su dominio: $\text{dom}(\text{cdu}) = \text{string } 10$
- * **Idioma:** este atributo se justifica ya que se considerará un tesauro bilingüe (aunque esta decisión es arbitraria, ya que no existe ninguna cuestión previa que impida que sea multilingüe). El atributo será denominado 'IDIOMA' y su dominio será la enumeración de los idiomas que lo constituyen, esto es: $\text{dom}(\text{Idioma}) = \{\text{ESP, INC}\}$

El tesauro es bilingüe o multilingüe en el sentido de que cada idioma actúa con su semántica, independientemente, disponiendo de un nexo de unión, relación semántica que nos permite pasar de un idioma a otro, de acuerdo con las equivalencias interlingüísticas.

La representación en GORDAS de todos los atributos, así como de sus correspon-

22 G. Van Slype. *Los lenguajes de Indización: Concepción, construcción y utilización en los sistemas documentales*. 1991.

23 CDU. *Clasificación Decimal Universal*. 1991, 462p.

dientes dominios vendrían a definirse mediante las sentencias:

```
DEFINE VALUESET NOMBRE STRING
255 DEFINE VALUESET CODIGO AA
INTEGER RANGE 0:99999
```

```
DEEM VALUESET CDUSTRING 10
DEFINE VALUESET IDIOMA AS {ESP,
ING}
```

Definición de tipos de entidad

Las entidades, representadas como tales en el caso del tesauro, van a ser los términos, o sea: Campos semánticos, Descriptores, No-Descriptores, Descriptores auxiliares y las notas explicativas; puesto que son los que participarán en el esquema, mediante las correspondientes relaciones. Éstos serán definidos como:

* Campos semánticos: de los aspectos que hay que considerar dentro de un campo semántico (los temas y las facetas) se descartarán las facetas por dos razones:

- Es una clasificación abstracta. Los usuarios están más acostumbrados a razonar en términos de disciplinas (temas), y no en términos de la naturaleza de las palabras.
- El número de facetas es limitado, y resulta insuficiente para distribuir el conjunto de descriptores de un tesauro en grupos suficientemente reducidos como para que sean aprehendidos cómodamente por el usuario.

El primer aspecto (temas) permite agrupar los términos dentro de un campo del saber (mediante una ordenación alfabética), se trata de un contexto que será más útil.

A partir de ahora, campo semántico sólo vendrá referido a 'tema', lo cual no quita que las facetas sean incluidas como un atributo de los descriptores y no-descriptores.

El conjunto o tipo de entidad va a ser denominado 'CS', y está descrito por los atributos NOM_CS, CDU-CS, COD_CS e ID_CS. Donde COD_CS actuará como clave única. Así, la declaración de conjunto de entidad estará definida, en un tiempo dado t , por:

$$CS^t = ((NON_CS, CDU_CS, COD_CS, ID_CS), (COD_CS))$$

los dominios son los correspondientes a los conjuntos de valores a partir de los cuales son descritos. En sentencias GORDAS, este conjunto quedaría expresado como:

```
DEFINE ENTITYTYPE CS
ATTRIBUTES NOM_CS VALUESET
NOMBRE, CDU_CS VALUESET CDU,
COD_CS VALUESET CODIGO
UNIQUE, ID_CS VALUESET IDIOMA
```

* Descriptores: éstos, como ya sabemos, son los elementos centrales de un tesauro, pues a partir de ellos será posible la recuperación de la información. El conjunto o tipo de entidad se denominará 'DES', y está definido por los atributos NON_DESC, CDU_DESC, COD_DESC, e ID_DESC, y actuando COD_DESC como clave única. Así, la declaración de conjunto de entidad se define en un tiempo dado t , por:

$$DES^t = ((NON_DESC, CDU_DESC, COD_DESC, ID_DESC), (COD_DESC))$$

Igual que en el caso anterior los dominios serán los correspondientes de los conjuntos valor, y en GORDAS quedaría expresado mediante:

```
DEFINE ENTITYTYPE DESC
ATTRIBUTES NOM_DESC VALUESET
NOMBRE, CDU_DESC VALUESET
CDU, COD_DESC VALUESET CODIGO
UNIQUE, ID_DESC VALUESET
IDIOMA
```

* No-Descriptores: éstos, como también se sabe, son términos que no se emplearán en la recuperación en forma directa. A pesar de ello, son de una valiosa utilidad dentro del tesauro, ya que a través de las relaciones nos remiten a sus descriptores correspondientes, lo que enriquece las posibilidades de búsqueda y de recuperación de la información. Denominaremos a su conjunto 'ND', y su definición, en un tiempo dado t , será:

$$ND^t = ((NON_ND, CDU_ND, COD_ND, ID_ND), (COD_ND))$$

donde COD_ND es la clave única. Su correspondiente representación en GORDAS vendría dada por:

```
DEFINE ENTITYTYPE ND
ATTRIBUTES NOM_ND VALUESET
NOMBRE, COD_ND VALUESET
CODIGO UNIQUE, ID_ND VALUESET
IDIOMA
```

* Notas explicativas: El hecho de tener en consideración a este conjunto o tipo de entidad, denominado 'SN', ha sido dado por la necesidad y la generalización de su uso en un tesauro, pues ofrece aspectos declarativos sobre el conjunto entidad 'DESC', de manera

que permiten enriquecer la semántica del tesauro, y consecuentemente del esquema. Para esta declaración de entidad sólo hemos considerado dos atributos NOM_SN y COD_SN. Su definición en un tiempo dado t , será en este caso:

$$SN^t = ((NOM_SN, COD_SN), (COD_SN))$$

Las notas explicativas se consideran como monolingües, en el sentido, de que normalmente vendrán escritas en el idioma del descriptor del cual dependen (aunque no existe ninguna restricción en el hecho de que una nota se escriba en un idioma diferente de su descriptor, o que si la longitud no sobrepasa a la dada por el dominio, se escriban en el idioma del descriptor y su traducción a otro/s idioma/s). No estarán adscritas a ningún campo semántico, pues su existencia dependerá del descriptor al cual se asocian, o sea, no existirá una nota explicativa si no tiene un descriptor al que asociarse, por lo que heredarán de él el atributo clave COD_DESC. En GORDAS se diría:

```
DEFINE ENTITYTYPE SN
ATTRIBUTES NOM_SN VALUESET
NOMBRE, COD_SN VALUESET
CODIGO UNIQUE,
```

Descriptores auxiliares: este tipo de términos no será considerado como conjunto de entidad, ya que no se relacionará puesto que la unión descriptor-descriptor auxiliar da como resultado al descriptor verdadero por el cual indizamos. Esto hace que no sea una entidad representable, lo cual puede verse como un caso de especificación o generalización (por ejemplo, si se describe el conjunto de entidad a partir del conjunto de descriptores unitérminos y el conjunto de descriptores auxiliares, se estaría definiendo una especialización del conjunto de descriptores compuestos). No obstante, la cuestión queda perfectamente resuelta en forma más simple, si se considera a los descriptores no sólo como unitérminos, sino también como términos compuestos.

Definición de categorías

En nuestro modelo, los conjuntos de entidades definidas no presentan distintos roles, tal y como señalamos, por razones de simplicidad del esquema, por tanto no se va a definir ningún predicado que restrinja la participación de las entidades en

una categoría. Por ello para todas las declaraciones de categorías se tendrá que:

$$C = E \text{ [TRUE]}$$

En GORDAS quedaría expresado por las SENTENCIAS:

```
DEFINE CATEGORY CS = CS
```

```
DEFINE CATEGORY ND = ND
```

```
DEFINE CATEGORY DESC = DESC
```

```
DEFINE CATEGORY SN = SN
```

Definición de relaciones

Antes de entrar a describir las diferentes relaciones semánticas se realizarán dos clasificaciones de las mismas, que sirven para clarificar mejor su naturaleza, representando en definitiva restricciones implícitas del esquema.

Una primera clasificación es la referida a los tipos de términos (tipos de entidad) que intervienen en la relación, así pues, las relaciones son:

- * Heterogéneas, esto es, cuando intervienen dos tipos de términos diferentes. Corresponden a este caso las relaciones de: Equivalencia intralingüística, Pertenencia y de Notas.
- * Homogéneas, cuando los términos implicados son todos del mismo tipo, aquí están incluidas las relaciones de: Equivalencia interlingüística, Jerarquía y Asociación.

La otra clasificación se puede hacer con base en la reciprocidad de las relaciones, esto es, si un término 'a' está relacionado con 'b' entonces, si existe el mismo tipo de relación del término 'b' con el término 'a', se dirá que la relación es simétrica, pero si no existe el mismo tipo, es asimétrica. Según este criterio las relaciones se clasifican en:

- * Simétricas: relación de equivalencia interlingüística y de asociación.
- * Asimétricas: relación de equivalencia intralingüística, pertenencia, jerarquía y notas.

Relaciones de pertenencia

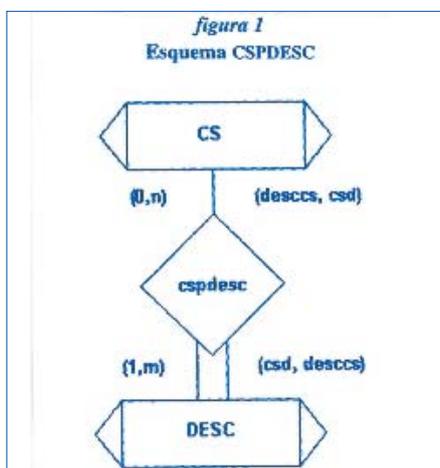
Relación heterogénea, pues los términos relacionados son: campos semánticos y descriptores, y asimétrica puesto que si un descriptor 'a' pertenece a un campo semántico 'c' no puede existir un campo semántico 'c' que pertenezca a un descriptor 'a'.

Las formas que presenta: monopertenencia y polipertenencia son consideradas como restricciones en la participación de las categorías implicadas.

La relación se define por la secuencia CSPDESC = (CS, DESC),

que carece de atributos propios, el grado de la relación es $\text{degree}(\text{CSPDESC}) = 2$.

De la representación diagramática (figura 1) se infiere que las restricciones estructurales de las categorías, son (0:n) ($i_1=0$, participación parcial de la categoría CS) para la conexión (descscs, csd), o lo que es lo mismo, a una entidad CS pueden pertenecer un mínimo de 0 entidades DESC y un máximo de n ; y (1:m) ($i_1=1$, participación total de la categoría DESC) para la conexión (csd, descscs), esto es, una entidad DESC pertenece como mínimo a 1 categoría CS y como máximo a n .



Esto da que el rango de la relación es (n:m). Ahora se puede interpretar la monopertenencia mediante el subrango (1:m) (un DESC pertenece a un sólo CS), y la polipertenencia con el rango de relación (n:m) (un DESC pertenece a dos o más CS), así pues, de la categoría DESC se puede definir el subconjunto de entidades que sólo pertenecen a una ocurrencia de entidad de la categoría CS, mientras que la polipertenencia quedaría representada por el rango de la relación.

En GORDAS sería expresado como:

```
DEFINE RELATIONSHIP CSPDESC
FROM CS (dese, es) MIN 1 MAX N, DESC
(es, desc) MIN 1 MAX M
```

La única restricción que va a ser definida es la referida al idioma, mediante la cual se establece que la categoría CS y DESC va a estar en el mismo idioma cuando entre ellas exista una relación de pertenencia CSPDESC, o dicho de otro modo:

```
DEFINE CONSTRAINT ON DESC:
```

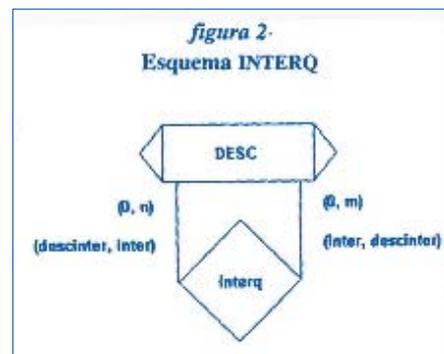
```
ID_DESC OF DESC EQUAL ID_CS OF csd
OF DESC
```

Relaciones de equivalencia interlingüística

Es una relación homogénea, ya que se da entre términos descriptores; y simétrica, pues si se tiene que un descriptor 'a' en una lengua A es equivalente a un descriptor 'b' en una lengua B, existe que 'b' en la lengua B es equivalente a 'a' en la lengua A.

Las formas que aparecen en esta relación, cierta, parcial y nula, son consideradas como restricciones en la participación de las categorías implicadas. La relación se define por la secuencia INTEREQ = (DESC, DESC') que carece de atributos y el grado de la relación es $\text{degree}(\text{INTEREQ})=1$.

En la representación diagramática (figura 2) se observa que la restricción estructural de la categoría, para ambas trayectorias (inter, descint) y (descint, inter), es (0:n) y (0:m) respectivamente ($i_1=0$, participación parcial, en ambos casos), esto quiere decir que una entidad DESC en una lengua (conexión inter, descint) tendrá como mínimo 0 entidades DESC' equivalentes en la otra lengua y como máximo n , de igual modo sucederá para la otra trayectoria (DESC' respecto a DESC).



Esto da como rango de la relación el par (n:m). Ahora se puede interpretar las diferentes formas que presenta la relación:

- * Equivalencia cierta: subrango (1:1).
- * Equivalencia parcial: subrango (Pm) o (n:1) (por la simeta).
- * Equivalencia nula: subrango (1:0) o (0:1) (por la simetría).

Así, se pueden definir dos subconjuntos de entidades en la categoría DESC, los que tienen una equivalencia cierta y los que la tienen nula, ya que para los que presentan una parcial, el rango es igual al de la relación (n:m).

En GORDAS se expresaría como:

```
DEFINE RELATIONSHIP INTEREQ
FROM DESC (inter, descint) MIN 0
MAX N, DESC' (descint, inter) MIN 0
MAX M
```

La única restricción que se definirá es la referida a la necesidad de que las entidades de la categoría implicadas en la relación INTEREQ deben tener el atributo idioma con valores diferentes, o lo que es lo mismo:

```
DEFINE CONSTRAINT ON DESC:
ID_DESC OF DESC NOT EQUAL
ID_DESC OF inter OF DESC
```

Relaciones de equivalencia intralingüística

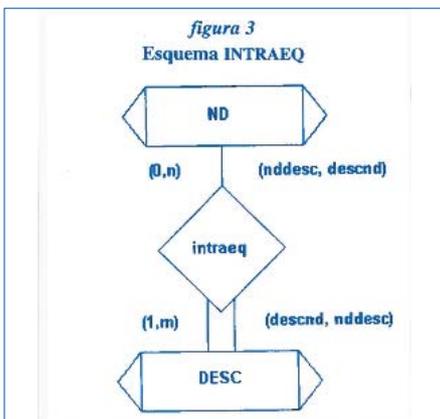
Es una relación heterogénea que relaciona no-descriptores y descriptores; y asimétrica, ya que si un no-descriptor 'a' es equivalente a un descriptor 'b', no existe la misma equivalencia del descriptor 'b' sobre el no-descriptor 'a'. La relación presenta dos formas: la monoequivalencia y pluriequivalencia, mismas que serán consideradas como restricciones de las categorías participantes.

Se define por la secuencia:

```
INTRAEQ = (ND, DESC),
```

la relación carece de atributos propios, el grado de la relación es degree (INTRAEQ)= 2.

La representación diagramática (figura 3) indica que las restricciones estructurales de las categorías, son (0:n) ($i_1=0$, participación parcial de la categoría DESC) para la conexión (nddesc,descnd), o sea una entidad DESC es equivalente a 0 entidades ND como mínimo y a n como máximo; y (1:m) ($i_1=1$, participación total de la categoría ND), esto es, una entidad ND es equivalente a 1 entidad DESC como mínimo y a n como máximo.



Esto significa que el rango de la relación es (n: m). Las entidades de la categoría ND que intervengan en monoequivalencia se podrán interpretar como un subconjunto, pues tendrían en subrango (1:n). Las dos variantes, que aparecen en la monoequivalencia, quedan perfectamente establecidas mediante las conexiones 'descnd' y 'nddesc', que como se sabe, indican el sentido de la relación. En cuanto a las dos variantes de la pluriequivalencia, no se han considerado relevantes pues pueden ser consideradas como subconjunto de la relación definida aquí, ya que la facultativa presenta el subrango (1:2), y la obligatoria el (1:m). Su representación en GORDAS sería:

```
DEFINE RELATIONSHIP INTRAEQ
FROM DESC (nddesc, descnd) MIN 0
MAX N, ND (descnd, nddesc) MIN 1
MAX M
```

La única restricción que se va a imponer es la referida a que tanto la categoría ND como DESC tengan el mismo valor en sus atributos COD_ND Y COD_DESC, o lo que es lo mismo

```
DEFINE CONSTRAINT ON DESC:
ID_DESC OF DESC EQUAL ID_ND
OF nddesc OF DESC
```

Relaciones jerárquicas

Se trata de relaciones homogéneas, pues se dan entre dos descriptores; y asimétricas, puesto que si un descriptor 'a' es jerárquico sobre un descriptor 'b', no existe la relación 'b' jerárquico sobre 'a', pero sí que 'b' está jerárquicamente bajo 'a'.

Las formas que presenta de monojerarquía y polijerarquía son tratadas como restricciones en la participación de la categoría participante.

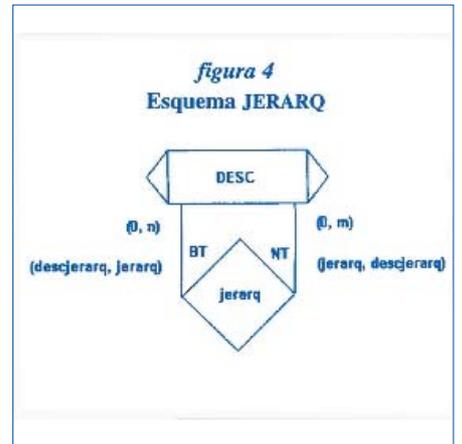
La relación se define como

```
JERARQ = (DESC, DESC'), JERARQ' = (DESC', DESC)
```

que carece de atributos, el grado de la misma es degree(JERARQ)=1. Por ello, y de la definición de asimetría se diferencian dos roles: NT (o rol específico) y BT (o rol genérico); así, en la definición de relación el orden de la secuencia es relevante, pues en la primera se dice que DESC es BT de DESC' y consecuentemente que DESC' es NT de DESC.

En la representación diagramática (figura 4) se observa que las restricciones de participación de la categoría son

(0:n) y (0:m), en ambos casos se tiene que $i_1=0$ por tanto la participación es parcial. Esto quiere decir que en ambos roles una entidad DESC estará relacionada jerárquicamente con 0 entidades DESC', como mínimo, y n entidades DESC' como máximo.



El rango de la relación es (n: m). Ahora se deben interpretar las dos formas que presenta la relación dependiendo de los roles:

rol	rango	forma
BT	(0:0)	TT (Top Term)
BT	(0:1)	monojerarquía
BT	(0:m)	polijerarquía
NT	(0:0)	DT (Down Term)
NT	(0:1)	
NT	(0:n)	

En GORDAS, se expresaría como:

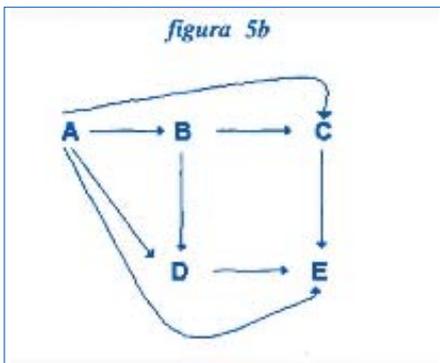
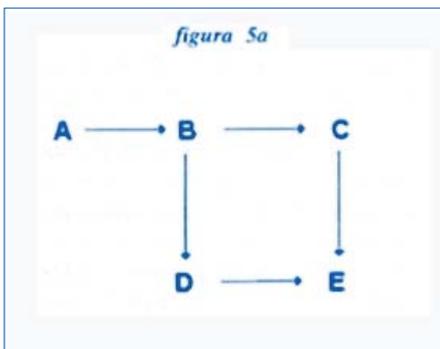
```
DEFINE RELATIONSHIP JERARQ FROM
DESC (jerarq, descjerarq) MIN 0 MAX N
DESC (descjerarq, jerarq) MIN 0 MAX M
```

La restricción que será impuesta, al igual que en los casos anteriores, es la del idioma, o sea, esta relación se dará entre entidades donde el valor del atributo ID_DESC sea igual, así se tiene:

```
DEFINE CONSTRAINT ON DESC:
ID_DESC OF DESC EQUAL ID_DESC
OF jerarq OF DESC
```

Aunque no se expresó de manera explícita en la definición de la relación, hay que entender que se definen sólo relaciones subordinadas inmediatas (figura 5a), esto se expresaría por la secuencia JERARQ(DESC_i, DESC_{i+1}, .., DESC_{j-1}, DESC_j)

donde $DESC_i \neq DESC_i$ y de longitud $n \geq 1$. Mediante funciones de composición y simplificación,²⁴ se puede obtener la relación derivada JERARQ* de JERARQ, que compondría todas las conexiones posibles (figura 5b), con ello se conseguiría la imposibilidad de crear una conexión, pues ésta ya estaría creada mediante la función composición. En la representación diagramática (figura 6), se aprecia que se han renombrado las conexiones con 'alljerarq' y 'alldescjerarq', para representar todas las conexiones (las reales y las generadas por la función composición), para esta relación no es relevante la restricción de participación de las categorías.

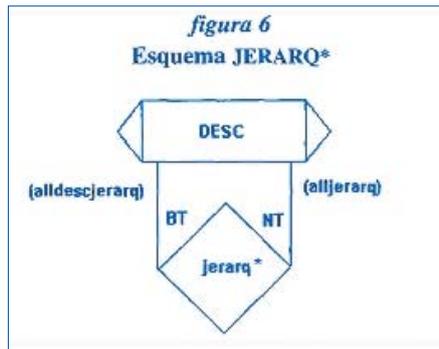


En GORDAS se expresaría del siguiente modo:

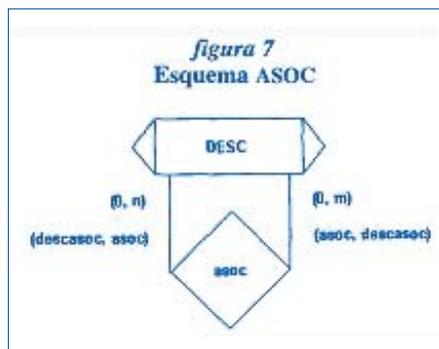
```
DEFINE DERIVED BINARY RELATIONSHIP
(Alljerarq, alldescjerarq) FROM DESC
TO DESC TO BE JERARQ* COMPOSE JERARQ
```

Relaciones asociativas

Son relaciones homogéneas, pues se dan entre dos descriptores: y simétricas, puesto que si un descriptor 'a' está asociado con otro descriptor 'b', entonces existe que 'b' está asociado con 'a'. FIGURA6



La relación se define por la secuencia ASOC = (DESC, DESC'), que no contiene atributos y su grado es $\text{degree}(\text{JERARQ}) = 1$. En la representación diagramática (figura 7), se observa que las restricciones de participación de la categoría DESC es (0:n) y (0:m) para las dos conexiones, por lo que existe una situación análoga a las relaciones interlingüísticas y jerárquicas.



Esto da que el rango de la relación es (n: m), y queda expresado en GORDAS como:

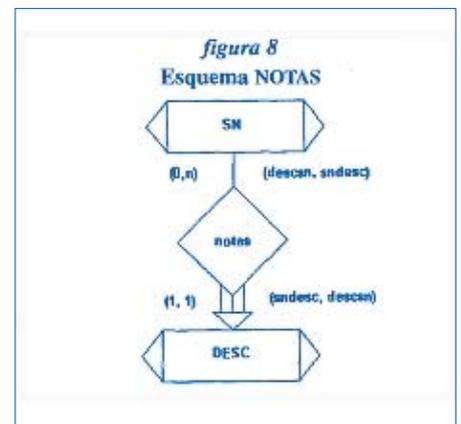
```
DEFINE RELATIONSHIP ASOC FROM
DESC (asoc, descasoc) MIN 0 MAX N
DESC (descasoc, asoc) MIN 0 MAX M
La única restricción viene referida al idioma, que obliga a que dos entidades DESC, tengan el mismo valor en su atributo ID_DFSC, así
DEFINE CONSTRAINT ON DESC:
ID_DESC OF DESC EQUAL ID_DESC OF
asoc OF DESC
```

Relaciones notas

Son relaciones heterogéneas, pues se dan entre un descriptor y una nota explicativa; y asimétricas, puesto que si una nota explicativa 'a' está relacionada con un descriptor 'b' no existe la misma relación entre 'b' y 'a'.

La relación se define como la secuencia NOTAS = (DESC, SN).

que no contiene atributos y su grado es $\text{degree}(\text{JERARQ}) = 2$. En su representación diagramática (figura 8) se observa que la restricción de participación de las entidades SN en la conexión (sn, descsn) es (1:1), esto es $i_1=1$ e $i_2=1$. Es un caso de relación específica, lo que quiere decir que la existencia de la categoría SN depende de la categoría DESC, mediante esta relación, o lo que es lo mismo existirá una entidad SN si y sólo si existe una entidad DESC con la cual podamos relacionarla. Mientras que la categoría DESC tiene una participación Parcial (0:n) ya que $i_j=0$.



```
En GORDAS se representaría mediante:
DEFINE RELATIONSHIP NOTES FROM
DESC (descn, sudesc) MIN_0 MAX_N
SN (sudescsn, descsn) SPECIFIC MIN 1
MAX 1
```

Una restricción implícita viene dada por tratarse de una relación específica, lo que produce que la categoría SN herede los atributos de la categoría.

Consideraciones adicionales al esquema

Se expondrán algunos aspectos no tratados en el desarrollo que se está siguiendo, pero que van a servir para completar el esquema.

Eliminar ambigüedades es una operación imprescindible para conseguir la univocidad de los descriptores. Los elementos distorsionadores de esta propiedad son: la sinonimia, cuasisinonimia y polisemia, a continuación se dan sus definiciones:

24 R. Elmasri y G. Wiederhold. Op. cit.

Sinonimia:

Dada la secuencia $(S_1, \dots, S_i, \dots, S_n)$, los S_n términos son sinónimos si $\forall S_i \exists$ una relación INTRAQ con un y sólo un DESC.

Cuasisinonimia:

Dada la secuencia $(S_1, \dots, S_i, \dots, S_n)$, los S_n términos son cuasisinónimos si $\forall S_i \exists$ una relación INTRAQ con dos o más DESC distintos.

Polisemia:

Dada la secuencia $(T_1, \dots, T_i, \dots, T_n)$, un término S es polisémico si $\forall T_i \exists$ que $S \in T$.

Cuando se presenta un caso de sinonimia o cuasisinonimia, la forma de operar para eliminarlas, es elegir, entre la secuencia el término que se asignará al conjunto de entidad DESC, los demás quedan asignados al conjunto de entidad ND, generándose las correspondientes relaciones INTERQ. En el esquema no se ha considerado la relación de pertenencia entre las categorías CS y ND (por tanto, ambos conceptos son iguales), pues esta relación está implícita, por lo que es innecesaria, ya que las entidades de la categoría ND tienen una pertenencia implícita a la categoría CS a través de las relaciones existentes de ambas con la categoría DESC.

La polisemia podrá ser eliminada mediante una relación JERARQ, que permitirá que el significado extenso de un término S se restrinja en sus significados por ubicación en diferentes jerarquías. Se podría pensar en las relaciones INTRAQ como eliminación de la polisemia, pero lo que se estaría, precisamente, generalizando es su significado, lo cual es contrario a lo que se pretende. Otra restricción que hay que considerar es que cuando entre dos entidades DESC existe una relación del tipo JERARQ, entonces no puede existir entre ellas una relación del tipo ASOC. Esta restricción no es posible expresarla con una sentencia DEFINE CONSTRAINT ON, debido a la sintaxis de GORDAS, por ello se definirá por medio de una expresión lógica para ser insertada en las operaciones de actualización, evitando así que el esquema pierda cohe-

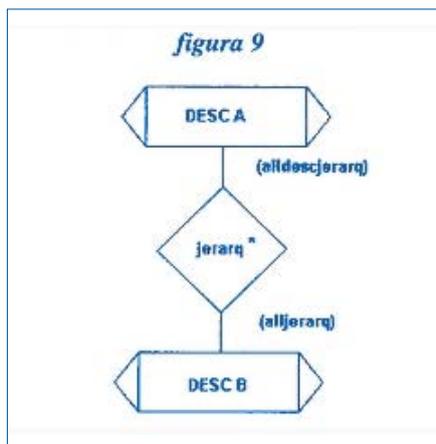
rencia. Para ello se partirá de la representación diagramática de la **figura 9**. La resolución vendría, respondiendo a la pregunta ¿qué atributo COD_DESC a través de la conexión *alljerardesc* (se preguntaría por códigos de DESC B) son iguales a CODIGO_DESC? pero tal que, estos códigos tienen a su vez que cumplir que COD_DESC a través de *alljearq* (códigos de DESC A) sean iguales a CODIGO_DESC'.

La pregunta en GORDAS quedaría expresada así:

```
(COD_DESC OF alljearqdesc:
COD_DESC OF alljearq EQUAL
CODIGO_DESC') EQUAL CODIGO_DES
```

Si la respuesta es TRUE, entonces existe una relación JERARQ (en realidad en JERARQ*) entre esas dos entidades DESC. En este caso, como lo que se pretende es que no exista, para que así se pueda dar la relación ASOC se tiene que poner la negación de la expresión:

NOT <expresión lógica>

**CONCLUSIONES Y APORTACIONES**

En los últimos años se han desarrollado numerosos sistemas de gestión automática de tesauros, tal y como se puede ver en “CoñYección Automática de Tesauros” “Thesauri facing new chachenges.”²⁵ Sin embargo el desarrollo de la inmensa mayoría está basado en técnicas tradicionales de construcción de software, sin llegar a plantearse la necesidad de un mo-

delado del sistema. En “Cambio automático del lenguaje pivote en un tesauro multilingüe informatizado”²⁶ se introduce el concepto de un lenguaje pivote con base en el cual se produciría un total desarrollo de un tesauro multilingüe. En *Aportaciones a la automatización de la normalización del tratamiento documental*²⁷ se sigue trabajando con el mismo enfoque tradicional, pero se hace especial hincapié en la necesidad de identificar las entidades y las posibles asociaciones que se pueden establecer entre ellas en todos los niveles de jerarquía de un tesauro.

Es evidente que existen otras posibilidades de tratar nuestra hipótesis de trabajo que en algunos casos son complementarias, y en otras alternativas. Nos referimos a las aportaciones de la inteligencia artificial, que se consideran complementarias a nuestra propuesta, de forma que ésta alcanzara un mayor grado de interactividad con el usuario, como lo muestra *Logic Based Approach to Semantic Query Optimization*²⁸ mediante la optimización semántica en la interrogación en bases de datos deductivas. Si bien, esto es en sí un buen objetivo, antes de tratarlo era necesario conocer y tener la experiencia tanto en, la estructura de los tesauros como en la respuesta que iban a dar éstos, aplicándoles un MD para conseguir un esquema conceptual.

Con todo ello se obtuvieron las siguientes conclusiones:

* En el marco de los sistemas de recuperación de la información, los lenguajes documentales se ofrecen como herramientas adecuadas. Su desarrollo ha hecho posible la sincronización de significados en los mismos términos, tanto cuando se trata de indizar para su almacenamiento como a la hora de su recuperación. Su estructura normalizada y normativa trata de evitar las ambigüedades propias del tratamiento del lenguaje natural, lo que permite la concordancia de ambos puntos del ciclo de la información.

25 E. Laguna Serrano y otros. “Confección Automática de Tesauros.” *Revista Española de Documentación Científica*. 1989, p. 129-140.

26 J. R. Pérez Alvarez-Osorio y otros. “Cambio Automático de lenguaje pivote en un tesauro multilingüe informatizado.” *I Jornadas Españolas de Documentación Automatizada*. 1986, p. 601-620.

27 F. J. Martínez Méndez. *Aportaciones a la Automatización de la normalización del tratamiento documental*. 1989, 116 p.

28 U.S. Chakravarthy et. al. “Logic-Based Approach to Semantic Query Optimization.” *ACM...* 1990, p. 162-207.

- * Frente a los primeros lenguajes de clasificación que empezaban a plantear problemas por su rigidez estructural, aparecieron los tesauros como herramientas flexibles de representación de conceptos codificados fácilmente combinables, con una estructura semántica importante, que lo han hecho el lenguaje documental más adecuado para la indización.
- * A pesar del aparente marco teórico en el que parece que se desenvuelve la construcción de tesauros, no es menos cierto que ésta carece de una base teórica conceptual que consiga, en mayor o menor grado, la eliminación de ambigüedades.
- * La creciente preocupación por desarrollar principios metodológicos que permitan que el trabajo de diseño se lleve a cabo en forma más eficiente y sistemática, es lo que lleva a concebir modelos teóricos, siendo éstos los mecanismos de abstracción que permiten describir objetos, sus relaciones, la semántica, restricciones, etcétera, que ofrecen además una estructuración flexible, de manera que ayuden a aquellas personas que se enfrenten con el modelado.
- * Una alternativa a los problemas de diseño mediante el uso de modelos, es la

introducción de una etapa intermedia de diseño: la definición del esquema conceptual, el cual nos da una representación del mundo real, independiente tanto de las condiciones de almacenamiento como de la eficiencia de los sistemas físicos.

- * El proceso de optimación es muy útil para el diseño lógico desde el punto de vista práctico, el modelo ER optimado por nosotros, proporciona una fuente de estructuras de datos estables, desde la cual se puede derivar una estructura lógica final adecuada para los requerimientos de acceso a los datos.
- * El modelo ER extendido responde a la búsqueda de una herramienta simple, pero completa, este modelo se encuentra fuertemente definido, y permite modelar de manera natural gracias a la clara distinción que hace entre objetos y sus relaciones, lo que le ha llevado a ser uno de los modelos más ampliamente aceptados.
- * Como se esperaba, el modelo ER extendido se muestra como un soporte adecuado para la construcción del esquema del tesoro. Una vez elegidos los elementos relevantes y más significativos constitutivos del mismo, queresponden a una estructura semántica consistente y completa, se logró

dar una definición de tesoro, proponiendo el “concepto tesoro” dentro de un marco teórico.

- * El propósito es pues, diseñar un esquema conceptual para definir el tesoro como un concepto en sí mismo, para obtener una herramienta que permita la creación de las bases de datos de tesauros específicos, pero con base en un esquema capaz de detectar las posibles incoherencias, lo que proporcionará la deseada integridad de los datos.
- * Se adoptó GORDAS como medio de especificación de los subesquemas ER por ser un lenguaje comprometido, por un lado, con las demandas que hacen referencia a la información y además se expresen en términos cercanos a las percepciones del usuario, y, por otro, por ser un lenguaje formal y no ambiguo. Así, en los subesquemas su implementación ha sido simple, incluso en aquellos casos donde las restricciones explícitas no estaban soportadas por el modelo ER.
- * De esta forma la elaboración de un producto requerirá de un mínimo esfuerzo, ya que se parte de un diseño robusto y consistente:

REFERENCIAS BIBLIOGRÁFICAS

- ABRIAL, J.R. “Data semantics.” En Data Base Management. Klimbie, J.W. and Koffeman, K.L. (eds.). *Proceedings of IFIP Conference on Data Base Management*, North-Holland, Amsterdam, 1974. p. 1-60.
- BACHMAN, C.W.; M. Daya. “The Role Concept in Database Models.” En *Proceedings of 3rd IEEE International Conference on Very Large Databases*, Tokio, 1977. p. 464-476
- BILLER, H.; E. Neuhoj,d. “Semantics of data bases: the semantics of data models.” *Information Systems*, Vol. 3, No. 1, 1978. p. 11-30.
- BRACCHI, G.; P. Paolini. and G. Pelagatti. “Binary Logical Associations in Data Modelling.” En I.M. NIJSSEN (ed) *Modelling in Database Management Systems, Proceedings of IFIOP TC2 Conference*, Freudenstadt, Amsterdam, 1976.
- CDU. *Clasificación Decimal Universal*. 6° ed. abreviada. Madrid: AENOR, 1991. 462 p.
- CODD, E. F. “Extending the Database Relational Model to Capture More Meaning.” *ACM Transactions on Database Systems*, Vol. 4, No. 4, Dec 1979, pp.397-434; IBM Research Report RJ2599, San Jose, (Calif.), Aug 1979.

- CHAKRAVARTHY, U.S.; J. Grant and J. Minker. "Logic-Based Approach to Semantic Query Optimization." *ACM Transactions on Databases Systems*, Vol. 15, No. 2, June 1990, p. 162-207.
- CHAMBERLIN, D.D. "SEQUEL 2: a Unifies Approach to Data Definition, Manipulation and Control." *IBM Journal of Research and Development*, Vol. 20, No. 6, Nov. 1976, p.560-575.
- CHEN, P.P. "The Entity-Relationship Model-Toward a Unified View of Data." *ACM Transactions on Databases Systems*, Vol. 1, No. 1, 1976. p. 9-36.
- DATE, C.J. *An Introduction to Database Systems*. 3rd ed., Addison-Wesley, 1981. 4rd. A-W, 1986.
- DAVIS, C.G. *et al.* (Eds.) *Entity-relationship approach to software engineering*. North-Holland, Amsterdam, 1983.
- ELMASRI, R.; G. Wiederhold. "Data model integration using the structural model." *Proceedings of ACM SIGMOD Conference*, 1979. p. 191-202.
- . "GORDAS: A formal high-level query language for the entity-relationship model." En *Entity-Relationship Approach to Information Modeling and Analysis*, P.P. Chen (ed.). North-Holland, 1983. p. 49-72.
- FALKENBERG, E. "Concepts for modeling information." *Modelling in Database Management System*, North-Holland, Amsterdam, 1976.
- FUGMANN, R. "An interactive Classaurus on the PC." *International Classification*, Vol. 17, No. 3/4, 1990. p. 133-137.
- GANZMANN, J. "Criteria for the evaluation of thesaurus software." *International Classification*, Vol. 17, No. 3/4, 1990. p. 148-157.
- HAMMER, M.; D. McLeod. "The semantic data model: A modeling mechanism for database applications." *Proceedings of ACM SIGMOD Conference*, 1978. p. 26-36.
- "Database Description with SDM: A Semantic Database Model." *ACM Transactions on Database Systems*, Vol. 6, No.3, Sep. 1981. p. 351-386.
- HOGSHEAD DAVIS, K. "Need for "Flexibility" in a Conceptual Model." *Information & Management*, Vol. 18, 1990. p. 231-241.
- KLUG, A. and D. Tsichritzs. (ed). *The ANSI/X3/SPARC Report of the study Group on Data Base Management Systems* (AFIPS Press, 1977).
- LAGUNA SERRANO, E.; A. de Irazabal Nerpell; A. Valle Bracero. "Confeción Automática de Tesauros." *Revista Española de Documentación Científica*, Vol. 12, No. 2, 1989. p.429-140.
- LANNING, R.W. *An Entity-Relationship Approach to Model Management*. Elsevier Science Publishers B.V. 1986. p. 65-72.
- LUM, V. *New Orleans Database Design Workshop Report*, IBM Report RJ2554, 1979.
- MARTÍNEZ MÉNDEZ, F.J. *Aportaciones a la automatización de la normalización del tratamiento documental*. Escuela Universitaria de Informática. Universidad de Murcia, 1989. 116 p.
- MINSKY, M. "Computer science and representation of knowledge." *Proceedings of National Computer Conference*. AFIPS 42, 1973.
- NAVATHE, S.B.; M. Schkolnick. "View representation in logical database design." *Proceedings of ACM SIGMOD Conference*, 1978. p. 144-156.

- PÉREZ ÁLVAREZ-OSSORIO, J.R.; A. Riudavets Montes; A. Valle Bracero. "Cambio automático de lenguaje pivote en un tesoro multilingüe informatizado." *Jornadas Españolas de Documentación Automatizada*, 1986. p.601-620.
- RADA, R. "Maintaining Thesauri and Metathesauri." *International Classification*, Vol. 17, No. 3/4, 1990. p.158-164.
- RITZLER, CI. "Comparative Study of PC-supported Thesaurus Software." *International Classification*, Vol. 17, No. 3/4, 1990. p. 138-147.
- ROBERTS, N. "The prehistory of the information retrieval thesaurus." *Journal of Documentation*. 40 (4), 1984. p. 271-285.
- ROUSSOPOULOS, N.; J. Mylopoulos. "Using semantic networks for data base management." *Proceedings of ACM International Conference on Very Large Data Bases*, 1975.
- SCHEUERMANN, P.; G. Schiffner; H. Weber. *Abstraction capabilities and invariant properties modelling within the entity-relationship approach*. Chen (ed.) North-Holland Publishing Company, 1980. p. 121-141.
- SCHMID, H.; J. Swenson. "On the semantic of the relational data model." *Proceedings of ACM SIGMOD Conference*, 1975.
- SCHMITZ-ESSER, W. "Thesauri facing new challenges." *International Classification*, Vol. 17, No. 3/4, 1990. p. 129-132.
- SHIPMAN, D.W. "The functional Data Model and the Data Language DAPLEX." *ACM Transactions on Database Systems*, Vol. 6, No. 1, March 1981. p. 140-173.
- SMITH, I.M. and D.C.P. Smith. "Database Abstractions: Aggregation and generalization." *ACM Transactions on Database Systems*, Vol. 2, No. 2, Jun 1977. p. 105-133.
- STONEBRAKER, M; E. Wong, P. Kreps, and G. Held. "The Design and Implementation of INGRES." *ACM Transactions on Database Systems*, Vol. 1, No. 1, Sep. 1976, p. 189-222.
- VAN SLYPE, G. *Los lenguajes de Indización: Concepción, construcción y utilización en los sistemas documentales*. Madrid: F. Germán Sánchez Ruipérez, 1991.
- WEELDREYER, S.A. *Structural aspects of the entity-category-relationship model of data*. Honeywell CCSC, technical report, HR-80-251. Bloomington, Minnesota. March, 1980.
- WIEDERHOLD, G. *Database Design*. Mac Graw Hill, N.Y. 2nd. ed. 1977.
- . R. ELMASRI. "The structural Model for Database Design." *Proceedings of International Conference on the Entity-Relationship*. Approach to Systems Analysis and Design, Los Ángeles (Calif.), Dec 1979