

Documentación periodística y bases de datos: elemento para su fundamentación como disciplina y propuesta de conjunto nuclear de diccionario de datos

Lluís Codina

Universitat Pompeu Fabra, Barcelona
lluis.codina@cpis.upf.es

M. Eulàlia Fuentes

Universitat Autònoma de Barcelona
eulalia.fuentes@uab.es

RESUM

S'estableixen els aspectes bàsics de cara a la confecció d'un diccionari de dades, document necessari de cara a crear i administrar una base de dades (en aquest cas les indicacions són per a la creació d'un diccionari de dades per a una base de dades de documentació periodística), així com exemples de camps concrets per al disseny d'aquesta base de dades. Finalment s'estudien els usos i aplicacions d'aquest model, així com les seves limitacions.

RESUMEN

Se establecen los aspectos básicos de cara a la confección de un diccionario de datos, documento necesario para la creación y administración de una base de datos (en este caso las indicaciones son para la creación de un diccionario de datos para una base de datos de documentación periodística), así como ejemplos de campos concretos para el diseño de esta base de datos. Finalmente se estudian los usos y aplicaciones de este modelo, así como sus limitaciones.

ABSTRACT

Basic aspects used to make a date dictionary, document needed to create and administrate a database (in this case indications are for the creation of a date dictionary for a journal documentation database) are established, as well as examples of concrete fields used to design this database. Finally, uses and applications of this modes, as well as its limitations, are studied.

1. Introducción

Hace tiempo que los ordenadores han penetrado en los centros de documentación y en los archivos de prensa. Pese a ello, estaba pendiente una propuesta formal de diccionario de datos, detallada y articulada, que fuera de utilidad general para las diversas clases de organismos e instituciones que pueden estar interesadas en desarrollar bases de datos de prensa.

Son muchos y muy variados los organismos y entidades que deben gestionar información periodística en sus centros de documentación además, como es obvio, de las empresas periodísticas.

Por citar solamente los casos más característicos:

- a) Casi todas las grandes instituciones de la Administración y del Estado, desde ministerios hasta ayuntamientos, pasando por gobiernos autonómicos, diputaciones o parlamentos, acostumbran a producir bases de datos de prensa, generalmente, de uso interno.
- b) Muchas corporaciones medianas y grandes, ya sean del sector público o privado, suelen mantener sus propios servicios de información basados en documentación periodística.
- c) Existe una amplia oferta de bases de datos comerciales cuyo contenido principal está formado por documentación periodística.
- d) Existe una variada constelación de organismos que desarrollan bases de datos de prensa para el propio consumo y con fines sumamente diversos. Por citar un ejemplo que puede resultar significativo, uno de los autores de este trabajo es colaborador del Observatorio de la Comunicación Científica (www.upf.es/occ), un centro especial de investigación de la Universidad Pompeu Fabra de Barcelona que apoya muchas de sus investigaciones y proyectos de I+D en una base de datos de prensa de creación propia que vacía las noticias de prensa que publican los principales diarios españoles.¹

Lo que esto significa es que, en cualquier momento dado, podremos encontrar a una amplia variedad de entidades y organismos, además, por supuesto, de las propias empresas periodísticas, que necesitan diseñar y producir bases de datos cuyo contenido principal o exclusivo son informaciones de prensa.

En este trabajo se ofrece, por tanto, una propuesta de diccionario de datos que denominamos conjunto nuclear de diccionario de datos para bases de datos de prensa, abreviadamente, CNDD(P), por Conjunto Nuclear de Diccionario de Datos para (P)rensa.

El propósito es el de ofrecer un conjunto de meta datos que puedan utilizarse en el diseño y producción de bases de datos de prensa tal cual se ofrecen aquí (*escenario A*, o caso ideal); o con los cambios que cada contexto requiera, en la confianza que estos cambios, o bien consistirán en utilizar una parte del conjunto de metadatos que aquí se ofrecen (*escenario B*), según interese en cada contexto o bien el conjunto consistirá en ampliar este conjunto (*escenario C*), pero utilizando todos o una buena parte del conjunto que se ofrece aquí. En los tres escenarios la idea es proporcionar todo o alguna parte significativa del trabajo resuelta.

2. Diccionarios de datos y metadatos

¿Qué es un diccionario de datos? El diccionario de datos es un documento escrito que actúa como una herramienta que ayuda al diseñador y administrador de una base de datos a garantizar la calidad, la fiabilidad, la consistencia y la coherencia de la información introducida en la base de datos, de tal manera que el diccionario de datos marcará decisivamente el rendimiento y la calidad global del sistema de información. Consiste en una descripción detallada, según se indicará más adelante, de todos los campos que forman (o formarán) parte del modelo de registro de una base de datos.

1. Puede verse una buena muestra de la variedad de la oferta en bases de datos relacionadas con la información periodística, y la información en general, en el trabajo de María del Valle Palma (1999).

En realidad, el diccionario de datos es una herramienta de doble significado. En primer lugar es una herramienta de diseño ya que proporciona una guía para que el diseñador de la futura base de datos pueda adoptar decisiones sobre el papel (literalmente), es decir, para poder tener toda la estructura teórica de la base de datos delante de él antes de comenzar la carga de datos y, por tanto, a tiempo de hacer rectificaciones sin que eso suponga tirar trabajo anterior por la ventana.

Una vez la base de datos pasa a la fase de producción, el diccionario de datos se transmuta en guía de estilo para garantizar la calidad del producto, es decir, la calidad de la base de datos.

A diferencia de otros sistemas de información, por ejemplo, los de tipo ofimático, las bases de datos documentales son proyectos de larga duración y basan su utilidad en la cantidad de información que acumula una sola base de datos a lo largo de meses, años e incluso décadas.

Pero esta ventaja se desvanece fácilmente si la larga duración implica numerosas y frecuentes inconsistencias. Estas inconsistencias se producen si no existe un libro de estilo claro que señale sin ambigüedades qué clase de información acepta cada campo, qué tipo de datos corresponde a cada campo, qué lengua, etc.

Si diversos equipos de personas han trabajado en una misma base de datos a lo largo de meses o años sin un buen libro de estilo, es inevitable que la coherencia, la consistencia y la calidad de la información de la base de datos se resienta, y puede resentirse hasta tal que punto que la haga, de facto, inutilizable en cuestión de poco tiempo.

Precisamente, lo que pretende evitar el diccionario de datos es este problema. Al indicar con toda claridad qué significado tiene cada uno de los campos de la base de datos y proporcionar otros tipos de indicaciones, el diccionario de datos se convierte en la mejor garantía de calidad de cualquier proyecto de información a largo plazo.

De hecho, antes de que normas como Dublin Core pusieran de moda el término «metadatos» en la comunidad documental, los diccionarios de datos han sido considerados uno de las clases de metadatos (datos sobre datos) imprescindibles en el mundo de los sistemas de información.

De hecho, creemos que nunca se insistirá lo suficiente que aquello que hace (aparentemente) inteligente a un sistema de información es la combinación de información sistematizada más metadatos. Cualquier sistema de información documental, por tanto, se beneficiará del uso de diccionarios de datos.

3. Conceptos de bases de datos

Antes de proponer una estructura de diccionario de datos y un conjunto de campos descritos con esa estructura, sería conveniente introducir algunos conceptos que ayudarán a entender esta propuesta.

En concreto, en la tabla siguiente definiremos algunas nociones que creemos necesario clarificar antes de seguir adelante.

TABLA 1: CONCEPTOS BÁSICOS PARA EL DISEÑO DE BASES DE DATOS

Concepto	Definición
Atributo	Una propiedad relevante de una entidad (p.e., los atributos relevantes de una noticia de prensa pueden ser el autor, la fuente, etc.).
Base de datos	Un sistema de información que mantiene registros sobre una parte del mundo real. En lugar de limitarnos a ver a una base de datos como una colección de informaciones (concepto que en sí mismo es correcto), con esta noción nos acostumbramos a pensar en una base de datos como un sistema que debe mantener una estrecha correspondencia de semejanza (de mapeo) con alguna parte determinada de la realidad. Esto ayuda a diseñar la base de datos: la base de datos será tanto más eficiente cuanto mejor represente a la realidad.
Campo	Una de las zona en las que se puede articular un registro. Cada campo en un registro corresponde a un atributo en una entidad, por tanto también puede decirse que un campo es una representación de un atributo de una entidad (p.e., campos como autor, fuente, etc., suelen formar parte de un modelo de registro para documentación de prensa).
Dominio	El conjunto de valores que puede admitir un campo.
Entidad	Las cosas del mundo real (cosas abstractas o concretas, cosas animadas o inanimadas) que están (o estarán) representadas la base de datos. Si podemos representar entidades es porque tienen propiedades reconocibles. Estas propiedades se denominan atributos.
Mundo real	Parte de la realidad, física o conceptual, que está representada en una base de datos. Se denomina mundo real por oposición al conjunto de ordenador más software, que es el sistema de información o mundo simbólico.
Registro	<p>La representación de una entidad en una base de datos. Un registro se compone de campos. En el mundo real identificamos entidades, su representación en una base de datos se llama registro.</p> <p>En realidad, por rigor, deberíamos hablar, respectivamente, de tipos de entidad y de tipos de registros, a fin de distinguir entre los tipos y sus ocurrencias. Un modelo de registro para representar noticias de prensa es, efectivamente, un tipo de registro; en cambio el registro n. 123 es una ocurrencia concreta de registro (la n. 123). El conjunto teórico de las noticias de prensa que pueden estar representadas en una base de datos forman el tipo de entidad noticia de prensa. La noticia de la fin de la guerra de Kosovo es una ocurrencia de noticia.</p>
Tipo de dato	<p>La naturaleza de los datos que admite un campo. Esta naturaleza determina el tipo de operaciones que admite un campo. Los tipos de datos más comunes son:</p> <ul style="list-style-type: none"> – Texto (o alfanuméricos) – Números – Fechas – Lógicos – Gráfico (o binario) – etc. <p>Un campo de texto admite operaciones de ordenación alfabética, pero no aritméticas, mientras que un campo numérico admite operaciones aritméticas, un campo lógico admite valores de verdadero y falso, etc.</p> <p>Las bases de datos manipulan declaraciones de tipos de datos para «saber» qué hacer con un campo. Ante un campo texto, la base de datos sabe que el conjunto de ceros y unos que contiene debe traducirlo a cadenas de caracteres, que puede ordenarlo o compararlo con otro campo de texto, etc. En cambio, ante un campo gráfico, sabe que el conjunto de ceros y unos que contiene el campo debe transformarlo en un mapa de bits, etc.</p>
Valor	El contenido concreto de un campo concreto.

4. Estructura de un diccionario de datos

Nuestra propuesta de estructura de diccionario de datos es la siguiente:

TABLA 2: ESTRUCTURA DE UN DICCIONARIO DE DATOS

Parámetro	Significado
Etiqueta	Nombre único con el cual se identifica cada uno de los campos. La etiqueta es el nombre interno que utiliza la base de datos para nombrar los campos y puede diferir del texto que aparece en las vistas de usuario de la base de datos. Algunos programas de gestión de bases de datos imponen grandes restricciones, por ejemplo, los nombres de etiqueta no pueden contener espacios o determinados caracteres (como «/», «—», etc.) y no pueden tener más de n caracteres (por ejemplo, no pueden tener más de 3, 8, 12 caracteres, etc.). En cambio, esos mismos programas pueden no presentar apenas restricciones a los nombres atribuidos a los campos en la vista de usuario.
Dominio	Definición, por extensión o por comprensión, del conjunto de valores que admite el campo.
Tipo de dato	Tipo de dato del campo, p.e., numérico, alfanumérico, lógico, etc. (ver la tabla n. 1 para una definición más amplia).
Lengua	La lengua del campo debe ser en cada caso, o bien la lengua del documento o bien la lengua del centro de documentación.
Indización	Señala si un campo debe o no estar indizado.
Tratamiento documental	Indica, si el caso, qué tratamiento documental debe recibir un campo. Las posibilidades básicas son: lenguaje libre o lenguaje controlado. En este último caso, puede ser un lenguaje controlado cerrado (una lista de valores predefinidos) o abierto (un campo de descriptores controlado por un tesaurus, por ejemplo). Si es el caso, el diccionario de datos puede indicar en este parámetro, o bien la lista de valores o bien la norma o lenguaje de indización que se debe utilizar.
Controles de integridad	Los controles de integridad puede ser muy variados y sus posibilidades reales de implementación en una base de datos depende del programa de gestión de bases de datos utilizados. Los controles más habituales son: – Campo obligatorio (debe contener algún valor obligatoriamente) – Campo único (no puede haber otro campo de otro registro con el mismo valor) – Valor único (el campo solamente puede tener un valor) etc.
Observaciones	Notas u observaciones si es el caso.
Ejemplos válidos	Ejemplos de valores válidos para el campo.

Como hemos indicado, el diccionario de datos es, antes que nada, un documento textual que puede o bien imprimirse en papel o bien administrarse mediante una base de datos.

En cualquier caso, buena parte de la eficiencia del diccionario de datos depende la posibilidad de implementarlo en el programa de gestión de la base de datos. Por ejemplo, la mayor parte de la eficacia de los controles de integridad se perderá si el programa de gestión de bases de datos no advierte al operador que un campo obligatorio está vacío, o si no advierte que un campo que únicamente debería admitir una lista de valores se ha llenado con un valor no declarado en la lista, etc.

En tales casos, es decir, cuando el programa de gestión de bases de datos presenta pocas o ninguna posibilidad de implementar las características del diccionario de datos, la eficacia del mismo queda limitada a la capacidad de autodisciplina o de rigor de los operadores humanos que se ocupen de la carga de datos. Pero aún en el mejor de los casos, se producirán muchos más errores que cuando el control se realiza de forma automática. Por ejemplo, sin control automático de integridad, inevitablemente, en un porcentaje de registros quedará vacío, por error, algún campo estratégico, por ejemplo el título de la noticia, la fuente, la fecha de publicación, etc.

La consecuencia que se desprende es obvia: los responsables de centros de documentación deberían elegir siempre que pudieran sistemas de gestión de bases de datos en los que se pueden implementar la mayor cantidad posible de parámetros del diccionario de datos.

También debería ser obvio que no todos los parámetros del diccionario de datos son implementables de forma directa en el sistema de gestión de bases de datos. En concreto, aspectos como el dominio o la lengua del campo no suelen ser fácilmente implementables, salvo como información para el usuario. Supóngase que el dominio de un campo se expresa por comprensión, salvo con el uso de técnicas avanzadas de inteligencia artificial, no existe forma de incorporar esa definición en la base de datos. En general, los sistemas de gestión de bases de datos que se pueden encontrar en el mercado no permitirán implementar de manera automática los dos parámetros indicados: el dominio y la lengua.

5. Conjunto Nuclear de Datos para Prensa

Nuestra propuesta de un conjunto nuclear de diccionario de datos estándar para prensa incluye los campos indicados en la Tabla n. 3, donde los campos se han separado, por conveniencia expositiva en dos grupos: i), atributos de la entidad (en nuestro caso, noticias de prensa); ii), campos de control administrativo.

TABLA 3: LISTA DE CAMPOS DEL CND(P)

<i>Campos de representación de atributos de la noticia</i>	<i>Campos de control administrativo</i>
(01) Título (02) Autor (03) Fuente (04) Lugar publicación (05) Lugar redacción (06) Fecha publicación (07) Portada (08) Idioma (09) Género (10) Clasificación (11) Descriptores temáticos (12) Descriptores geográficos (13) Descriptores biográficos (14) Descriptores entidades (15) Descriptores cronológicos (16) Recursos gráficos (17) Resumen	(18) Fecha de alta (19) Fecha de modificación (20) Operador (21) Número de registro

Como se puede ver, nuestra propuesta incluye campos para representar el núcleo de los atributos del tipo de entidad *noticia de prensa* (campos 1-17) más campos para establecer un mínimo sistema de control administrativo (18-21).

Sobre estos últimos cuatro campos cabe señalar que su utilidad puede considerarse casi universal o, dicho de otro modo, puede llegar a pagarse un precio absurdamente alto si no se utilizan en una base de datos. La razón es la siguiente: en el ciclo de vida de toda base de datos documental es probable que, por razones de mantenimiento, de corrección de errores, de recuperación de datos por fallos diversos, etc., alguna vez la recuperación de información dependa de alguno de los cuatro parámetros indicados: ¿qué registros se han entrado desde la fecha x?, o bien ¿qué registros ha entrado Fulano entre tal y tal fecha?, o aún, ¿qué registros fueron modificados entre tal y tal fecha?, etc.

5.1. Diccionario de datos del CND(P)

01

Campo Título	
Etiqueta	Título
Dominio	Título completo del documento incluyendo antetítulo y subtítulo si los tuviera. Se entra en el siguiente orden: Título: antetítulo: subtítulo
Tipo de dato	Alfanumérico
Lengua	Del documento
Indización	Sí
Tratamiento documental	NC*
Controles de integridad	Obligatorio
Observaciones	–
Ejemplos válidos	Del imperio al caos: un informe de la ONU afirma que la transición económica en la ex URSS ha sido letal para millones de personas

*: NC = No corresponde

02

Campo Autor	
Etiqueta	Autor
Dominio	Nombre del autor del documento. Se entra en este orden: «Apellidos, Nombre» en el caso de autores personales
Tipo de dato	Alfanumérico
Lengua	NC
Indización	Sí
Tratamiento documental	Diccionario de autoridades
Controles de integridad	Obligatorio
Observaciones	Si no aparece nombre de autor, se indicará «Redacción». Si es agencia, se indicará el nombre de la agencia como autor, p.e., «Agencia EFE»
Ejemplos válidos	Sabater, Fernando

03

Campo Fuente	
Etiqueta	Fuente
Dominio	Identificación del medio, incluyendo cabecera, fecha y paginación
Tipo de dato	Alfanumérico
Lengua	Del documento
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	–
Ejemplos válidos	El País, 19 de septiembre de 1999, p. 1

04

Campo Lugar publicación	
Etiqueta	LugarPub
Dominio	Topónimo de la sede del medio
Tipo de dato	Alfanumérico
Lengua	Del documento
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	-
Ejemplos válidos	Barcelona

05

Campo Lugar redacción	
Etiqueta	LugarRed
Dominio	Lugar de redacción del documento
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Diccionario de autoridades para nombres geográficos
Controles de integridad	-
Observaciones	El lugar de redacción y el lugar de publicación no tienen por qué coincidir. Un medio con sede en Barcelona puede publicar una noticia de su redacción en Madrid
Ejemplos válidos	Madrid

06

Campo Fecha publicación	
Etiqueta	FechaPub
Dominio	Fecha de publicación del documento
Tipo de dato	Fecha, con el formato dd/mm/aaaa
Lengua	NC
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	En este campo radica la posibilidad de hacer operaciones de recuperación por fechas concretas o por rangos de fechas
Ejemplos válidos	12/09/1999

07

Campo Portada	
Etiqueta	Portada
Dominio	Valores de verdad (V, F; o bien su equivalente Sí, No) por Verdadero, Falso, en relación a la pregunta: ¿Es noticia de portada?
Tipo de dato	Lógico
Lengua	NC
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	Este campo permite recuperar, de un tema dado, todos los documentos que han sido portada, o bien saber si un determinado tema a generado portadas, etc.
Ejemplos válidos	Sí, No

08

Campo Idioma	
Etiqueta	Idioma
Dominio	Idioma del documento
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Diccionario de autoridades de nombres de lenguas
Controles de integridad	Obligatorio
Observaciones	-
Ejemplos válidos	Inglés, Castellano, etc.

09

Campo Género	
Etiqueta	Género
Dominio	Género periodístico del documento
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Lista cerrada de valores (descriptores de género)
Controles de integridad	Obligatorio
Observaciones	-
Ejemplos válidos	Crónica, Entrevista, Editorial, etc.

10

Campo Clasificación	
Etiqueta	Clasificación
Dominio	Ámbito sectorial de la noticia
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Lista cerrada de valores (nombres de ámbito o de sección)
Controles de integridad	Obligatorio
Observaciones	-
Ejemplos válidos	Economía, Internacional, Cultura, etc.

11

Campo Descriptores temáticos	
Etiqueta	DescTema
Dominio	Conceptos relevantes del documento, expresados mediante descriptores, según el principio N temas relevantes = N descriptores
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Lenguaje controlado de tipo asociativo, p. e., tesauro
Controles de integridad	Obligatorio
Observaciones	Deben asignarse, como se indica en el dominio tantos descriptores como temas relevantes contenga el documento, y siempre aplicando las normas ISO/UNE de determinación de los temas de un documento y de construcción de tesauros. En cuanto al número, se recomienda asignar entre 3, como mínimo, y 20 descriptores como máximo
Ejemplos válidos	Cine, Música, Procesos de desconolización, Conflictos bélicos, Inflación, Juegos Olímpicos, etc.

12

Campo Descriptores geográficos	
Etiqueta	DescGeo
Dominio	Nombres geográficos de lugares (países, ciudades, regiones, continentes, etc.) afectados de manera significativa por la noticia (lugar del suceso, si es el caso, ámbito geográfico de influencia, nombres de estados afectados, etc.)
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Diccionario de autoridades de nombres geográficos
Controles de integridad	-
Observaciones	-
Ejemplos válidos	España, Cataluña, Estados Unidos, Hospitalet de Llobregat, París, Europa, Asia, Mediterráneo, etc.

13

<i>Campo Descriptores biográficos</i>	
Etiqueta	DescBio
Dominio	Nombre propios de personas involucradas de manera significativa en la noticia. Siempre que sea posible, es decir, cuando se posean datos suficientes, se indicará Nombre Primer Apellido. En el caso de personajes del mundo de la política, sindicatos, miembros del Gobierno y la Administración se indicará el cargo entre paréntesis
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Diccionario de autoridades de nombres de personas
Controles de integridad	–
Observaciones	–
Ejemplos válidos	Jordi Pujol (presidente de la Generalitat de Cataluña); José María Aznar (presidente del Gobierno español); etc.

14

<i>Campo Descriptores entidades</i>	
Etiqueta	DescEnt
Dominio	Nombres de entidades públicas o privadas, corporaciones, empresas, entidades de la Administración, ONG, etc., involucradas de manera significativa en la noticia
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Diccionario de autoridades para nombres de organismos y entidades
Controles de integridad	–
Observaciones	Se utilizará la forma mejor conocida en cada caso: acrónimo o forma desarrollada; en caso de duda, la forma desarrollada
Ejemplos válidos	ONU, Unión Europea, UNESCO, etc.

15

<i>Campo Descriptores cronológicos</i>	
Etiqueta	DescCron
Dominio	Año(s) de ocurrencia de la noticia o de influencia de la noticia o de la época histórica (décadas, siglos) afectada por la noticia
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	Se establecerá una forma estándar para representar años, décadas, rangos de años, siglos, etc.
Controles de integridad	–
Observaciones	El año o la fecha del hecho objeto de la noticia no tiene por qué corresponder con la fecha de publicación, de aquí la ne-

	cesidad de este campo. Por ejemplo, en 1999 abundaron las noticias referidas a la época de la guerra fría (1945-1989), por lo cual una noticia publicada en 1999 podía tener como descriptor cronológico el año 1950, o los años 60, etc.
Ejemplos válidos	Años 60, Siglo xx, 1999, 1945, 1914-1918, etc.

16

Campo Recursos gráficos	
Etiqueta	RecGráficos
Dominio	Recursos gráficos que acompañan a la noticia. La sintaxis es la siguiente: NombreRecurso: Título o pie del recurso
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	NC
Controles de integridad	-
Observaciones	-
Ejemplos válidos	Fotografía: El canciller alemán, Gerhard Schröder Infografía: Estación espacial MIR Tabla: Distribución del crecimiento económico por comunidades autónomas, 1990-1995

17

Campo Resumen	
Etiqueta	Resumen
Dominio	Resumen del documento. En las noticias con lid, puede consistir en una transcripción del lid
Tipo de dato	Alfanumérico
Lengua	Del centro de documentación
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	-
Ejemplos válidos	«Habrán camillas y bombonas de oxígeno en la cumbre, mientras los equipos deciden qué desarrollo deben usar los corredores para subir rampas tan duras»

18

Campo Fecha de alta	
Etiqueta	Fecha Alta
Dominio	Fecha de alta del registro
Tipo de dato	Fecha
Lengua	NC
Indización	Sí
Tratamiento documental	NC

Controles de integridad	Obligatorio
Observaciones	Algunas bases de datos tienen previsto la carga automática de este campo
Ejemplos válidos	13/09/1999

19

<i>Campo Fecha de modificación</i>	
Etiqueta	FechaModificación
Dominio	Fecha de modificación, si es el caso, del registro
Tipo de dato	Fecha
Lengua	NC
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	En un registro que no ha sido modificado, la fecha de alta y la fecha de modificación es la misma. Algunas bases de datos tienen previsto la carga automática de este campo.
Ejemplos válidos	13/09/1999

20

<i>Campo Operador</i>	
Etiqueta	Operador
Dominio	Identificador de la persona que ha creado el registro
Tipo de dato	Alfanumérico
Lengua	NC
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio
Observaciones	-
Ejemplos válidos	Fulano de Tal, FT, etc.

21

<i>Campo Número de registro</i>	
Etiqueta	NumReg
Dominio	Número de registro
Tipo de dato	Numérico
Lengua	NC
Indización	Sí
Tratamiento documental	NC
Controles de integridad	Obligatorio. Valor único. Valor no repetido.
Observaciones	Algunas bases de datos tienen previsto la carga automática de este campo
Ejemplos válidos	1, 123, 6789, etc.

5.2. Modelo de registro

Con el grupo de campos indicados en la tabla n. 3, y atendiendo al diccionario de datos desarrollado en el apartado anterior (5.1.) podemos tener un modelo de registro con la siguiente estructura y contenidos:

TABLA 5: TIPO DE REGISTRO PARA DOCUMENTACIÓN DE PRENSA
CON DATOS DE UNA NOTICIA (EJEMPLO REAL)

Título	Indonesia cede a la presión internacional: el ministro de defensa recomendará al presidente Habibie que acepte una fuerza de pacificación
Autor	García, Javier
Fuente	El País, 12 de septiembre de 1999, p. 2
Lugar publicación	Madrid (España)
Lugar redacción	Yakarta (Indonesia)
Fecha Publicación	12/09/1999
Portada	Sí
Idioma	Castellano
Género	Reportaje
Clasificación	Internacional
Desc. Temáticos	Relaciones internacionales, Conflictos bélicos, Matanzas, Independencia, Procesos de pacificación, Milicias populares
Desc. Geográficos	Timor Oriental, Indonesia, Dili
Desc. Biográficos	Wiranto (general ejército Indonesia), Bill Clinton (presidente EU), Yusuf Habibie (presidente de Indonesia)
Desc. Entidades	Gobierno de Indonesia, ONU, Unión Europea, Fondo Monetario Internacional
Desc. Cronológicos	1999
Recursos gráficos	Fotografía: Unas 10.000 personas se manifestaron ayer en Sydney en apoyo a la independencia de Timor Oriental
Resumen	Las presiones internacionales parecen haber surtido efecto sobre el Gobierno indonesio. Ayer, el general Wiranto, hombre fuerte del régimen de Yakarta, abrió la puerta a la aceptación por su Gobierno de una fuerza multinacional de paz que detenga la matanza de la población de Timor Oriental a manos de las milicias proindonesias. EE UU anunció que suspende toda la venta de armas a Indonesia, y la Unión Europea, que estudiará mañana una medida similar, expresó su respaldo a la interrupción del programa de ayudas del Fondo Monetario Internacional (FMI) a Yakarta.
Fecha alta	13/09/1999
Fecha modificación	13/09/1999
Operador	FT
Núm. registro	007788

6. Uso, alcance y limitaciones

Como ya se indicó en la introducción, no es necesario utilizar este diccionario de datos en su totalidad. Según la conveniencia del proyecto, este conjunto de datos puede ser solamente una parte de un conjunto mayor o, al revés, puede utilizarse solamente una parte de los 21 campos que forman esta propuesta.

Además, el lector observará que algunas especificaciones de este diccionario de datos son claramente discrecionales. Por ejemplo, la etiqueta. Se ha indicado aquí una etiqueta para cada campo por coherencia, puesto que todo diccionario de datos debería tener especificada claramente la etiqueta, pero es indudable que, en vez de una etiqueta como «Desc-Tema» puede elegirse «Descriptores_Temáticos» o bien «Descriptores Temáticos», etc., si el programa de gestión de bases de datos lo permite.

Por otro lado, el alcance de este diccionario se limita a la gestión de bases de datos sobre publicaciones impresas, es por ello que no se ha incluido ningún campo específico para expresar atributos propios de recursos digitales, tales como direcciones de Internet (URL). Sin embargo, no debería ser difícil adaptar este conjunto de meta datos para su aplicación a otros soportes, puesto que en lo esencial, las propiedades de la entidad noticia son similares.

Podemos señalar también que esta propuesta no se ocupa de algunos aspectos clásicos en la gestión documental de la información periodística y que suelen recibir tratamientos muy diferentes según los contextos.

Por ejemplo, siempre ha habido problemas para identificar y relacionar las diversas piezas que forman una misma unidad documento/noticia. Para exponer este problema debemos recordar que, una cosa es el hecho noticiable, que puede prolongarse a lo largo de días y semanas y que es el mismo hecho noticiable a través de diversos medios. Otra cosa es el documento, ya que podemos considerar que cada unidad narrativa con un título y un autor es un documento. Por tanto, un mismo diario puede publicar n documentos sobre el mismo hecho noticiable. Por ejemplo, la noticia de las agresiones armadas de la milicia a ciudadanos en Timor Oriental (hecho noticiable) ocupaba tres páginas en un mismo diario: en la portada y en las dos páginas siguientes y en cada una de las páginas existía por lo menos un documento distinto. Por tanto, aquí tenemos 1 hecho noticiable: 3 documentos.

¿Debe existir alguna convención para identificar y relacionar los n documentos que en un mismo diario se refieren a la misma noticia? Más aún, ¿debería existir alguna convención para relacionar los n documentos publicados por distintos medios que se refieren a desarrollo en el tiempo a un mismo hecho noticiable de largo alcance? Nosotros no entramos en esta clase de cuestiones, cuya respuesta dependerá mucho del contexto. Como en el caso anterior, sea cual la sea la solución que se adopte, el grueso del diccionario de datos que se presenta quedará sin experimentar cambios.

Por último, tampoco queda afectada esta propuesta por la posibilidad de relacionar los registros de la base de datos con el texto completo de los documentos o incluso con un facsímil de la noticia, por lo que tampoco nos pronunciamos en este trabajo sobre tal cuestión, salvo señalar, como aviso para navegantes, que disponer del texto completo de la noticia no es lo mismo que disponer de una representación estructurada en campos de la misma. Lo mejor, naturalmente, es tener ambas cosas.

7. Conclusiones

Dada una problemática de tratamiento documental recurrente, como es el caso de la gestión de información periodística en distintas clases de entornos, es posible hacer «economías de escala» con el uso del concepto de meta datos y de diccionario de datos, como hemos intentado hacer aquí.

Gracias a tales conceptos, los esfuerzos de análisis y diseño de diversos proyectos relacionados con la documentación periodística, y que se han intentado condensar aquí, podrían ser útiles para que futuros equipos de trabajo no se vean obligados a «reinventar la rueda» sino, que bien al contrario, por lo menos una parte del trabajo lo encuentren solucionado. Esta ha sido nuestra mejor esperanza al preparar esta propuesta.

8. Anexo: ilustraciones

Las ilustraciones que siguen consisten en capturas de pantalla de los programas documentales Idealist (distribuido en España por Technostock <www.technostock.es>) y FileMaker (FileMaker <www.filemaker.com>).

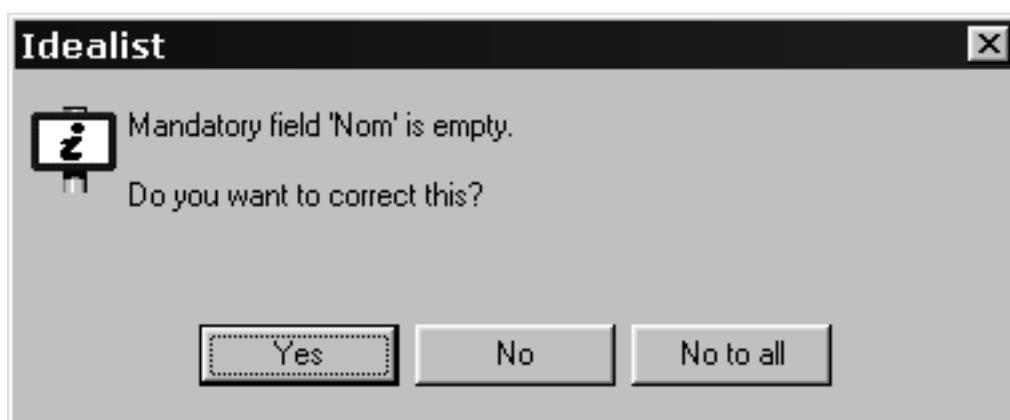
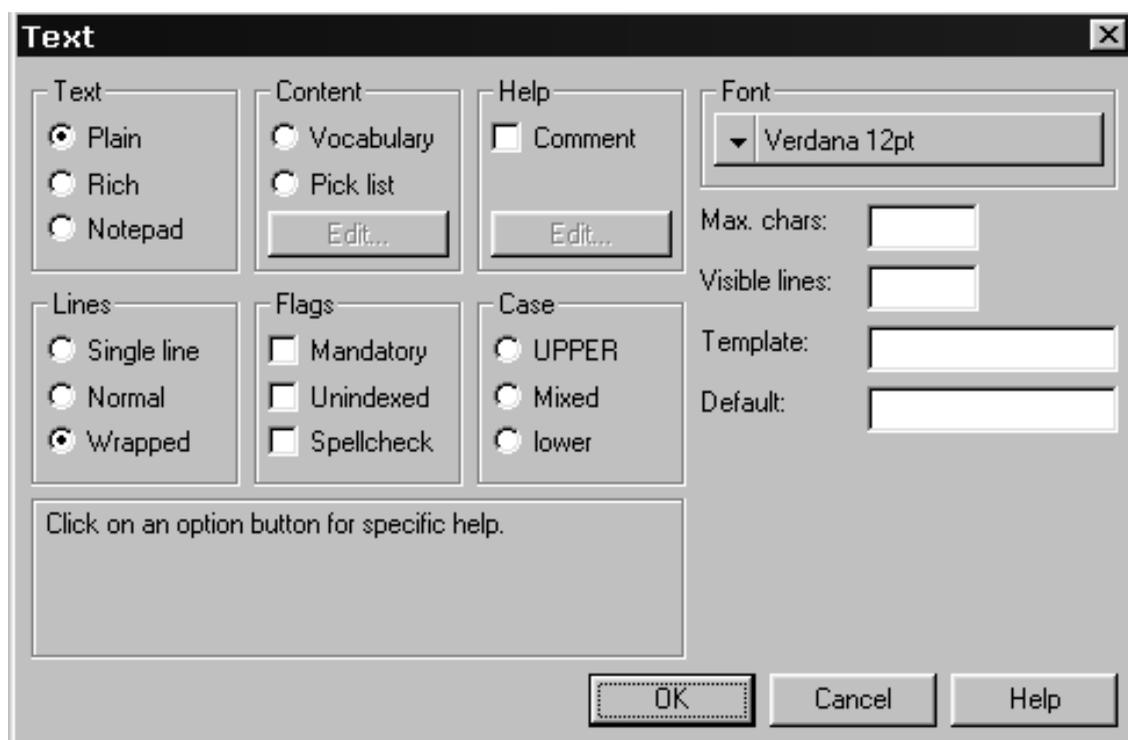
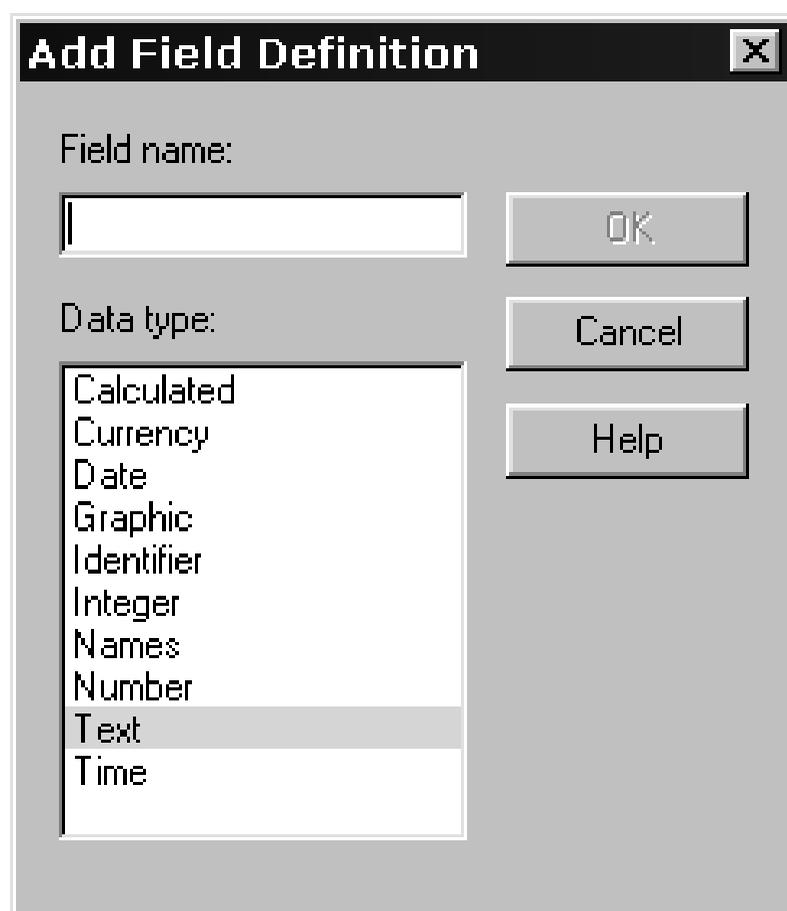
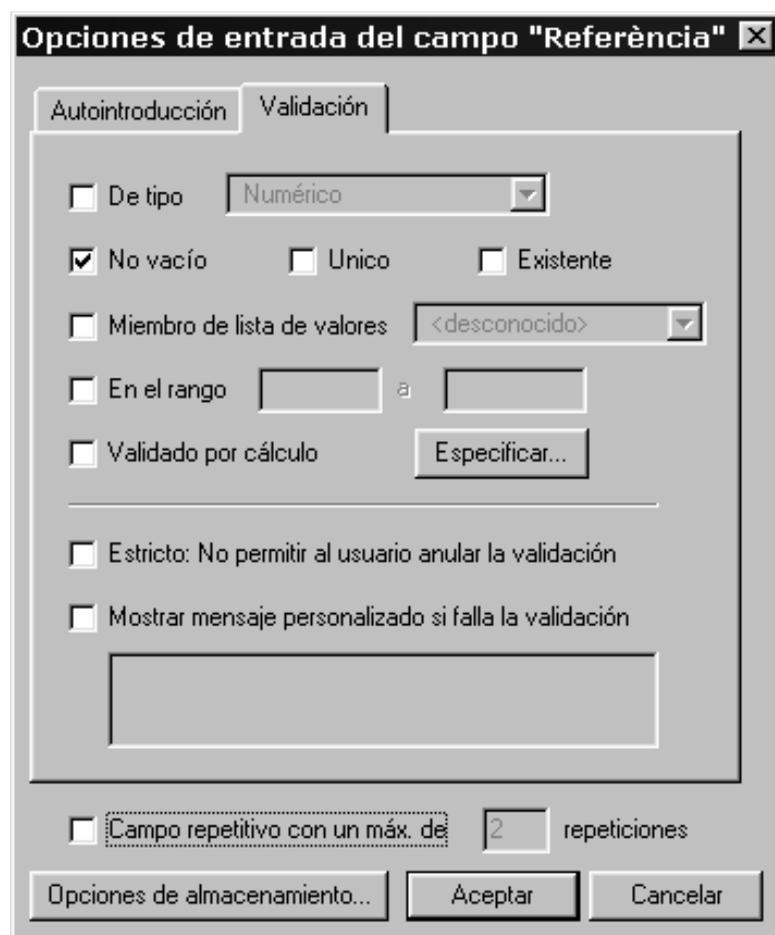
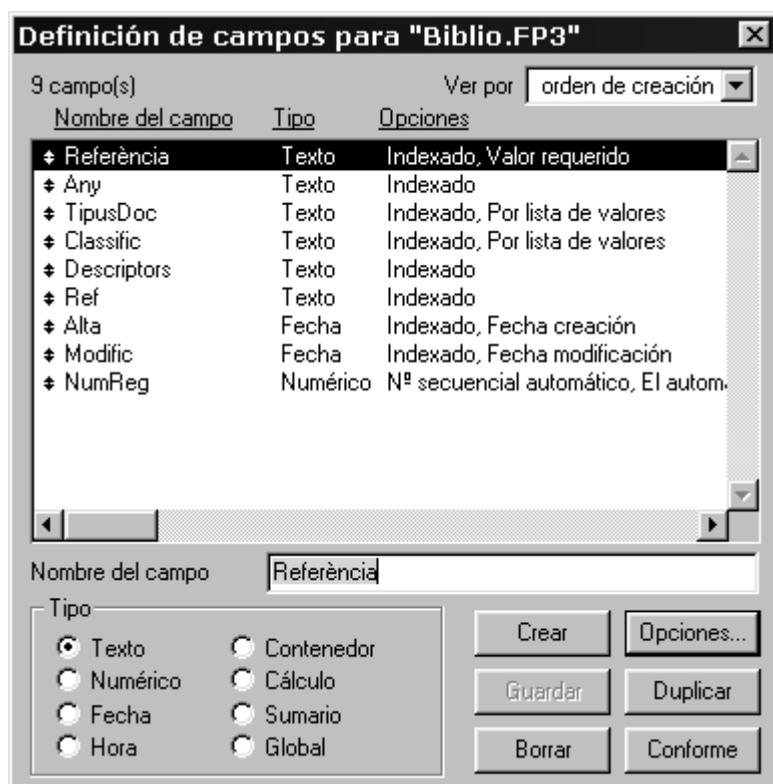


FIGURA 1: Un típico mensaje de error de un sistema de gestión de bases de datos advirtiendo que se ha dejado vacío un campo obligatorio. En este caso (programa Idealist) nos permite rectificar si se trata de un error o dejarlo vacío. Con ello se evita que quede vacío inadvertidamente.



FIGURAS 2 Y 3: Los tipos de datos que soportan algunos programas (Idealist) pueden ser numerosos (figura 2). En este caso tenemos 10 tipos de datos. Además podemos ver (figura 3) otras especificaciones del diccionario de datos que podemos implementar en los campos de tipo alfanumérico.



FIGURAS 4 Y 5: OTRA FORMA DE IMPLEMENTAR DICCIONARIOS DE DATOS, EN ESTE CASO DE FILEMAKER. VEMOS EN ESTE CASO LA POSIBILIDAD DE UTILIZAR 8 TIPOS DE DATOS. EN LA FIGURA SIGUIENTE ADEMÁS VEMOS DIVERSAS OPCIONES DE INTEGRIDAD TAL COMO LAS ENTIENDE FILEMAKER.

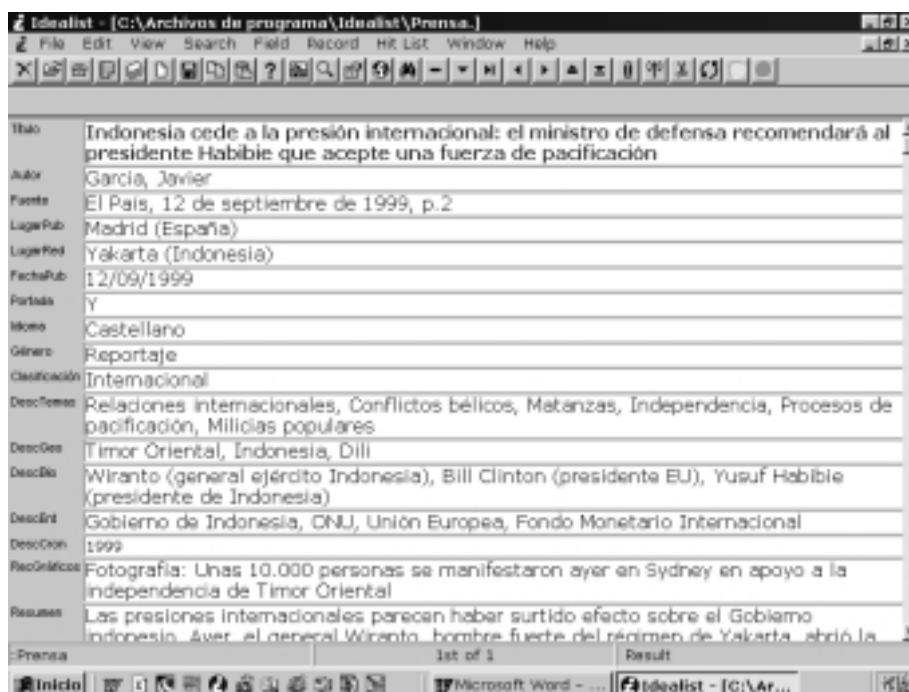


FIGURA 6: IMPLEMENTACIÓN EN IDEALIST DEL MODELO DE REGISTRO COMENTADO EN ESTE ARTÍCULO CON UNA NOTICIA REAL.

Bibliografía

- ABADAL FALGUERAS, Ernest. «Diseño y creación de una base de datos en un medio de comunicación». En: FUENTES, M. Eulàlia (ed.). *Manual de Documentación periodística*. Madrid: Síntesis, 1995, p. 19.
- AENOR. *Norma UNE 50-106-90. Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: AENOR, 1990, p. 47.
- BAIGET, T. *Análisis de sistemas de información*. Barcelona: Institut Català de Tecnologia, 1986, p. 64 (documento reprografiado).
- CHECKLAND, P. B. *Systems thinking, systems practice* Chichester: Wiley, 1981.
- CHECKLAND, P. B.; SCHOLLES, J. *Soft systems methodology in action*. Chichester: Wiley, 1990.
- CHEN, P.P-S. «The entity-relationship model: towards a unified view of data». *ACM transactions on databases systems*, v. 1, n. 1, 1976, p. 9-36.
- CODINA, L. *Sistemes d'informació documental: concepció, anàlisi i disseny de sistemes de gestió documental amb microordinadors*. Barcelona: Pòrtic, 1994, p. 224.
- CODINA, L. «Metodología de análisis de sistemas de información y diseño de bases de datos documentales: aspectos lógicos y funcionales». En: Baró, J.; Cid, P. (eds.). *Anuari SO-CADI de Documentació i Informació 1998*. Barcelona: SOCADI, 1998, p. 195-210.
- CURRAS, E. *La información en sus nuevos aspectos*. Madrid: Paraninfo, 1988, p. 307.

- FUENTES, M. Eulàlia; CONESA, Alícia. *La documentació periodística: Catalunya, Espanya i altres experiències europees*. Barcelona: Generalitat de Catalunya. Centre d'Investigació de la Comunicació, 1994, p. 183.
- FUENTES, M. Eulàlia (ed.). *Manual de documentació periodística*. Madrid: Editorial Síntesis, 1995, p. 230. FUENTES, M. Eulàlia. *Documentación y periodismo*. Pamplona: Eunsa, 1997, p. 132. JACKSON, G. A. *Introducción al diseño de bases de datos relacionales*. Madrid: Anaya, 1990, p. 203.
- LEWIS, P. *Information systems development*. London: Pitman, 1994, p. 260.
- PALMA, María del Valle. «Bases de datos y servicios de información disponibles en Internet». Documentación digital (CD-ROM). Barcelona: Universitat Pompeu Fabra, 1999 (versión en soporte óptico del artículo en línea en la web del curso telemático de Documentación Digital: (<http://docdigital.upf.es>)).
- UNDERWOOD, P. G. *Soft Systems Analysis and the management of libraries, information services and resource centres*. London: Library Association, 1996, p. 198.
- VAN SLYPE, Georges. *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1991, p. 198.
- WALKER, D.W. *Sistemas de información basados en ordenador*. Barcelona: Marcombo, 1991.
- YOURDON, E. *Análisis estructurado moderno*. México: Prentice-Hall Hispanoamericana, 1993, p. 735.