

BASES DE DATOS DOCUMENTALES: ESTRUCTURA Y PRINCIPIOS DE USO

Luis Rodríguez Yunta
CINDOC-CSIC

INDICE

- 1. Introducción.**
- 2. Definición de base de datos documental bibliográfica**
- 3. Tipología de las bases de datos documentales**
 - 3.1. Tipología según el organismo productor**
 - 3.2. Tipología según el modo de acceso**
 - 3.3. Tipología según su cobertura temática**
 - 3.4. Tipología según su cobertura documental**
 - 3.5. Tipología según el modelo de tratamiento documental**
- 4. Estructura interna de una base de datos**
- 5. La recuperación de información en bases de datos documentales**
 - 5.1. Opciones de búsqueda**
 - 5.2. Herramientas de búsqueda**
 - 5.3. Campos de búsqueda y control del vocabulario**
- 6. Otras operaciones de la consulta de una base de datos: visualización, impresión, exportación,...**
- 7. ¿Cómo realizar búsquedas en bases de datos?**
- 8. ¿Qué debe conocerse sobre una base de datos antes de utilizarla?**
- 9. Bases de datos y acceso al documento original**
- 10. Bibliografía**

1. INTRODUCCIÓN

El objetivo de este capítulo es presentar las características generales de las bases de datos bibliográficas y ofrecer algunas informaciones básicas para un uso más eficaz de las mismas. Bajo la denominación genérica de bases de datos documentales se engloban diferentes modelos que pueden diferenciarse tanto en el contenido (tipos de documentos que describen) como en la estructura de las mismas. Los primeros apartados de este capítulo intentan establecer una tipología.

Sin embargo, existen otras diferencias entre bases de datos marcadas por los distintos programas informáticos con los que se realiza el acceso a su interrogación. La segunda parte de este capítulo sistematiza las funciones fundamentales que pueden encontrarse en dichos programas.

2. DEFINICION DE BASE DE DATOS DOCUMENTAL BIBLIOGRÁFICA

Una base de datos es un conjunto de información estructurada en *registros* y almacenada en un soporte electrónico legible desde un ordenador. Cada registro constituye una unidad autónoma de información que puede estar a su vez estructurada en diferentes *campos* o tipos de datos que se recogen en dicha base de datos. Por ejemplo, en un directorio de miembros de una asociación, un registro será la ficha completa de cada uno de los socios. En cada registro se recogerán determinados datos, como el nombre, la profesión, la dirección o el teléfono, cada uno de los cuáles constituye un campo.

Hay tres grandes modelos de bases de datos según el contenido de sus registros:

- **Bases de datos con información factual:** recogen informaciones muy concretas y actuales, generalmente de carácter numérico: datos estadísticos, series históricas, resultados de encuestas, convocatorias de becas o subvenciones, convocatorias de eventos, ofertas de empleo,...
- **Directorios:** recogen datos sobre personas o instituciones especializadas en una actividad o materia concreta. Hay directorios de profesionales, de investigadores, de centros de investigación, de bibliotecas, de revistas científicas, de empresas, de editoriales,...
- **Bases de datos documentales.** En este último grupo, cada registro se corresponde con un documento, sea éste de cualquier tipo: una publicación impresa, un documento audiovisual, gráfico o sonoro, un documento de archivo, un documento electrónico,...

Una base de datos se crea y mantiene de forma continuada con el objetivo de resolver necesidades de información concretas de un colectivo, una empresa o el conjunto de la sociedad. Estos recursos electrónicos pueden ser consultables directamente en formato electrónico o ser utilizados para elaborar productos impresos: bibliografías, directorios, informes.

Los registros de las bases de datos documentales pueden incluir o no el contenido completo de los documentos que describen, según lo cuál se distinguen tres categorías:

- **Bases de datos de texto completo:** son aquellas que estén constituidas por los propios documentos en formato electrónico, por un volcado completo de su texto. Pueden incorporar además campos en los que se contiene la información fundamental para facilitar su descripción y recuperación. En estos sistemas la operación de búsqueda (que puede abarcar la totalidad del texto) y la consulta del documento se producen sin salir del propio sistema de información.
- **Archivos electrónicos de imágenes:** están constituidos por referencias que permiten un enlace directo con la imagen del documento original, sea éste un documento iconográfico (fotografías, imágenes de televisión,...) o un documento impreso digitalizado en formato de imagen. En estas bases de datos normalmente la búsqueda está limitada a los campos de la referencia bibliográfica y no se pueden localizar otros términos presentes en el texto completo del documento original.
- **Bases de datos referenciales:** sus registros no contienen el texto original sino tan sólo la información fundamental para describir y permitir la localización de documentos impresos, sonoros, iconográficos, audiovisuales o electrónicos. En estos sistemas de información sólo se puede obtener referencias sobre documentos que habrá que localizar posteriormente en otro servicio (archivo, biblioteca, fototeca, fonoteca,...) o solicitar a un servicio de suministro de documentos. Sin embargo, una base de datos referencial puede incluir campos que faciliten la localización del documento (bibliotecas, signaturas, direcciones en Internet,...) o incluso enlaces directos para obtener directamente el original a través de otro programa (tratamiento de texto, navegador de Internet,...).

Las *bases de datos bibliográficas* son generalmente bases de datos documentales referenciales, cuyos registros contienen referencias de documentos impresos o de texto.

3. TIPOLOGIA DE LAS BASES DE DATOS DOCUMENTALES

Se pueden establecer diferentes tipologías para caracterizar los diferentes modelos de bases de datos documentales. Aquí se han establecido cinco clasificaciones alternativas:

- Según el organismo productor.
- Según el modo de acceso.
- Según su cobertura temática y el destinatario.
- Según su cobertura documental.
- Según el modelo de tratamiento documental.

Las tres primeras son aplicables a cualquier tipo de base de datos, mientras que las otras dos son específicas de las bases de datos documentales.

3.1. TIPOLOGÍA DE LAS BASES DE DATOS SEGUN EL ORGANISMO PRODUCTOR

3.1.1. Bases de datos de organismos públicos y de la administración: Las bibliotecas y centros de documentación de los ministerios, instituciones públicas, universidades y organismos públicos de investigación elaboran gran cantidad de recursos de información. Estos sistemas pueden ser:

- *Bases de datos de acceso público*, sean gratuitas o no.
- *Bases de datos de uso interno*, con información de acceso restringido.

3.1.2. Bases de datos de instituciones sin ánimo de lucro: Fundaciones, asociaciones, sindicatos y organizaciones no gubernamentales elaboran frecuentemente sus propios sistemas de información especializados.

3.1.3. Bases de datos de entidades privadas o comerciales: Los centros de documentación, bibliotecas y archivos de las empresas pueden elaborar distintos tipos de sistemas de información:

- *Bases de datos de uso interno* para facilitar la circulación de información dentro de la empresa
- Bases de datos de uso interno que ocasionalmente ofrecen servicio hacia el exterior (usuarios particulares u otras instituciones).
- *Bases de datos comerciales*, diseñadas específicamente para ser utilizadas por usuarios externos.

3.1.4. Bases de datos realizadas por cooperación en red: Se trata de sistemas de información cuya elaboración es compartida por diversas instituciones. Algunas

bases de datos internacionales se elaboran a través de este sistema de trabajo, con diversos centros nacionales responsables de la información perteneciente a cada país.

3.2. TIPOLOGÍA DE LAS BASES DE DATOS SEGÚN EL MODO DE ACCESO

3.2.1. Bases de datos de acceso local. Para consultarlas es necesario acudir al organismo productor, a su biblioteca o centro de documentación. Pueden ser consultables en monopuesto o en varios puntos de una red local.

3.2.2. Bases de datos en CD-ROM. Pueden adquirirse por compra o suscripción bien directamente por un particular o por una biblioteca o centro de documentación que permita su consulta a sus usuarios. En algunas instituciones se instalan diferentes CD-ROM en una red local para permitir su consulta desde cualquier ordenador conectado a la misma.

3.2.3. Bases de datos en línea. Pueden consultarse desde cualquier ordenador conectado a través de un modem a la línea telefónica o a través de un servidor a Internet. La consulta puede ser libre (gratuita) o exigir la solicitud previa de un clave personal de entrada (denominada comúnmente con el término inglés password). Para obtener un password puede exigirse la firma de un contrato.

Hay diferentes tipos de acceso en línea:

- **acceso vía telnet o mediante conexión telefónica:** el usuario realiza una conexión estable al host (gran ordenador) en donde se halla la base de datos, bien a través de la línea telefónica ordinaria, de redes especiales de transmisión de datos o a través de Internet. La interfaz de usuario instalada en dicho ordenador remoto determinará si la interrogación debe realizarse por menús o por comandos o expresiones de un lenguaje determinado. Cuando un usuario entra en una base de datos vía telnet establece una sesión de trabajo interactiva con el programa que gestiona la base de datos, que le permite aplicar todas las posibilidades de interrogación que tenga el sistema: selección, combinación y visualización o impresión de resultados. En cualquier momento podrá visualizar todas las búsquedas realizadas hasta ese instante y establecer combinaciones entre ellas.
- **acceso vía web:** conexión a través de un formulario existente en una página web de Internet, diseñado para lanzar preguntas a una base de datos. Este medio de acceso se realiza, en ocasiones, con la restricción de tener que definir la estrategia de búsqueda en un único paso, sin que sea posible combinar búsquedas realizadas anteriormente.

Una misma base de datos puede tener acceso local y además una edición en CD-ROM y un sistema de acceso en línea. Sin embargo, puede haber diferencias en el contenido presente en cada uno de estos formatos o en el grado de actualización de la información. Por ejemplo, el productor de una base de datos puede ofrecer la conexión en línea a la base de datos completa con actualización diaria y, en cambio, editar un CD-ROM que tan sólo contenga los últimos cinco años de información y se actualice semestralmente.

3.3. TIPOLOGÍA DE LAS BASES DE DATOS SEGÚN SU COBERTURA TEMÁTICA

3.3.1. Bases de datos científico-tecnológicas: contienen información destinada a los investigadores de cualquier ámbito científico o técnico. A su vez, este grupo puede dividirse en:

- *Bases de datos multidisciplinares*: abarcan varias disciplinas científicas o técnicas.
- *Bases de datos especializadas*: recopilan y analizan documentos pertinentes para una disciplina o subdisciplina concreta: investigación biomédica, farmacéutica, química, agroalimentaria, social, humanística, etc.

3.3.2. Bases de datos económico-empresariales: contienen información de interés para empresas, entidades financieras,...

3.3.3. Bases de datos de medios de comunicación: contienen información de interés para los profesionales de medios de comunicación de masas: prensa, radio, televisión,...

3.3.4. Bases de datos del ámbito político-administrativo y jurídico: contienen información de interés para los organismos de la administración y los profesionales del Derecho: legislación, jurisprudencia,...

3.3.5. Bases de datos del ámbito sanitario: además de las propias del primer grupo especializadas en ciencias de la salud, existen otros sistemas con información de interés sanitario: historiales médicos, archivos hospitalarios,...

3.3.6. Bases de datos para el gran público: contienen información destinada a cubrir necesidades de información general, de interés para un gran número de usuarios.

3.4. TIPOLOGÍA DE LAS BASES DE DATOS SEGÚN SU COBERTURA DOCUMENTAL

3.4.1. Bases de datos centradas en un único tipo de documento. Su objetivo es recopilar y permitir la localización de un tipo documental muy concreto. Hay bases especializadas en patentes, tesis doctorales, informes, artículos de revista,...

Como ejemplos de este modelo pueden citarse:

- La base de datos *NTIS* (National Technical Information Service), especializada en documentos técnicos oficiales de Estados Unidos), que puede consultarse en <http://www.ntis.gov> o en <http://library.dialog.com/bluesheets/html/bl0006.html>
- Las bases de datos especializadas en tesis doctorales, *Dissertation Abstracts* (de carácter internacional, cuya descripción puede consultarse en <http://library.dialog.com/bluesheets/html/bl0035.html>) o *Teseo* (tesis españolas, consultable en <http://www.mcu.es/TESEO/index.html>).

3.4.2. Bases de datos que reúnen varios tipos de documentos. Su objetivo es dar información sobre una disciplina, incorporando para ello diferentes tipologías documentales.

Como ejemplo de este modelo puede citarse la base de datos *ERIC*, especializada en una temática concreta (Educación) sobre la cual recoge referencias bibliográficas de artículos, monografías, informes, tesis... Puede consultarse en <http://askeric.org/Eric/> o en <http://library.dialog.com/bluesheets/html/bl0001.html>

3.5. TIPOLOGÍA DE LAS BASES DE DATOS SEGÚN EL MODELO DE TRATAMIENTO DOCUMENTAL

3.5.1. Bases de datos de sumarios o sin análisis de contenido: se componen de referencias bibliográficas sencillas, en las cuáles el productor se limita a grabar los datos de la propia fuente y no realiza ningún análisis del contenido. Incorporan solamente los datos descriptivos fundamentales para localizar el documento, presentes en el sumario: autor, título y datos de la fuente. En ocasiones pueden incorporar el resumen editado con el propio documento original, generalmente elaborado por el propio autor. En la mayoría de estas bases de datos la búsqueda por materias sólo puede realizarse a través de las palabras del título del artículo. En muchas áreas temáticas los títulos pueden resultar poco significativos. Los resultados obtenidos en una búsqueda tienen cierta pertinencia, pero la escasez de puntos de acceso para la búsqueda por materias no garantiza la exhaustividad.

Como ejemplos de este modelo pueden citarse las bases:

- *Currents Contents* del Institute for Scientific Information
<http://www.isinet.com/isi/products/index.html>
<http://library.dialog.com/bluesheets/html/bl0440.html>
- *Uncover* de la Universidad de Tennessee
<http://www.lib.utk.edu:90/uncover>
<http://library.dialog.com/bluesheets/html/bl0420.html>
- *Article@INIST*
<http://services.inist.fr/public/fre/conslt.htm>
- *Compludoc*
<http://www.ucm.es/BUCM/complu/frame.htm>

Por la similitud de contenido, podrían incluirse dentro de este modelo los servicios de suscripción a revistas electrónicas establecidos por empresas de distribución de publicaciones periódicas, como por ejemplo el servicio *SWETSCAN* (<http://www.swetsblackwell.com/custss-about.htm>).

3.5.2. Catálogos de bibliotecas: Son bases de datos que responden estrictamente a los fondos contenidos en una biblioteca o en una red de bibliotecas (catálogos colectivos). Tienen una alta homogeneidad, gracias a la aplicación de normas internacionales de Catalogación. En estos sistemas se puede garantizar la búsqueda por ejemplar conocido. Por el contrario, la búsqueda por materias es en ocasiones dificultosa y con algunas variaciones entre diferentes catálogos. Puede realizarse a través de la clasificación o de los encabezamientos de materias.

Como ejemplos de este modelo puede destacarse en España el Catálogo de la Biblioteca Nacional *Ariadna* (<http://www.bne.es/esp/cat.htm>) y los catálogos

colectivos, de la *red de bibliotecas del CSIC* (<http://sauco.csic.es:4505/ALEPH>) o de la red de bibliotecas universitarias *REBIUN* (<http://www.uma.es/rebiun/>).

Ejemplo de registro del catálogo colectivo de la red de bibliotecas del CSIC

Autor: Luque Sánchez, Ángela de
Ontoria Peña, Antonio
Título: Personalismo social
Subtítulo: hacia un cambio en la metodología docente
Responsabilidad: Ángela de Luque Sánchez, Antonio Ontoria Peña
Publicación: Córdoba : Universidad. Servicio de Publicaciones , 2000
Colación: 254 p. ; 24 cm.
ISBN: 84-7801-520-5
Materias: Pedagogía - Aspectos sociales
Psicología de la educación
Idioma: spa
País: esp
Fecha: 2000
Notas: Bibliog. : p. 241-254
Número CIRBIC : 0868461
Localización : M-EstSoc Sig.: 2/218

3.5.3. Bases de datos con análisis documental más completo: Son sistemas de información que incorporan un mayor número de puntos de acceso para facilitar la localización por materias. Cada registro bibliográfico incluye bien un resumen del contenido del documento original y/o un conjunto de conceptos o términos representativos de los temas tratados en el mismo. Se denomina indización al proceso de asignación a un documento de estos términos que describen su contenido y que constituyen entradas en un índice de materias (palabras clave o descriptores).

Estas bases de datos pueden ser independientes de un fondo documental concreto, ya que su objetivo no es describir una colección sino proporcionar el mayor número de referencias potencialmente relevantes para cubrir unas necesidades de información.

Dentro de este modelo se incluyen numerosas bases de datos especializadas, de gran importancia y prestigio internacional como:

- *Medline* (<http://www.ncbi.nlm.nih.gov/PubMed/>)
- *ERIC* (<http://askeric.org/Eric/>)
- *Chemical Abstracts* (<http://www.cas.org/casdb.html>)
- *INSPEC* (<http://www.iee.org.uk/publish/inspec/>)
- *Sociological Abstracts* (<http://www.csa.com/csa/factsheets/socioabs.shtml>)
- *Francis* (<http://www.inist.fr/francis/francis.htm>)
- *Pascal* (<http://www.inist.fr/pascal/pascal.htm>).

Dentro de las bases de datos españolas puede destacarse por su mayor antigüedad las bases de datos del CSIC (*Índice Médico Español*, *ICYT* e *ISOC*, consultables en <http://www.cindoc.csic.es/prod/dbsconx.html>) y *Teseo* del Consejo de Universidades (<http://www.mcu.es/TESEO/index.html>).

Dentro de este grupo de bases de datos no hay un modelo único, sino una gran variedad de posibilidades, según los criterios de análisis documental adoptados por el productor del sistema de información:

- *Bases de datos con clasificación y resúmenes.* En estos sistemas de información la búsqueda por materias se realiza generalmente a través de las palabras del título y del resumen. Tienen una alta cantidad de puntos de acceso, pero en ocasiones la búsqueda puede ser muy difícil.
- *Bases de datos con clasificación e indización* por descriptores o palabras clave (para la definición de estos conceptos véase el apartado 5.3.). Se caracterizan por permitir una mayor pertinencia de la recuperación. Sin embargo para sacarles un óptimo rendimiento es conveniente que el usuario conozca el listado de términos utilizados para la descripción del contenido de los documentos.
- *Bases de datos con clasificación, indización y resúmenes:* teóricamente reúnen las ventajas de los dos grupos anteriores. Dentro de este modelo conviven bases de datos con diferentes criterios de indización (genérica o específica) y diferentes modelos de resúmenes (informativos o indicativos).

3.5.4. Índices de citas. Son sistemas de información en los que, además de extraer los datos de descripción de los documentos, se vacían sistemáticamente las referencias bibliográficas citadas en los artículos de las revistas científicas. Su mayor utilidad radica en servir de apoyo para la elaboración de estudios bibliométricos y de productividad científica: autores, instituciones y publicaciones científicas más citadas; pervivencia de la literatura científica,...

Ejemplos de este modelo son las bases de datos producidas por el Institute for Scientific Information: *SciSearch*, *Social SciSearch* y *Arts & Humanities Search* (<http://www.isinet.com/isi/products/index.html>).

4. ESTRUCTURA INTERNA DE UNA BASE DE DATOS

Las bases de datos se componen de diferentes *registros* con su correspondiente número de identificación. En una base de datos bibliográfica cada registro se corresponde con una *referencia de un documento original*, que puede ser una publicación independiente (monografía, tesis, informe,...) o una parte con autonomía de contenido dentro de otra publicación (artículo incluido en una revista, una serie o una compilación, ponencia o comunicación en las actas de una reunión científica,...)

La información contenida en cada registro bibliográfico se estructura en diferentes *campos* para facilitar su control sistemático y su recuperación individualizada. Algunos campos reflejarán la descripción formal del documento y otros van destinados a reflejar su contenido temático. Según el modelo de base de datos y el tipo de documentos vaciado en la misma se establece una estructura de campos particular.

Por ejemplo, en una base de datos donde se incorporan referencias de artículos de revista se encuentran campos como los siguientes:

- *Autor o autores* del documento original.
- *Título del artículo.*
- *Título de la revista* en la que aparece.
- *Año de publicación.*

- *Datos fuente*: volumen, número, páginas.
- *Idioma* en el que está escrito el documento.
- *Clasificación* que describe el tema central tratado en el documento.
- *Palabras clave* que caracterizan cada uno de los temas tratados con suficiente relevancia en el artículo (véase el punto 5.3.).
- *Resumen*: texto breve que sintetiza el contenido del documento.

En algunas de estas bases de datos el campo de *palabras clave* se sustituye por un conjunto de diferentes campos que informan sobre distintos aspectos del contenido del documento. Con ello se pretende mejorar la pertinencia de la recuperación. Según la cobertura temática se pueden encontrar campos como los siguientes:

- *Descriptores* (véase el punto 5.3.): términos que describen conceptos teóricos tratados de forma especialmente relevante en el artículo.
- *Identificadores*: nombres propios de personas, obras o instituciones tratadas en el artículo.
- *Topónimos* o identificadores geográficos: nombres propios de lugar.
- *Sustancias químicas* (en bases de datos de esta temática u otras afines).
- *Magnitudes numéricas* (en bases de datos de ingeniería)
- *Coordenadas geográficas* (en bases de datos sobre geología o geografía)
- *Legislación y/o Jurisprudencia* (en bases de datos de Derecho)
- *Período histórico*: fechas, décadas o siglos a los que se refiere la temática tratada en el documento (en bases de datos sobre Historia).

También pueden existir otros campos para la descripción formal:

- *Afiliación institucional* de los autores.
- *ISSN* de la revista.
- *País de edición* de la revista.
- *Biblioteca o fondo* donde se localiza el documento.

A continuación se muestran varios ejemplos de una misma base de datos que incluyen una amplia diversidad de campos para diferentes tipos de documentos, extraídos de la base de datos *ISOC* del CINDOC (<http://www.cindoc.csic.es/prod/dbsconx.html>):

Base de datos ISOC: Registro de artículo de revista:

Núm. registro: 411476

Autores: Camacho Mene, María Teresa

Título del artículo: La inmigración en la comunidad de Murcia: Una experiencia de acogida

Lugar de trabajo: Cáritas Murcia, España

ISSN de la revista: 0417-8106

Título Revista: DOCUMENTACION SOCIAL

Datos fuente: 2000, (121): 295-307

Tipo documento: AR

Lengua: ES

Clasificación: 751407;751602

Localización: ISOC

Descriptores: Inmigrantes; Política social; Acción social; Integración social; Centros de acogida; Acogimiento familiar; Necesidades básicas; Formación profesional; Protección de menores

Identificadores: Cáritas

Topónimos: Región de Murcia

Resumen: El fenómeno de la inmigración responde a causas estructurales, muchas de las cuales están motivadas por la injusticia, que hay que buscar tanto en la situación de los países de origen como de los de acogida. La inmigración no es un fenómeno nuevo en Europa, ni en España y ni siquiera en Murcia. La novedad está en que la Región ha pasado de ser emisora a receptora de emigrantes, y además en mayor número que la del resto de España, lo que supondría unos 45.000 inmigrantes..... (A.)

(resumen cortado para elaborar este ejemplo)

Base de datos ISOC: Registro de comunicación o ponencia:

Núm registro: 408788

Autores: Ordóñez Vergara, María José

Título del artículo: Información y desarrollo industrial

Lugar de trabajo: Univ. Coruña, España

ISBN/ISSN: 84-7585-919-4

Datos fuente: 2000: 487-497, 6 ref

Tipo documento: AC

Modo del documento: PO

Lengua: ES

Clasificación: 200105

Localización: ISOC

Congreso: Jornadas Españolas de Documentación. VII. 2000. Bilbao

Editor: Bilbao: Universidad del País Vasco, 2000

Notas: Gráficos, anexo con revistas solicitadas más de 15 veces

Descriptores: Demanda de información; Farmacia; Estudios de usuarios; Industria farmacéutica; Revistas científicas

Identificadores: CINDOC

Topónimos: Comunidad de Madrid

Resumen: En este artículo se exponen los principales resultados de un estudio pormenorizado a partir de los datos obtenidos de una investigación, basada en métodos indirectos de recogida de datos, donde se corroboran las conclusiones anteriormente obtenidas sobre las empresas farmacéuticas ubicadas en la Comunidad Autónoma de Madrid (CM).

(resumen cortado para elaborar este ejemplo)

Base de datos ISOC: Registro de informe o documento de trabajo:

Núm. Registro: 379077

Autores: Maudos, Joaquín; Pastor, José Manuel; Pérez García, Francisco; Quesada Ibáñez, Javier

Título del Artículo: Eficiencia en costes y beneficios en la banca europea

Título en Inglés: Cost and Profit Efficiency in European Banks

Lugar de Trabajo: Univ. Valencia, España

Datos Fuente: 1999, (Wp-Ec 99-12): 36 P., 23 ref.

Tipo Documento: IN

Lengua: EN

Clasificación: 303120

Localización: ISOC

Editor: Valencia: Instituto Valenciano de Investigaciones Económicas, 1999

Serie: Documento de Trabajo

Signatura: DT-ECO-9912

Notas: Resumen (ES, EN), cuadros, gráficos

Descriptores: Sistema bancario; Análisis coste-beneficio; Eficiencia económica; Modelo Econométrico

Topónimos: Países de la Unión Europea

Periodo histórico: 1993: 1996

Resumen: El análisis de ratios contables como indicadores de competitividad ha sido completado y ocasionalmente revisado por la creciente utilización de indicadores más sofisticados de eficiencia. En los últimos años, más de un centenar de trabajos han analizado la eficiencia de las instituciones financieras, centrándose mayoritariamente en la vertiente de los costes... (A.) (resumen cortado para elaborar este ejemplo)

Para facilitar la rapidez en la recuperación de información, los sistemas de gestión documental permiten la elaboración de *diccionarios o índices alfabéticos*. Algunos programas trabajan con un índice único formado por palabras procedentes de todos los campos de cada registro. Otros manejan diferentes índices para cada campo. El productor de la base de datos deberá configurar cómo se construyen estos diccionarios según las necesidades de información que se desea cubrir y según las posibilidades que ofrece el programa o software utilizado. La información de algunos campos puede no ser interrogable si el diseño del procedimiento de elaboración del diccionario no lo incluye. Los índices pueden estar formados por palabras (unitérminos) o por frases (texto completo de un campo o de cada ocurrencia de un campo: cada autor, cada descriptor,...)

5. LA RECUPERACIÓN DE INFORMACIÓN EN BASES DE DATOS DOCUMENTALES

Un mismo sistema de información puede estar accesible desde distintos soportes: acceso vía telnet, acceso desde una página web y edición en CD-ROM. Cada uno de estos formatos puede tener su propio sistema de interrogación y una forma de presentación o interfaz de usuario diferente. En el acceso tradicional en línea (telnet) la consulta puede realizarse a través de un *formato de menús* en los que el usuario va seleccionando sucesivas opciones o mediante complejos *lenguajes de recuperación* formados por comandos y operadores. Por el contrario, en la consulta en páginas web y en la mayoría de los CD-ROM el usuario debe completar *formularios* diseñados en entorno Windows.

Algunos sistemas permiten que la consulta pueda realizarse en varias bases de datos simultáneamente. Esta opción es posible a través de algunos distribuidores de bases de datos y en determinadas ediciones en CD-ROM.

5.1. OPCIONES DE BUSQUEDA


En general todos los sistemas de recuperación de información (se presenten en forma de lenguaje de comandos, por menús o en formularios) permiten realizar diferentes modalidades de búsqueda:

5.1.1. Búsqueda directa: se teclea directamente una o varias palabras en el espacio reservado para ello por el sistema de interrogación en la base de datos. Pueden distinguirse a su vez dos tipos dentro de esta modalidad:

- **Interrogación en texto libre:** el usuario realiza su consulta sin tener en cuenta la estructura de campos de los registros de la base de datos. Generalmente, el sistema busca el término o los términos solicitados en la totalidad de los campos de la base. En algunas bases de datos esta búsqueda se realiza solamente en determinados campos, aquellos que el administrador de la misma ha establecido.


Ejemplo extraído de la pantalla de búsqueda de la base ICYT:

Búsqueda en campos básicos: (Título, descriptores, identificadores, topónimos, resumen)

alguna palabra 
microbiología

- **Interrogación en campos individuales:** el usuario realiza su consulta directamente sobre el campo o campos que ha seleccionado previamente: autor, revista, descriptores, etc...

Ejemplo extraído de la pantalla de búsqueda de la base ICYT:

Búsqueda por campos: Autores 
la frase 
Ruiz Hernández

5.1.2. Búsqueda a través de índices: en vez de teclear un término, el usuario visualiza un diccionario o índice alfabético de las entradas de todos los campos o de uno concreto (índice de autores, descriptores,...) y selecciona las más adecuadas a su búsqueda. El proceso puede parecer más lento pero es más seguro, ya que la búsqueda se adapta mejor al contenido real de los registros de una base de datos.

En esta posibilidad hay que tener en cuenta si el sistema trabaja con uno o varios índices y cómo se realizan los mismos:

- *Índices por palabras:* incluyen solamente unitérminos (conceptos expresados en una sola palabra). Se elaboran, por ejemplo, con todas las palabras extraídas del título o del resumen de los documentos. La mayoría de los productores de bases de datos establecen una lista de "palabras vacías" (preposiciones, artículos,...) para reducir el tamaño de estos índices. Los términos presentes en esta lista de palabras vacías no serán consultables.

Ejemplo extraído de la consulta en la base de datos ICYT:

Mostrar índice

del campo: Título - Índice

comenzando desde:

microbi

Nuevo comienzo

Documentos	Términos
1	<input type="checkbox"/> MICROBENTOS
1	<input type="checkbox"/> MICROBIALES
86	<input type="checkbox"/> MICROBIANA
22	<input type="checkbox"/> MICROBIANAS
32	<input type="checkbox"/> MICROBIANO
32	<input type="checkbox"/> MICROBIANOS
3	<input type="checkbox"/> MICROBICIDA

- *Índices por frase:* incluyen entradas compuestas por varias palabras. Por ejemplo: apellidos y nombre del autor, descriptores o encabezamientos de materias,...

Ejemplo extraído de la consulta en la base de datos ICYT:

Mostrar índice del campo: Autores (frase) - Índice
comenzando desde:
alvarez Nuevo comienzo

Documentos	Términos
1	<input type="checkbox"/> ALVARES, F.
2	<input type="checkbox"/> ALVAREZ ACEVEDO Y ALONSO, M.
2	<input type="checkbox"/> ALVAREZ AGUAYO, F.
2	<input type="checkbox"/> ALVAREZ AGUILAR, J.
3	<input type="checkbox"/> ALVAREZ AHEDO, J.
3	<input type="checkbox"/> ALVAREZ ALBA, J.
3	<input type="checkbox"/> ALVAREZ ALCON, M.

5.1.3. Búsqueda jerarquizada: la interrogación se realiza a través de una estructura jerárquica. Es decir, a partir de un concepto genérico se pueden localizar, no sólo los registros en los que aparece dicho término, sino además todos aquellos en los que figure algún concepto más específico de su campo semántico. Esta opción de búsqueda sólo será posible en aquellos sistemas de información que cuenten con campos codificados o que utilicen lenguajes estructurados jerárquicamente en la clasificación o en los descriptores.

5.1.4. Búsqueda a través de códigos: en determinados campos de algunas bases de datos la interrogación no se realiza a través de un texto sino a través de códigos numéricos o alfanuméricos. Pueden ser campos codificados la clasificación, el idioma, la tipología documental,...

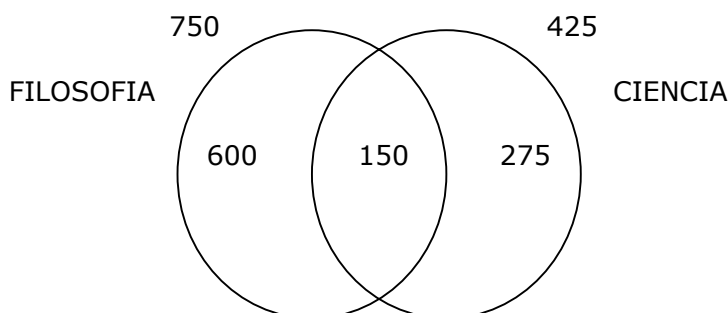
5.2. HERRAMIENTAS DE BÚSQUEDA

Para realizar una búsqueda compleja en una base de datos es preciso utilizar un número elevado de conceptos. Hay que precisar los diferentes aspectos temáticos implicados y además contar con la posibilidad de que cada uno de ellos pueda estar expresado de diferentes maneras (sinónimos, temas relacionados,...). Los sistemas de recuperación de información cuentan con diferentes herramientas para construir una estrategia y relacionar de forma clara los diferentes términos utilizados en una búsqueda.

5.2.1. Operadores lógicos o booleanos. Permiten la combinación de conceptos en una misma búsqueda. Los sistemas de recuperación utilizan las operaciones propias de la teoría de conjuntos o álgebra de Boole:

- *Suma o unión* (A or B, A o B): selecciona todos los registros en los que aparece tanto A como B como ambos a la vez.
- *Resta o negación* (A and not B, A no B): selecciona sólo los registros en los que se encuentre el término A sin estar acompañado del término B. Es una operación que debe realizarse con la máxima precaución para no dejar de seleccionar documentos que sí pueden ser de interés para la consulta deseada.
- *Intersección* (A and B, A y B): selecciona solamente los registros en los que aparece simultáneamente los conceptos A y B.
- Algunos sistemas permiten también la operación contraria: A xor B: registros donde está A o B pero no A y B simultáneamente.

Por ejemplo, en una base de datos existen 750 registros en los que se encuentra la palabra 'Filosofía' y 425 en los que aparece el término 'Ciencia'. Según se combinen ambos conceptos en una misma búsqueda se obtienen estos resultados:



- 'Filosofía' and 'Ciencia' = 150 registros en los que figuran ambos términos. Este valor corresponde a la intersección entre los conjuntos anteriores.
- 'Filosofía' or 'Ciencia' = 1025 registros en los que figura alguno de estos conceptos o bien ambos. Hay que señalar que el resultado de una suma de dos conjuntos puede no coincidir con la suma aritmética de los valores de cada uno ($750 + 425$) ya que aquellos registros que forman parte de la intersección de ambos conjuntos se contabilizan una única vez.

- 'Filosofía' and not 'Ciencia' = 600 registros en los que figura el primer concepto sin estar acompañado del segundo.
- 'Ciencia' and not 'Filosofía' = 275 registros en los que figura el primer concepto sin estar acompañado del segundo.

La combinación de diferentes operadores en una misma búsqueda puede exigir la necesidad de utilizar paréntesis para marcar claramente el orden del proceso de búsqueda. Por ejemplo, si se desea buscar documentos sobre artes plásticas, escultura o pintura en Italia no sería correcta la siguiente fórmula:

Italia AND Artes plásticas OR Escultura OR Pintura

En este caso la intersección con el conjunto "Italia" afectaría solamente a "artes plásticas". Por el contrario, se debería utilizar un paréntesis para conseguir que la intersección se realice con los tres términos siguientes y no sólo con el primero:

Italia AND (Artes plásticas OR Escultura OR Pintura)

Generalmente estas combinaciones booleanas pueden realizarse también a posteriori entre conjuntos establecidos en búsquedas ya realizadas. Por el contrario, algunas sistemas de recuperación vía web no permiten combinar búsquedas anteriores.

Debe tenerse en cuenta que en algunos sistemas de interrogación se configura uno de estos operadores booleanos (and, or o not) para que se ejecute por defecto cuando se teclean dos términos. Obviamente el resultado de la consulta variará si se interpreta que debe buscar registros en los que aparezcan ambos a la vez (intersección) o bien cualquiera de los dos (suma booleana).

5.2.2. Operadores sintácticos, de proximidad o de adyacencia.

Frecuentemente el operador de intersección (and) no es suficiente para precisar una búsqueda. Por ejemplo, si se desea localizar registros que traten sobre el escritor 'Juan Madrid' la expresión 'Juan and Madrid' dará como resultado todas las referencias en las cuáles aparecen ambos términos pero no necesariamente relativos a este autor. Para mejorar la pertinencia de la búsqueda, los sistemas de recuperación de información incorporan frecuentemente la posibilidad de combinar conceptos relacionados entre sí por criterios sintácticos: presencia en la misma frase, el mismo párrafo o el mismo campo, aparición en determinado orden o con una separación mínima de palabras entre ellas, etc.

Por ejemplo, si se selecciona el operador de *búsqueda por frase* y se teclea la expresión 'Juan Madrid' el sistema seleccionará sólo aquellos documentos en los que estas dos palabras aparezcan juntas y en este orden. Este tipo de operadores son especialmente necesarios cuando se interroga sobre campos con textos extensos (resumen o incluso el texto del documento completo).

5.2.3. Truncado: permite la interrogación conjunta de conceptos que comparten una misma raíz. Mediante un signo de truncado (*, ? o \$, según el programa) los sistemas de recuperación permiten realizar una búsqueda conjunta de todas las palabras de un índice que comparten un mismo comienzo o raíz.

Por ejemplo, con la expresión: 'Econom*' se recuperan todos los documentos en los que figure cualquier término de esta raíz: economía, económico, económica,...

La mayoría de los sistemas de recuperación sólo permiten el truncado por la derecha (sufijos), otros pueden permitirlo también por la izquierda (prefijos) o incluso el enmascaramiento interno (truncado limitado a uno o varios caracteres dentro de una palabra).

5.2.4. Operadores numéricos y de rango. En campos numéricos el sistema de recuperación puede permitir la búsqueda de todos los valores mayores o menores a una cifra o comprendidos en un intervalo determinado.

Por ejemplo, si se desea localizar solamente documentos publicados a partir de 1995, habrá que seleccionar en el campo *año de publicación* los valores mayores que esta cifra o bien utilizar la expresión 1995:1999 si se prefiere acotar un periodo.

La existencia de esta opción en los sistemas de recuperación facilita la consulta en los campos de carácter numérico (número de registro, año de publicación, códigos de clasificación, periodo histórico,...)

5.2.5. Acotar o limitar conjuntos. La interrogación se ejecuta sólo sobre los registros resultado de una búsqueda anterior, en lugar de sobre toda la base de datos.

5.3. CAMPOS DE BÚSQUEDA Y CONTROL DEL VOCABULARIO

La búsqueda por materias se realiza generalmente a través de dos tipos muy diferentes de campos:

- a) **Campos de texto libre:** título del documento, resumen, el texto completo (en bases de datos que lo incluyan).
- b) **Campos formados por entradas de un índice alfabético de materias.** Son campos sin la estructura sintáctica de un texto, formados por grupos nominales de una sola palabra (por ejemplo "Filosofía") o más de una ("Filosofía de la Ciencia"), que definen un tema de búsqueda por sí mismas o en combinación con otras entradas del índice. Estas entradas se denominan términos de indización y puede haber diferentes tipos:
 - **Encabezamientos y subencabezamientos de materia.** Es el sistema utilizado en los catálogos de las bibliotecas para describir las materias tratadas en un documento. Se trata normalmente de términos procedentes de un **lenguaje controlado** establecido previamente (Listas de encabezamientos de materia). Cada entrada del índice se compone de una materia principal (encabezamiento), que puede estar acompañada de facetas de materia, lugar, tiempo o forma (subencabezamientos). Por ejemplo, la entrada 'Mujeres – Historia – Recursos bibliográficos – Cádiz' sería un único encabezamiento con tres subencabezamientos.

- **Palabras clave:** términos simples o compuestos extraídos de los documentos, sin que estén presentes en ningún lenguaje controlado preestablecido. En un sistema de indización por palabras clave pueden utilizarse tantas formas de expresar el mismo concepto como sean utilizadas por los autores de los documentos originales, por ejemplo "Política ambiental" y "Política de medio ambiente".
- **Descriptores:** son términos simples o compuestos extraídos de un **lenguaje controlado** que traducen los conceptos utilizados por los autores, para evitar la ambigüedad del lenguaje natural. Estos lenguajes documentales controlados pueden ser simples léxicos o listados alfabéticos o tratarse de **tesauros** con una estructura jerárquica entre sus términos. Los productores de una base de datos pueden elaborar su propio tesoro o adoptar uno elaborado por otra institución.
- **Otros campos de descripción del contenido.** En algunas bases de datos existen diferentes campos destinados a los términos de indización para facilitar el control del vocabulario y reducir las ambigüedades existentes entre nombres comunes (campo de descriptores), nombres propios (campo de identificadores), topónimos (identificadores geográficos), o cualquier otro definido para resolver necesidades específicas de información (compuestos químicos, coordenadas geográficas, legislación y jurisprudencia, periodo histórico). El contenido de estos campos también puede tener un control del vocabulario a través de listados de autoridades. Además puede existir un campo para reflejar aquellos términos que no se han incorporado aún al tesoro (campo de palabras candidatas).

La búsqueda debe adaptarse a las características de esta estructura de campos, adoptando dos tipos de estrategias complementarias:

- a) Cuando se realiza una búsqueda en campos de texto libre (título, resumen,...) o de palabras clave (*lenguajes no controlados*), se deben tener en cuenta las diferentes formas en las que puede aparecer un concepto en las referencias bibliográficas: sinónimos, conceptos afines, variaciones gramaticales,...
- b) Por el contrario, cuando la búsqueda se realiza en campos con control del vocabulario, debemos intentar localizar la forma seleccionada en el *lenguaje controlado* utilizado por los productores de la base de datos. Para ello, es conveniente consultar previamente el tesoro o léxico de indización o bien realizar la búsqueda a través de la visualización de los índices de cada campo.

6. OTRAS OPERACIONES DE LA CONSULTA DE UNA BASE DE DATOS: VISUALIZACIÓN, IMPRESIÓN, EXPORTACIÓN, ETC.

Una vez realizada la interrogación en la base de datos, el sistema de recuperación puede permitir alguna o todas estas opciones:

- **Visualización** en pantalla de las referencias. Puede realizarse en un formato standard o bien permitir la *selección de un formato* más completo o más sencillo, según el número de campos que se incluyan.

- **Ordenación** de las referencias de acuerdo al contenido de un campo concreto: autores, fecha de publicación, etc.
- **Selección** de las referencias concretas que resulten de interés para el usuario, que marcará las mismas.
- **Envío** de la totalidad de las referencias o sólo de las marcadas al soporte elegido por el usuario: impresora, fichero texto en el propio ordenador o en un disquete, fichero de exportación de datos o mensaje de correo electrónico.
- Realización de una *petición a un servicio de suministro de documentos*, solicitando la recepción de fotocopias de los artículos seleccionados por correo ordinario o electrónico.
- Lanzamiento de una segunda interrogación a partir de *enlaces hipertexto*.
- Visualización del *historial de las búsquedas* realizadas hasta el momento.
- Grabación o reutilización de *estrategias de búsqueda*.

7. ¿CÓMO REALIZAR BUSQUEDAS EN BASES DE DATOS?

Para realizar una consulta eficaz en una base de datos se debe intentar encontrar la estrategia más adecuada en cada sistema de información. Para ello se recomienda seguir estos pasos:

- 1) *Seleccionar la base de datos* más adecuada a la consulta. La elección tendrá en cuenta la cobertura temática, la homogeneidad del lenguaje utilizado y la tipología documental analizada por un sistema de información.
- 2) *Familiarizarse* con las características de la base de datos y del sistema de recuperación.
- 3) *Seleccionar los conceptos de búsqueda* que representen de forma precisa el tema de interés. Deben expresarse con todas sus posibles variantes (sinónimos, conceptos afines), aunque conviene evitar los términos innecesarios.
- 4) *Determinar la estrategia de búsqueda*: en qué campos puede encontrarse de forma más pertinente la información buscada y cómo combinar los diferentes conceptos, con operadores booleanos o de adyacencia. Las herramientas como el truncado son muy útiles, pero siempre que se compruebe su alcance para evitar la selección de otros términos no deseados. Cuando sea necesario se consultarán directamente los índices de los campos seleccionados u otros recursos (tesauros, listados,...).
- 5) *Realizar la búsqueda y visualizar los resultados*, valorando la adecuación de los mismos y la eficacia de la estrategia utilizada.

- 6) Si el resultado es poco satisfactorio o existen dudas sobre su idoneidad, se debe *repetir la búsqueda* ampliando o limitando la estrategia utilizada.
- 7) *Seleccionar los registros* adecuados a la necesidad de información y realizar una petición de suministro de documentos o bien obtener un listado de referencias por impresora o en fichero.

8. ¿QUÉ DEBE CONOCERSE SOBRE UNA BASE DE DATOS ANTES DE UTILIZARLA?

Para utilizar eficazmente las posibilidades de recuperación en un sistema de información es recomendable adaptarse lo más posible a sus características particulares. Para ello es necesario estar informado previamente de su contenido y de cómo debe realizarse la búsqueda. Se debe tener en cuenta:

- Si el sistema permite la consulta de una única base de datos o de varias simultáneamente. En este caso habrá que valorar si se trata o no de bases de datos de características similares.
- Cuál es su cobertura temática. En las bases de datos especializadas se pueden realizar búsquedas prescindiendo de aquellos conceptos que definen el propio ámbito temático del sistema.
- Si se trata de un catálogo o de una base de datos documental con descriptores y/o con resúmenes.
- En el caso de que haya descriptores o encabezamientos de materias, si estos proceden o no de un léxico controlado accesible.
- Cuál es la estructura de campos y en cuáles se realiza la búsqueda por defecto cuando no se selecciona uno concreto.
- Cuando se teclean dos términos se debe saber si el sistema interpreta que debe buscar registros en los que aparezcan ambos a la vez (intersección) o bien cualquiera de los dos (suma booleana). A esto se le denomina operador por defecto. El resultado obviamente será distinto.

9. BASES DE DATOS Y ACCESO AL DOCUMENTO ORIGINAL

En algunos sistemas de información, como en las bases de datos con información factual, las documentales con texto completo y los directorios, el usuario obtiene directamente los datos o los documentos que precisa. Por el contrario, las bases de datos referenciales sólo cubren un primer paso en la búsqueda de información, ya que el usuario tan sólo obtiene la noticia de la existencia de un documento.

Por ello, estos sistemas pueden tener asociado un servicio de acceso al documento original. A través de él, el usuario que ha recibido el resultado de una consulta

bibliográfica puede solicitar directamente aquellos documentos que son de su interés. El servicio de acceso al documento puede realizar las siguientes funciones:

- Préstamo interbibliotecario de monografías. Las grandes bibliotecas nacionales y las redes de bibliotecas universitarias han desarrollado los sistemas internacionales y nacionales del préstamo interbibliotecario.
- Envío de tesis, informes o documentos de archivo en microfilm o en microficha.
- Envío de artículos en fotocopia, a través del correo ordinario.
- Envío de artículos a través del correo electrónico. En muchos centros se está utilizando el programa ARIEL para el envío de copias de los artículos a través del correo electrónico. Este sistema consigue una alta calidad de las copias que supera claramente al envío por fax, y al mismo tiempo permite una gran rapidez en el suministro de documentos frente al correo ordinario.

En la actualidad, muchos documentos están ya en formato electrónico. Un valor añadido de las bases de datos bibliográficas puede ser el enlace directo con la copia electrónica del documento original. El usuario puede decidir imprimirlo o realizar una copia en su propio ordenador personal.

10. BIBLIOGRAFIA RECOMENDADA PARA AMPLIAR ESTE TEMA

CORDÓN GARCÍA, J.A.; LÓPEZ LUCAS, J.; VAQUERO PULIDO, J.R.: *Manual de búsqueda documental y práctica bibliográfica*. Madrid: Pirámide, 1999.

LANCASTER, F.W.: *El control del vocabulario en la recuperación de información*. Valencia: Universidad, 1995.

LÓPEZ YEPES, J. (coord.): *Manual de Información y Documentación*. Madrid: Pirámide, 1996.

SALVADOR OLIVÁN, J.A.; ANGÓS ULLATE, J.M.: *Técnicas de recuperación de información. Aplicación con Dialog*. Gijón: Trea, 2000