

Tema 8

BASES DE DATOS DOCUMENTALES

8.1 INTRODUCCIÓN

¿Qué entendemos por documento? Un conjunto de información o conocimiento humano: un libro, una revista, una factura, un carnet de identidad, una conferencia editada, una cinta de vídeo, un disco, etc.

En consecuencia, un fondo documental será un conjunto homogéneo de documentos que pueden encontrarse en cualquier soporte (papel, magnético, óptico, ...)

Una SGBDD (Sistema de Gestión de Bases de Datos Documentales) se ocupa de la gestión de documentos optimizando el almacenaje y facilitando su recuperación. A diferencia de cualquier otro SGBD, un SGBDD **no realiza ningún tratamiento sobre la información**. Simplemente la almacena y posibilita su recuperación.

8.2 ANÁLISIS DOCUMENTAL

Un SGBDD ofrece herramientas para realizar dos funciones básicas:

- almacenar la información documental y
- facilitar su recuperación.

Para tal fin se emplean técnicas de Análisis Documental. Sus conceptos básicos son:

- La indexación
- El lenguaje documental

El Análisis Documental se realiza a tres niveles de detalle:

- **Asiento:** Se determinan los identificadores de la información (título, autor, ...)

- **Descriptores:** Se extraen las palabras clave más representativas (indexación)
- **Resumen (o abstract):** Resumen analítico del texto íntegro para hacer más fácil la consulta.

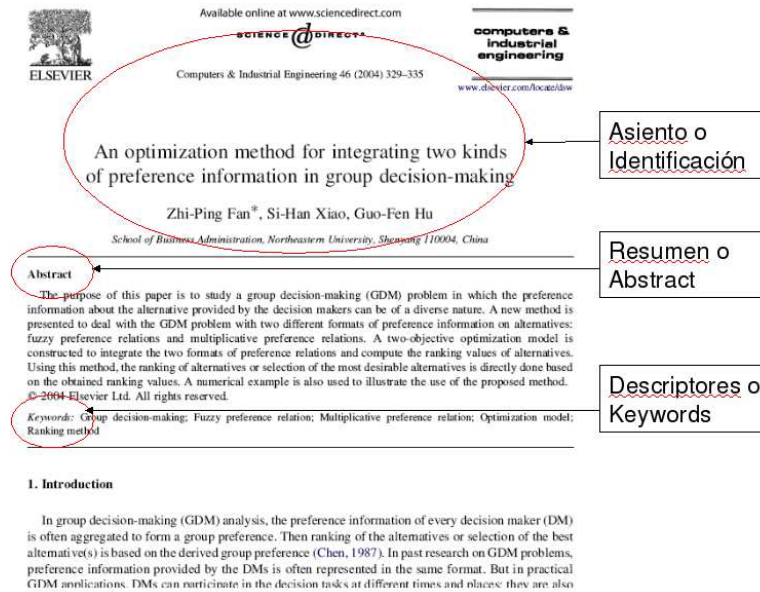


Figura 8.1: Análisis documental

8.3 LENGUAJES DOCUMENTALES

Es un lenguaje artificial que permite realizar una representación formal de documentos considerando para ello la manera en que se realizan las demandas o consultas por parte de los usuarios.

En la práctica, un lenguaje documental viene a ser un conjunto estructurado de términos mediante los cuales se indexan los documentos.

TESAUROS Es un lenguaje documental que incluye relaciones semánticas de tres tipos:

- Relaciones de equivalencia (sinónimos)
- Relaciones jerárquicas (término general y término específico)
- Relaciones asociativas (términos relacionados, vea también, ...)

Láminas
 LT Fuentes de información
 < Documentos gráficos
 .< Materiales no librarios
 ..< Fuentes de información
 - Comics

Lápices ópticos
 LT Tecnol. de la información y las comunicaciones
 < Periféricos
 .< Equipos informáticos
 ..< Informática

Lectores
 LT Profesionales de la información y usuarios
 < Perfiles de usuarios
 .< Usuarios
 ..< Profesionales de la información y usuarios
 - Acceso al documento bibliotecario
 - Clubes de lectura
 - Demanda de información
 - Hábitos de lectura
 - Historia de la lectura
 - Lectura pública

Lectores de códigos de barras
 LT Tecnol. de la información y las comunicaciones
 < Periféricos
 .< Equipos informáticos
 ..< Informática

Lectura pública
 LT Profesionales de la información y usuarios
 Utilícese para indicar la actividad lectora
 de los usuarios de unidades de información.
 < Usuarios
 .< Profesionales de la información y usuarios
 > Clubes de lectura
 > Hábitos de lectura
 - Fomento de la lectura
 - Historia de la lectura
 - Lectores

Figura 8.2: Ejemplo de lenguaje documental

<http://www.ucm.es/info/DAP/tesauro.htm> TESAURO

Término: <input type="checkbox"/>	Biblioteconomía
Término(s) relacionado(s): <input type="checkbox"/>	Bibliometría
<input type="checkbox"/>	Biblioteca
<input type="checkbox"/>	Procesamiento de la información
Término amplio: <input type="checkbox"/>	Documentación científica
Término(s) específico(s): <input type="checkbox"/>	Catalogación bibliográfica
<input type="checkbox"/>	Servicios bibliotecarios
<hr/>	
Término: <input type="checkbox"/>	Economía
Término(s) relacionado(s): <input type="checkbox"/>	Actividades económicas
<input type="checkbox"/>	Economía de la educación
<input type="checkbox"/>	Política económica
Término amplio: <input type="checkbox"/>	Ciencias Económicas
Término(s) específico(s): <input type="checkbox"/>	Economía del trabajo
<hr/>	
Término: <input type="checkbox"/>	Economía de la educación
Término(s) relacionado(s): <input type="checkbox"/>	Economía
<input type="checkbox"/>	Presupuesto
Término amplio: <input type="checkbox"/>	Política educativa
Término(s) específico(s): <input type="checkbox"/>	Costo de la educación
<input type="checkbox"/>	Economía universitaria
<input type="checkbox"/>	Financiación de la enseñanza
<input type="checkbox"/>	Gastos de educación
<input type="checkbox"/>	Presupuesto de centro
<input type="checkbox"/>	Presupuesto de educación

Figura 8.3: Ejemplo de tesauro

8.4 INDEXACIÓN

Operación orientada a poner de manifiesto los temas del documento entre-sacando los elementos que lo representen para su posterior localización. Dos fases:

- Extracción de descriptores. Disponemos de tres métodos:
 - Estadístico (análisis de frecuencias)
 - Por asignación (listas normalizadas)
 - Sintáctico (análisis morfológico y semántico)
- Traducción a lenguaje documental (Tesauro)

Métodos de indexación Todos los métodos persiguen dos objetivos:

- mínimo espacio de almacenamiento y
- máxima rapidez en indexación y respuesta

Tipos de indexación

- Por palabras: se indexa cada palabra del documento (excepto las incluidas en el diccionario de palabras vacías)
- De "string": Se pueden indexar palabras o frases completas. Este método da lugar a índices menos extensos.

Diccionario de palabras vacías Conjunto de palabras que no deseamos que formen parte de los índices como pueden ser los artículos, las preposiciones, y todas aquellas palabras que no son relevantes en la recuperación de la información y que solo contribuyen a ocupar espacio y disminuir la velocidad de acceso a la información.

Filtros Herramientas que permiten igualar caracteres para que el sistema de búsqueda los considere iguales:

- Mayúsculas y minúsculas (A=a,...)
- Vocales acentuadas y sin acentuar (á=a,...),
- etc.

