

# Bases de Datos Textuales

Bases de datos avanzadas

20/04/2012

Instituto Tecnológico de Costa Rica

Alonso Vargas Astúa

## Introducción

Las bases de datos textuales cumplen como objetivo la búsqueda y recuperación de información, la funcionalidad de estas bases de datos se extiende a múltiples áreas como la educación, la industria de software, medicina, automatización de oficinas entre otras y en este documento se detallaran algunas de las características de estas bases de datos.

## Bases de datos textuales.

Las bases de datos textuales tienen como objetivo ayudar al usuario a encontrar lo que buscan como un sistema de recuperación de archivos.

Una de las características principales en las bases de datos textuales es que las consultas se realizan de forma fácil quiere decir que en vez de realizar una consulta compleja como se hace en un sistema gestor de bases de datos se puede realizar una consulta simple en lenguaje natural, por ejemplo si se quiere obtener información sobre la luna se podría encontrar información con frases como “ciclo lunar”, “exploración de la luna” u otros que pueden relacionarse, a estas consultas que se realizan por lenguaje natural se les llama también Conocimiento lingüístico, en ocasiones suele pasar que no se halle una relación de lo que el usuario consulta con la información, este problema puede ser resuelto con el uso de etiquetas, lemas, frases comunes y sinónimos.

Las bases de datos textuales se pueden dividir en modelos según su estructura, estos modelos sirven para recuperar información, y además los modelos deberían deducir lo que el usuario realmente quiso preguntar, algunos modelos especializados para obtener información son:

- **Modelo Booleano:** este modelo usa operadores lógicos y un sistema de relevancia binaria (relevante o no relevante).
- **Modelo Vectorial:** utiliza un conjunto de palabras claves para descartar las opciones en todo el texto que no se relacionan con la consulta, este proceso puede ser ayudado por las etiquetas y los lemas, funciona como la búsqueda de un conjunto de términos en un conjunto de documentos.
- **Modelo probabilístico:** Existe un conjunto de documentos que son relevantes a la consulta, para cada documento se evalúa la posibilidad de que el usuario lo encuentre relevante, al final se recuperan los documentos que tengan mayor probabilidad de relevancia.

Otro aspecto importante en el tema de las bases de datos textuales es la indización, en un sistema de índices estándar cada palabra se almacena con una referencia numérica que especifica a que documento pertenece. Otros tipos de índices son:

- **Índices analíticos:** reúne frases y palabras clave en orden alfabético junto a las páginas donde pueden encontrarse, puede ser más eficiente si se recurre al uso de conocimiento lingüístico.

- **Índices invertidos:** puede estar formado por el vocabulario que son el conjunto de términos del texto, y el posteo que es el conjunto de documentos donde aparece el termino, los índices invertidos funcionan de la siguiente manera se recorren los textos siguiendo una secuencia, cada termino que es leído es buscado en el vocabulario, si no existe en el vocabulario se agrega a una lista de posteo vacía, y el documento que se está leyendo se agrega al final de la lista de posteo, cuando se termina la lista esta se agrega al disco.

El truncamiento en la gestión de las bases de datos ayuda a extraer las raíces de las palabras para poder encontrar mejor las relaciones entre las consultas y los documentos.

### **Sistemas de recuperación de Información.**

Estos sistemas son herramientas de búsqueda y recuperación de datos, para los sistemas de bases de datos textuales se utiliza la recuperación de documentos (information retrieval), para este tipo de recuperación de datos necesita de un Sistema de tratamiento y recuperación de información documental (STRID), estos sistemas manipulan únicamente documentos.

Algunos de estos sistemas de tratamiento y recuperación son:

- **Exploradores:** estos sistemas realizan las búsquedas en los documentos sin crear ficheros inversos o diccionarios.
- **Sistemas de gestión de bases de datos documentales:** automatizan las características de sistemas tradicionales de gestión y creación de bases de datos.
- **Gestión bibliográfica:** para el tratamiento de bibliografías especializadas.
- **Recuperación de información:** permite consultas en bases de datos que no son modificables.
- **Sistemas de gestión electrónica de documentos.**
- **Indizadores, Sistemas Hipertextuales, Gestion de Información personal,**
- **Sistemas Integrados para unidades de Información:** ofrece soporte a todas las operaciones que tienen lugar en un archivo.
- **Sistemas de gestión de bases de conocimiento.**

Un Sistema de tratamiento y recuperación de información documental basa en algunos aspectos básicos como lo son la creación de bases de datos textuales, edición de documentos, recuperación y salida de información.

## Conclusión.

Las bases de datos textuales o documentales funcionan como una herramienta para la recuperación de documentos y para simplificar procesos de búsqueda.

Las consultas que se realizan son más simples.

Estas bases de datos ayudan de modelos que tienen como objetivo lograr que la consulta de los clientes sea óptima.

Los índices simplifican el proceso de búsqueda de información y hacen que el proceso sea más rápido.

## Bibliografía:

UCR – ECCI, Bases de Datos Textuales, Una Herramienta en el Trabajo de la Recuperación de Información, Prof. Bach. Kryscia Daviana Ramírez Benavides.

ACM SIGMOD. Publicly Available Database software.

<http://www.sigmod.org/databaseSoftware/>

BLAIR, D.C. Language and Representation in Information Retrieval. Amsterdam: Elsevier, 1990 [Volver]

CMS-Spain.com <http://www.ecm-spain.com/home.asp>

CODINA, Lluís. Bases de Datos Documentales: Talleres de Sistemas de Gestión de Bases de Datos Documentales. <http://www.lluiscodina.com/metodos.htm#ii>

LANCASTER, F.W. y WARNER, A.J.: Information Retrieval Today. Arlington: Information Resources Press, 1993. [Volver]

PÉREZ, Chantal. Bases de datos y Bases de conocimiento. <http://elies.rediris.es/elies18/522.html>

TRAMULLAS, Jesús. "Sección 2: Los sistemas de bases de datos y los SGBD". En Introducción a la Documática. <http://tek.docunautica.com/>

TRAMULLAS, Jesús. "Sección 3: La recuperación de información". En Introducción a la Documática. <http://tek.docunautica.com/>

TRAMULLAS, Jesús. "Sección 4. Sistemas informáticos de tratamiento y recuperación de información documental". En Introducción a la Documática. <http://tek.docunautica.com/>