



La sociedad de la información debe concebirse no como un **MEDIO** distinto de hacer lo mismo, sino como un nuevo **ENTORNO** en el que debemos saber desenvolvernos.

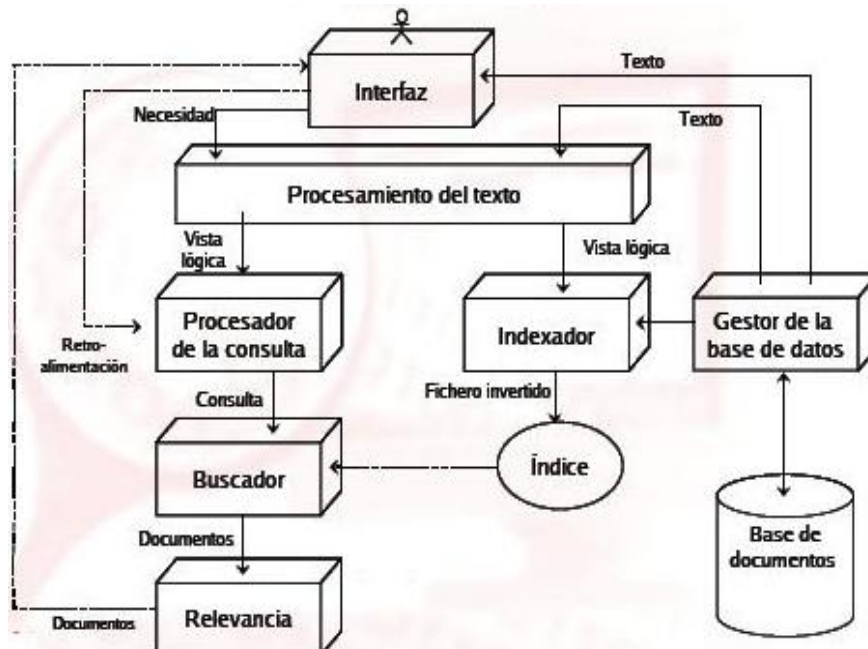
OBJETIVO

Crear una página web que trate sobre un término relacionado con la asignatura de [Sistemas avanzados de Recuperación de la Información \(S.A.R.I\)](#), de la [Universidad Carlos III de Madrid](#). El término que he desarrollado es **SQL**.

Una vez realizada se debiera "subir" a un servidor, en este caso he utilizado el de mi proveedor de servicios internet en Oviedo, que es [Telecable](#).

Luego darse de alta en un motor de búsqueda y tratar de lograr un buen posicionamiento.

El título hace referencia, a lo que desde mi punto de vista, es el proceso complejo de **recuperación de información**.



En la actual sociedad de la información en la que estamos inmersos, la **información** se ha convertido en un **recurso** esencial para la **creación de valor** en las empresas, instituciones y administraciones públicas.

De ahí mi **interés** en **conocer las técnicas mínimas** que permiten controlar y manejar la información, así como de **dotarse de las habilidades** imprescindibles para la **búsqueda y selección** de dicha información, a fin de poder tomar **decisiones acertadas** o de obtener nueva información.

Si teneis dudas o sugerencias escribirme a: beatricheboliche@hotmail.com

STRUCTURED QUERY LENGUAJE



Hasta la década de los 80, las personas que preparaban las consultas e informes de una base de datos debían ser programadores. Al aparecer las bases de datos con lenguajes de consulta sencillos y estandarizados, semejantes al lenguaje natural, el proceso de consulta puede hacerlo cualquier usuario mediante un lenguaje escrito asequible.

El lenguaje de gestión de bases de datos más conocido en la actualidad es el SQL, Structured Query Language, que es un lenguaje estándar internacional, comúnmente aceptado por los fabricantes de generadores de bases de datos.

El SQL permite:

- Definir una **base de datos** mediante tablas
- Almacenar **información** en tablas.
- **Seleccionar** la información que sea necesaria de la base de datos.
- Realizar **cambios** en la información y estructura de los datos.
- Combinar y calcular datos para **conseguir** la información necesaria.

El lenguaje SQL está compuesto por comandos, cláusulas, operadores, funciones de agregado.

Comando DLL	Permite crear y definir nuevas bases de datos, campos e índices.
CREATE	Utilizado para crear nuevas tablas, campos e índices
DROP	Empleado para eliminar tablas e índices
ALTER	Utilizado para modificar las tablas agregando campos o cambiando la definición de los campos.
Comando DML	Permite generar consultas para ordenar, filtrar y extraer datos de la base de datos.
SELECT	Utilizado para consultar registros de la base de datos que satisfagan un criterio determinado.
INSERT	Utilizado para cargar lotes de datos en la base de datos en una única operación.
	Utilizado para modificar los valores de los campos y

UPDATE	registros especificados.
DELETE	Utilizado para eliminar registros de una tabla de una base de datos.

[Subir](#)

Cláusula	Modificar los datos que desea seleccionar
FROM	Utilizada para especificar la tabla de la cual se van a seleccionar los registros Utilizado para consultar registros de la base de datos que satisfagan un criterio determinado.
WHERE	Utilizada para especificar las condiciones que deben reunir los registros que se van a seleccionar.
GROUP BY	Utilizada para expresar la condición que debe satisfacer cada grupo.
HAVING	Utilizado para eliminar registros de una tabla de una base de datos.
ORDER BY	Utilizada para ordenar los registros seleccionados de acuerdo con un orden específico.

[Subir](#)

Operador	Efectuan operaciones con los datos
AND	Es el "y" lógico. Evalúa dos condiciones y devuelve un valor de verdad sólo si ambas son ciertas.
OR	Es el "o" lógico. Evalúa dos condiciones y devuelve un valor de verdad si alguna de las dos es cierta.
NOT	Negación lógica. Devuelve el valor contrario de la expresión..
<	Menor que.
>	Mayor que.
<>	Distinto de.
<=	Menor ó Igual que.
>=	Mayor ó Igual que.
=	Igual que.
BETWEEN	Utilizado para especificar un intervalo de valores.
LIKE	Utilizado en la comparación de un modelo.
IN	Utilizado para especificar registros de una base de datos.

[Subir](#)

--	--

Función	Devolver un único valor que se aplica a un grupo de registros.
AVG	Utilizada para calcular el promedio de los valores de un campo determinado.
COUNT	Utilizada para devolver el número de registros de la selección.
SUM	Utilizada para devolver la suma de todos los valores de un campo determinado.
MAX	Utilizada para devolver el valor más alto de un campo especificado .
MIN	Utilizada para devolver el valor más bajo de un campo especificado.

[Subir](#)

Si teneis dudas o sugerencias escribirme a: beatricheboliche@hotmail.com

S.A.R.I.

[Recuperación de Información en la Web](#) | [Metadatos](#) | [XML](#).

[directorios](#) , los [motores de búsqueda](#) y los [metabuscadore](#)s.

- Google: <http://www.google.es/>
- Yahoo: <http://www.yahoo.es>
- Argus Clearinghouse: <http://www.clearinghouse.net/>
- AskJeeves: <http://www.ask.com/>
- Copernic: <http://www.copernic.com/>
- MetaCrawler: <http://www.metacrawler.com>
- Teoma: <http://www.teoma.com/>
- LookSmart: <http://search.looksmart.com/>
- WiseNut : <http://www.wisenut.com>
- [Animación explicativa \(EL PAIS\)](#)

Directorios

Es el sitio web que gestiona una **base de datos** que se ha creado de forma manual , clasificando las url **en categorías** Estos servicios reúnen una menor cantidad de sitios (es decir: sus bases de datos son más selectivas) pero se destacan porque ofrecen información más detallada sobre los documentos que indizan.

Muchas veces se redactan notas que describen con bastante minuciosidad el contenido o características de cada sitio, gracias al trabajo de un equipo de evaluadores, redactores o de los responsables de cada página incluida.

Casi siempre responden a una temática más acotada que puede ser un país, una región, un tópico, etc.

Algunos de estos servicios combinan dos o tres variables haciendo difícil su encasillamiento en una sola categoría; además se modifican constantemente ofreciendo cada vez más facilidades.

La tendencia actual es permitir dos opciones: la clasificación temática a través de un esquema de categorías de materia (Humanidades, Salud, Ciencias Sociales, etc.) y la búsqueda libre a través del motor de búsqueda (formulario que hay que completar).

Sus características son:

- Selección y **clasificación manual de recursos**.
- **Poca actualización** de los datos
- **Poca exhaustividad**
- **Relevancia** de los resultados y calidad

- Por lo general suelen ser **temáticos**.

[Subir](#)

Motores de búsqueda

Sitio web por el cual se tiene acceso a una base de datos automatizada. Se caracterizan por la potencia del motor de búsqueda que usan: suelen ser muy sofisticados así como su capacidad de rastrear información rápidamente entre millones de documentos.

La indización de las páginas se realiza por medio de una categoría de software llamada spider (araña en inglés: se usa para figurar el recorrido por la Web o telaraña) o también robot. El robot rastrea en forma automática las páginas web (puede ser la totalidad de la página o sólo el título y las primeras frases) y luego incorpora la información relevada a una base de datos propia. Por último, estos sistemas agregan un motor de búsqueda que se presenta al usuario como un formulario en blanco en el que hay que ir completando datos.

En los casilleros vacíos pueden ingresarse una o varias palabras y una vez formulado el pedido, el sistema recorre su base de datos para ver qué respuestas puede ofrecer.

Dado que la información se multiplica constantemente, estos buscadores incorporan técnicas de búsqueda complejas (operadores booleanos, de proximidad, búsqueda por frase, indicadores de relevancia, truncamiento, etc.) para contar con elementos u opciones suficientes para afinar la recuperación de la información lo más posible. Suelen poseer bases de datos muy grandes.

- Las **url** son recopiladas e **indizadas automáticamente**.
- Son exhaustivos y muy actualizados.
- **Manipulables**
- Dan problemas de calidad en los resultados y **ambigüedad** terminológica

[Subir](#)

Metabuscadores

Son buscadores que utilizan los motores de diversos sistemas en forma simultánea.

Igual que los servicios individuales se valen de diferentes técnicas de búsqueda que pueden variar de acuerdo con el metabuscador elegido.

Se presenta un formulario que una vez completado, se ejecuta en forma simultánea en varios buscadores individuales.

Los resultados se presentan indicando en cuál de cada buscador utilizado fue hallado ese documento.

La ventaja reside en el hecho de que una búsqueda de este tipo ahorra el trabajo de tener que ir probando suerte con cada uno separadamente.

Limitaciones: no se pueden hacer búsquedas demasiado refinadas, se recuperan hasta 50 resultados de cada buscador incluido (mucho menos de lo que podría obtenerse utilizando un buscador único), y ciertas estrategias de búsqueda pueden no ejecutarse correctamente.

Son tres los tipos de metabuscadore que existen:

- **Metabuscadore**: corresponden exactamente a lo que se ha descrito como tales.
- **Multibuscadore**: no realizan la combinación de los resultados, simplemente lanzan la consulta a varios buscadores.
- **Agentes** de búsqueda: son metabuscadore que han de instalarse localmente.

[Subir](#)

Metadatos

"datos estructurados sobre los datos."

Los metadatos son información descriptiva sobre un objeto o recurso tanto si éste es físico [formato tradicional tangible] como electrónico.

Las fichas de catálogo de biblioteca representan un tipo de metadatos perfectamente establecido, que ha servido como herramientas de gestión de colecciones y de recuperación de recursos durante décadas.

Los metadatos pueden tanto generarse a "a mano", como obtenerse automáticamente a través de un software..

Sus características son:

- **Capacidad de Relación** o de establecer enlaces.
- Imprescindibles en la **recuperación global** de la información en Internet.
- **Indizar y clasificar** información de diversos tipos.
- **Integrar de forma heterogénea** fuentes de información muy diversas, así como diferentes formatos de Bases de Datos Relevancia de los resultados y calidad

[Subir](#)

XML

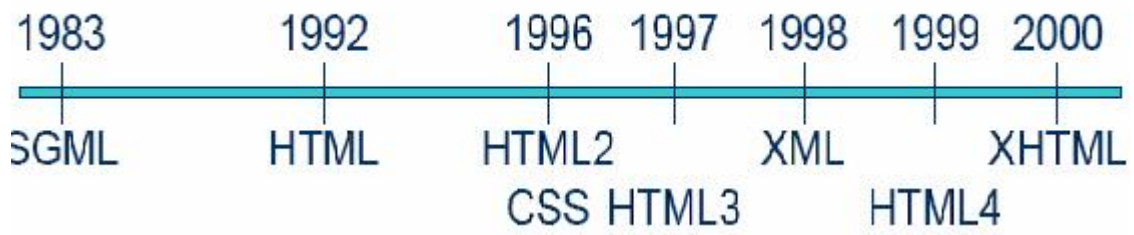
Extensible Markup Language, no es más que un conjunto de reglas para definir **etiquetas semánticas** que nos organizan un documento en diferentes partes.

Es un **metalenguaje** que define la sintaxis utilizada para definir otros lenguajes de etiquetas estructurados.

En teoría HTML es un subconjunto de XML especializado en presentación de documentos para la Web, mientras que XML es un subconjunto de SGML especializado en la gestión de información para la Web.

En la práctica XML contiene a HTML aunque no en su totalidad.

Desde su creación, XML ha despertado encontradas pasiones, y como para cualquier tema en Internet, hay gente que desde el principio se deja iluminar por sus expectativas, mientras otras muchas lo han ignorado.



[Subir](#)

Si teneis dudas o sugerencias escribirme a: beatricheboliche@hotmail.com

LA AUTORA



Valladolid 2 abril 1983

Soy Beatriz Gonzalez Seco, tengo 23 años y estudio Documentación en la Universidad Carlos III de Madrid. Anteriormente realicé la Diplomatura en Biblioteconomía y Documentación en la Universidad de Salamanca. En la actualidad estoy trabajando como documentalista en el Principado de Asturias.

Documento en .pdf [HACIENDO VISIBLE LO INVISIBLE](#).

Si teneis dudas o sugerencias escribirme a: beatricheboliche@hotmail.com

Diferentes artículos

- <http://informationr.net/ir/8-2/paper148.html>

Análisis crítico de la evaluación de los sistemas de recuperación de información en el contexto de la web.

- <http://www.ugr.es/~benjamin/Enfoques.pdf>

Descripción de implementaciones prácticas a partir del modelo cognitivo de recuperación de información.

- <http://www.fing.edu.uy/inco/cursos/tsi/TSI2/Trabajos/2002/TSI2-2002-Grupo2.pdf>

Cómo lograr una visión unificada de múltiples fuentes de datos independientes, buscando sus puntos en común y estableciendo correspondencias entre ellas?

- Artículo: [Posicionamiento en buscadores: una metodología práctica de optimización de sitios web](#).
- Artículo: [Propuesta y desarrollo de un modelo de la recuperación de la información en la Web: Capítulo 2: LA RECUPERACION Y LOS MODELOS DE RECUPERACION DE INFORMACION](#).

Bitácoras

Todo lo relacionado con bibliotecas, archivos, gestión de la información,...





[Si teneis dudas o sugerencias escribirme a: beatricheboliche@hotmail.com](mailto:beatricheboliche@hotmail.com)

Conoce las páginas de otros compañeros!!

- [SIDAR: Acceso Universal](#)
- [Accesibilidad](#)
- [CSS: Hojas de estilo](#)
- [HTML](#)
- [Bombing](#)
- [Metabuscadores](#)
- [Lenguajes de marcado Web: SGML](#)
- [Indización](#)
- [Novedad en la recuperación de información](#)
- [Palabras Clave](#)
- [Buscador Quaero](#)
- [Tesauros](#)
- [Internet Invisible](#)
- [Usabilidad](#)
- [Precisión como medida de recuperación de la información](#)
- [DHTML Páginas Dinamicas](#)
- [Modelo Vectorial](#)
- [JavaScript](#)
- [TREC](#)
- [Navegabilidad](#)
- [Agente de Búsquedas](#)
- [Cobertura en Recuperación](#)
- [Posicionamiento Web](#)
- [Bases de Datos](#)
- [Robots de Búsqueda](#)
- [Cloaking](#)
- [Directorio Abierto](#)
- [Clustering](#)
- [Silencio en la recuperación](#)
- [Datamining](#)
- [Frecuencia de Actualización](#)
- [Modelos Booleanos](#)
- [Optimización Web](#)
- [RSS Guia Facil](#)
- [Directorios](#)
- [Recuperación de la Información](#)
- [Topics Maps](#)
- [Interfaz de Búsqueda](#)
- [W3C](#)
- [Dublin Core](#)
- [Páginas Doorway](#)
- [Multibuscadores](#)
- [Yahoo](#)

- [Frecuencia de Actualización de contenidos Web](#)
- [Metadatos](#)
- [DoorWay](#)
- [Exhaustividad Relativa](#)
- [Multibuscadores](#)
- [Metadatos](#)
- [Yahoo](#)
- [Motores de Búsquedas](#)
- [URI](#)
- [SANDBOX](#)
- [Posicionamiento Web](#)
- [Fichero Inverso](#)
- [Dataware Housing](#)
- [Lycos](#)
- [Exhaustividad Recall](#)
- [Google](#)
- [Esfuerzo en la Recuperacion](#)
- [Guia de XML](#)
- [Acceso Sari](#)
- [Directorio SARI](#)
- [Directorio SARI](#)
- [Multibuscadores](#)

Si teneis dudas o sugerencias escribirme a: beatricheboliche@hotmail.com