

# HERRAMIENTAS INFORMÁTICAS PARA LA DOCUMENTACIÓN

## Tema 2. Métodos de búsqueda

### 1. Introducción

Como hemos visto a lo largo de los temas previos, una manera de documentarse consiste en buscar información relevante en grandes repositorios de documentos electrónicos como la *world wide web*, bases de datos en línea o discos ópticos. Estos almacenes de información o repositorios se componen de una enorme cantidad de documentos, donde cada documento constituye una unidad de información que se recupera por separado. Dependiendo de cada caso, un documento puede tener distintas formas, pudiendo ser una foto, un texto, una ficha bibliográfica, un vídeo, etc. Por otra parte, cada repositorio de información nos ofrece sus propios mecanismos de acceso a los documentos que almacena. La principal función de estos mecanismos es ayudar al usuario a recuperar la información más relevante a sus necesidades de entre la gran cantidad de documentos disponibles.

En este tema vamos a estudiar los mecanismos de acceso más comúnmente utilizados hoy en día. La eficiencia de cada herramienta de recuperación de información no sólo depende de los mecanismos de acceso que proporcione, sino que también depende en gran medida de la habilidad del usuario para manejarlas. Es por lo tanto muy importante saber muy bien cómo funcionan, para poder utilizar estas herramientas de la mejor manera posible.

### 2. Buscadores en Internet

Dada la ingente cantidad y variedad de información accesible a través de Internet, la dificultad con la que podemos encontrarnos ya no es tanto que la información no se encuentre disponible, sino que no sepamos encontrarla o extraerla de entre los miles de documentos y datos accesibles. Para solventar este problema han ido surgiendo toda una serie de herramientas de mayor o menor calidad que nos permiten buscar información en la red.

Hasta no hace muchos años existían un número bastante reducido de buscadores que podíamos utilizar. Sin embargo, en la actualidad hemos llegado a un punto en el que el número y variedad de buscadores es tan grande que lo que comienza a resultar difícil es decidirse por uno o varios de ellos. De hecho, en la red han comenzado a aparecer buscadores de buscadores que clasifican miles de ellos.

En este apartado vamos a intentar diferenciar los distintos tipos de buscadores. Para ello vamos a establecer cuatro criterios de clasificación en función de la respuesta a varias preguntas básicas sobre los buscadores:

¿Qué tipo de información buscan?

¿Cómo encuentran la información en Internet?

¿Cómo se estructura y presenta la información al usuario?

¿Qué mecanismo utiliza el usuario para encontrar la información en el buscador?

La respuesta a estas cuatro preguntas nos llevará a diferenciar distintos tipos de buscadores en cada uno de los cuatro apartados siguientes.

## **2.1. Tipos de información**

Tal y como vimos en un tema anterior, Internet esta constituida por toda una serie de servicios distintos que incluyen la web, el correo electrónico, las listas de correo, los grupos de noticias y los repositorios FTP entre otros. En la red podemos encontrar buscadores especializados en encontrar información asociada a distintos servicios. Así podemos encontrar:

- Buscadores de páginas web
- Buscadores de listas de correo o foros web
- Buscadores de grupos de noticias
- Buscadores en repositorios FTP

Además podemos encontrar otros buscadores especializados en encontrar distintos tipos de información (ver figura 1), tales como:

- Noticias en prensa
- Direcciones postales o de correo electrónico de personas
- Programas de ordenador
- Canciones o sonidos en general
- Imágenes

Por último, podemos encontrar en la red páginas y portales que incorporan sus propios buscadores que limitan los resultados al contenido de las mismas, tales como:

- Bases de datos
- Catálogos de bibliotecas
- Universidades

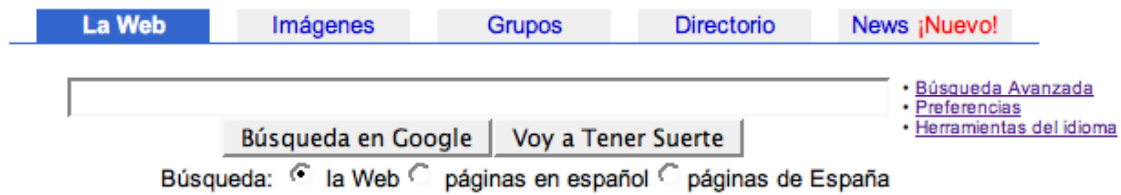


Figura 1. Distintos tipos de información que se pueden encontrar con Google España.

En los siguientes apartados nos centraremos fundamentalmente en los buscadores de páginas web, pero muchas de las ideas introducidas son aplicables a buscadores en otros servicios o de otros tipos de información.

## 2.2. Mecanismo de recopilación

En este apartado vamos a diferenciar distintos tipos de buscadores de páginas web en función de cómo encuentran la información en Internet. En primer lugar, es necesario desmentir una creencia ampliamente extendida respecto a dónde se encuentra la información que el buscador nos presenta como resultado cuando realizamos una búsqueda en el mismo. Muchos usuarios creen que cuando rellenan el formulario del buscador e inician la búsqueda, este busca las páginas que cumplen el criterio establecido en Internet y nos presenta los resultados obtenidos por pantalla. Si esto fuese así, para encontrar los miles de resultados que suelen presentar algunos buscadores, serían necesarias varias horas, o incluso varios días, cuando lo normal es que los resultados se presenten al usuario en pocos segundos.

En realidad el buscador extrae la lista de páginas web presentada al usuario de un almacén donde se guarda la información sobre las mismas. Este almacén suele constar de varios discos duros y se encuentra en el mismo lugar que el ordenador donde se ejecuta el buscador.

Para entender esto, es necesario diferenciar dos procesos distintos llevados a cabo por los buscadores de páginas web (ver figura 2). Por un lado el buscador debe recopilar la información sobre las páginas web de Internet y almacenarla localmente. Por otro lado, el buscador debe atender las peticiones realizadas por los usuarios y presentar los resultados encontrados en el almacén local. Esto es, hay que diferenciar el entorno de búsqueda que la herramienta presenta a los usuarios cuando se accede a su página web junto al mecanismo utilizado para encontrar la información en el almacén local, de la forma en que el buscador ha podido recoger la información sobre todas las páginas web entre las que realiza la búsqueda.

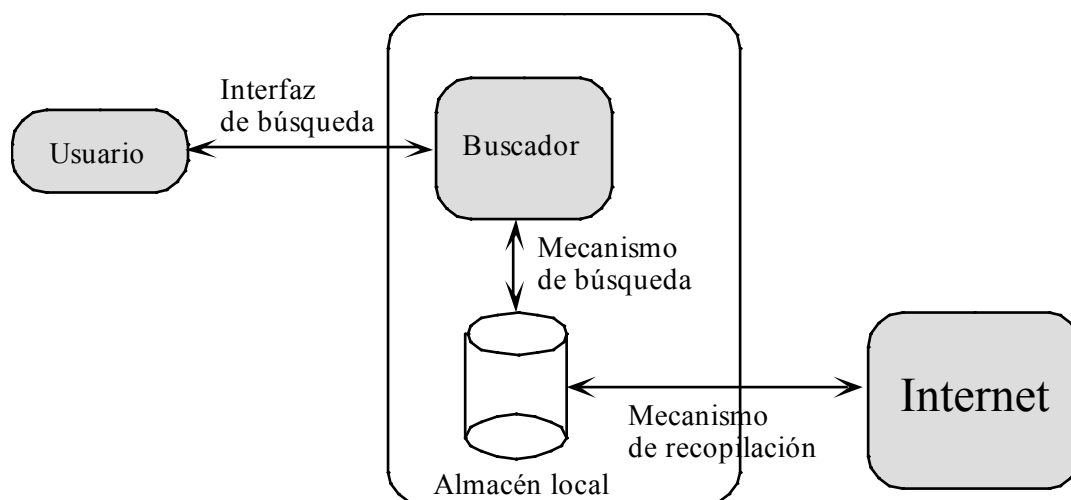


Figura 2. Esquema básico de funcionamiento de los buscadores en Internet.

En este apartado vamos a diferenciar dos tipos de buscadores básicos en función del modo en que recopilan la información sobre las páginas web en el almacén local.

Uno de los métodos utilizados por los buscadores para recopilar la información es recorrer las distintas páginas de Internet e ir almacenando la información sobre las mismas. Este tipo de buscadores utilizan un programa de ordenador especializado en recorrer la red, extraer la información sobre las distintas páginas que encuentra y recopilarla en el almacén local. Este tipo de programas que recorren la red se denominan robots de búsqueda o arañas (debido a que recorren la web, telaraña en inglés). Los robots de búsqueda se dedican a realizar un recorrido sistemático de todas las páginas web que encuentran y a acceder a los distintos enlaces que contienen, cuando terminan de realizar este recorrido vuelven a comenzar, de modo que intentan mantener la información almacenada lo más actualizada posible.

Es tan inmensa la cantidad de páginas web a recorrer que algunos robots de búsqueda pueden tardar varias semanas en revisarlas todas. Dado que las páginas que se nos presentan cuando realizamos una búsqueda son las del almacén local y no han sido accedidas en ese momento en Internet por el buscador, puede ocurrir perfectamente que hayan cambiado desde el momento en que el robot las almacenó, e incluso que hayan desaparecido. De ahí que en ocasiones cuando pulsamos los enlaces que nos presentan los buscadores, pueden producirse errores por intentar acceder a páginas inexistentes.

El otro método de recopilación de páginas utilizado por los buscadores se basa en las contribuciones de los usuarios. Muchos buscadores proporcionan a los usuarios la posibilidad de incluir su página web en el almacén local en el que realizan las búsquedas o de sugerir páginas de otros usuarios. Para ello, el usuario sólo tiene que acceder a la opción adecuada y proporcionar la información relacionada con la página a añadir que se le solicite. Por ejemplo, la figura 3 presenta el formulario de *Yahoo! España* para sugerir un nuevo sitio web.

**Sugerir un sitio web**

**Paso 2 de 2: sitio web e información de contacto**

**Información sobre el sitio web** \* Datos obligatorios

Añada la información de la página web. No escriba TODAS las letras en mayúsculas. No incluya superlativos ni frases publicitarias. Tenga en cuenta que Yahoo! España se reserva el derecho de modificar las descripciones presentadas.

\* **Título:**   
 · El nombre del sitio web (ej. Museo Picasso)

\* **URL:**   
 · Es la dirección del sitio web y empieza por "http://"

**Ubicación geográfica:**   
 · Ciudad, provincia, país (si es aplicable)

**Comentario:**   
 · Escriba una frase breve y sencilla de menos de 25 palabras.  
 · No repita el título del sitio web o el nombre de la categoría.  
 · Ponga en mayúsculas sólo la primera letra de la oración y no incluya código HTML.  
 · Por favor, no use frases publicitarias como por ejemplo "Somos los mejores!" o "El mejor sitio web de Internet!"

Figura 3 Parte del formulario de Yahoo España para sugerir un nuevo sitio.

### 2.3. Catálogos y motores de búsqueda

El tercero de los criterios de clasificación de los buscadores de páginas web se basa en el modo en que estos estructuran la información almacenada y cómo presentan los resultados de las búsquedas al usuario. En base a este criterio podemos diferenciar dos tipos de buscadores fundamentales: catálogos y motores de búsqueda.

Los catálogos, también denominados índices y directorios, guardan las páginas web en su almacén local clasificadas en una serie de categorías en función del tema que abordan. Por ejemplo, podemos diferenciar una serie de categorías básicas tales como Artes, Educación, Ciencia y Tecnología, Juegos, ... y clasificar las distintas páginas recopiladas en alguna de ellas o en varias a la vez. A su vez, dentro de cada categoría podemos diferenciar distintas subcategorías que refinan más el contenido de las páginas. Por ejemplo, dentro de la categoría Arte, podemos diferenciar las subcategorías Música, Literatura, Cine, ... Las distintas subcategorías pueden a su vez dividirse, y podemos construir un gran árbol jerárquico que abarque todos los temas posibles para los que podemos encontrar páginas en Internet. La figura 4 presenta las categorías en las que se clasifican las páginas del catálogo incluido en *google.es*.



Figura 4. Catálogo incluido en google.es.

Cuando se realiza una búsqueda en un catálogo, las páginas resultantes se muestran al usuario indicando alguna de las categorías en las que se encuentran clasificadas, tal y como puede comprobarse en la figura 5.

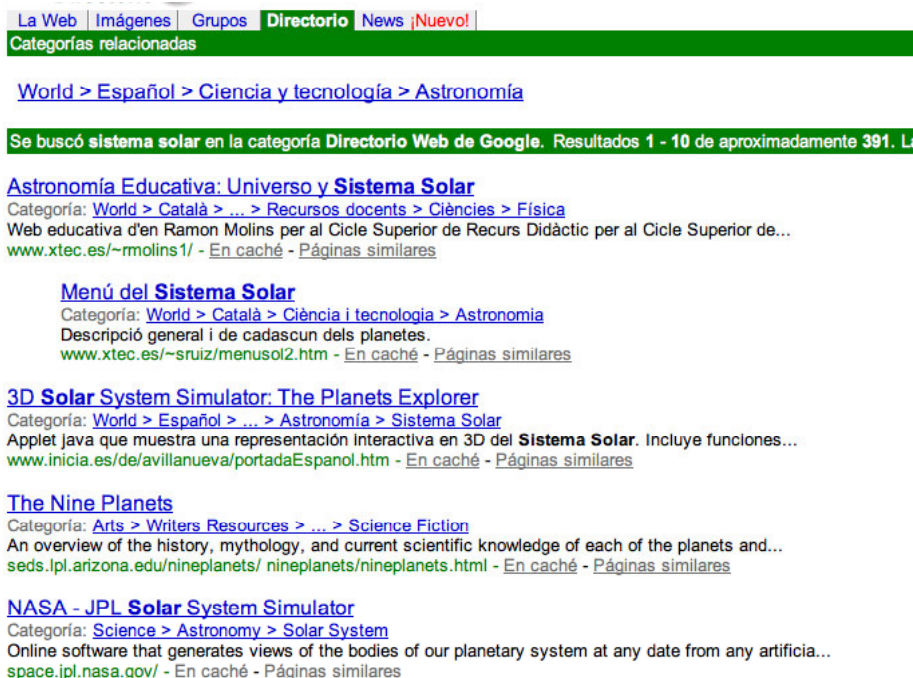


Figura 5. Resultado de una búsqueda en el catálogo incluido en google.es.

La segunda gran clase de buscadores de páginas web son los motores de búsqueda. En este caso, la información almacenada localmente en el buscador no se clasifica en función de su contenido, sino que se encuentra toda almacenada al mismo nivel. Cuando

se realiza una búsqueda en un motor de búsqueda, las páginas resultantes se muestran al usuario como una lista sin ningún tipo de clasificación temática. Por ejemplo, la figura 6 muestra el resultado presentado por el motor de búsqueda *Altavista* cuando buscamos páginas conteniendo las palabras `sistema solar`.



Figura 6. Resultado de una búsqueda en el motor de búsqueda Altavista.

Por otro lado, es importante saber, que algunos de los mejores sitios web con buscadores combinan ambos tipos de buscador, incluyendo tanto un catálogo como un motor de búsqueda. Así, el buscador Google, incluye un motor de búsqueda en *www.google.com* y un catálogo en *directory.google.com*.

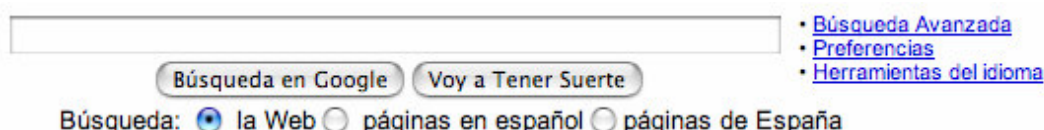
Por último, es posible relacionar los catálogos y motores de búsqueda con los dos modos de funcionamiento que hemos distinguido en el apartado anterior. Así, la mayoría de los catálogos obtienen la información que contienen en base a la suscripción de las páginas por parte de los usuarios. Normalmente la información almacenada sobre las distintas páginas se reduce a su título y a una breve descripción de su contenido. Las páginas suscritas son posteriormente filtradas por un grupo de personas que deciden si se van a incluir y en qué categorías van a incluirse. Debido a este modo de funcionamiento, es normal que el número de páginas que podemos encontrar en un catálogo sea mucho menor que el que contienen los mejores motores de búsqueda. En cuanto a los motores de búsqueda, normalmente se apoyan en un robot de búsqueda para recopilar las distintas páginas que almacenan. Este proceso es automático y no requiere ningún proceso de clasificación por parte de personas. Además, la información que se suele almacenar sobre las páginas se basa en todo el texto que contienen y no sólo en una breve descripción de las mismas como ocurre en los catálogos.

## 2.4. Mecanismo de búsqueda

Los distintos buscadores proporcionan al usuario una página web que permite realizar búsquedas de páginas web incluidas en el almacén local. En función del tipo de buscador con el que trabajemos podremos utilizar diferentes mecanismos de búsqueda. Fundamentalmente se distinguen dos métodos de búsqueda: por categorías y por palabra clave.

La búsqueda por categorías, también denominada búsqueda jerárquica, sólo puede emplearse en los catálogos, pues son estos los únicos buscadores que clasifican las páginas en función de su contenido. Este tipo de buscadores presentan al usuario la lista de las categorías que diferencian en forma de enlaces, tal y como puede verse en la figura 4. El usuario puede seleccionar una de las categorías y ver desplegadas, de nuevo en forma de enlaces, las distintas subcategorías que incluye. Podemos continuar navegando por los distintos niveles del árbol de categorías o catálogo hasta encontrar aquella que nos interese. A partir de un determinado nivel del catálogo, se muestran al usuario no sólo las subcategorías existentes, sino también las páginas clasificadas dentro de la categoría que se está visualizando. De esta forma, es posible encontrar las páginas que nos interesa sin más que navegar por las categorías del catálogo hasta encontrar la adecuada.

La búsqueda por palabra clave se basa en elegir un criterio o condición de búsqueda que define las características de las páginas que queremos encontrar. Normalmente este criterio de búsqueda consiste en una palabra o en una combinación de palabras mediante símbolos especiales como +, -, "", etc. Tal y como podemos ver en la figura 7, el criterio de búsqueda se teclea en un formulario presentado por el buscador y este busca en su almacén local las páginas que lo cumplen y las presenta por pantalla.



The image shows a portion of the Google search page. At the top is a search input field. Below it are two buttons: 'Búsqueda en Google' and 'Voy a Tener Suerte'. Underneath these buttons is the text 'Búsqueda:' followed by three radio buttons. The first radio button is selected and is labeled 'la Web'. The second is labeled 'páginas en español' and the third is labeled 'páginas de España'. To the right of the search bar, there are three blue links: 'Búsqueda Avanzada', 'Preferencias', and 'Herramientas del idioma'.

Figura 7. Formulario de búsqueda simple de Google.

En el caso de los motores de búsqueda este es el único método que podemos utilizar, y el resultado obtenido es una lista de páginas web. En el caso de los catálogos, también podemos realizar búsquedas por palabra clave. En este caso, el buscador puede listar las páginas que cumplen el criterio clasificadas por categorías, o puede mostrar además una lista de categorías que lo cumplen.

El establecimiento de un criterio adecuado de búsqueda por palabra clave puede ser bastante complicado. En muchos casos, el criterio utilizado no es lo suficientemente restrictivo y obtenemos demasiadas páginas, o puede que las páginas que encontremos no se correspondan con los resultados esperados. Debido a esto, los distintos buscadores permiten establecer criterios de búsqueda de una gran variedad y complejidad. De



hecho, la mayor parte de los buscadores permiten usar dos niveles de búsqueda en función del grado de complejidad del criterio a utilizar: búsqueda simple y búsqueda avanzada. Por ejemplo, la figura 8 muestra el formulario de búsqueda avanzada del motor de búsqueda Google.

Figura 8. Formulario de búsqueda avanzada del motor Google.

En muchas ocasiones, los buscadores no sólo permiten establecer criterios de búsqueda relacionados con el contenido de las páginas, sino también respecto a otros factores, tales como su ubicación geográfica, la fecha de creación, los enlaces que contiene, etc.

## 2.5. Otros tipos de buscadores

Además de los cuatro criterios descritos al principio del apartado 2, existen otras formas de clasificar los distintos buscadores. En este apartado revisaremos brevemente alguna de ellas.

### Por especialización temática

Aunque los buscadores más utilizados permiten encontrar páginas web dedicadas a cualquier tema, existen herramientas de búsqueda que se especializan en determinados temas de modo que tan sólo encuentran páginas relacionadas con los mismos. Podemos denominar a los primeros buscadores genéricos, y a los segundos buscadores temáticos. Así, podemos encontrar buscadores especializados en ciencias sociales como *Sosig* ([www.sosig.ac.uk](http://www.sosig.ac.uk)) o en otros temas más concretos como derecho ([www.findlaw.com](http://www.findlaw.com)) o arqueología ([archnet.asu.edu](http://archnet.asu.edu)), entre otros muchos. La figura 9 muestra un buscador temático centrado en el ámbito del derecho.



Figura 9. Buscador temático sobre derecho Jurisweb.

### Por ámbito geográfico

De nuevo, la mayoría de los buscadores permiten encontrar páginas web en todo el mundo. Sin embargo, existen otros que restringen la búsqueda a determinados países o regiones. Así por ejemplo en el buscador *La nueva España* ([www.lanuevaespana.es/buscador](http://www.lanuevaespana.es/buscador)), que podemos ver en la figura 10, podrás encontrar tan sólo direcciones de Internet en Asturias. Por otra parte, los buscadores más potentes como Altavista o Google, permiten restringir el ámbito de búsqueda, o bien a un país determinado o bien a un idioma concreto. Así, podemos encontrar páginas en español, situadas en cualquier país del mundo, o podemos encontrar páginas situadas en España.



Figura 10. Buscador de ámbito geográfico asturiano.

Por otro lado, podemos encontrar los multibuscadores y los metabuscadores. Aunque existe una cierta confusión entre ambos, podemos decir que ambos permiten realizar una búsqueda en varios buscadores a la vez utilizando un solo formulario. La diferencia entre ellos radica en cómo muestran los resultados. Los multibuscadores suelen mostrar por separado los resultados obtenidos en cada buscador utilizado. Los metabuscadores,

por otro lado, realizan un proceso previo de filtrado de los resultados en el que, por ejemplo, eliminan las páginas repetidas y, finalmente, muestran los resultados en una sola lista.

Un ejemplo de multibuscador puede ser *www.multicrawl.com*, mientras uno de los metabuscadores más conocidos es *www.metacrawler.com*. Este último devuelve las páginas encontradas por ejemplo en *Google*, *Yahoo* y *Altavista*, entre otros buscadores. La figura 11 muestra el formulario de búsqueda del multibuscador *Supercrawl*, y en ella podemos ver la lista de buscadores que utiliza.



Figura 11. Formulario de búsqueda del multibuscador Supercrawl.

## 2.6. ¿Cuándo usar un catálogo y cuándo un motor?

Una de las cuestiones básicas a plantearse cuando queremos encontrar alguna página web es si debemos utilizar un catálogo o un motor de búsqueda. La decisión a tomar depende del tipo de información que queremos encontrar y de nuestros conocimientos previos sobre la materia.

Los directorios son por definición selectivos en lugar de inclusivos. Normalmente su objetivo es almacenar páginas de calidad sobre distintos temas, y no almacenar todas las páginas de Internet.

Cuando queremos realizar una búsqueda general sobre una materia relativamente amplia, suele ser mejor utilizar un catálogo que un motor de búsqueda. Por ejemplo, si estamos buscando bibliotecas, es mejor acceder a un buen catálogo y buscar la categoría con ese nombre, bien navegando a través del árbol de categorías o bien utilizando el formulario de búsqueda por palabra clave. Las bibliotecas son un tema que prácticamente todos los catálogos diferencian como una categoría. De este modo, si accedemos a la misma, tenemos la garantía de encontrar un listado más o menos amplio de bibliotecas, e incluso estas pueden estar clasificadas por tipos tales como bibliotecas públicas, universitarias o por países. El que el tema que estemos buscando esté incluido en una categoría del buscador, nos garantiza que todas las páginas que encontremos en la misma tratan sobre el mismo como uno de sus temas fundamentales y muy probablemente respondan a nuestras expectativas cuando realizamos la búsqueda. Esto no es así en el caso de los motores de búsqueda. Si utilizamos la palabra biblioteca como criterio de búsqueda en un motor, podemos encontrar decenas de miles de páginas

que lo cumplen. Muy probablemente la inmensa mayoría de las páginas encontradas no serán páginas de bibliotecas y tendrán poco que ver con el tema. Esto es así porque los motores de búsqueda intentan encontrar las palabras utilizadas en cualquier parte de todo el texto de las páginas que almacenan. El hecho de que una palabra esté incluida en una página no es indicativo de que trate sobre un tema relacionado con la misma. Imaginemos por ejemplo, el tipo de página en el que podría aparecer una palabra tan genérica como biblioteca. Para evitar este problema en cuanto a la "calidad" de los resultados obtenidos con motores de búsqueda, estos suelen utilizar algún tipo de clasificación para presentar en primer lugar las páginas más relevantes sobre el tema buscado. En un apartado posterior detallaremos algo más este aspecto.

Por el contrario, si queremos realizar una búsqueda sobre un tema muy concreto, como una película, una empresa, una receta de cocina o un medicamento, es mejor utilizar un motor de búsqueda. Este tipo de temas tan concretos no suelen constituir categorías de un catálogo, y dado que los catálogos almacenan un número mucho menor de páginas, es mejor utilizar un motor de búsqueda para encontrar información sobre ellos.

Por otro lado, el uso de catálogos requiere un conocimiento, aunque sea mínimo, sobre el tema sobre el que se realiza la búsqueda. Esto nos permitirá navegar por las categorías del catálogo hasta encontrar las adecuadas al tema buscado, o al menos reconocer las categorías adecuadas cuando el catálogo nos las presente como resultado de una búsqueda mediante palabra clave. Sin embargo, para utilizar un motor de búsqueda, tan sólo es necesario saber describir el tema sobre el que queremos buscar (cosa que a veces no es tan fácil), es decir, encontrar la palabra o combinación de palabras que lo describen, aunque no sepamos nada sobre el mismo.

Para finalizar, algunos ejemplos de búsquedas en los que deberíamos utilizar un tipo u otro de buscador, puede que nos ayude a saber cuándo utilizarlos.

<b>Catálogo</b>	<b>Motor de búsqueda</b>
Un buen sitio sobre música o sobre un género particular de cine.	Información sobre una canción o película en particular
Sitios relacionados con un tipo de cocina.	Una receta de un plato determinado.
Sitios relacionados con la medicina alternativa.	Información sobre un medicamento.
Sitios relacionados con un género literario.	Información sobre un libro determinado.
Sitios relacionados con un cantante o actor.	Información sobre una persona a partir de su nombre.

### 3. Portales

Aunque no son un tipo de buscador, los portales son un mecanismo ideal para el acceso a la información en Internet. Podemos definir un portal como un sitio web que ofrece al usuario, de forma fácil e integrada, el acceso a una serie de recursos y servicios a los que se puede acceder a través de Internet. Este tipo de sitios se denominan portales porque constituyen una puerta de acceso a Internet. El objetivo final de las empresas que publican los principales portales es que los usuarios los utilicen como página de acceso por defecto de su navegador.

Debido a la enorme publicidad que se ha dado a portales tales como *Terra*, *Ya* o *Wanadoo*, muchos sitios web han pasado a autodenominarse portales, cuando no pasan de constituir unas páginas más o menos completas. Aunque no existe una definición exacta de en qué momento un sitio web puede denominarse portal, sí podemos distinguir una serie de servicios que debería proporcionar todo portal que se precie:

- Un buscador que integre un motor de búsqueda y un catálogo y que permita buscar tanto en Internet como en las distintas páginas que constituyen el portal. De hecho, algunos de los principales buscadores aparecidos hace unos años, tales como *Yahoo!*, *Lycos* o *Altavista*, han tendido que transformarse en portales que permiten el acceso a muchos más servicios
- Acceso a buscadores específicos de otros tipos de información, tales como software, música, personas, etc.
- Un servicio de noticias que recopile las últimas noticias aparecidas en prensa y que permita realizar búsquedas entre ellas.
- Un servicio de listas de correo electrónico, un servicio de foros web o grupos de charla (*chats*) donde se puedan plantear discusiones sobre distintos temas entre los usuarios.
- Acceso a información de utilidad, tal como información meteorológica, directorios de empresas y personas, mapas y callejeros, carteleras de cine, ofertas de empleo, etc.
- Acceso a tiendas para poder realizar compras de distintos tipos de artículos o para planificar un viaje o adquirir los billetes de tren o avión.
- Cuentas de correo electrónico gratuitas para los usuarios, así como la posibilidad de publicar páginas web.

Por otro lado, podemos diferenciar dos tipos fundamentales de portales:

- Los **portales horizontales** o generales son aquellos que permiten el acceso a información y servicios sobre cualquier tema en el que estemos interesados. Entre ellos podemos citar *Ya (ya.com)*, *Terra (terra.es)* o *wanadoo (wanadoo.com)*.
- Los **portales verticales** o temáticos son aquellos que centran sus contenidos o los servicios ofrecidos en un determinado tema. Así, podemos encontrar portales especializados en derecho, como *derecho.org* o en temas relacionados con la administración del estado, como *administración.es*.

El acceso a la mayoría de los servicios ofrecidos a través de los portales web es gratuito, sin embargo, algunos servicios, tales como las cuentas de correo adicionales, el espacio para publicar páginas web de gran tamaño, etc. pueden requerir del pago de unas cuotas periódicas. No obstante, la principal fuente de financiación actual de los portales es la financiación que obtienen de los anunciantes que insertan sus anuncios (*banners*) en el portal y que suele ser proporcional a la cantidad de usuarios que hayan accedido al mismo.

#### 4. Búsquedas basadas en condiciones sobre los documentos

La mayoría de las herramientas de búsqueda de información nos permiten definir una o varias condiciones que todos los documentos recuperados deben cumplir. Por medio de estas condiciones, los usuarios describen algunas de las propiedades de la información requerida. En esta sección vamos a estudiar los diferentes tipos de condiciones que se pueden utilizar para hacer una búsqueda de información.

En primer lugar hemos de decir que en muchas ocasiones, cuando realizamos una búsqueda, sobre todo en un almacén de información tan grande y heterogéneo como es la web, una gran parte de los documentos devueltos como respuesta no nos interesan. Estos documentos, aun cumpliendo las condiciones que hemos establecido no contienen el tipo de información que nosotros esperamos encontrar cuando las definimos. Denominaremos a estos documentos irrelevantes, mientras que aquellos que sí responden a nuestras expectativas serán documentos relevantes.

Es necesario considerar que cuanto más restrictivas sean las condiciones de recuperación, menor será el número de documentos recuperados. Del mismo modo, unas condiciones demasiado débiles producirán una respuesta con una cantidad demasiado elevada de documentos, siendo la mayor parte de ellos irrelevantes para el usuario. Es por lo tanto importante, identificar cuáles son las propiedades más importantes de la información requerida, y de esa manera recuperar el mayor número posible de documentos relevantes y el menor número posible de documentos irrelevantes.

Por ejemplo, si consultando una base de datos con bibliografía se deseara encontrar una referencia de una guía de viajes que esté muy actualizada para llevar en nuestras próximas vacaciones en Morella, hipotéticamente se podrían utilizar las siguientes condiciones:

- tipo\_documento = "Guía" Y año\_publicación = 2005 Y tema = "Morella".

Con estas condiciones el sistema nos va a devolver muy pocas referencias ya que las guías de viajes no se publican anualmente, y la mayoría de ellas no cubren una sola ciudad. Bajo ese criterio sería mejor utilizar las siguientes condiciones:

- tipo\_documento = "Guía" Y año\_publicación > 1990 Y tema = ("Castellón" O "Comunidad Valenciana").

En este caso la amplitud de fechas ya es demasiado grande puesto que aquellas guías publicadas a principios de los noventa ya estarán muy desfasadas. Además, el sistema nos devolverá referencias de guías de zonas de Valencia que no nos interesan por ser posiblemente demasiado amplias o alejadas. Por lo tanto, lo mejor es encontrar una formulación intermedia que no nos devuelva tantas referencias, como por ejemplo:

- tipo\_documento = "Guía de viajes" Y año\_publicación > 1998 Y tema = ("El Maestrazgo" O "Morella").

#### 4.1. Eficiencia de una consulta

En la siguiente figura se trata de ilustrar la relación que hay entre los documentos que hay en la base de datos adecuados a nuestras necesidades de información (relevantes), los que no deberían ser recuperados porque no interesan (no relevantes o irrelevantes) y los que se recuperan según las condiciones expresadas en la consulta. El objetivo es conseguir en la respuesta el máximo número de documentos relevantes, y el mínimo de documentos irrelevantes.

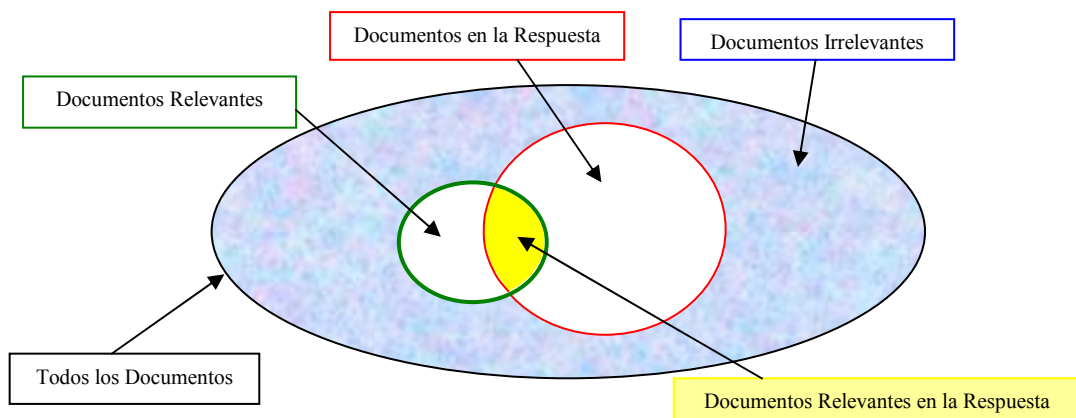


Figura 12. Grupos de documentos que definen la eficiencia de una consulta.



La eficiencia de una consulta indica la proporción que hay entre los documentos relevantes de la base de datos, los documentos relevantes que se recuperan en la consulta, y los documentos irrelevantes que se recuperan en la consulta. Aunque en la mayoría de los sistemas es imposible obtener un 100% de eficiencia, en un buscador con las suficientes herramientas de búsqueda, un usuario con habilidad para utilizarlas puede obtener eficiencias muy elevadas.

La falta de eficiencia de una consulta se debe a diversos factores, entre los que destacan la imposibilidad de utilizar determinados criterios de búsqueda en el buscador, la falta de precisión con que están descritos los contenidos del sistema, y la poca habilidad de los usuarios al formular sus consultas. Por lo tanto, es muy necesario conocer las claves del funcionamiento de cada sistema concreto, para así poder formular consultas y condiciones de búsqueda muy eficientes.

#### 4.2. Indicadores de relevancia

Algunos sistemas de recuperación de información presentan a los usuarios los documentos de las respuestas junto con unos números o porcentajes asociados. Estos valores, denominados indicadores de relevancia, se calculan durante la recuperación de los documentos utilizando métodos estadísticos, y sirven para indicar al usuario cuáles son probablemente los documentos que más le van a interesar en la respuesta.

En general, los documentos en los que las palabras clave especificadas aparezcan próximas entre sí tienen más posibilidades de ser relevantes que aquellos en los que aparezcan alejadas unas de las otras. Alternativamente, cuanto las palabras clave aparecen en un lugar estratégico, como el título o el resumen del documento, entonces es más probable que este documento sea más relevante que los demás en la respuesta. También suele ser más relevantes aquellos documentos en los que las palabras especificadas aparecen un mayor número de veces.

Páginas Web	Vista en orden de Google PageRank	Mostrar en orden alfabético
<a href="http://www.ciencia-ficcion.com/">Ciencia Ficción</a>	<a href="http://www.ciencia-ficcion.com/">http://www.ciencia-ficcion.com/</a>	Reúne la información que circula por los distintos grupos de noticias especializados.
<a href="http://www.terra.es/personal/p.agusti/hgwells.htm">La máquina del tiempo</a>	<a href="http://www.terra.es/personal/p.agusti/hgwells.htm">http://www.terra.es/personal/p.agusti/hgwells.htm</a>	Texto completo de la novela de H.G. Wells.
<a href="http://www.quintadimension.com/">QuintaDimension.com - Revista virtual de ciencia-ficcion, terror y fantasia</a>	<a href="http://www.quintadimension.com/">http://www.quintadimension.com/</a>	Revista Virtual sobre ciencia ficcion, terror, fantasía, comic y rol. Noticias, artículos, foros de dscusion, postales virtuales...
<a href="http://www.canalficcion.com">Canal ficción</a>	<a href="http://www.canalficcion.com">http://www.canalficcion.com</a>	Un portal de ciencia ficción, terror y superheroes. fantasía, superhéroes y dibujos animados.
<a href="http://personal3.iddeo.es/masters/KingKong/">King Kong rescata el Titanic</a>	<a href="http://personal3.iddeo.es/masters/KingKong/">http://personal3.iddeo.es/masters/KingKong/</a>	Relato de ciencia-ficción animado, con sonidos y fotografías, sobre el viaje de King Kong hasta el Titanic.
<a href="http://www.ttrantor.org">Términus trantor</a>	<a href="http://www.ttrantor.org">http://www.ttrantor.org</a>	Recopilación y base de datos literaria de ciencia ficción, fantasía y terror, con posibilidades de búsqueda por autor.
<a href="http://www.cienciaficcion.org/">Canal #Cienciaficción</a>	<a href="http://www.cienciaficcion.org/">http://www.cienciaficcion.org/</a>	Canal IRC, listas de correo, noticias y otras actividades relacionadas con la ciencia ficción.

Figura 13. Resultados ordenados por índice de relevancia de una búsqueda en el Directorio Google



Siguiendo estos criterios, los documentos de la respuesta pueden ordenarse por su indicador de relevancia cuando se presenten al usuario, tal y como puede verse en la figura 13. El tamaño de las barras verdes a la izquierda de cada página representa el grado de relevancia que le otorga el buscador a la página con respecto al tema buscado, en este caso “literatura de ciencia ficción”. Si la lista es larga, el sistema puede presentar sólo los más importantes. En otros casos el usuario puede especificar junto con las condiciones de la consulta, el índice mínimo de relevancia que deben tener los documentos recuperados.

#### **4.3. Frases o palabras clave**

En este caso la condición de recuperación consiste en una palabra o una frase que debe aparecer en el texto de los documentos recuperados. Sólo se recuperan los documentos que contengan una o varias ocurrencias de la palabra clave o de la frase especificada. Normalmente para indicar que queremos encontrar una frase y no varias palabras aisladas, es necesario escribirla entre comillas. Por ejemplo, los buscadores no interpretan igual "La guerra de las galaxias" que La guerra de las galaxias. En un apartado posterior veremos que es posible combinar varias palabras o frases para establecer condiciones de búsqueda más precisas, o incluso restringir la ubicación en el documento de las mismas.

El usuario debe elegir una o varias palabras que describan bien la información requerida, teniendo cuidado de que no sean demasiado genéricas. Esto significa que ciertas palabras no son adecuadas para describir la información de los documentos. Este es el caso de muchos adjetivos, adverbios, preposiciones, verbos, pronombres o algunos sustantivos como los números. La razón es que estas palabras no dicen nada acerca de los documentos requeridos ya que pueden aparecer en cualquier documento independientemente del tema que trate. En consecuencia, estas palabras devolverían un número elevado de documentos irrelevantes. A estas palabras se las llama *stop-words*.

Por ejemplo, palabras como *verde*, *antes*, *estos* o *siete*, dicen poco sobre el tema del que trata el documento que las contiene, mientras sustantivos como *papiroflexia*, *astrología*, *librería* o *diccionario* caracterizan mejor el contenido del documento.

No obstante, la aparición de determinada palabra o frase en todo el contenido de un documento, que puede tener una longitud considerable, no siempre identifica el tema sobre el que trata, sino que puede aparecer en un contexto totalmente distinto al que define el contenido del mismo. Por ejemplo, el hecho de que aparezcan las palabras *paella* y *plato* en un documento no significa que trate sobre la cocina valenciana, sino que puede tratar sobre OVNIS y describirlos usando esas mismas palabras.

#### **4.4. Condiciones sobre atributos y campos**

Algunos sistemas de información permiten especificar condiciones sobre ciertos atributos asociados a los documentos. Por ejemplo, es posible indicar cuál debe ser su autor o su fecha de publicación. En este caso, estos datos no deben aparecer

necesariamente en el texto de los documentos recuperados, sino que cada documento de la base de datos tiene asociados unos atributos que lo describe y sobre los que se evalúan estas condiciones.

- Autor= 'Juan Campos'
- Fecha-Publicación= '12/04/1992'

En otros casos, el texto de los documentos esta dividido en campos, de tal manera que es posible indicar sobre qué campo se quiere evaluar la condición. Por ejemplo, es posible indicar condiciones como que el documento contenga cierta palabra en su título, o en su resumen, etc.

En algunos buscadores de páginas web podemos indicar además condiciones sobre la ubicación del documento, restringiendo la información encontrada a un determinado servidor (*anubis.uji.es*) o a un dominio más amplio como *.com*, *.fr* o *.uji.es*.

Por ejemplo, el buscador Altavista permite restringir las búsquedas realizadas a determinados campos de la página web mediante las denominadas funciones especiales de búsqueda. Para ello se utilizan etiquetas especiales al definir el criterio de búsqueda para buscar títulos de páginas o dominios. Las etiquetas admitidas actualmente son:

<b>Etiqueta</b>	<b>Función</b>
<b>anchor:texto</b>	Busca aquellas páginas que contienen la palabra o expresiones especificada en el texto de un enlace. <code>Anchor:"garden.com"</code> buscará las páginas que contengan el enlace "garden.com".
<b>applet:clase</b>	Busca las páginas que contienen un subprograma Java específico. Utilice <code>applet:transformación</code> si desea encontrar páginas utilizando subprogramas llamados transformación.
<b>domain:nombre de dominio</b>	Busca las páginas que se encuentran dentro del dominio especificado. Utilice <code>domain:uk</code> si desea encontrar páginas del Reino Unido o <code>domain:com</code> si desea encontrar páginas de sitios comerciales.
<b>host:nombre</b>	Busca las páginas en un ordenador específico. La búsqueda <code>host:www.shopping.com</code> buscará las páginas en el ordenador Shopping.com y <code>host:dilbert.com</code> buscará las páginas en el ordenador llamado dilbert en unitedmedia.com.
<b>image:nombre de archivo</b>	Busca las páginas con imágenes que tengan un nombre de archivo específico. Utilice <code>image:playas</code> si desea encontrar páginas que contengan imágenes llamadas playas.
<b>like:texto URL</b>	Busca páginas parecidas a la URL específica o relacionadas con ella. Por ejemplo, <code>like:www.abebooks.com</code> busca sitios web que venden libros de segunda mano o libros poco comunes parecidos al sitio <code>www.abebooks.com</code> . <code>like:sfpl.lib.ca.us/</code> busca sitios de bibliotecas públicas o universitarias.
<b>link:texto URL</b>	Busca las páginas que contienen un vínculo a una página con el texto URL especificado. Utilice <code>link:www.zip2.com</code> si desea encontrar todas las páginas vinculadas a Zip2.com.
<b>text:texto</b>	Busca las páginas que contienen el texto especificado en cualquier parte de la página excepto una imagen, un enlace o una URL. El texto <code>text:graduación</code> buscará todas las páginas con el término graduación.
<b>title:texto</b>	Busca las páginas que contienen la palabra o expresiones especificada en el título de la página (en la mayoría de los navegadores aparece en la barra de título)
<b>url:texto</b>	Busca las páginas que contengan una palabra o expresiones específica en la URL. Utilice <code>url:zip2</code> si desea buscar todas las páginas en todos los servidores que contengan la palabra zip2 en el nombre de host, la ruta o el nombre del archivo, es decir, en toda la URL.

#### 4.5. Condiciones enlazadas con operadores

Los operadores nos permiten expresar condiciones más complejas que unas simples palabras o frases. De esta forma los usuarios pueden indicar a las herramientas de acceso más características de los documentos que desean recuperar, lo que contribuye a ajustar las respuestas a los documentos que realmente se necesitan. Cada operador tiene su significado, y se puede aplicar a un tipo de condición. Según su significado se clasifican en lógicos, aritméticos y textuales. Las condiciones con operadores aritméticos y textuales se pueden combinar con los operadores lógicos como veremos al final de este apartado.

##### 4.5.1. Operadores Aritméticos

Los operadores aritméticos se aplican sobre números y fechas almacenados en algún atributo o campo de los documentos. Son los siguientes:

$a > b$	a mayor que b
$a < b$	a menor que b
$a = b$	a igual a b
$a \geq b$ o $a \geq b$	a mayor o igual que b
$a \leq b$ o $a \leq b$	a menor o igual que b

Por ejemplo, se puede pedir que el año de publicación de un documento sea anterior a 2000: año < 2000.

##### 4.5.2. Operadores Textuales

Algunas herramientas de acceso permiten utilizar operadores textuales para indicar la posición relativa de las palabras entre sí, o para truncar las mismas. Los operadores textuales más comunes son los siguientes:

1. **Existencia.** Estos operadores permiten forzar la aparición o no aparición de una determinada palabra o frase en el documento. Podemos distinguir dos operadores de este tipo:
  - **Presencia.** Obligan a que la palabra indicada aparezca en el documento. Normalmente se antepone sin espacios el signo + a la palabra, p.e. +*cine* fuerza la aparición de la palabra *cine* en el documento.
  - **Ausencia.** La palabra indicada no debe aparecer en el documento. Normalmente se antepone sin espacios el signo - a la palabra, p.e. -*baloncesto* devuelve todos los documentos que no contienen esa palabra.

2. **Adyacencia.** Las dos palabras deben ser consecutivas: `fibra ADJ óptica`. Se recuperan los documentos en los que aparezca por lo menos una vez estas dos palabras consecutivamente.
3. **Proximidad.** Ambas palabras no deben distar en más de un número preestablecido de palabras: `fibra NEAR óptica`. En este caso, ambas palabras deben aparecer en los documentos pero además en posiciones cercanas. La mayoría de las veces no podrá haber más de 10 o 20 palabras entre ellas. De esta manera se puede intentar que las palabras aparezcan en la misma frase o párrafo.

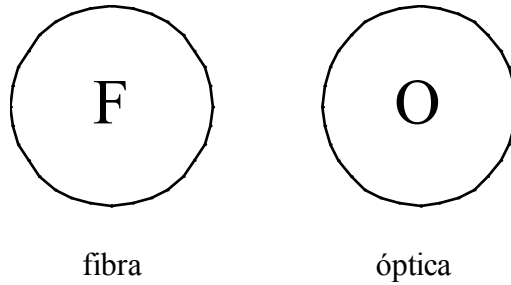
El hecho de que dos palabras aparezcan próximas entre sí permite suponer una mayor relación entre las mismas y con el tema tratado por el documento que si aparecen separadas en cualquier parte del mismo.

4. **Truncamiento.** Hay tres posibles operadores de truncamiento para texto:
  - **Prefijo.** Se buscará cualquier palabra que termine con el texto especificado. Por ejemplo, `*ordenador` buscará documentos en los que aparezca la palabra `ordenador, microordenador O multiordenador`.
  - **Sufijo.** Se buscará cualquier palabra que empiece con el texto especificado. Por ejemplo, `proces*` buscará documentos en los que aparezca la palabra `proceso, procesamiento O procesador`.
  - **Centro.** Se buscará cualquier palabra que empiece y termine con el texto especificado. Por ejemplo, `proces*s` buscará documentos en los que aparezca la palabra `procesos, procesamientos O procesadores`.

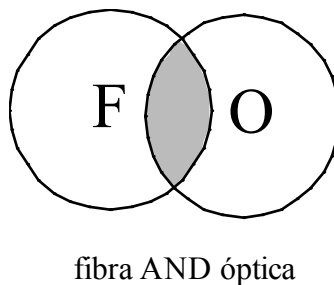
#### 4.5.3. Operadores Lógicos

Los operadores lógicos (también llamados operadores *booleanos*) más importantes son tres: AND, OR y NOT. Actualmente se encuentran disponibles en la mayoría de los sistemas de recuperación de información. No obstante, no todos los buscadores utilizan las mismas palabras o símbolos para estos operadores. Así, el operador lógico AND puede ser el signo + o incluso el símbolo &. En todo caso, la función de estos operadores consiste en enlazar condiciones simples y otras condiciones con operadores aritméticos y textuales, de tal manera que el usuario puede crear expresiones mucho más complejas y definir con mayor precisión las características de los documentos buscados. A continuación vamos a ver cómo se utilizan a través de un ejemplo.

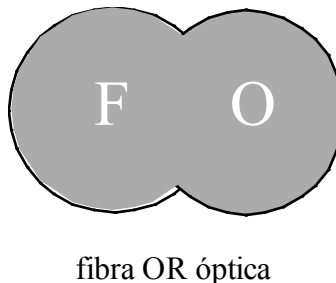
Supongamos que representamos de la siguiente manera el conjunto de documentos que contienen la palabra `fibra` (F) y el conjunto de documentos con la palabra `óptica` (O).



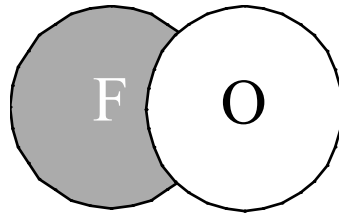
- El operador AND permite indicar que sólo se desea recuperar aquellos documentos que contengan ambas palabras. A veces también se simboliza con el operador  $\cap$ . Gráficamente se corresponde con los documentos que se encuentran en la intersección de ambos conjuntos.



- El operador OR permite indicar que se desea recuperar los documentos que tienen una de esas palabras (una o bien la otra, o bien ambas). A veces también se simboliza con el operador  $\cup$ . Gráficamente se trata de la unión de ambos conjuntos de documentos:

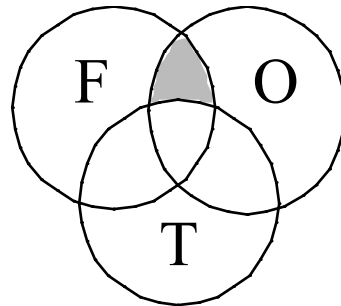


- El operador NOT permite eliminar de la respuesta los documentos que contengan cierta palabra. Es recomendable utilizarlo siempre junto con más condiciones AND, pero nunca sólo o con condiciones OR, ya que recuperaría una cantidad enorme de documentos de la base de datos que no cumplen esa condición. A veces también se simboliza con el operador  $-$ . Gráficamente se selecciona la parte oscura de la imagen.



fibra AND NOT óptica

Los operadores AND y OR son utilizados muy frecuentemente, mientras que el operador NOT se encuentra en los métodos de acceso en raras ocasiones, ya que es difícil de utilizar. La causa de esta dificultad está en que si pusiéramos la condición NOT *fibra* sin ir acompañada de ningún AND, se tratarían de recuperar todos los documentos que no contienen la palabra *fibra*, lo que abarcaría a la inmensa mayoría de los incluidos en el repositorio. Lo mejor es encadenar varios operadores lógicos separados con paréntesis como muestra el siguiente ejemplo:



(fibra AND optica) AND (NOT telecomunicaciones)

Hay que tener en cuenta que dependiendo de cómo se pongan los paréntesis, el significado de la consulta final puede cambiar. Para terminar la sección vamos a ver algunos ejemplos de cómo combinar los operadores aritméticos y textuales con los lógicos. Para poder describir nuestros requerimientos con precisión es importante aplicar bien el uso de los operadores y de los paréntesis.

- ('seguridad' ADJ 'social') AND fecha > 10/1/99 AND fecha < 15/1/99
- (\*ordenador' OR 'proces\*') NEAR ('automatic\*')
- (fecha > 1/1/99 AND fecha < 1/2/99) OR (fecha > 1/3/99 AND fecha < 1/4/99)
- (autor='Cela') OR (autor='Lorca') AND (NOT tipo\_libro='Teatro')

## 5. Formularios

Los formularios son un método de acceso que facilita mucho la búsqueda a sus usuarios. Por ejemplo, la figura 14 muestra el formulario de búsqueda avanzada incluido en el motor Altavista España. El usuario recibe una pantalla gráfica dividida en varios campos donde se solicita toda la información requerida para realizar la búsqueda. De esta manera es más sencillo indicar todas las condiciones de la búsqueda sin que se olvide ninguna, ya que cada campo se utiliza para especificar un tipo de condición. En algunos campos el usuario puede escribir cualquier valor coherente con la información solicitada, mientras que en otros casos sólo es posible escoger entre un número restringido de valores. Es este último caso, el usuario puede desplegar una lista de posibles valores para el campo y escoger uno.

Además de las condiciones sobre los documentos que se quieren extraer, los campos pueden solicitar otro tipo de información como contraseñas de acceso, datos personales o económicos. Normalmente los formularios también incluyen unos botones que sirven para lanzar la ejecución de cierta acción como puede ser ‘buscar’, ‘enviar’, ‘borrar’, ‘anular’, ‘ayuda’, etc.



The image shows the Altavista advanced search interface. At the top left is the Altavista logo. The main heading is "Búsqueda Web avanzada" with an "Ayuda" link on the right. Below this, there are two main sections: "Construya una consulta con..." and "Buscar con...".

The "Construya una consulta con..." section has four radio buttons and corresponding input fields: "todas estas palabras", "esta secuencia exacta", "cualquiera de estas palabras", and "y ninguna de las siguientes palabras". To the right of these fields is an "ENCONTRAR" button and a "Búsqueda básica" link.

The "Buscar con..." section has two radio buttons and input fields: "esta expresión booleana" and "ordenar por". To the right of the "esta expresión booleana" field, there is a note: "Utilice [términos](#) como AND, OR, AND NOT, NEAR. Las páginas que contienen estas palabras aparecerán en las primeras entradas del listado."

At the bottom of the form, there are two groups of radio buttons: "BUSCAR:" with options "Internacional" and "España" (selected), and "RESULTADOS EN:" with options "Todos los idiomas" and "español, inglés" (selected).

Below the radio buttons, there are two sections: "Fecha:" with options "por intervalos de tiempo:" (selected) and "por intervalo de fechas:". The "por intervalos de tiempo:" option has a dropdown menu showing "En cualquier momento". The "por intervalo de fechas:" option has two input fields for dates and the text "(dd/mm/aa)".

Finally, there is a "Tipo de archivo:" section with a dropdown menu showing "Todo tipo de archivos".

Figura 14. Formulario de búsqueda avanzada del motor Altavista España.



## 6. Lenguajes de interrogación

Un lenguaje de interrogación es un conjunto de órdenes o instrucciones que permiten comunicarse con el sistema para que seleccione los documentos de un repositorio. Por medio de estas instrucciones no sólo podemos especificar las condiciones de búsqueda, sino que también podemos manipular los resultados y llevar control de la evolución de la sesión de trabajo. Estos lenguajes son el método de búsqueda de información más antiguo, aunque su uso es recomendable sólo para los profesionales ya que requiere un proceso de aprendizaje mucho más complicado que el de los formularios. La razón de esto es que para utilizar un lenguaje de interrogación, primero se debe conocer una serie de órdenes con una sintaxis más o menos difícil, y el usuario no suele recibir ningún tipo de ayuda ni orientación interactiva durante su utilización.

<b>ORDEN</b>	<b>ACCIÓN</b>
<b>BEGIN:</b>	El usuario abre una sesión de trabajo en una base de datos.
<b>EXPAND:</b>	Lista el diccionario de palabras clave disponibles junto con su número de ocurrencias en los documentos almacenados. Las palabras con muchas ocurrencias son poco discriminantes, y viceversa.
<b>SELECT:</b>	Selecciona una respuesta formada por el conjunto de documentos que cumplen las condiciones que se indican a continuación.
<b>COMBINE:</b>	Realiza operaciones con los conjuntos de las respuestas (intersección (AND), unión (OR), resta (NOT), etc.)
<b>TYPE:</b>	Visualiza los documentos de la respuesta.
<b>DISPLAY SETS:</b>	Visualiza la secuencia de respuestas obtenidas hasta el momento. Cada consulta formulada tiene su respuesta asociada.
<b>PRINT:</b>	Imprime los documentos de una respuesta.
<b>SAVE:</b>	Graba la secuencia de respuestas obtenidas hasta el momento para que la podamos volver a utilizar en una sesión de trabajo futura.
<b>LOGOFF:</b>	El usuario cierra una sesión de trabajo en la base de datos.

Cada sistema de recuperación de información puede tener su propio lenguaje de interrogación pero todos ellos efectúan básicamente las mismas operaciones, ya que la estrategia de búsqueda se fundamenta en los operadores lógicos. La tabla anterior muestra algunos de los más comunes.

Como puede verse, el rango de operaciones que permiten los lenguajes de interrogación es mayor que el de los formularios, pero su utilización no es tan simple. Las diferencias entre diversos lenguajes de interrogación estriban en el procedimiento y en el nombre de las órdenes a emplear.

### **Ejemplo de uso del lenguaje anterior:**

Supongamos que queremos encontrar información sobre "el diseño de edificios en áreas con inestabilidad sísmica". Para ello utilizaremos la base de datos codificada como 8.

```

B   8
S   edificio OR construccion
      37513 edificio
      101101 construccion
S1  129716 edificio OR construccion
S   diseño
S2  393508 diseño
S   seism?
S3  17856 seism?
S   S1 AND S2 AND S3
      129716 S1
      392508 S2
      17856 S3
S4  1645 S1 AND S2 AND S3
S   S4 AND PY=1999
      1654 S4
      43069 PY=1999
S5  29 S4 AND PY=1999
T   S5/8/1

```

## 7. Una posible estrategia de búsqueda de información

Dado que no existe una estrategia única para buscar información, es importante que el usuario elabore una antes de acceder al sistema de recuperación de información. Buscar una estrategia consiste en seleccionar un conjunto de condiciones de búsqueda que recuperen un conjunto de documentos adecuados a las necesidades de información, no deben ser ni demasiados documentos (nos hacen perder tiempo y esfuerzo), ni pocos ya que se nos escaparía mucha información interesante. Durante el proceso de búsqueda, esta estrategia se deberá ir ajustando dependiendo de los resultados parciales obtenidos.

Antes de iniciar la consulta con el buscador podemos plantearnos algunas cuestiones previas:

En primer lugar intentaremos plantear la consulta a realizar en lenguaje natural, es decir, como si fuésemos a preguntar a un humano experto en el tema.

A continuación es necesario extraer los términos básicos de la consulta, esto es, las palabras más importantes que caracterizan los documentos a encontrar. Si seleccionamos conceptos ambiguos es posible que obtengamos demasiados resultados, muchos de ellos no relacionados con la información requerida. Por otro lado, si seleccionamos términos demasiado precisos, es posible que obtengamos muy pocos resultados, por lo que en estos casos puede ser interesante utilizar sinónimos o palabras relacionadas.

En tercer lugar podemos plantearnos si queremos restringir los términos buscados a algún campo en particular, tal como el título del documento, el resumen, el autor, etc.

En cuarto lugar nos podemos preguntar si existe alguna limitación temporal en los documentos deseados, es decir, si queremos que sean posteriores a una fecha dada o anteriores a la misma, o que se encuentren en un rango de fechas. En el mismo sentido podemos restringir los resultados obtenidos a documentos en un idioma determinado.

Una vez planteadas las cuestiones anteriores u otras relacionadas con la clarificación de la consulta a realizar, podemos pasar a utilizar la herramienta de búsqueda correspondiente.

El siguiente organigrama trata de dar una orientación acerca de cómo proceder durante la búsqueda de información.

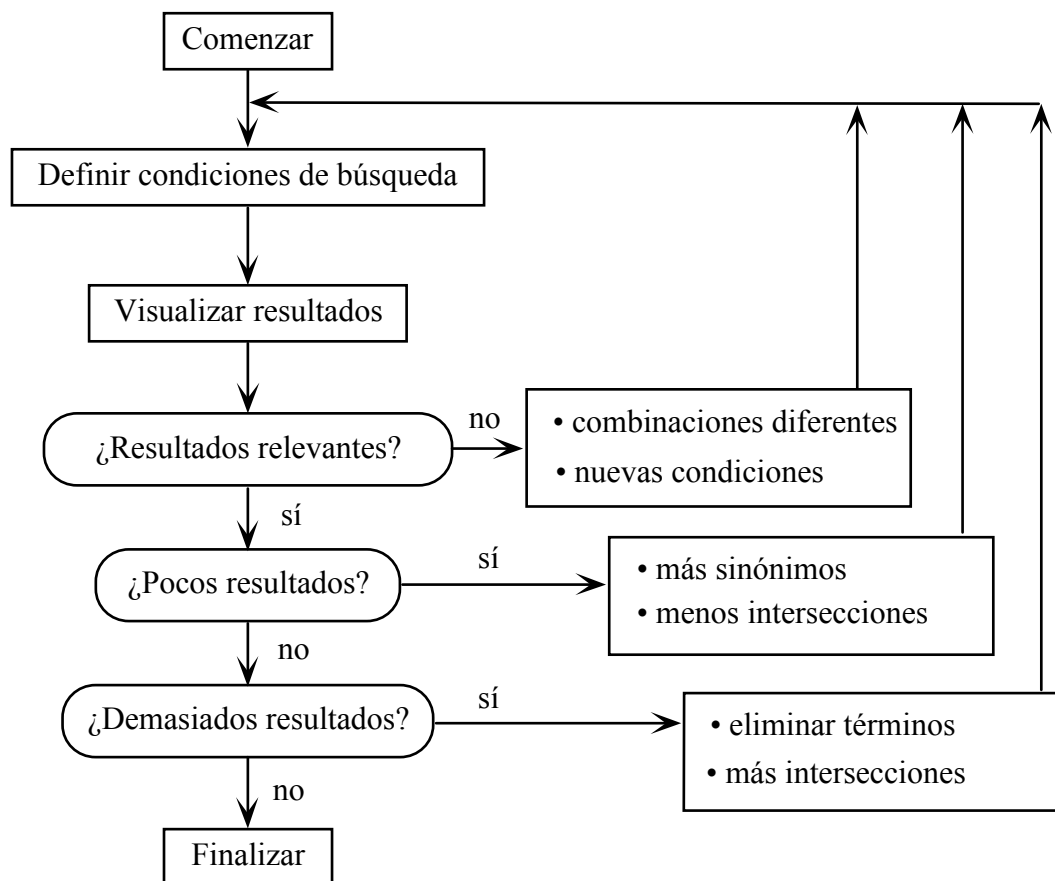


Figura 15. Esquema de una posible estrategia de búsqueda de documentos.

Existen otras técnicas generales que nos pueden permitir mejorar los resultados de nuestra búsqueda.

Si se obtienen pocos resultados

- Utilizar diferentes buscadores y acceder a buscadores que almacenan un mayor número de páginas.
- Eliminar cualquier restricción relativa a la localización geográfica de los documentos o a su idioma. De hecho, el número de páginas en inglés es mucho mayor que en cualquier otro idioma, por lo que puede resultar conveniente buscar páginas en inglés.
- Poner todas las palabras de búsqueda en minúsculas y sin acentos. La mayoría de los buscadores indexan los documentos utilizando palabras clave en minúsculas y sin acentos, por lo que si las buscamos de la misma forma es posible que podamos obtener más documentos.

Si se obtienen demasiados resultados

- Utilizar buscadores temáticos sobre la materia que queremos encontrar.
- Utilizar un buen catálogo e intentar encontrar una categoría relacionada con el tema deseado.
- Restringir la búsqueda usando frases en lugar de palabras sueltas o combinando varias de ellas con intersecciones (AND).
- Restringir la búsqueda a un determinado idioma, forzar que las palabras o frases aparezcan en el título.
- Restringir la búsqueda a un determinado servidor o a un dominio concreto.

## **8. Evaluando los resultados de las búsquedas**

Si hay algo que debería haber quedado claro a estas alturas del tema es la enorme cantidad de información que puede encontrarse en Internet. Sin embargo, una par de máximas que nunca deberíamos perder de vista son que “cantidad no es lo mismo que calidad” y que “no porque esté en internet tiene que ser verdad”.

En último término será la experiencia y el sentido común el que nos ayudará a determinar la veracidad y la calidad de la información encontrada. No obstante, transcribimos aquí el denominado criterio AAOCC (Authority, Accuracy, Objectivity, Currency and Coverage):

### **Authority (Autoridad)**

- ¿Quiénes son los autores? ¿Están cualificados? ¿Son creíbles?
- ¿De quién son afiliados? ¿Afecta su afiliación a su credibilidad?
- ¿Quién publica la información? ¿Cuál es su reputación?

### **Accuracy (Exactitud)**

- ¿Es la información exacta? ¿Es fiable y sin errores?
- ¿Son razonables las interpretaciones e implicaciones?
- ¿Hay evidencias que sustenten las conclusiones? ¿Se pueden verificar?
- ¿Citan los autores sus fuentes y referencias?

### **Objectivity (Objetividad)**

- ¿Cuál es el propósito de la información? ¿Qué pretenden conseguir los autores?
- ¿Afecta el propósito a la forma de la presentación?
- ¿Existe alguna predisposición implícita o explícita?

- ¿La información refleja hechos u opiniones?

**Currency (Actualidad)**

- ¿Es la información actual? ¿Todavía es válida?
- ¿Ha sido modificada por alguna investigación posterior?

**Coverage (Cobertura)**

- ¿Es la información relevante para nuestro tema o tarea?
- ¿A qué audiencia va dirigida?
- ¿Se presenta el material a un nivel adecuado?
- ¿Es completa la información? ¿Es exclusiva?