



# Cómo buscar y encontrar información en Internet

Álvaro Ibáñez  
(alvy@idg.es)

Se calcula que en Internet hay cerca de 60 millones de páginas web de información y que circulan por la Red unos 100.000 mensajes públicos diarios con todo tipo de referencias. Esto crea una situación en la que hay buenas y malas noticias: las buenas, que cada vez hay más información pública sobre todos los temas imaginables; las malas, que encontrarla no siempre es tan fácil como a muchas personas les gustaría, sobre todo a los principiantes. Nunca lo de “encontrar la aguja en el pajar” tuvo un equivalente más real, ni era posible imaginar un pajar tan grande, creciendo día a día y que se llena de agujas cada vez más y más interesantes.

Cuando se habla de Internet, se suele describir en primer lugar la gran cantidad de páginas con información disponible que hay en ella, pero la parte más complicada (las búsquedas) queda siempre en un segundo plano, como desagradable sorpresa para quienes se conectan por primera vez. Es el navegante el que tiene que ingeniárselas para dar con la información deseada. En muchas ocasiones, la información y la búsqueda en sí están llena de paradojas y situaciones extrañas, muchas veces frustrantes.

En cualquier caso, es normal cierta confusión entre los principiantes a la hora de buscar algo en Internet: la Red está llena de información valiosa, pero también hay muchísima información poco relevante (denominada “ruido”) que hace perder tiempo a la hora de buscar algo. Los propios sistemas de búsqueda no son perfectos, y el desconocimiento de las herramientas empeora las cosas.

El dicho popular “todo lo que puedas soñar que existe está en Internet” tiene mucho más de cierto que de falso, pero no dice nada sobre lo fácil o difícil que pueda ser encontrar lo que uno busca. Cualquiera que lleve tiempo circulando por la Red le podrá confirmar que seguramente existe eso que por razones de trabajo o afición está buscando, pero también debe advertirle que necesitará algo de tiempo para aprender a usar los buscadores y filtrar el ruido recibido hasta dar en la diana. Tal vez sea la temperatura de una remota ciudad, una descripción del penacho de Moctezuma o una foto de un carnero merino... La página más interesante sobre Kenneth Branagh, por ejemplo, puede estar escondida entre quinientas otras con referencias sobre el actor, sus películas, citas y referencias a Shakespeare, pero se puede terminar encontrando con un



poco de paciencia. [Incidentalmente: todos los ejemplos anteriores y los que verá en este artículo son reales, pero se necesitaron entre 2 minutos y una hora para encontrarlos en los lugares adecuados.]

Muchas veces, los usuarios terminan padeciendo lo que se denomina “angustia de la información”: hay demasiados datos revoloteando por el ciberespacio, imposibles de procesar (o encontrar) por un ser humano por mucho tiempo que le dedique al día, y la tarea de encontrar “lo mejor sobre...” se convierte en una misión imposible. En ocasiones esto hace que la persona se desespere al buscar una página o documento sobre un tema determinado, y que olvide lo más importante: dedicar previamente unas horas (o días) a comprender exactamente cómo funcionan los sistemas de búsqueda de Internet. Si se mira de otro modo, es como aprender el funcionamiento de un nuevo software para sacarle el mayor partido: todo el tiempo empleado en esta formación es extremadamente valioso e importante.

---

## UN POCO DE HISTORIA

Hasta la llegada de Internet, pocas personas consideraban factible poder localizar toda la información publicada sobre un tema de su interés en periódicos, revistas, agencias, notas de prensa e informes corporativos de empresas u organizaciones. Y mucho menos podían imaginar que algún día podrían acceder al instante a información detallada sobre esos temas en debates, y ver opiniones y aportaciones de otras personas con intereses similares a los suyos.

Con la llegada de Internet, todo esto ha cambiado: las posibilidades de la Red global han hecho que empresas y organizaciones se preocupen de hacer pública su información en forma de páginas web para promocionar productos, servicios o ideas, sean unas zapatillas deportivas, un programa político o un proyecto para la conservación ecológica de algún paraje remoto.

Por otra parte, la red Usenet (englobada en Internet) ofrece más de 13.000 grupos vivos de noticias y charla entre usuarios en los que quedan publicados miles de mensajes a diario de

todas partes del mundo, sobre cualquier tema imaginable: ciencia, política, series de televisión, discusiones técnicas y científicas y partidas de ajedrez, entre otros.

Afortunadamente, desde que la Red fue consciente de su crecimiento y de la dificultad a la hora de catalogar la información, aparecieron diversas empresas que basan su negocio en recopilar de un modo u otro toda esa información y hacer que esté a disposición de cualquier navegante.

Lo que verdaderamente ha cambiado con la llegada de Internet es que toda la información de la Red existe en formato digital y que miles de ordenadores de todo el mundo la almacenan públicamente. Gracias a ello, se puede buscar cualquier palabra o concepto y acceder a ella en cuestión de segundos. Lo mejor de todo es que no hace falta ningún programa especial: existen “buscadores de Internet”, en forma de páginas de la World Wide Web a las que se puede acceder con cualquier programa navegador desde un ordenador personal.

Con una nueva página publicada en algún punto de la World Wide Web de Internet cada dos o tres segundos, y más de un mensaje de texto por segundo en Usenet haciendo comentarios o dando una opinión, ningún ser humano podría mantener un catálogo de todo ello. Sin embargo, muchas empresas se dedican a realizar esta tarea clasificando la información más relevante de forma temática y manual: son los llamados índices de Internet, simbolizados por el precursor de todos ellos, Yahoo.

Con la llegada de nuevas tecnologías y mejores conexiones, también aparecieron nuevos sistemas más potentes que recopilaban toda la información de la World Wide Web y de Usenet: los motores de búsquedas. Encabezados por AltaVista, de Digital, los servicios de este tipo emplean “robots” inteligentes que saltan de una página a otra de la Web (a través de los enlaces de hipertexto) recogiendo páginas y almacenando toda la información en una gigantesca base de datos.

Cada sistema tiene sus ventajas y desventajas, y ninguno es perfecto ni completo ni está exento de fallos de precisión. Teniendo en cuenta lo

cambiante de la Red (se dice que una dirección URL tiene una vida media de menos de dos meses antes de que desaparezca o cambie de sitio), esto es inevitable. La propia idiosincrasia de los usuarios y organizaciones que publican la información hace que encontrar lo que se busca sea a veces una árdua tarea.

## ANTES DE EMPEZAR

Hay que dejar claras algunas premisas y conceptos antes de dedicarse a buscar información en Internet, que ayudan a hacerse una idea de la situación.

- **Es muy probable que exista en Internet la información que usted busca.** Esta es la premisa básica, que resulta cierta casi siempre. (Si no lo cree, haga una prueba en serio y sorpréndase con los resultados). Naturalmente, para comprobar hasta que punto es cierto hay que hacer un pequeño ejercicio mental al respecto: calcular qué posibilidades hay de que gente con intereses similares haya publicado esa información, si existen empresas, organismos o grupos que trabajen sobre ese tema en concreto, o si hay aficionados suficientes que compartan esa misma dedicación y hablen sobre ello o recopilen información. Algunos ejemplos: si usted busca una lista de películas en las que participe un actor determinado, a buen seguro que habrá diversas bases de datos de cine y aficionados que habrán recopilado esa información. Si necesita una fotografía de un animal, es factible que la pueda encontrar en algún archivo dedicado (sitios dedicados a la fotografía, archivos históricos, informes de biólogos, zoológicos, etc.) Si quiere ampliar información sobre un fenómeno astronómico para un trabajo de clase, seguramente encontrará información publicada por eminentes científicos, profesores y alumnos en cualquier universidad del mundo, o transcripciones de entrevistas, conferencias o artículos en revistas especializadas.

- **No todo el mundo publica toda la información en Internet.** Este primer problema hace referencia, especialmente, a entidades o empresas cuyos negocios se basan precisamente en vender información. El mejor ejemplo son los

periódicos: cualquiera de ellos puede leerse casi en su totalidad en Internet, pero en general no mantienen ni todos los archivos de su hemeroteca (tal vez, sólo una semana o un mes) ni la información tiene el mismo lujo de detalles (secciones, fotos, gráficos, actualizaciones, etc.) Con las editoriales que publican libros y enciclopedias sucede algo parecido: podrá consultar listas de publicaciones, tal vez leer algunos índices e incluso capítulos de algunos libros, pero no espere poder leer la totalidad de un libro en Internet. [Una curiosa excepción son los libros antiguos, libres de derechos de autor.]

- **No todo el mundo está en Internet, o no está al cien por cien, o no publica la información como se debería publicar.** Aunque cada vez son más las empresas y entidades en la Red, no todas están allí. Si usted busca información sobre un partido político español determinado, puede encontrarse con que todavía no tiene página web. O que la página web es puramente testimonial (sin contenidos interesantes). O incluso que publica sólo una pequeña parte de lo que podría mostrar en formato digital, a pesar de editar su programa electoral y emitir notas de prensa y comunicados. El Boletín Oficial del Estado, por ejemplo, publica únicamente los resúmenes diarios, no el boletín completo (que se debe comprar por separado). El Gobierno, por otro lado, publica todas las notas de prensa oficiales desde el web de La Moncloa, aunque no siempre están al día ni son de fácil acceso.

- **Prepárese a encontrar información en otros idiomas.** No se asuste: es lo más normal. La Red es un fenómeno mundial, y el castellano sólo supone una mínima parte (entre un uno y un tres por ciento, se calcula) del total. A menos que lo que usted busque sea algo realmente interesante y popular en España (y alguien lo haya tratado con dedicación) es probable que sólo lo encuentre en algún recóndito lugar y generalmente en inglés. Los interesantes documentos de preguntas y respuestas sobre temas concretos (FAQ) suelen estar sólo en inglés, como el resto de la documentación "oficial" de Internet. El hecho de que las universidades americanas estén mucho más avanzadas en todos los aspectos técnicos hace que tanto alumnos como profesores publiquen trabajos e informes, que naturalmente



están en inglés. El inglés es también el principal idioma de las agencias y los medios de comunicación escritos. Los web oficiales de las grandes multinacionales suelen estar en inglés, y las versiones españolas son generalmente pobres, con poca información y menos mantenimiento. Por lo tanto, esos datos astronómicos o esa explicación sobre una nueva técnica quirúrgica las encontrará seguramente sólo en inglés. Las biografías de toreros son (por pura lógica) más difíciles de encontrar aunque no sólo por el idioma: sencillamente, menos gente a nivel mundial está interesadas en ellas y dependerá de que alguien haya hecho ese trabajo en España o no y lo haya publicado en Internet (los documentos sobre Béisbol, en cambio, son mucho más numerosos). La única excepción a esto es que la información en castellano engloba no sólo a España sino también a todos los países de Latinoamérica más integrados en Internet, especialmente México, Argentina y Perú, donde suele haber datos interesantes.

• **No toda la información está en la red de la World Wide Web.** Aunque la WWW es la mayor red multimedia de la actualidad, y la de más fácil acceso, hay información que no está publicada en la Web, sino que debe buscarse en otros lugares de Internet, principalmente Usenet (la zona de debate a base de mensajes de texto, con más de 13.000 grupos internacionales). Esto incluye bases de datos especializadas (a las que a veces se puede acceder desde la Web y a veces no, y hay que usar otros programas clientes como Telnet, en modo texto) o redes privadas o comerciales (como las que hay a disposición de médicos, abogados o prensa). Si su trabajo gira en torno a una especialidad, lo mejor es localizar primero los principales sitios de la WWW que mantengan colegas de trabajo o investigación, y comenzar por ahí. En ocasiones, la información que se encuentra en las páginas web lleva a servidores de archivos (FTP) o listas de correo (e-mail) que generalmente “escapan” a las labores de catalogación de los buscadores. Ejemplo: si los datos de una empresa que usted busca están en las Páginas Amarillas, también estarán en la base de datos del mismo nombre que hay en la web ([www.paginas-amarillas.es](http://www.paginas-amarillas.es)), a la que se puede acceder con un navegador... pero el texto en sí (información, dirección, teléfono, etc.) no está almacenado en los buscadores, hay que lle-

gar a él a través de la página de esa base de datos.

• **No pida imposibles.** Alguien preguntó una vez “¿en qué lugar de Internet puedo encontrar una previsión meteorológica detallada de los próximos 30 días?”. Aunque la Red está llena de páginas con previsiones, fotografías de satélites, mapas y datos sobre temperaturas de todas las ciudades del mundo, la pregunta en sí es un imposible: nadie puede publicar esa información porque sencillamente el tiempo no es predecible a tan largo plazo. Piense dos veces antes de intentar buscar algo que no va a encontrar en Internet ni encontraría en la mejor biblioteca o archivo del mundo. No encontrará los números de la lotería de la próxima semana ni el sentido de la vida, aunque tal vez sí muchas teorías al respecto (sobre ambas cosas).

• **Prepárese a filtrar la información.** Parte de lo que hay en Internet son datos objetivos, pero una parte más importante todavía son opiniones. Cuando se localiza la información deseada, hay que usar un filtro racional para valorar si lo que ha encontrado tiene sentido: si se trata de un trabajo con datos objetivos, una teoría o simplemente un punto de vista. Por ejemplo: si usted investiga sobre la Alemania Nazi, se encontrará con páginas web perfectamente diseñadas y con todo lujo de detalles y referencias que explican que en realidad no murió ningún judío en los campos de concentración y que todo es una enorme conspiración (!). En cambio, otras páginas le llevarán a informes, documentos y libros que explican con precisión todo lo que sucedió durante la Segunda Guerra Mundial. Lo mismo sucede en los terrenos de la ciencia (especialmente en el de las nuevas teorías) y los humanísticos (con opiniones... sobre opiniones). No se crea nada a la primera, investigue un poco las fuentes tras haber localizado los documentos. Imagine que Internet es como recibir un libro que habla sobre los planetas, pero del que no sabe si estaba en la sección de Física o Astrología de la biblioteca.

• **No desespere.** En ocasiones no se encuentra la información que se busca, pero esa misma información puede aparecer con el paso del tiempo, a medida que más gente y entidades llegan a la Red. También se puede hacer una



pregunta en algún lugar de Internet (Usenet o las listas de correo) y esperar que alguien responda.

Las búsquedas en Internet son muchas veces cuestión de filosofía: hay navegantes que consideran que es incluso más divertido el hecho de buscar que el encontrar las cosas en sí, dado que durante el "viaje de búsqueda" se descubren sitios y lugares que algo tienen que ver con lo buscado aunque sea tangencialmente, pero que aportan algo que se había pasado por alto. En último caso, el consuelo de todo cibernauta: aunque lo que busque no esté en ese momento en Internet, tampoco lo estaría si la Red no existiera... y al menos puede que aparezca en el futuro. Y, aunque parezca obvio, en cualquier caso toda esa información es pública y gratuita.

## TIPOS DE BUSCADORES

Desde el punto de vista práctico, los buscadores de Internet se dividen principalmente en dos categorías: los índices y los motores de búsqueda. En los índices, hay personas detrás de ellos que clasifican inteligentemente las páginas web y los recursos de la red en categorías predeterminadas, a modo de gran árbol. Los motores de búsqueda son, en cambio, bases de datos que almacenan, una versión "reducida" de todas las páginas web y recursos que encuentran en la Red y que, aunque almacenan más información, son ligeramente menos precisos.

Otra forma de clasificar los buscadores es hacerlo por el lugar en que se encuentran. Se puede trabajar con buscadores globales, a nivel mundial, o con buscadores específicos de un país, como España. Actualmente hay más de una decena de índices españoles, que contienen información únicamente sobre recursos disponibles en España (o en castellano y otras lenguas oficiales), y también motores de búsqueda que recopilan únicamente información sobre páginas en castellano. Elegir uno u otro tipo de buscador (global o local) depende tanto de lo que cada cual puede procesar (información en varios idiomas, cantidad de páginas) como de lo interesante que puedan ser los resultados (ej. disposiciones administrativas, noticias locales o el teléfono

de un restaurante al que pedir una pizza... mejor cuanto más cerca.)

El número de buscadores de Internet aumenta cada día, de modo que sería imposible y poco práctico explicar el funcionamiento detallado de cada uno de ellos en un solo artículo. En vez de ello, y dado que todos operan de forma muy parecida, a continuación se explican dos de los más importantes (Yahoo y AltaVista) por ser en los que están basados todos los demás, y finalmente se darán algunos consejos para realizar las búsquedas que son válidos para éstos y otros buscadores con ligeras adaptaciones. También encontrará un cuadro con los principales buscadores de todos los tipos.

### índices: Yahoo

[www.yahoo.com](http://www.yahoo.com)

Nacido de la mano de dos estudiantes en abril de 1994 como una iniciativa para catalogar los recursos más atractivos de la World Wide Web (cuando la Web todavía era pequeña), Yahoo se ha convertido en una fructífera empresa que ofrece su servicio de forma gratuita a cambio de que el usuario vea algo de publicidad cada vez que accede para realizar una consulta. Las ventajas principales de Yahoo son su completitud, su sencilla y atractiva interfaz y su facilidad de uso.

Los empleados de Yahoo examinan personalmente páginas Web y recursos de todo el mundo y las incluyen en una "guía jerárquica universal." Una vez examinado el material, incorporan cada una de las páginas en una categoría predeterminada (con referencias cruzadas a otras si es necesario), hacen un pequeño resumen de su contenido y lo publican en el catálogo general, un árbol que recoge "todo lo que pueda existir" completamente organizado en más de 20.000 categorías y cientos de miles de páginas documentadas.

El árbol de Yahoo comienza con unas ramas de generalidades: Arte, Ciencia, Noticias, Informática, Juegos... y se despliega a medida que el usuario va avanzando pantallas: Arte lleva a Pintura, Escultura, Fotografía; Fotografía lleva a Fotógrafos, Biografías, Escuelas... y así sucesivamente.



*Yahoo es el índice por excelencia. Con más de 20.000 categorías, clasifica toda la información de los recursos de Internet (páginas web, grupos de noticias y sitios FTP) de forma estructurada e inteligente.*

Los punteros finales del árbol de categorías de Yahoo no son más que enlaces a recursos de Internet: grupos de noticias de Usenet, sitios FTP, y, principalmente, páginas web. Yahoo en sí no almacena esas páginas: simplemente proporciona los enlaces HTML adecuados. Un solo clic con el ratón es suficiente para saltar a ellos gracias a la magia del hipertexto de Internet.

Una primera forma de usar Yahoo es conectarse y navegar por el árbol de categorías (por ejemplo: Entretenimiento: Cine: Películas: Directores: Quentin Tarantino: Pulp Fiction) hasta que se encuentra lo que se desea. A la derecha de cada categoría aparece siempre un número entre paréntesis que indica cuantas referencias y subcategorías contienen las siguientes subcategorías.

Inteligentemente, Yahoo cataloga de forma prioritaria los FAQ (documentos de preguntas frecuentes) y los recursos que contienen índices de cualquier tipo. En el caso del Cine, por ejemplo, aparecen la espléndida Internet Movie Database, ([www.imdb.com](http://www.imdb.com)) y los FAQ de los grupos de noticias [art.rec.movies](mailto:art.rec.movies).\* (muchas preguntas comunes están contestadas allí). Cuando una categoría tiene el signo arroba (@) en su nombre quiere decir que se trata de una referencia cruzada a otra categoría similar que se encuentra en otro nivel de Yahoo.

Otra forma de navegar por Yahoo es buscar palabras clave que aparezcan en el nombre de

una categoría o en la descripción de los propios recursos que se están buscando. Esta búsqueda se puede hacer en todo Yahoo o sólo en una categoría determinada (de un nivel hacia abajo). El resultado: al buscar "Umberto Eco" (por ejemplo) aparecen páginas de categorías sobre el autor, en distintas áreas (Autores/Ficción literaria; Literatura/Semiótica) y también algunas de las páginas concretas, enlazadas hacia la WWW.

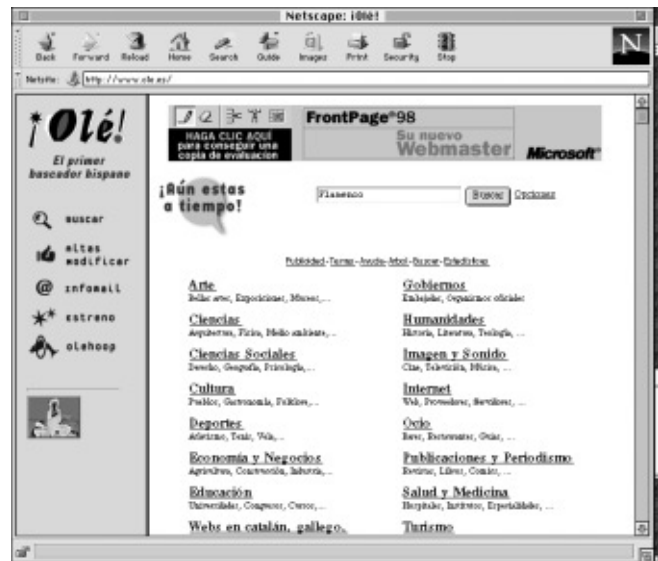
Yahoo ofrece algunos detalles más que son de interés: junto a las nuevas incorporaciones a la base de datos aparece un icono destacado ("New") durante unas semanas, así como otro icono ("Cool", unas gafas rojas) si el sitio es especialmente interesante. (Como en todas las clasificaciones de Internet, la valoración depende del gusto de los catalogadores, pero dado el gran trabajo que realizan en Yahoo, un icono de este tipo suele ser sinónimo de calidad).

Desde el menú principal de Yahoo también se puede acceder a noticias y novedades procedentes de agencias, dar saltos aleatorios por la Red en los momentos de aburrimientos o acceder a servicios personalizados, como el Yahoo para niños o el Yahoo personalizado ([my.yahoo.com](http://my.yahoo.com)), que se puede configurar según gustos personales.

Los consejos principales para usar Yahoo o un buscador de tipo índice son fáciles de recordar:



*Olé, el más popular de los índices españoles ha conseguido ser un lugar en el que encontrar con facilidad recursos sobre España y en castellano. Para dar de alta una nueva página basta conectarse a Olé y rellenar un formulario.*



- **Examine el árbol de categorías detalladamente para saber cómo está estructurado el índice.** Esto le ayudará a encontrar más fácilmente páginas sobre un tema determinado sin “irse por las ramas”. Dedique algo de tiempo a pedir las ampliaciones completas de los árboles de Yahoo (opción “Subcategory Listing” en cada sección) para hacerse una idea global del buscador.

- **Las páginas regionales están todas bajo la denominación “Regiones, Países” (en el caso de España, en [www.yahoo.com/regional/countries/spain](http://www.yahoo.com/regional/countries/spain), con subcategorías para autonomías, provincias, ciudades, etc.).** En otras palabras: es mucho más fácil encontrar los medios de comunicación españoles a partir de ahí que haciéndolo a través del índice general. A efectos prácticos, imagine que Yahoo guarda una especie de copia del árbol global bajo la cabecera de cada país.

- **Para una búsqueda rápida, considere la ventana de búsquedas e introduzca algunas palabras, pero en ese caso cuente con recibir algo de ruido.** La opción “búsquedas avanzadas” permite usar frases lógicas (como AltaVista), para forzar o eliminar palabras no deseadas. No desprecie tampoco la oportunidad de ampliar la búsqueda conectando con AltaVista desde Yahoo, una opción que se ofrece al final de cada búsqueda.

Los índices españoles [ver cuadro] tienen todos un funcionamiento muy similar a Yahoo, tanto en la estructura de sus árboles como en el método de búsquedas. La ventaja es que los recursos que almacenan suelen estar más detallados y en castellano. Algunos también ofrecen búsqueda mediante expresiones lógicas avanzadas. Por desgracia, en general resultan más lentos que los servidores internacionales, pues las máquinas sobre las que están situados tienen menos potencia y el estado de la red suele ser peor.

La paradoja de los índices es que sólo muestran información que previamente haya sido facilitada a sus administradores por algún usuario (quien creó una página o alguien de la Red). Si un sitio web no ha sido “dado de alta” en el índice, puede no aparecer nunca por mucho que se busque (a pesar de existir ahí fuera).

Decidirse por uno u otro índice (internacional o nacional, en sus distintos “sabores”) es cuestión de experimentar un poco y valorar los resultados. Si quien ha publicado la información en Internet ha sido cuidadoso, habrá incluido referencias en los principales índices españoles y también en Yahoo. Un problema de los índices españoles es que al ser tantos (y a veces inestables) los usuarios muchas veces no dan de alta sus páginas en todos ellos, sino sólo en los más populares. En general, si un recurso merece la pena, o si la información y el trabajo que contienen es espe-



*El motor de búsquedas y la base de datos de AltaVista pueden almacenar una copia casi completa de la World Wide Web y localizar cualquier palabra que haya sido escrita en cualquier remoto lugar. Su velocidad es sorprendente.*

cialmente interesante, los autores se habrán preocupado de darlo a conocer adecuadamente a través de Yahoo y el resto de buscadores.

**Motores de búsqueda: AltaVista**  
altavista.magallanes.net

La información existente en las páginas de Internet es tan voluminosa que a veces requiere otro tipo de buscadores más potentes: los “motores de búsqueda” de contenidos. Se trata de gigantescas bases de datos que almacenan un índice con todas las palabras que aparecen en cualquiera de las más de 60 millones de páginas de la WWW y de los 100.000 mensajes públicos diarios de Usenet. Esto incluye información como la de periódicos y revistas, bibliotecas, trabajos de investigación y, por supuesto, todo tipo de temas mundanos y de interés personal que la gente conectada a Internet publica de una forma u otra. Son una buena forma de localizar información si no se tiene muy claro en qué categoría puede enmarcarse, dónde pueden estar los recursos que la contengan o si el autor se ha molestado en comunicar su existencia a los buscadores.

AltaVista puede, por así decirlo, encontrar referencias a cualquier “palabra” que se encuentre en una página Web o mensaje público de Usenet, y dar acceso a ese recurso en cuestión de segundos mediante un enlace desde la pantalla del navegador. Para llevar a cabo esta tarea, que en el caso de AltaVista comenzó en diciembre de 1995 y que ya ha indizado más de 30 millones de páginas, Digital cuenta con diversas

tecnologías punta: un robot de búsqueda de páginas en Internet y motores de base de datos, indizado y clasificación; todo ello funcionando en servidores Alpha con varios gigabytes de RAM y discos RAID. El índice de la Web ocupa más de 40 GB, y el ancho de banda de los centros AltaVista supera los 100 Mbps. Los diversos “sitios espejo” de AltaVista (servidores regionales), repartidos por todo el mundo, reciben más de 10 millones de consultas cada día. España cuenta desde mediados de 1997 con uno de esos centros locales, con instrucciones en castellano y búsqueda por idiomas, gracias a un acuerdo entre AltaVista y Telefónica (altavista.magallanes.net).

De cara al usuario, el funcionamiento de un motor de búsqueda como AltaVista no puede ser más sencillo: basta teclear una frase y revisar los resultados: el motor de búsqueda analiza la base de datos en base a conceptos tales como cuántas veces aparecen las palabras buscadas, si el contexto es el mismo (cercanía), si aparece lo buscado en el título de la página o al principio de ésta, si son palabras comunes o no, etc.

Así, por ejemplo, usando la búsqueda simple (la que aparece al conectarse a la página) y tecleando una frase como “películas de Pedro Almodovar” aparecen al instante cientos de referencias, clasificadas de mejor a peor aproximación. Lo que hace AltaVista es buscar en su índice, extraer las páginas que contienen cualquiera de esas palabras o varias de ellas y presentarlas según un criterio de aciertos respecto a la frase





*Lycos, desde hace poco en castellano, se caracteriza por obtener un mayor número de aciertos significativos. Admite búsquedas lógicas (con las partículas and y not), y realiza derivación de palabras, como “metálico” a partir de “metal.”*



de búsqueda. Un clic con el ratón sobre los resultados dirige el navegador web a la página de destino.

La potencia de AltaVista es tremenda, pero, como todos los buscadores, no siempre acierta al cien por cien. Al buscar “naturaleza y medio ambiente en España”, por ejemplo, aparecen páginas sobre naturaleza y artículos de boletines y revistas (como “Estructura y función del biopolímero suberina”) que poco tienen que ver con lo que se busca, excepto que contienen alguna de las palabras.

A pesar de su aparente sencillez, AltaVista cuenta con opciones avanzadas para eliminar parte del “ruido” inherente a su método de almacenamiento. Esto es así porque si se busca únicamente la palabra “radio”, por ejemplo, aparecerá información tanto sobre emisoras de radio como sobre matemáticas (el radio de una circunferencia) o el radio como elemento químico: AltaVista no puede hacer mucho más... la palabra “radio” aparece más de 3 millones de veces en la WWW.

Entre lo que se podrían considerar “fallos” en una búsqueda (digamos, “naturaleza y medio ambiente en España”) estarían páginas que provienen de México, como el web de su Ministerio de Medio Ambiente (nada que ver con España) u otros fallos más graves se producen simplemente porque algunas páginas contienen de algún modo las palabras “medio” y “naturaleza”. Incluso si hubiera una página con un texto como “El cine se diferencia de la televisión por ser un medio...” sería encontrada y calificada como relativamente buena.

Por esta razón, las búsquedas se pueden realizar de forma más estricta, incluyendo algunas instrucciones especiales, para evitar que aparezcan páginas no deseadas y concretar lo que se está buscando. Hacerlo es tan sencillo como comprender la sintaxis de las frases de búsqueda y emplear ciertos operadores lógicos, mejorando así los “aciertos” de la base de datos.

Siguiendo con el ejemplo anterior, para encontrar páginas sobre protección del medio ambiente en España, se podría usar una frase de búsqueda como:

naturaleza “medio ambiente” protección + España -Mex\*

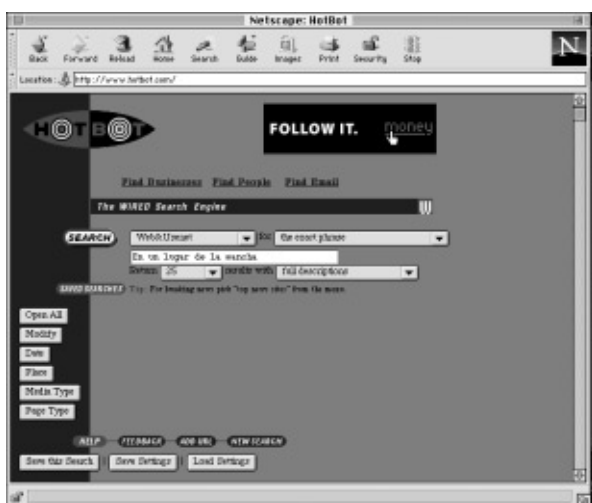
Los operadores básicos que se han usado en esta frase son:

**Comillas.** Permiten “pegar” varias palabras en una expresión, de modo que si en una página no aparecen exactamente, sino en otro orden (“el ambiente en el medio rural”) se desestime la información encontrada.

**Mayúsculas y minúsculas.** Al diferenciar mayúsculas y minúsculas se fuerza a que esas palabras aparezcan literalmente (ej. “España”).

**Incluida (+).** El signo más delante de una palabra obliga a que la palabra que va a continuación esté incluida en la información encontrada.

**Eliminar (-).** El signo menos elimina páginas que contengan la palabra que va a continuación. [Si se emplea este guión para un término que



*HotBot, asociado con la revista Wired y con motor de búsquedas Inktomi, está basado en superordenadores escalares e indiza automáticamente páginas de toda la Web, como AltaVista. Cuenta con más de 50 millones de páginas.*

contenga guiones, escriba las dos palabras sin separarlas con un espacio (“deja-vu”).]

**Raíz (\*).** El asterisco sirve para indicar raíces de palabras (por ejemplo, “méxico” o “mexicano”). En general, el asterisco reemplaza a una serie indeterminada de letras: “col\*r” es tanto “color” como “colour”.

Con la construcción del ejemplo (que no es la mejor, pero sirve para el propósito explicativo), se fuerza a que aparezca siempre la expresión literal “medio ambiente”, la información de España y se eliminan las referencias que procedan de México. Esto da lugar a un número más reducido de aciertos: artículos sobre España y el medio ambiente, catalogadores con información sobre naturaleza, el programa de algún partido político (que debe incluir su punto de vista al respecto), información sobre campos de trabajo, una asociación de Boy Scouts y una orden ministerial del BOE.

Una novedad reciente de AltaVista es la selección por idiomas, haciendo una selección previa en la ventana de búsquedas. Aunque es interesante, depende en gran medida del correcto diseño de las páginas HTML que han sido publicadas en la red (que identifiquen el idioma de lo que contienen o no, mediante un código HTML). Como muchas de ellas no incluyen los códigos de idioma adecuados, hay que usar esta opción con precaución para no perder páginas útiles, que AltaVista clasifica como de idioma “indeterminado”.

El modo de “búsquedas avanzadas” de AltaVista (al que se accede a través del icono del mismo nombre de la ventana principal) proporciona otro método de seleccionar la información. Puede acotar los aciertos por fechas, con los operadores AND y OR (similares a + y -) y con frases de búsqueda más complejas. Incluso se puede especificar la cantidad de palabras máximas de separación que debe haber entre dos palabras determinadas.

Una última forma de localizar información con AltaVista es dirigir la búsqueda a Usenet. La red de grupos de noticias y debate contiene información más al día que la Web, y es fácil localizar información de usuarios particulares haciendo preguntas y dando respuestas sobre un tema concreto.

Uno de los secretos de AltaVista, que queda fuera de la vista del usuario, es Scooter. El robot inteligente de Digital rastrea la Web constantemente, buscando enlaces en las páginas y saltando a ellos para consultarlos, indizarlos y buscar nuevos enlaces. Su trabajo no termina nunca, y es lo que permite que cada día se añadan y actualicen cientos de miles de nuevas páginas sin que los usuarios o administradores tengan que solicitarlo personalmente. Adicionalmente, AltaVista es tremendamente respetuoso y “ecológico” con la Red: Scooter viaja de página en página procurando no sobrecargar ningún servidor con consultas repetidas. Además, los usuarios pueden añadir código HTML a sus páginas personales o de empresa para evitar que sean indizadas



*DejaNews es un buscador específico de Usenet (grupos de noticias). Permite localizar cualquier mensaje publicado en Usenet por autor, tema, o palabras claves, desde hace muchos años hasta el momento actual.*



por el robot, en caso de que sean privadas. Scooter también actúa de forma inteligente: las páginas que más a menudo cambian (como las de noticias) son las más visitadas y actualizadas. También, automáticamente, las que no pueden ser revisitadas en varios días son eliminadas para evitar enlaces que no lleven a ningún lugar.

## CONSEJOS GENERALES

Tanto si usa motores de búsqueda como índices, hay algunos consejos relativos a las palabras o frases de búsqueda que pueden serle de utilidad. A continuación se indican algunos de los más generales, cuyo funcionamiento puede variar según cada buscador. Lo primero que debe hacer es leer y revisar con calma la Ayuda o el FAQ (Preguntas Frecuentes) de su buscador favorito para adaptar estos consejos o comprobar si funcionan del mismo modo en unos que en otros. Recuerde que leer las instrucciones del buscador durante unos minutos le ahorrará horas de búsquedas infructuosas y le permitirá aprender nuevos trucos.

- **Use minúsculas y no use tildes, al menos al principio.** Esto aumenta las posibilidades de encontrar una o varias palabras, dado que la mayor parte de los buscadores son sensibles a mayúsculas/minúsculas sólo si se emplean letras en mayúsculas, o a las tildes sólo si se escriben las palabras con tilde. Es decir: "leon"

dará por válidas palabras como "leon", "león", "León" y "LEON", mientras que "León" hace que deba aparecer la palabra exactamente así, en mayúsculas y acentuada.

- **Busque pistas y siga los enlaces.** Aunque una búsqueda parezca difícil porque lo que aparece no es profesional ni completo, como las páginas personales de muchos usuarios, no pierda los enlaces que ha encontrado y vaya guardándolos, por si necesita hacer una revisión exhaustiva. En muchas ocasiones, un pequeño enlace en una página personal puede darle palabras o términos interesantes sobre los que continuar la búsqueda, especialmente cuando se trata de páginas en otros idiomas.

- **Utilice primero los buscadores en español.** Si lo que busca puede estar disponible en castellano, comience por ahí, aunque la probabilidad sea pequeña. Tal vez un voto de confianza a los buscadores locales (índices y motores de búsqueda) le sorprenda por los resultados y le ahorre tiempo. Algunos de estos buscadores, como Olé ([www.ole.es](http://www.ole.es)), son lugares habituales en los que los internautas españoles dan de alta sus páginas, haciéndolas más fáciles de localizar. Otros, como Sol ([www.sol.es](http://www.sol.es)), buscan con gran precisión contenidos únicamente sobre España, aunque estén bajo otros dominios (.com, .org, etc.)

- **Recuerde buscar en inglés.** Si no encuentra lo que busca en castellano, comience a hacer las búsquedas en inglés. Necesitará saber



no solo la traducción exacta de las palabras de búsqueda, sino si tienen sentido en el contexto de la información que busca. Shakespeare es Shakespeare en todas partes, pero buscar información sobre el nacimiento de Jesús en Belén en inglés es más fácil si se usa “Bethlehem” como el nombre del pueblo; el genial Miguel Ángel tiene más referencias en inglés como “Michelangelo”, y Cristóbal Colón (sólo 100 en castellano) como “Christopher Columbus” (4.000 en inglés). Respecto al contexto, no se moleste en encontrar información sobre el viejo Volkswagen modelo escarabajo traduciendo literalmente la palabra “escarabajo”: la denominación internacional es “water-cooled Volkswagen” (la seria) o “Volkswagen bug” (la informal).

- **Vigile la ortografía.** Parece una tontería, pero muchas veces una búsqueda infructuosa se debe a errores de ortografía que pasan desapercibidos. Utilice los contadores de aciertos (por ejemplo, en AltaVista) que indican cuántas veces aparece cada palabra en la búsqueda: si aparece ninguna o muy pocas veces tal vez sea que ha cometido algún error al escribir el término. Esto es muy común al escribir nombres propios extranjeros (como “Tchaikovsky”). Tenga especial cuidado con las palabras en inglés: su ortografía no siempre es fácil ni “como suena”. Y no olvide que a veces hay diferencias entre el inglés americano y el británico: lo que en un lado del charco es “color”, “organize” y “elevator” en el otro es “colour”, “organise” y “lift”, por poner algunos ejemplos. En caso de duda, procure encontrar una raíz de la palabra y añada el asterisco (“Tchaik\*” sirve para “Tchaikovsky”).

- **Sea genérico.** Según la búsqueda, a veces conviene no usar nada más que las raíces de las palabras (“españ\*” para España, por ejemplo) para poder encontrar expresiones similares (españa, español, españolas), variaciones de la palabra en plural, género, tiempo de un verbo, etc.

- **Vaya al grano.** Si sabe cómo debe aparecer una expresión, tecléela como tal (encerrada entre comillas, con todas las palabras). Es mejor “Miguel de Cervantes” (si busca al célebre escritor) o “física cuántica” que “miguel cerv\*” (más genérica, pero que dará algunas sorpresas como páginas sobre Miguel que hablen de cerveza) o

“+ física + cuántica” (las palabras podrían aparecer no contiguas en el resultado).

- **Lea el FAQ de un tema concreto.** Intente localizar (si existe) el FAQ (Documento de Preguntas Frecuentes) del tema sobre el que está buscando información, sean juegos, películas, ciencia o política. Hay muchos documentos de este tipo de Internet, mantenidos por voluntarios. Aunque no encuentre a la primera la respuesta a su pregunta, hallará enlaces a sitios muy interesantes.

- **Opte por las páginas de calidad.** Las páginas de calidad tienden a ser más completas y estar desarrolladas por personas y grupos interesados en recopilar información de forma completa y cuidadosa, manteniéndola actualizada. Esté principalmente atento a las “páginas oficiales sobre...” (también las “no-oficiales”, que muchas veces son incluso mejores), las páginas de “recursos sobre...” y similares. Los sitios web bien diseñados tienen menos errores, cambian menos a menudo de sitio y en general son mejor referencia que las páginas personales o de aficionados, aunque siempre hay excepciones. Las páginas de calidad también son más mencionadas y mejor catalogadas por buscadores y páginas sobre temas similares.

- **Intente localizar empresas y organizaciones representativas.** ¿Quién mejor que Bandai ([www.bandai.com](http://www.bandai.com)) para responder a todas sus preguntas sobre los Tamagotchi? ¿O que la NASA ([www.nasa.gov](http://www.nasa.gov)) para encontrar las fotografías de Marte? ¿O que Greenpeace ([www.greenpeace.org](http://www.greenpeace.org)) para resolver dudas ecológicas? Cuando busque información y datos sobre productos, empresas u organizaciones, intente primero encontrar un web principal por su nombre de dominio. Recuerde que los de empresas españolas y universidades acaban en .es (España); los de empresas a nivel internacional en .com (Comercial); los de organizaciones sin ánimo de lucro en .org (organización); los del gobierno y organismos oficiales americanos en .gov (Gobierno) y los educativos (Americanos) en .edu (Educación). Las excepciones a esto son muchos servidores españoles (comerciales o no) acaban en .com (y no en .es). También conviene recordar que muchas entidades abrevian su nombre: El Ministerio de Educación y Cultura es [www.mec.es](http://www.mec.es) y la Universidad Autónoma de



Madrid, [www.uam.es](http://www.uam.es). En muchos casos, aunque la información no esté en las propias empresas o entidades, encontrará punteros a otros sitios representativos de la Red.

- **Limite los dominios y servidores.** Este truco sirve para AltaVista, y permite limitar el dominio (dirección de Internet) en que deben aparecer los resultados. Es útil para buscar información sólo dentro de una determinada entidad (empresa, universidad) o incluso un país. Una búsqueda con algunas palabras y además "+ host:\*.es" hace que sólo aparezcan páginas cuya dirección termine en .es (España). [Advertencia: hay páginas españolas alojadas bajo servidores terminados en .com, por lo que este método es sólo relativamente efectivo.] Ejemplo: `loto* loter* + host:*.es` muestra páginas sobre "loto" y "lotería" únicamente de España (evitando las de otros países).

- **Busque enlaces directamente.** También en AltaVista, la expresión "link:www.paginaweb.es" permite encontrar todas las páginas web que contengan enlaces a una página determinada. Es una forma rápida de encontrar sitios que apunten a una página que usted considere interesante, y generalmente esos sitios tendrán información de temas similares.

- **Piense en el resultado y adivine la frase correcta.** A veces es casi imposible dar con una frase que delimite correctamente lo que se está buscando: sencillamente aparecen demasiados aciertos se busque cómo se busque. Pruebe a pensar a la inversa: suponga que la página que usted busca realmente existe... ¿Cómo habrá escrito la persona que la ha creado la información que usted busca? Usando la lógica, a veces hay resultados sorprendentes. ¿Cómo llamaría a una página que explica cómo atarse el nudo de una corbata? Con la frase exacta entrecomillada ("How to tie a tie") puede encontrar esa y otras 53 páginas parecidas. "Erase una vez" le llevará a páginas de cuentos (muchas con relatos para niños), al igual que "Once upon a time" (en inglés, unas 14.000).

- **Si no lo encuentra, pregunte.** Si nada da resultado y no consigue encontrar lo que busca, utilice un último recurso: preguntar. Dese una vuelta por los grupos en castellano de Usenet, o por cualquiera de los grupos internacionales, busque el más adecuado para hacer su pre-

gunta y láncela. A buen seguro alguien le contestará con algunas referencias o al menos pistas. En este sentido son muy útiles las listas de correo que existen sobre los más diversos temas: aunque su contenido suele escapar a la acción de los buscadores, sus participantes suelen ser amables y cuentan con más datos y sentido común que los motores de búsquedas. [A este respecto, una dirección interesante: [www.findout.com](http://www.findout.com). Si formula aquí una pregunta (¡cualquier pregunta!) un grupo de voluntarios dedicará parte de su tiempo a investigarla para usted y darle una respuesta y referencias. ¿Se puede pedir más?]

Un consejo final: si no existe, invéntelo. Definitivamente, esto es sólo para los más emprendedores, pero es a la vez la mejor aportación que se puede hacer a la Red. Si no ha encontrado lo que busca, o sólo lo ha encontrado en inglés, considere publicar usted mismo esa información creándola desde cero (o traduciéndola) e incluyendo todos los enlaces posibles y referencias a temas relacionados. Recuerde siempre que la Red existe porque los usuarios la han llenado de contenidos y enlazado en una formidable maraña de enlaces, y que todo usuario debería aportar su granito de arena. Su página será visitada y otras personas podrán usar de la información que usted ha recopilado. Piense que les estará haciendo el mismo favor que le han hecho a usted todos los que de forma desinteresada publicaron los materiales que usted encontró en búsquedas anteriores. No está de más devolver el favor a la Red de vez en cuando.

## CONCLUSIONES

Este artículo ha pretendido servir a la vez de introducción a la búsqueda de información en Internet y a la vez de manual práctico con los trucos más comunes, indicando por dónde y cómo comenzar. Además de todo lo mencionado, existen muchos más recursos en la Red para buscar personas, direcciones de e-mail, teléfonos, empresas y la información disponible en miles de bases de datos... e incluso páginas que buscan en varios buscadores de Internet a la vez (los llamados "metabuscadore"). A buen seguro que irá descubriendo todas ellas a medida que circule por Internet.



Para acabar, recuerde siempre que todo el tiempo que emplee en investigar y leer sobre el funcionamiento de los buscadores de Internet, sobre las principales páginas de recursos de cada tema de interés y sobre el funcionamiento en general de la Red le supondrá un ahorro de tiempo enorme cuando vuelva a buscar información: aunque el pajar es grande, existen herramientas adecuadas para encontrar la aguja, por muy escondida que esté.

## BUSCADORES DE INTERNET

Este cuadro resume los principales buscadores a nivel internacional y también los que desde España proporcionan información y contenidos en castellano. Se ha dividido en índices (árboles de categorías) y Motores de búsqueda (buscadores de contenidos), aunque en algunos casos (como Excite o Lycos) el mismo buscador puede ofrecer ambos servicios bajo el mismo o distintos nombres.

También se han incluido algunas páginas de buscadores especiales (metabuscadors que realizan búsquedas en varios servicios a la vez) y temáticos (como DejaNews, orientado sólo a Usenet). En cualquier caso, esta clasificación debe entenderse no como algo exhaustivo, sino como un punto de partida de los lugares más populares por los que comenzar a buscar: cada día hay nuevos servicios similares en la World Wide Web. Usted mismo podrá valorarlos haciendo algunas pruebas y comprobando su precisión, velocidad y sencillez de uso.

## ÍNDICES

### *Internacionales*

Yahoo	www.yahoo.com
Excite	www.excite.com
InfoSeek	www.infoseek.com
a2z	a2z.lycos.com
Magellan	www.mckinley.com
Galaxy	galaxy.einet.net

### *En castellano*

Olé	www.ole.es
Ozú.com	www.ozu.com
Ozú.es	www.ozu.es
Navegador Intercom	www.intercom.es/navegador
¿Dónde?	donde.uji.es
El Índice	www.globalcom.es/indice
Enxebre (Galicia)	www.enxebre.com
Web Tour	www.webtour.net
Inspector de Telépolis	www.telepolis.com
Infor-Acceso	www.infor.es
Eliana	www.eliana.com

## MOTORES DE BÚSQUEDA

### *Internacionales*

AltaVista	altavista.digital.com
Lycos	www.lycos.com
HotBot	www.hotbot.com
WebCrawler	www.webcrawler.com
HotBot	www.hotbot.com

### *En castellano*

AltaVista	altavista.magallanes.net
Lycos	www-es.lycos.com
Sol	www.sol.es
Trovator	trovator.combios.es
Hispavista	www.hispavista.com
Ugabula	www.ugabula.com
Inspector de Telépolis	www.telepolis.com

## VARIOS

MetaCrawler	www.metacrawler.com
Ciudad Robot	www.ciudadrobot.com
Search.com	www.search.com
Starting Point	www.stpt.com
DejaNews	www.dejanews.com
NewsBot	www.newbot.com

*Álvaro Ibáñez (alvy@idg.es; bbs.seker.es/~alvy) es Director de iWorld, la revista de Internet de IDG Communications. Este artículo está disponible en Internet para su consulta en la dirección [www.idg.es/iworld/especial/buscar.html](http://www.idg.es/iworld/especial/buscar.html).*