
LOCALIZACIÓN DE INFORMACIÓN EN MOTORES DE BÚSQUEDA EN INTERNET. ANÁLISIS DE LA EFECTIVIDAD.

.....
JOSÉ LUIS SALMERÓN SILVERA
SALVADOR BUENO ÁVILA

Departamento de Economía y Empresa
Universidad Pablo de Olavide de Sevilla

PAULA LUNA HUERTAS

Departamento de Economía Financiera y Dirección de Operaciones
Universidad de Sevilla

CON LA EXCEPCIONAL EXPANSIÓN QUE HA SUFRIDO INTERNET EN LOS ÚLTIMOS AÑOS, SE HA PUESTO A DISPOSICIÓN DE LOS USUARIOS UN INGENTE VOLUMEN DE INFORMACIÓN. EL TAMAÑO DE LA WEB ES GIGANTESCO, Y CONTINÚA CRE-

173

ciendo según un modelo exponencial (Kobayashi *et al.*, 2000, p. 148). Por ello, uno de los grandes retos profesionales de la actualidad, dado la inmensidad de la red y su rápido crecimiento, es conocer el método idóneo para localizar la información que deseamos rápidamente (Lang, 2001, p. 44).

Webster y Paul (1996) definen el acto de navegar como buscar sin ayudas. Se trata de una acción reservada a aquellas personas con disponibilidad de tiempo, pero no es adecuada para aquellas que necesitan una información específica rápida-

mente. Por esta razón, ha surgido un importante número de herramientas para buscar y descubrir la información depositada en la red más eficientemente.

Por tanto, estos buscadores surgen como herramientas de ayuda para buscar en la Web en respuesta a la dificultad, por parte del hombre, de mantener un catálogo con toda la información que contiene la red (Dassler, 2001: p. 123), debido fundamentalmente a la volatilidad de ésta (Ibáñez, 1997) por su volumen. De ahí que las herramientas de búsqueda no pueden encontrarlo todo, por el simple hecho de

no poder abarcar la *web* en su totalidad (Fisher *et al.*, 2000, p. 43).

Según la herramienta de búsqueda Inktomi (<http://www.inktomi.com/>), en febrero de 2000 se estimaba la presencia de mil millones de páginas web (Dassler, 2001, p. 122). Si, además, Sullivan afirma que las herramientas de búsqueda sólo son capaces de abarcar el 35% de toda la WWW (Dassler, 2001, p. 122), podemos hacer una aproximación del tamaño de la red.

Uno de los estudios donde se pudo demostrar la efectividad de los buscadores es

páginas de sus bases de datos y las palabras clave, para posteriormente devolverle al usuario los resultados (Clarke *et al.*, 2000, pp. 291-293).

Estos buscadores disponen de un pequeño lenguaje que permite personalizar las búsquedas utilizando diversos operadores booleanos y realizando con ellos, búsquedas avanzadas, pudiendo, asimismo, limitar los aciertos por fechas, según la localización geográfica de las páginas, amén de otras opciones.

METODOLOGÍA PARA UTILIZAR LAS HERRAMIENTAS DE BÚSQUEDA

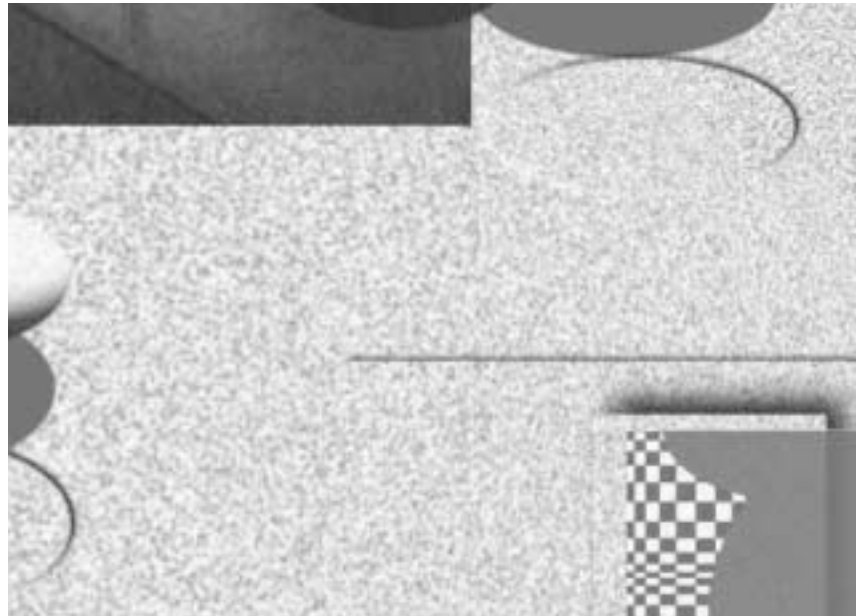
No existe una metodología de búsqueda ideal. Su continuo uso es el mejor aprendizaje, sin embargo proponemos una metodología básica y bastante eficaz, desarrollada a partir de nuestra experiencia profesional. Ésta se divide en cuatro pasos, y son los siguientes:

DEFINICIÓN DE LA BÚSQUEDA

Es fundamental comenzar acotando la búsqueda a realizar tan estrechamente como sea posible, identificando los parámetros siguientes: el tema de búsqueda; los límites geográficos, con objeto de elegir el tipo de buscador (internacional, nacional o regional); el idioma; los límites temporales; el objetivo de la búsqueda; los conceptos claves que permiten identificar esa información; el tipo de fuente donde puede encontrarse (página Web, grupos de noticias, bases de datos, etc.), entre otros.

Concretamente, para identificar el área temática y los conceptos claves, se deberá utilizar una terminología que no sea ambigua, revisar los posibles sinónimos buscando palabras que realmente definan el concepto y traducirlos al inglés en el caso de usar buscadores internacionales.

Así mismo, debemos identificar todos los elementos antes de realizar la búsqueda, porque a menudo, la parte más difícil de una búsqueda consiste en saber cómo



otras personas pueden haber denominado el objeto buscado.

DEFINICIÓN DE UNA O MÁS ESTRATEGIAS DE BÚSQUEDA

Una vez definido el objetivo de la búsqueda, se deberán analizar las distintas estrategias de búsqueda y ordenarlas según la esperanza de encontrar objetivos. Estas estrategias dependerán del objeto a buscar, pero a *grosso modo* apuntamos una serie de estrategias genéricas en el orden en que deberían analizarse.

Si el área temática a la que pertenece el objeto de la búsqueda está claramente identificada, en primer lugar puede acudir a una o más herramientas de búsqueda por índice temático, del tipo de Yahoo!, o de las especializadas en ese tema, si es que existen. Ese tipo de herramientas permitirá obtener información genérica sobre el objeto, es decir, nos ofrecerá una orientación de los posibles resultados.

Posteriormente puede utilizarse alguna herramienta de búsqueda por contenido o motores de búsqueda, del tipo de AltaVista, y más concretamente de su opción de búsqueda avanzada, para obtener información más específica y actualizada sobre el objeto en cuestión. Como complemento de las dos estrate-

gias anteriores se pueden utilizar uno o más Metabuscaadores que realicen la búsqueda simultáneamente con distintas herramientas (motores de búsqueda, directorios y buscadores especializados), teniendo en cuenta que estos buscadores en paralelo no suelen tener acceso a toda la potencialidad que esas herramientas ofrecen individualmente, lo que puede dar lugar a búsquedas menos precisas que si se hubiese utilizado en cada herramienta por separado (Dassler, 2001, p. 128).

Por último, si fuese necesario para completar la información ya obtenida, se podrían utilizar otros recursos de Internet, tales como los servidores FTP, Archie, Bases de datos, etc.

Una vez que sabemos que los buscadores proporcionan importantes ventajas ante la búsqueda sería interesante analizar cuál es la estrategia de búsqueda que prefieren los usuarios. Con respecto a este tema, White *et al.* (2000) desarrollaron un estudio sobre la estrategia que se adopta al iniciar la búsqueda entre un conjunto de tres posibilidades: directorios, motores de búsqueda y buscar directamente en las páginas Web. Se realizó el estudio sobre una población de 54 miembros familiarizados con estas estrategias de búsqueda. El trasfondo era identificar la estrategia inicial según la naturaleza de las preguntas planteadas

(cerrada/predecible, cerrada impredecible, abierta/predecible, cerrada/predecible). La elección de una u otra estrategia está influenciada por la cantidad de información o el número de sitios que tenían que ser buscados, la probabilidad de relevancia y el esfuerzo que se debía realizar para diseñar la pregunta con la que realizar la búsqueda.

Los resultados fueron los siguientes:

✓ El 80% utilizaba las tres estrategias, un 19% dos de ellas y tan sólo un 2% desarrollaba una. En cómputo general, para buscar información sobre las preguntas, el 43% optó como primera opción por los motores de búsqueda, un 30% accedió directamente a las páginas y el resto acudió a los directorios.

✓ Para las cuestiones de origen impredecibles se utilizaron los motores de búsqueda entre un 60%-66% de los casos. El uso de los directorios estaba en segundo lugar, pero más pronunciado en el caso de las cuestiones abiertas.

✓ Para las cuestiones predecibles, la estrategia preferida era buscar directamente en las direcciones. Si eran abiertas, la siguiente opción era los motores de búsqueda. Para el caso de preguntas cerradas, la elección como segunda estrategia era igual tanto para los directorios como para los motores de búsqueda.

Esto nos pone de manifiesto que ante búsquedas de información predecibles (se saben donde están ubicadas) es mejor acudir directamente a las páginas. Por el contrario, para cuestiones impredecibles, es mejor optar por los motores de búsqueda.

DISEÑO DE LA(S) SOLICITUD(ES) DE BÚSQUEDA PARA CADA ESTRATEGIA

Para cada estrategia de búsqueda planteada en el paso anterior deberá diseñarse la solicitud de búsqueda de acuerdo a las posibilidades de la herramienta a utilizar. Para ello, se partirá de cada uno de los conceptos que conforman el tema sobre el que se quiere buscar información con nuestro lenguaje, que se denominan palabras clave.



Aunque se puede utilizar una sola palabra para definir una búsqueda, el usuario tiende a incluir al menos dos, con el fin de incrementar la relevancia de los sitios localizados (Fisher *et al.*, 2000, p. 45). Los términos de la búsqueda (palabras clave) pueden ser, fundamentalmente:

- Una o más palabras sueltas.
- Una o más frases (grupo de palabras que deberán aparecer seguidas), que pueden resultar muy útil para localizar un determinado texto.
- Cualquier combinación de palabras sueltas y frases.

Como paso siguiente se buscarán los posibles sinónimos o términos que se relacionen con cada uno de los conceptos o palabras claves.

Existen unos buscadores especializados en la búsqueda de palabras clave relacionadas, como, por ejemplo, el buscador en inglés que reside en la dirección <http://www.jimtools.com/keywords/index.html>.

Hay que tener cuidado con el buscador en cuestión a la hora de escribir las palabras clave en mayúsculas o en minúsculas, ya que en algunos casos resulta indiferente pero en otros no. Algo similar ocurre con las tildes de los acentos.

En casi todos los buscadores, si las palabras sueltas y/o frases se separan por espacios en blanco y no se enlazan con ningún operador, se estará indicando que se muestren aquellos resultados en donde aparece cualquiera de esas palabras y/o frases.

En la mayoría de las herramientas de búsqueda por índice y en todas las herramientas de búsqueda por contenido podrán utilizarse para acotar o ampliar la búsqueda ciertos operadores. Con estos operadores booleanos se procederá de la siguiente manera: una vez que se tenga la lista con todos los términos o palabras clave, se establecerán las relaciones lógicas que se darán entre ellos. Para ello, se utiliza el operador booleano OR para relacionar sinónimos o conceptos relacionados, AND para relacionar términos que corresponden a conceptos distintos, pero que queremos que aparezca en el mismo documento y AND NOT para eliminar los documentos que contengan un término no deseado. Se utilizarán los paréntesis para indicar la operación a efectuar en primer lugar, normalmente se utilizan para ejecutar en primer lugar una unión (OR), y su resultado combinarlo con otro término mediante AND. También se deberá utilizar el carácter * para localizar una subcadena de caracteres, y para recuperar palabras en singular y en plural además de adverbios.

Así, por ejemplo, si se desea buscar las subvenciones que existen para las empresas de reducidas dimensiones, la ecuación de búsqueda que se puede construir es: (ayuda* OR subvención* OR financiación*) AND (empresa* OR pyme OR «pequeña y mediana empresa»).

También en algunas herramientas de búsqueda se puede utilizar determinada sintaxis para restringir la búsqueda a determinados campos de información, por ejemplo: a los títulos de las páginas *Web* (title:), a los URL (url:), a los dominios (host:), a los hipervínculos (link:), a los nombres de imágenes (image:), etc.

Para ello, se deberán registrar los posibles resultados obtenidos con cada una de ellas. Éstos deberán analizarse leyendo el resumen que se ofrece para ver si lo que nos interesa se encuentra en ese lugar; los primeros siempre suelen ser las coinciden-

cias que muestran un mayor parecido con la búsqueda.

APLICACIÓN DE LAS DIFERENTES ESTRATEGIAS DISEÑADAS

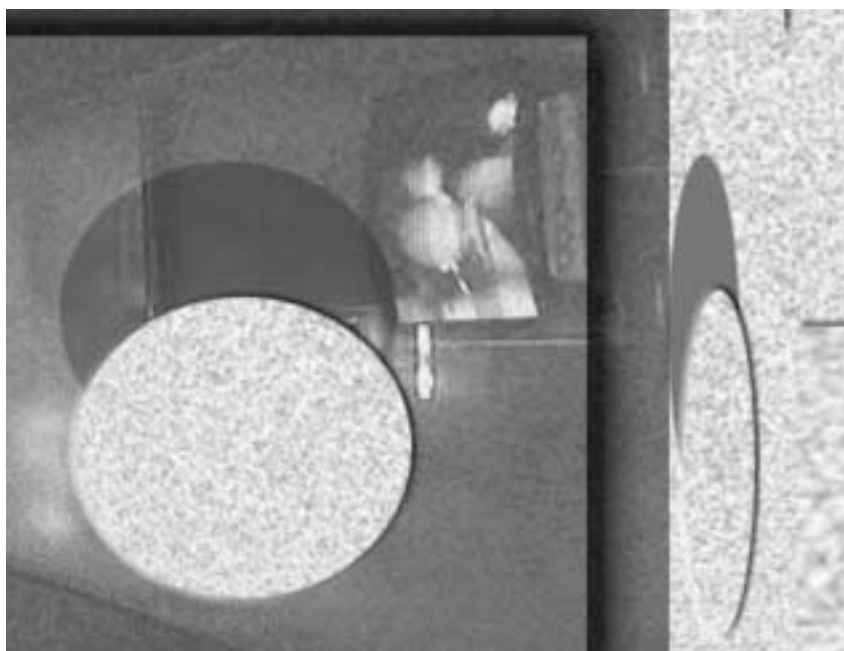
A partir de los primeros resultados obtenidos en la búsqueda se podrán obtener otros términos que puedan utilizarse para acotar o ampliar la lista de coincidencias con la búsqueda, siendo recomendable volver a realizar el 3º y el 4º paso.

Los navegadores permiten grabar estos resultados, al igual que cualquier otra página Web, en un fichero con formato de texto ASCII, que podrá ser leído con cualquier aplicación informática o en su formato original HTML, que también puede ser leído por las últimas versiones de los procesadores de texto. También existen programas, generalmente *shareware*, especialmente diseñados para bajarse de la red páginas Web con todos sus elementos multimedia y las que enlazan con ella, con objeto de imprimirlas en otro momento, conectar con algunas de sus conexiones sin necesidad de volver a ellas en línea o cualquier otra operación. Algunos de estos *software* son: WebWhacker, Webstripper, WebVcr, Teletport Pro y MemoWeb (estos programas *shareware* se pueden buscar en la red a través de buscadores especializados en localizar *software*).

Además, recomendamos utilizar la lista de marcadores (Bookmark) del navegador Netscape o la agenda de páginas favoritas del navegador Explorer o la lista interactiva (Hotlist) de Mosaic para guardar aquellas direcciones (URL) que nos interesen, evitando así tener que memorizarlas y reescribirlas.

FACTORES CLAVES EN LA BÚSQUEDA

Uno de los factores fundamentales para conseguir éxito en la búsqueda, entendiendo por tal encontrar la información deseada en el menor tiempo posible, es conocer el correcto uso de las herramientas de búsqueda. El tiempo empleado en comprender estas herramientas se asemeja a aprender el funcionamiento de un



nuevo *software*. Todo este tiempo de formación es extremadamente valioso e importante (Ibáñez, 1997).

Antes de nada, tenemos que tener en cuenta que cada herramienta de búsqueda, a pesar de que son muy parecidas, tiene su propio funcionamiento. Por ello, es recomendable leer siempre la ayuda que incluye el buscador y probar distintos buscadores, utilizando siempre que se pueda la opción de búsqueda avanzada (donde se utilizan los operadores booleanos y de posición), que se encuentra en la totalidad de los buscadores por contenidos y en algunos de los buscadores por índice, pues con ella se conseguirá resultados más ajustados.

Otra de las cuestiones más importantes es el idioma. Es aconsejable utilizar primero los buscadores en el idioma deseado, si lo que busca puede estar disponible en éste. Si no lo encuentra se deberá acudir a buscadores en inglés, que es el idioma predominante en la red. Si no entiende suficientemente una página Web en inglés o en otro idioma recuerde que en la red puede encontrar múltiples servicios donde traducen de manera gratuita el contenido de la página Web que se indique. Por ejemplo, algunos de estos sitios están en las siguientes direcciones: <http://www.freetranslation.com/>, <http://www.elmundo.es/traductor>, [\[do .es/traductor/\]\(http://www.systransoft.com/\) y <http://www.systransoft.com/>](http://www.elmun-</p>
</div>
<div data-bbox=)

Con el fin de que tengamos la garantía de que el buscador que hemos elegido contenga la información que buscamos es preciso conocer el campo de información que abarca el buscador (páginas Web, mensajes de grupos de noticias, etc.).

Otro de los factores clave para realizar una correcta búsqueda es cuidar la ortografía, ya que, por ejemplo, no es lo mismo escribir CDROM o CD-ROM. Los resultados que se obtendrán de escribirlo de una u otra forma serán distintos. Lo mismo ocurre con las mayúsculas y minúsculas, pues en algunos casos resulta indiferente pero en otros no.

También es el caso de las tildes de los acentos, ya que algunos buscadores serán sensibles y otros no. Para ello es recomendable utilizar OR, o los operadores de familias de palabras ?, * o \$.

En el caso de que se haya obtenido mucho ruido (resultados que no buscábamos) puede ser útil restringir la búsqueda a determinados campos de información: a los títulos de las páginas Web (t:), a los URLs (u:), a los dominios (h:), a los nombres de imágenes (i:), etc. Siempre es recomendable verificar la fiabilidad de la información encontrada, comprobando,

por ejemplo, su autoría, actualidad, propósito, etc.

Por último cabe comentar cómo en sitios Web del tipo de <http://www.learnthenet.com/spanish/html/19tips.htm> y <http://coverage.cnet.com/Content/Features/Dlife/Search/ss06.html> se recogen consejos para agilizar la búsqueda de información en Internet y en <http://www.searchenginewatch.com/> se pueden encontrar muchos trucos para perfeccionar las búsquedas en los principales buscadores internacionales.

OPERADORES DE CONSULTA

Los motores de búsqueda se caracterizan por la capacidad de realizar consultas mucho más específicas gracias a la utilización de los operadores de consulta. Cuando se realizan búsquedas por palabras clave, el buscador suele permitir relacionar dos o más palabras clave y frases mediante los operadores, y de esta manera realizar ajustes en la búsqueda para ampliarla, reducirla o dirigirla según nuestras previsiones iniciales o los resultados obtenidos en las búsquedas efectuadas anteriormente.

Para demostrar la importancia de la utilización de estos operadores, y más concretamente de los booleanos, Hölscher y Strube (2000) en su estudio sobre el comportamiento de expertos (bibliotecarios, consultores, autores de libros sobre búsqueda en Internet, etc.) y principiantes ante la búsqueda, añadieron una comparativa entre los expertos analizados y aquellas personas que utilizaban el buscador alemán Fireball. Se obtuvieron unos resultados muy interesantes, al demostrarse que la longitud media de la búsqueda por parte de los usuarios de Fireball era de 1,66 palabras, mientras que los expertos utilizaban una media de 3,64, es decir, suelen utilizar dos palabras más para definir la búsqueda.

Este resultado tiene como explicación que los expertos suelen utilizar en la búsqueda operadores de consulta en mayor medida, a excepción del operador +, que es igualmente de popular para ambos grupos. El hecho de que los expertos utilicen estos operadores al definir la búsqueda nos debería hacer reflexionar sobre la importancia de utilizarlos.

CUADRO 1
UTILIZACIÓN DE LOS OPERADORES
BOOLEANOS DE LOS EXPERTOS
Y USUARIOS DE FIREBALL

	Expertos del estudio %	Fireball %
AND	34,57	2,40
OR	8,02	0,09
NOT	-	0,06
()	7,41	0,10
+	29,01	24,82
-	-	0,48
« »	24,69	8,62

FUENTE: Hölscher y Stenbe (2000).

En el cuadro 1 encontramos el resumen comparativo de la utilización de estos operadores.

Cada herramienta de búsqueda utiliza su propio conjunto de operadores con una sintaxis específica. En las direcciones <http://lisweb.curtin.edu.au/staff/gwpersonal/compare> se recoge una comparativa de los interfaces de los mejores buscadores internacionales (AltaVista, Excite, FAST Search, Google, HotBot, Infoseek/GO.com, Northern Light y WebCrawler). Los operadores de consulta +/- son utilizados en todos los buscadores. El uso de los operadores booleanos está extendido en el conjunto de estos buscadores pero con ciertas diferencias. Así, por ejemplo, el buscador FAST Search no incluye el operador AND y AND NOT, Google excluye AND NOT e Infoseek no utiliza AND y AND NOT. En cuanto al operador de cercanía (NEAR), tan sólo lo incluyen AltaVista y Google. En todos los casos, para realizar una frase de búsqueda, ésta deberá colocarse entre comillas (« »). Excite y WebCrawler no incluyen otros campos de búsqueda como título (title:). Imagen (image:), texto (tex:), dirección (url:), etc.

En la página web de la Universidad de Northumeria, de Newcastle (<http://www.unn.ac.uk/central/isd/features.htm>) se llega a las mismas conclusiones, pero añadiendo en su estudio a Lycos. Las características de este buscador referente a estos operadores de consulta son: admite el uso de +/-; AND, OR y NOT; la búsqueda por campos, las « » para el uso de frases como palabras claves y NEAR.

En cambio no se puede utilizar el operador de truncamiento *.

También en la página Web <http://library.albany.edu/internet/choose.html> podemos encontrarnos con una explicación bastante acertada de la tipología de los operadores de consulta.

Operadores de exactitud, máscara o truncamiento. En primer lugar nos encontramos con las familias de palabras. Añadiendo al final de la palabra el carácter comodín *, en algunos casos es \$ o ? y en otros la herramienta tiene una opción específica que se puede activar. Se utiliza para indicar raíces o partes de palabras y buscar palabras que comienzan con esa raíz. Sería la manera de proceder si no se sabe cómo escribir la palabra a buscar exactamente.

El otro conjunto de estos operadores son las palabras completas. Se utiliza para encontrar sólo aquellas referencias que tengan la palabra exacta y no extensiones. Suele venir por defecto en la mayoría de los buscadores, pero en algunos hay que añadir al final de la palabra el carácter . para darle a entender al buscador que se trata de una palabra completa y no de una cadena de caracteres, y en algunas herramientas se deberá activar una opción específica.

Operadores de existencia. Son dos los tipos de operadores de existencia. Por un lado, los de presencia (añadiendo delante de la palabra el signo +). Exige la presencia de alguna(s) palabra(s) clave(s) en la información encontrada, lo que permitirá precisar más la búsqueda. Una búsqueda definida como la que sigue: +publicidad +internet, mostrará como resultados las páginas que contengan la palabra publicidad además de la palabra Internet. Por otro lado, los de ausencia (añadiendo delante de la palabra el signo -). Exige la ausencia de alguna(s) palabra(s) clave(s) en la información encontrada, lo que permitirá precisar más la búsqueda. Por ejemplo: +publicidad -banner que es lo mismo que publicidad AND NOT banner sólo mostrará como resultados las páginas que contengan la palabra publicidad y no contengan la palabra banner.

Operadores lógicos o booleanos. Son los siguientes:

Y lógico (AND) (también se puede poner &). Exige que aparezcan todas las palabras indicadas independientemente del orden. Es un operador de reducción que permite realizar consultas más específicas. Se utiliza principalmente para concretar más la búsqueda y reducir así el número de direcciones que se obtengan como resultado. Por ejemplo: publicidad AND internet sólo recogerá como resultados las páginas que contengan ambas palabras.

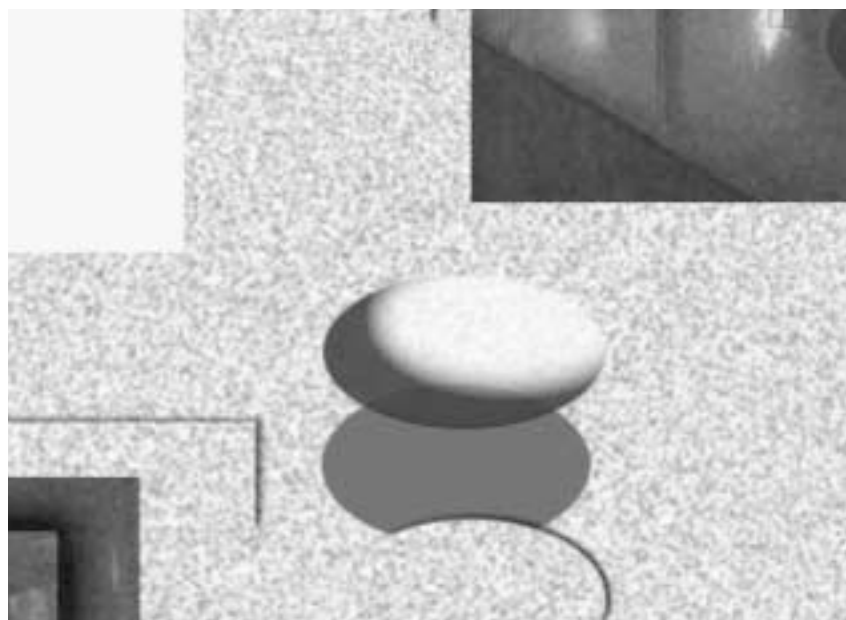
O lógico (OR) (también se puede poner !). Exige que aparezca al menos una de esas palabras indicadas. Es un operador de ampliación que permite, por ejemplo, buscar distintos sinónimos o una o más palabras claves acentuadas o no. Cuando no se pone operador alguno entre palabras, los buscadores suelen interpretarlo como si hubiéramos colocado OR.

NO lógico (NOT) (también se puede poner ^ y en algunas herramientas deberá utilizarse BUT NOT o AND NOT). Exige la ausencia de una o más palabras indicadas. Es un operador de reducción que permite eliminar posibles palabras parecidas pero sin interés para la consulta. Por ejemplo: publicidad AND NOT banner sólo recogerá como resultados las páginas que contengan la primera palabra y no la segunda.

Para una mayor información sobre el uso de las búsquedas booleanas, puede consultarse la siguiente dirección en inglés <http://library.albany.edu/internet/boolean.html>.

Operadores de proximidad o posicionales. Permiten especificar la distancia entre las palabras o términos que se intentan localizar con el buscador. Son los siguientes:

Proximidad (NEAR) (en algunas herramientas se deberán utilizar los corchetes [] o el carácter ~). Es muy parecido al Y lógico, pero además exige que entre ambas palabras no haya más de 10 palabras (esta cantidad puede variar dependiendo del buscador). Algunas herramientas pueden indicar la cantidad de palabras máximas de separación que debe haber entre dos palabras determinadas con el operador NEAR/n, siendo n el número de palabras máximas de separación que se desee. En esos casos, el operador NEAR



sólo obliga a que las palabras estén seguidas. Por ejemplo: publicidad NEAR Internet sólo recogerá como resultados las páginas que contengan las palabras publicidad e Internet relativamente juntas.

Frases. Se parece al Y lógico pero, además, exige que la secuencia de palabras sea idéntica. En la mayoría de los buscadores se suele poner la frase entre comillas o unirla mediante guiones sin dejar espacio en blanco. Por ejemplo: «publicidad en Internet» o publicidad-en-Internet sólo recogerá como resultados las páginas que contengan la frase «publicidad en Internet». En algunos pocos buscadores se deberá utilizar el operador FOLLOWED BY o ADJ entre las palabras en vez de entrecomillarlas.

Cabe destacar cómo el conjunto de todos estos operadores no es excluyente, pudiéndose utilizar conjuntamente para enlazar palabras y frases (aunque en la explicación detallada de cada operador sólo hemos utilizado palabras); en tal caso se suelen aplicar en el orden en que se especifican o rompiendo ese orden utilizando paréntesis. Con los paréntesis se logra agrupar las expresiones de búsqueda para evitar posibles confusiones con la inclusión de operadores y palabras clave.

Siempre nos puede quedar la duda de si estos operadores de consultas dan mejores resultados que si utilizamos el lenguaje

natural. Un estudio realizado por William Hersh *et al.* (2001) para comparar los resultados obtenidos utilizando los operadores booleanos y el lenguaje natural dio resultados bastante concluyentes. Se realizó el estudio sobre 24 personas (bibliotecarios) altamente experimentadas en las tareas de búsqueda. Los resultados mostraban que aunque estaban altamente preparadas en una interfaz con operadores booleanos, existían pocas diferencias de éxito entre el lenguaje natural y los operadores booleanos.

Efectividad de los operadores de consulta. Con este apartado pretendemos demostrar que con el uso de los operadores de consulta podemos obtener en mayor medida información útil (relevante y actualizada). Para ello, hemos efectuado un conjunto de nueve consultas, cada una de ellas sobre nueve de los buscadores más utilizados (ver cuadro 2), seleccionados a partir de directorios sobre buscadores de las páginas <http://lisweb.curtin.edu.au/staff/gwpersonal/compare> y <http://www.unn.ac.uk/central/isd/features.htm>, para así comprobar si los resultados se asemejan en cada uno de ellos. Nos centraremos en los motores de búsqueda y los directorios o buscadores por índice sobre la base de Dempsey (2000), ya que según este autor, los servicios de búsqueda en la Web giran en torno a dos paradigmas básicos: los motores de búsqueda y los directorios.

De los nueve buscadores, la mayoría de ellos son motores de búsquedas, concretamente ocho. Esto es así porque los motores se consideran las herramientas más potentes de búsqueda, aunque con algunos inconvenientes, como ya analizamos en la introducción de este artículo. El único de los directorios es Yahoo!, considerado como uno de los buscadores por índice más importantes y utilizados internacionalmente.

Para proceder al análisis hemos desarrollado las consultas sobre una materia extensa, los EIS o Executive Information System. El motivo por el que elegimos este tema es garantizar, *a priori*, un elevado número de páginas de resultado con la utilización de búsquedas simples, y así comprobar si con las mismas consultas, pero planteadas con los operadores booleanos, se obtenían menores páginas y si éstas son más relevantes. Las consultas fueron las siguientes:

1] En primer lugar, buscaremos información sobre los Executive Information System definiendo la consulta con sus siglas. La consulta quedaría del siguiente modo: EIS.

2] A continuación, plantearemos la misma consulta pero formulada como una sucesión de caracteres única. La consulta quedaría así: «executive information system».

3] En tercer lugar, realizamos la consulta con operadores de presencia. La consulta se formuló como sigue: +executive +information +system.

4] Seguidamente, se quiso plantear la misma consulta que la (3), pero con los operadores booleanos. Así se formuló la consulta: executive AND information AND system.

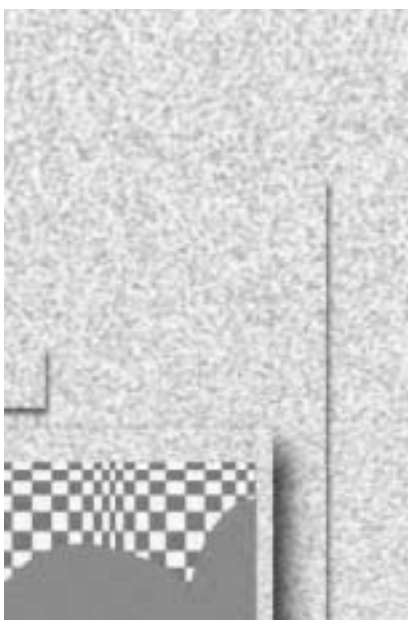
A continuación nos interesamos por la implantación de los EIS, y planteamos las consultas de las siguientes formas:

5] +implementation +«executive information system».

6] implementation AND «executive information system».

7] implementation AND executive AND information AND system.

8] implementation AND EIS.



Por último, quisimos saber la información que trataba sobre la implementación pero no sobre el control de los EIS. La consulta quedó como sigue:

9] implementation AND «executive information system» AND NOT control.

Una vez ejecutadas todas estas consultas en todos los motores de búsqueda, obtuvimos los resultados, en número de páginas, que se resumen en el cuadro 2.

En general se obtienen menores resultados utilizando una sucesión única de caracteres (2) que planteándola con las siglas (1). Esto es debido a que las siglas (EIS) no sólo hacen referencia a los Executive Information System, sino que pueden hacer referencia, por ejemplo, a una empresa, como se nos ha dado el caso. Por tanto, planteando la consulta de la forma (2) se aumentan las posibilidades de obtener información sobre los Executive Information System.

Es aconsejable utilizar « » cuando la consulta se refiera a un concepto que se denomine con una cadena de caracteres única, es decir, cuando la consulta haga alusión a un objeto, acción o acontecimiento con una única forma de identificación, como, por ejemplo, la «Torre Eiffel» o la «Teoría de la Relatividad». En nuestro caso es así, por ello se obtienen resultados más relevantes buscando por

la denominación completa de los EIS y no por sus siglas.

El motivo de plantear la consulta (3) como +executive +information +system era comprobar si entre las 10 primeras páginas de los resultados se contenía una información más relevante que en la (2). Al ejecutar la consulta (3) obtuvimos mayores o iguales resultados que la (2), a excepción de Excite y WebCrawler, cuyos resultados se vieron reducidos drásticamente (de 4.197.945 a 9.800 páginas). Además, podemos observar que la tónica general para aquellos buscadores que experimentaban un considerable aumento de los resultados, la relevancia se veía mermada (Fast y Lycos), a excepción de Northern Light, que mantiene una relevancia alta en ambos resultados. Para AltaVista, que aumentó sus resultados más moderadamente, observamos una mayor relevancia de la información devuelta con la consulta (3). Para aquellos buscadores que siguieron manteniendo los mismos resultados, la relevancia fue la misma. En Excite y WebCrawler, que vieron fuertemente reducidos sus resultados, se produjo un aumento de la relevancia de la información.

Con la pregunta (4) pretendimos observar si una consulta planteada con operadores booleanos se obtenían resultados distintos a los obtenidos en la consulta simple (3). Como se puede observar en el cuadro 2, todos los buscadores devolvieron los mismos resultados, a excepción de Fast, que multiplicó por 10 sus resultados, y de Lycos, con una reducción muy leve. Por tanto, en este caso hubiera sido indiferente utilizar los operadores de presencia o los lógicos.

A partir de la consulta (5) se hace referencia a la implantación de los EIS, una vez llegada a la conclusión de que para hacer consultas sobre los Executive Information System es mejor utilizar operadores posicionales, de ahí que planteemos las consultas (7) y (8), con el fin de confirmar esta conclusión.

Las consultas (5) y (6) prácticamente devuelven los mismos resultados, menos Excite, que los aumenta (de 2.920 pasa a 24.330), y de Northern Light, que los reduce (de 4.467.125 a 5.651). Por tanto, en este caso tampoco existen diferencias sig-

CUADRO 2
PÁGINAS OBTENIDAS DE CADA CONSULTA EN CADA UNO DE LOS MOTORES DE BÚSQUEDA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Altavista									
http://www.altavista.com	347.011	8.184	26.681	26.681	2.059	2.059	568.689	17.331	940
Excite									
http://www.excite.com	11.460	4.197.945	9.800	9.800	2.920	24.330	180	110	23.390
Fast									
http://www.alltheweb.com	485.916	9.305	1.689.246	10.683.404	2.872	2.851	447.394	447.394	1.251
Google									
http://www.google.com	30.000	12.700	12.700	37.700	5.040	5.060	47.900	25.100	2990
HotBot									
http://www.hotbot.com	+1.000	+1.000	+1.000	+1.000	935	935	+1.000	+1.000	585
Lycos									
http://www.lycos.com	466.836	9.264	1.781.712	1.682.496	2.835	2.815	441.557	25.960	1.235
Northern L.									
http://www.northernlight.com	454.864	15.634	2.432.230	2.432.230	4.467.125	5.651	495.948	23.795	2.309
Webcrawler									
http://www.webcrawler.com	11.460	4.197.945	9.800	9.800	2.920	179	180	8.046	23.390
Yahoo!									
http://www.yahoo.com	44	7.260	63	63	1.370	1.370	1	1	1.170

FUENTE: Elaboración propia.

nificativas con la utilización de operadores booleanos.

Comparando las consultas (5) y (7) o (6) y (7) vemos cómo no es adecuado tratar por separado las palabras que definen a los EIS, ya que se obtienen resultados muy exagerados con un alto componente de ruido. Lo mismo ocurre entre las consultas (5) y (8) o (6) y (8), donde definir a los Executive Information System por sus siglas proporciona resultados con mucho ruido. Excite es el único buscador que reduce los resultados tanto en la consulta (7) y (8), con una alta relevancia de la información para esta última.

Con la consulta (9) pretendíamos demostrar que al ser una matización de la (6), se obtenían menores resultados, como así se puede demostrar en el cuadro 2.

Por último, y con respecto a Yahoo!, obtuvimos los resultados que esperábamos de un buscador por índice, ya que con éste se obtuvo menores resultados que con los motores de búsqueda (ver cuadro 2). Sin embargo, es sorprendente la relevancia de la información que se devolvió en las nueve consultas planteadas.

CONCLUSIONES

El volumen de información en Internet crece a pasos agigantados debido a que se incorporan diariamente cientos de páginas y se actualizan otras tantas, entre las que se incluyen las de carácter científico. Y es más, las tradicionales fuentes de información, tales como bibliotecas y bases de datos, se han sumado a la red con la publicación de sus catálogos, pudiendo ser consultados la mayoría de ellos en su totalidad. Así pues, el investigador en un mismo lugar puede llegar a controlar todo el conjunto de fuentes. Por ello, cada vez más, se ven obligados a acudir a Internet para consultar y seleccionar su contenido.

Ahora bien, Internet posee el inconveniente de la ausencia de un catálogo donde se recoja la situación exacta de toda la información publicada, debido a la imposibilidad de abarcar la totalidad de la red, de ahí que buscar directamente en las páginas se convierta en una estrategia poco adecuada, a menos que se sepa con certeza dónde se sitúa la información deseada,

como hemos visto en el estudio de White (2000).

Por tanto, se hace imprescindible el uso de las herramientas diseñadas para la búsqueda como estrategia más eficaz por la capacidad de encontrar información útil (relevante y actualizada) en el menor tiempo posible. Además, para que la búsqueda de información sea lo más precisa posible, es aconsejable la utilización de los operadores o términos de consulta. Con ellos tendremos la posibilidad de obtener resultados los más limpios posible (sin ruidos).

BIBLIOGRAFÍA

- CLARKE CHARLES, L. A.; CORMACK, G. V. y TUDHOPE, E. A. (2000): «Relevance ranking for one to three term queries», *Information Processing & Management*, vol. 36, pp. 291-311.
- DASSLER WUKOVITZ, L. (2001): «Using internet search engines and library catalogs to locate toxicology information», *Toxicology*, vol. 157, pp. 121-139.
- DEMPSEY, B. J.; VREELAND, R. C.; SUMMER Jr., R. G. y YANG, K. (2000): «Desing and empirical evaluation of search software for

- legal professionals on the WWW», *Information Processing & Management*, vol. 36, pp. 253-273.
- FISHER INGID, E. y MURPHY SMITH, L. (2000): «The role for web search engines. The CPA», *Journal*, enero, pp. 43-75.
- HERSH, W.; TURPIN, A.; PRICE, S.; KRAEMER, D.; OLSON, D.; CHAN, B. y SACHEREK, L. (2001): «Challenging conventional assumptions of automated information retrieval with real users: Boolean searching and batch retrieval evaluations», *Information Processing & Management*, vol. 37, pp. 383-402.
- HÖLSCHER, CH. y GERHARD, S. (2000): «Web search behaviour of Internet experts and newbies», *Computers Networks*, vol. 33, pp. 337-346.
- KHOO, C. S. G.; MYAENG, S. H. y ODDY, R. H. (2001): «Using cause-effect relations in text to improve information retrieval precision», *Information Processing and Management*, vol. 37, pp. 119-145.
- KOBAYASHI, M. y TAKEDA, K. (2000): «Information Retrieval on the web», *ACM Computing Surveys*, vol. 32, nº 2, pp. 144-172.
- LANG, R. y STACEY, P. (2001): *Speed up your Internet searches. Occupational Hazards*, abril, pp. 44-46.
- WEBSTER, K. y PAUL, K. (1996): «Beyond Surfing: Tools and Techniques for Searching the Web», *Information Technology*, enero.
- WHITE, M. D. y IIVONEN, M. (2000): «Questions as a factor in Web search strategy», *Information Processing & Management*, vol. 37, pp. 721-740.
- ZAMIR, O. y ETZIONI, O. (1999): «Grouper: a dynamic clustering interface to Web search results», *Computer Networks*, vol. 31, pp. 1361-1374.

RECURSOS ELECTRÓNICOS:

<http://www.bib.uc3m.es/~jroble/documentac/buscar.htm#aprender> de la Universidad Carlos III de Madrid.

<http://hipatia.uc3m.es/~jroble/documentac/3a.htm>, también de la Universidad Carlos III. Artículo «Cómo buscar y encontrar información en Internet», de Álvaro Ibáñez y publicado en la revista *Iworld*, noviembre de 1997.