

Motores de búsqueda y Álgebra Booleana

2017-10-06

Motores de búsqueda y Álgebra Booleana

Documento que explica con claridad la mejor forma de utilizar los motores de búsqueda y el papel que en ellos desempeña el Álgebra Booleana. Localizar rápida y efectivamente información en la Red, implica utilizar las funciones avanzadas del motor de búsqueda.

Documento que explica con claridad la mejor forma de utilizar los motores de búsqueda y el papel que en ellos desempeña el Álgebra Booleana. Localizar rápida y efectivamente información en la Red, implica utilizar las funciones avanzadas del motor de búsqueda.

MOTORES DE BÚSQUEDA Y ÁLGEBRA BOOLEANA

Por: Bernie Dodge y Biblioteca de la Universidad de Albany

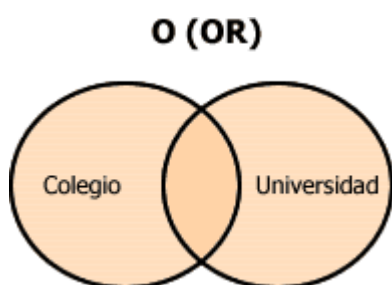
Internet es una inmensa base de datos. Como tal, sus contenidos deben buscarse de acuerdo con las reglas establecidas para realizar búsquedas en bases de datos. Gran parte de la búsqueda en las bases de datos, se apoya en los principios de la lógica Booleana. Estos principios hacen referencia a las relaciones lógicas existentes entre los términos de búsqueda a las cuales se les dio el nombre del matemático británico George Boole.

En los [motores de búsqueda](#) por Internet, las opciones para construir relaciones lógicas entre los términos de búsqueda se extienden más allá de la práctica tradicional de la búsqueda Booleana. Esto último se cubrirá en este mismo documento en la sección llamada "Búsqueda Booleana por Internet".

La lógica Booleana consiste en tres operadores lógicos:

O (or) Y (and) NO (not)

Cada uno de estos operadores se puede describir visualmente usando un diagrama de Venn, como se muestra a continuación



PREGUNTA: Deseo Información sobre colegios.

-

En esta búsqueda, vamos a recuperar registros en los cuáles **POR LO MENOS UNO** de los términos de búsqueda esté presente. Estamos buscando los términos **Colegio** y también **Universidad** porque los documentos que contienen una de estas dos palabras pueden ser relevantes.

-

Esto se ilustra por:

El operador lógico **O (OR)** se usa más comúnmente para buscar términos sinónimos o conceptos.

A continuación presentamos un ejemplo de cómo funciona el operador lógico **O (OR)**

- El círculo sombreado con la palabra colegio representa todos los registros que contienen la palabra "**colegio**"

El círculo sombreado con la palabra universidad representa todos los registros que tiene la palabra "**universidad**".

- El área sombreada en la que se entrelazan los dos círculos representa todos los registros que contienen las dos palabras "colegio" y "universidad"

|

TERMINO BUSCADO

|

RESULTADOS ALTAVISTA

|

RESULTADOS GOOGLE

|

|

Colegio

|

378.678

|

289.000

|

|

Universidad

|

1'140.371

|

3'250.000

|

|

Colegio OR Universidad

|

1'220.917

|

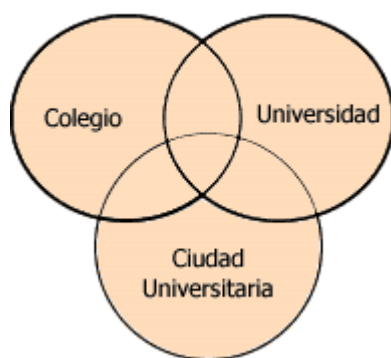
425.000

|

El operador lógico O (OR) ayuda a tamizar los resultados para recuperar todos los registros únicos que contienen uno de los términos, el otro o ambos.

En la medida en que se combinen más términos o conceptos en una búsqueda con el operador lógico O, mayor será la cantidad de registros que se van a encontrar.

O (OR)



|

TERMINO BUSCADO

|

RESULTADOS ALTAVISTA

|

RESULTADOS GOOGLE

|

|

Colegio

|

378.678

|

289.000

|

|

Universidad

|

1'140.371

|

3'250.000

|

|

Colegio OR Universidad

|

1'220.917

|

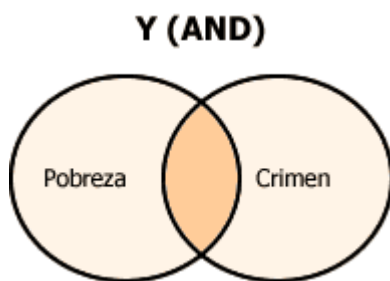
425.000

|

Colegio OR Universidad OR (Ciudad AND Universitaria)

1'245.846

155.000



Pobreza Y Crimen

PREGUNTA: ¿Qué relación existe entre Crimen y Pobreza?

- En esta búsqueda se recuperan registros en los cuáles **AMBOS** términos están presentes.
- Lo anterior se ilustra por el área sombreada en la que se entrelazan los dos círculos que representan a todos los registros que contienen tanto la palabra "Pobreza" como la palabra "Crimen"
- Es necesario anotar que no se recuperó ningún registro que contuviera únicamente la palabra "Pobreza" o únicamente la palabra "Crimen"

A continuación presentamos un ejemplo de cómo funciona el operador lógico **Y (AND)**

TERMINO BUSCADO

|

RESULTADOS ALTAVISTA

|

RESULTADOS GOOGLE

|

|

Pobreza

|

178.387

|

584.000

|

|

Crimen

|

179.369

|

73.400

|

|

Pobreza AND Crimen

|

120

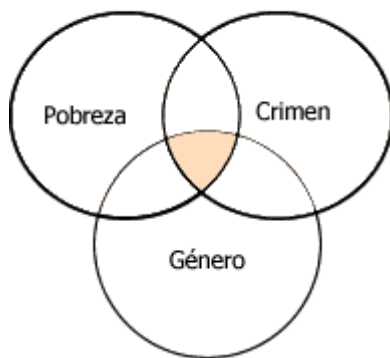
|

21.100

|

En la medida en que se combinen una mayor cantidad de términos y conceptos con el operador lógico **Y (AND)**, se van a recuperar una menor cantidad de registros.

Y (AND)



Por Ejemplo:

|

TERMINO BUSCADO

|

RESULTADOS ALTAVISTA

|

RESULTADOS GOOGLE

|

|

Pobreza

|

178.387

|

584.000

|

|

Crimen

|

179.369

|

73.400

|

|

Pobreza AND Crimen

|

120

|

21.100

|

|

Pobreza AND Crimen AND Genero

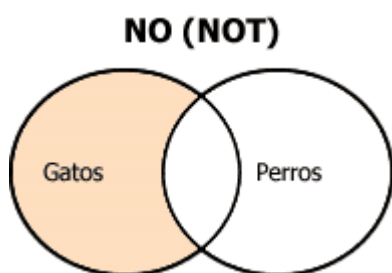
|

|

4.770

|

Solamente algunos de los motores de búsqueda hacen uso del operador de Proximidad (cercanía) en temas relacionados **CERCA** (Near). El operador de proximidad determina la cercanía en la que están situados los términos al interior de un documento fuente. Cerca (Near) es un Y (AND) restrictivo. La proximidad de los términos de búsqueda se determina de manera particular para cada uno de los motores de búsqueda. Por ejemplo Cerca en Alta Vista (Búsqueda Poderosa) es 10 palabras. Google tiene esta función en forma predeterminada.



Perros **NOT** gatos

PREGUNTA: Se requiere información sobre gatos pero se desea evitar cualquier información sobre perros.

- En esta búsqueda, se recuperan registros en los cuáles **SOLAMENTE UNA** de estas palabras está presente.
- Lo anterior está ilustrado por el área sombreada con la palabra Gatos que representa todos los registros que contienen la palabra "gatos"
- No se recuperan registros en los que aparece la palabra "perros" aún cuando la palabra "gatos" aparezca en ellos también.

A continuación presentamos un ejemplo del funcionamiento del operador lógico **NO (NOT)**.

|

TERMINO BUSCADO

|

RESULTADOS ALTAVISTA

|

RESULTADOS GOOGLE

|

|

Gato

|

132.806

|

370.000

|

|

Perro

|

184.674

|

307.000

|

|

Gato NOT Perro

|

10.605

|

1.610

|

El operador lógico NO (NOT) excluye registros o registros de los resultados de búsqueda. Tenga cuidado al usar NO, ya que el término que se quiere buscar puede estar presente de manera importante en documentos que también contienen el término que se desea excluir.

N. del T: la mayoría de los motores de búsqueda requieren que se ingresen los términos lógicos en inglés (AND, OR, NOT), así esté buscando con palabras claves en español.

BÚSQUEDA BOOLEANA EN INTERNET

Cuando se utiliza un motor de búsqueda en Internet, el uso de la lógica booleana puede manifestarse de tres maneras diferentes:

1. Lógica booleana completa con el uso de los operadores lógicos.
2. Lógica booleana implícita con la búsqueda por palabras claves.
3. Lenguaje predeterminado en una plantilla que llena el usuario.

1. LÓGICA BOOLEANA COMPLETA CON EL USO DE LOS OPERADORES LÓGICOS

Muchos de los motores de búsqueda ofrecen la opción de hacer una búsqueda Booleana completa que requiere la utilización de los operadores Boléanos lógicos.

|

PREGUNTA

|

OPERADOR

|

TÉRMINOS DE BÚSQUEDA

|

|

Necesito información sobre gatos

|

O (OR)

|

gatos OR felinos

|

|

Estoy interesado en Dislexia en Adultos

|

Y (AND)

|

dislexia AND adultos

|

|

Estoy interesado en radiación pero no en radiación nuclear

|

NO (NOT)

|

radiación NOT nuclear

|

|

Deseo aprender sobre el comportamiento de los gatos

|

O (OR),

Y (AND)

|

(gatos OR felinos) AND comportamiento

|

Nota: El uso de los paréntesis en la búsqueda se conoce como forzar el orden de procesamiento. En este caso incluimos la palabra OR en el paréntesis para que el motor de búsqueda proceda primero a realizar ésta parte de la búsqueda.

Seguidamente el motor de búsqueda combinará este resultado con la última parte de la búsqueda. Haciendo uso de éste método nos podemos asegurar de que los términos

que están con OR se mantienen unidos como una unidad lógica.

2. LÓGICA BOOLEANA IMPLÍCITA CON LA BÚSQUEDA POR PALABRAS CLAVES

La búsqueda por palabras claves se refiere al tipo de búsqueda en la cuál se ingresan los términos que representan el concepto que se quiere recuperar o buscar. En ésta no se utilizan operadores Boléanos.

La lógica Booleana implícita se refiere a la búsqueda en la que se utilizan símbolos, que representan los operadores Boléanos lógicos. En este tipo de búsqueda por Internet , la ausencia de un símbolo también es significativa, por ejemplo el espacio entre las palabras claves predeterminadas (default), tanto para el operador lógico OR (O) como para el operador lógico AND (Y). Muchos de las motores de búsqueda bien conocidos tradicionalmente aplican por defecto (default) el operador lógico O (OR), pero como regla general se están alejando de ésta práctica y utilizando por defecto (default) el operador lógico Y (AND).

La lógica Booleana implícita se ha vuelto tan común en la búsqueda por la Red que pude considerarse actualmente como una práctica estándar.

|

PREGUNTA

|

OPERADOR

|

TÉRMINOS DE BÚSQUEDA

|

|

Necesito información sobre gatos

|

O (OR)

|

gatos felinos(1)

|

|

Estoy interesado en Dislexia en Adultos

|

Y (AND)

|

+dislexia +adultos

|

|

Estoy interesado en radiación pero no en radiación nuclear

|

NO (NOT)

|

radiación -nuclear

|

|

Deseo aprender sobre el comportamiento de los gatos

|

O(OR),

Y (AND)

|

(gatos felinos) +comportamiento

|

(1) Este ejemplo es cierto para los motores de búsqueda que interpretan el espacio entre las palabras claves como el operador lógico Booleano O. Para poder saber que lógica tiene predeterminada (default) un motor de búsqueda, consulte los archivos de Ayuda que ofrece el motor.

3. LENGUAJE PREDETERMINADO EN UNA PLANTILLA QUE LLENA EL USUARIO

Algunos motores ofrecen una plantilla de búsqueda que le permite al usuario escoger el operador Booleano de un menú. Con frecuencia el operador lógico se encuentra expresado en lenguaje sustituto en lugar de estar presentado como en el operador mismo.

|

PREGUNTA

|

OPERADOR

|

TÉRMINOS DE BÚSQUEDA

|

|

Necesito información sobre gatos

|

O(OR)

|

Cualquiera de éstas palabras.

Puede contener las palabras.

Debe contener las palabras.

|

|

Estoy interesado en Dislexia en Adultos

|

Y (AND)

|

Todas estas palabras.

Debe contener las palabras.

|

|

Estoy interesado en radiación pero no en radiación nuclear

|

NO (NOT)

|

No debe contener las palabras.

Sin las palabras.

|

|

Deseo aprender sobre el comportamiento de los gatos

|

O(OR),

Y (AND)

|

Combine las opciones como se ha hecho en las preguntas anteriores, si la plantilla permite la inclusión de varias ordenes de búsqueda.

|

|

Cuadro de Comparación Rápida :

Lógica Booleana completa versus Lógica Booleana Implícita versus Plantillas

|

|

|

Booleana Completa

|

Booleana Implícita

|

Terminos en Plantilla

|

|

O (OR)

|

colegio OR universidad

|

colegio universidad

- ver nota abajo

|

Cualquiera de éstos términos.

Puede contener los términos.

Debe contener los términos.

|

|

Y (AND)

|

pobreza AND crimen

|

+pobreza +crimen

|

Todas estas palabras.

Debe contener las palabras.

|

|

NO (NOT)

|

gatos NOT perros

|

gatos -perros

|

Puede no contener las palabras.

Debe no contener las palabras.

|

|

CERCA,etc

(NEAR)

|

gatos NEAR perros (*)

|

N / A (*)

|

Cerca, Próximo.

- Esta afirmación de búsqueda se resolverá como Y lógico en los motores de búsqueda que utilicen Y por defecto (default). Ejemplos de estos incluyen AOL.com Search, Google y Lycos. Se debe consultar el archivo de ayuda en el

sitio de cada motor para averiguar que lógica utilizan en forma predeterminada (default).

|

****CUATRO CLAVES PARA UNA MEJOR BÚSQUEDA**

ESTRECHAR - PRECISAR - RECORTAR - ASEMEJAR**

La página perfecta con seguridad está en alguna parte. Es la página que tiene la información precisa que usted está necesitando y para usted es deseable e inalcanzable como una estrella lejana. Si solamente contará usted con una malla (Red) inmensa que le permitiera capturarla.

La mayoría de la gente hace uso de los motores de búsqueda simplemente escribiendo unos pocos términos en el campo de palabras claves del motor para dedicarse luego a mirar con detenimiento los resultados obtenidos. En algunos casos las palabras que se escogen dan por resultado una reducción indebida de la búsqueda que impide encontrar lo que se están buscando. Con mayor frecuencia el resultado es una pila de páginas web que no tienen mucha relación con lo que se busca y que deben ser filtradas por el usuario para encontrar algo. Pero la situación antes descrita puede mejorar y esa es la razón de ser de este documento.

El motor de búsqueda más comprehensivo que existe en el momento es Google y en este nos vamos a enfocar. El primer paso para convertirse en "hábil pescador" de páginas de Internet consiste en dominar la opción de Búsqueda Avanzada que ofrece Google, información que se encuentra en

http://www.google.com/advanced_search?hl=es.

Le sugerimos adicionar esta dirección a sus sitios Favoritos.

https://www.google.com/advanced_search?hl=es

Si usted convierte en hábito el uso de las cuatro claves que a continuación describimos, usted se volverá mejor investigador que el 90% de los usuarios que utilizan Internet. Se trata solamente de recordar cuatro elementos, cada uno de los

cuáles le suministrará una mejor "Red" para capturar información.

PRIMERA CLAVE : Comience a ESTRECHAR

Talvez el mayor problema que tiene la gente con los motores de búsqueda es que son ¡muy buenos! Usted puede escribir una palabra y en una fracción de segundo puede tener 20.000 páginas para revisar. Pero la mayoría de ellas no corresponderán exactamente a lo que está buscando y usted tendrá que gastar un montón de tiempo " buceando" entre ellas, para desechar las 19.993 respuestas inadecuadas.

Si usted sabe lo que necesita, por que no empieze por preguntar con la mayor precisión posible?

Piense en todas las palabras que deben aparecer siempre en la página "precisa". Póngalas en la casilla que indica : **CON TODAS LAS PALABRAS**. (With All the Words).

Piense también en todas las páginas distractoras que pueden aparecer porque uno o más de los términos utilizados en la búsqueda, tiene varios significados. ¿En que palabras puede usted pensar que le ayuden a eliminar esas páginas? Escríbalas en la casilla que indica **SIN LAS PALABRAS** (Without the words).

Si existen sinónimos del término que está buscando y estos pueden aparecer en la página que se desea encontrar, escríbalos en la casilla que indica : **CON ALGUNA DE LAS PALABRAS** (With Any of the Words).

Ensaye ahora cada una de las formas de búsqueda anteriores y registre cuantos sitios encontró.

|

PREGUNTA

|

RESULTADOS

|

|

Imagine que usted está interesado en conocer acerca del legendario continente perdido de Atlántida. Hay varias películas con la palabra Atlántida en el título, pero usted no desea esa información. Además, usted no está interesado en agencias de viajes que se llamen Atlántida.

|

Escriba en número de “aciertos para cada opción de búsqueda.

|

|

Esta es una forma inadecuada para buscar:

CON TODAS LAS PALABRAS:Atlántida

| |

|

Puede afinar la búsqueda intentando esta opción:

CON TODAS LAS PALABRAS:Atlántida continente perdido

SIN LAS PALABRAS:película cine filmes entretenimiento viajes

| |

|

Ahora intente lo siguiente:

CON TODAS LAS PALABRAS:Atlántida continente perdido

CON ALGUNA DE LAS PALABRAS:Atlantes

SIN LAS PALABRAS:película cine filmes entretenimiento viajes

| |

Cuando ensaye cada una de las formas de búsqueda, anote que clase de registros encuentra. Observe que mientras más específicos sean los términos que usted incluye o excluye, más enfocada es su búsqueda.

SEGUNDA CLAVE : Encuentre las Frases, PRECISAR

Las palabras se unen en forma predecible. Si usted escribe un frase en la casilla de la opción de Búsqueda Avanzada de Google que dice **CON LA FRASE EXACTA** (with the exact phrase), usted podrá encontrar páginas en las que las palabras escritas aparezcan juntas y en el orden en que se escribieron. Resulta obvio que esto facilita la búsqueda de cosas que tienen nombre propio compuesto por varias palabras (eje: lugares, títulos de libros, personas)

También es útil cuando usted puede recordar una frase particular de algo que usted ha leído y que ahora debe localizar. Cómo es el resto del cuento que empieza "Simón el Bobito llamó al pastelero".

La habilidad de hacer búsquedas por frases puede resultar sorprendentemente útil. Sospecha usted que un trabajo que entregó uno de sus estudiantes puede ser plagiado si no en su totalidad, si en forma parcial, sin que éste diera los créditos al autor original? Escriba una o dos frases del trabajo y observe que aparece. Así mismo puede usted darse cuenta si su trabajo se está copiando sin su autorización.

Otro uso de esta característica: descubrir la veracidad de cierto tipo de información. La próxima vez que reciba notificación respecto a la aprobación de una ley muy controversial o sobre un nuevo y peligroso virus de computador, compruébela antes de trasmitirla a otros. Para esto escriba cualquier frase poco usual o singular que vea en el correo electrónico y fíjese si existen otros comentarios sobre este rumor en particular.

|

PREGUNTA

|

RESULTADOS

|

|

Imagine que usted está interesado en conocer acerca del legendario continente perdido de Atlántida. Hay varias películas con la palabra Atlántida en el título, pero usted no desea esa información. Además, usted no está interesado en agencias de viajes que se llamen Atlántida.

|

Escriba en número de “aciertos para cada opción de búsqueda.

|

|

Esta es una forma inadecuada para buscar:

CON TODAS LAS PALABRAS:Atlántida

| |

|

Puede afinar la búsqueda intentando esta opción:

CON TODAS LAS PALABRAS:Atlántida

CON LA FRASE EXACTA:patria de los atlantes

| |

|

Ahora intente lo siguiente:

CON TODAS LAS PALABRAS:Pombo

| |

|

Afine la búsqueda de la siguiente manera:

CON TODAS LAS PALABRAS:Rafael Pombo

CON LA FRASE EXACTA:Simón el Bobito llamó al pastelero

| |

TERCERA CLAVE : RECORTAR en Forma Gradual la Dirección de un Sitio en Internet (URL)

Esta opción no es específica de Google, aunque usted va a utilizarla con frecuencia cuando este usando más eficazmente ese buscador. Con frecuencia usted encuentra una página excelente incrustada en una carpeta que está contenida en otra carpeta que a su vez está metida en otra carpeta. Usted sospecha que existen otras páginas interesantes cerca de ella, pero ¿cómo se pueden encontrar? Una forma sencilla es ir recortando paso a paso la dirección del sitio.

Al hacer esto, en ocasiones usted encuentra un mensaje que dice PROHIBIDO! Otras veces accede usted a una lista de archivos y directorios, y algunas veces llega a una página que tiene más enlaces. Es importante saber que cada paso en retroceso le indica a usted con mayor claridad cuál es el origen de la página.

Además es una buena estrategia a utilizar cuando no se puede localizar una página (esto es cuando usted recibe como respuesta de la búsqueda un mensaje 404). En estos casos es posible que el encargado del sitio haya movido la página, la haya llevado a una carpeta nueva o haya cambiado el nombre de la carpeta. Recorra el camino retrocediendo hasta el inicio de la dirección y vuélvalo a recorrer hacia delante a ver si de ésta manera puede encontrar la página perdida.

|

Usted ha encontrado un Web Quest realmente bueno y desea saber que más hay es ese sitio Web puede empezar a recortar la dirección.

Empiece aquí:

<http://www.ree.es/es/exporee/prepara-tu-visita/webquest>

Ahora recorte la última parte:

<http://www.ree.es/es/exporee/prepara-tu-visita/>

¿Qué pudo notar?

Continué recortando y observe que sucede:

<http://www.ree.es/es/exporee/>

<http://www.ree.es/es/>

|

CUARTA CLAVE : Busque Páginas que se ASEMEJEN

Una vez que usted haya encontrado algo que le gusta o requiere utilizando a Google, es muy fácil y además útil encontrar páginas semejantes o similares. ¿Cómo? Debajo del campo destinado a la Búsqueda Avanzada que usted ha estado utilizando hasta ahora, existen dos campos nuevos. Estos le permiten encontrar páginas que Google cree son similares a la que usted ya encontró o que tienen enlaces con la dirección que usted escribió: Encontrar páginas similares a la página y Encontrar páginas con enlaces a la página.

¿Cómo sabe Google que dos páginas se asemejan? Los detalles del funcionamiento interno de los motores de búsqueda son secretos del oficio, pero podemos asumir con cierto grado de certeza, que los motores se basan en la semejanza de las palabras y los enlaces externos, existentes en las páginas. Lo que si es cierto es que esta opción funciona admirablemente bien, especialmente cuando usted no sabe que palabras claves debe buscar.

Use esta herramienta para encontrar más información sobre algo que para usted es bueno. Úselo para encontrar páginas que están enlazadas a la página que usted encontró y le es útil. Es muy probable que las nuevas páginas también le sirvan.

También sáquele brillo a su Ego: si usted ha puesto una página hecha por usted en un servidor público y en el ha estado durante un tiempo prudencial, entérese de quién más ha puesto enlaces en sus paginas a su sitio.

|

PREGUNTA

|

RESULTADOS

|

|

Suponga que usted ha encontrado un sitio como Eduteka que le proporciona contenido para los docentes que desean integrar la tecnología a la educación y desea que otras páginas similares puede encontrar.

|

Escriba en número de “aciertos para cada opción de búsqueda.

|

|

Use la característica que tiene Google para encontrar páginas que se asemejen a una que a usted le guste mucho:

ENCONTRAR PÁGINAS SIMILARES A LA PÁGINA: www.maseducativa.com

| |

|

Otra forma de explorar un sitio Web es encontrar quién tiene enlaces a ese sitio.

ENCONTRAR PÁGINAS CON ENLACES A LA PÁGINA:www.maseducativa.com

| |

|

Pruebe lo siguiente:

SIMILARES:kids.msfc.nasa.gov

| |

|

ENLACES:kids.msfc.nasa.gov

| |

Finalmente para recapitular, recuerde las cuatro claves:

****ESTRECHAR**

PRECISAR

RECORTAR

ASEMEJAR**

que le permitirán realizar una búsqueda no solo mejor sino más efectivamente.

CRÉDITOS

"Búsqueda Booleana por Internet", documento creado por la Biblioteca de la Universidad de Albany

"Cuatro claves para una mejor búsqueda" escrito por Bernie Dodge
(bdodge@mail.sdsu.edu)