



Módulo 2

2.1 LAS ECUACIONES DE BÚSQUEDA Y LOS OPERADORES

Por Jose Antonio Senso Ruiz

Profesor del Departamento de Información y Comunicación. Grupo HUM-466 Información Científica. Universidad de Granada.

1. LAS ECUACIONES DE BÚSQUEDA

Una vez que tenemos claras cuáles serán las palabras que vamos a usar, porque son aquellas que mejor representan el significado de lo que queremos buscar, hay que construir una ecuación de búsqueda. Aunque tenga un nombre que adelanta algo complicado en realidad, dentro de la recuperación de información, las ecuaciones de búsqueda se emplean para expresar de la forma más exacta posible la necesidad de información.

Una ecuación de búsqueda está constituida por una serie de conceptos, expresados en forma de palabras clave, unidos entre sí por operadores de búsqueda, que son los que le dan sentido a la consulta.

En realidad, no es más que una forma de convertir nuestra necesidad de información en algo que pueda entender el sistema de recuperación de información. Hay que tener en cuenta que no podemos preguntarle como si estuviéramos hablando con un amigo, en lenguaje coloquial. Hay que hacer las preguntas en un lenguaje que la máquina pueda entender y procesar.

Dependiendo del sistema de recuperación de información, y de la interfaz que tenga, nos encontraremos con formularios que permiten la introducción de datos de manera amigable. En otras ocasiones, sin embargo, nos encontraremos con que el sistema nos obligará a tener que construir la ecuación de manera más “artesanal”.

Interfaz de búsqueda avanzada del catálogo de la Biblioteca Nacional. Gracias a esta interfaz es posible realizar ecuaciones de búsqueda muy completas de manera amigable. Sin embargo, no todos los Sistemas de Recuperación de Información cuentan con este tipo de interfaces, por lo que será necesario elaborar las ecuaciones de búsqueda de manera más artesanal.



Inicio Colecciones especiales Autoridades Bibliografía Española Recursos electrónicos

Volver Ayuda Enlace permanente Desconexión

Descubre una nueva forma de acceso a nuestras colecciones y recursos a través de **DATOS·BNE·ES**

Búsqueda avanzada

Todos los campos	<input type="text"/>	Y
Autor	<input type="text"/>	Y
Título	<input type="text"/>	Y
Materia	<input type="text"/>	Y
Colección/Serie	<input type="text"/>	Y
Notas	<input type="text"/>	Y
ISBN, ISSN, etc.	<input type="text"/>	Y
Datos de publicación	<input type="text"/>	

Idioma:

Tipo de documento:

Año de pub:

Ordenar por:

No te preocupes. Lo realmente importante es que conozcas la lógica, la filosofía, que se emplea. Si la dominas, dará lo mismo cómo sea la interfaz de búsqueda, ya que conocerás cómo funciona cada operador y qué implica su uso concreto.

Y para ello vamos a comenzar con el primer tipo de operador que encontraremos en todos los sistemas de recuperación de información: los booleanos.

2. LOS OPERADORES BOOLEANOS

También se les conoce como operadores lógicos, y se trata de palabras o símbolos que permiten conectar de forma lógica los conceptos que se han seleccionado para la búsqueda con el fin de ampliar o limitar las búsquedas. Todos los sistemas de recuperación de información los emplean, por lo que conocerlos es fundamental para hacer búsquedas precisas.

Cada operador booleano realiza una función concreta, pero el símbolo que se emplea para “activar” dicha función puede cambiar. Puede resultar lioso decirlo así, por lo que espero que estos ejemplos lo dejen más claro:



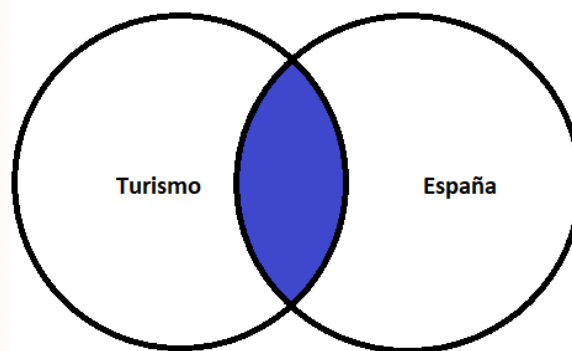
Operador de intersección

Función: se emplea para que el sistema muestre los documentos que tengan todos los términos de búsqueda especificados, independientemente del orden y de la posición que ocupen esos términos en dichos documentos.

Símbolo: habitualmente suele ser AND, aunque hay sistemas de recuperación que emplean otros, como "+", "y", "et", ".Y", "&"... Lo mejor es buscar la ayuda de la base de datos que estés consultando para ver cómo se emplea ahí.

Resultado: se suele usar para hacer búsquedas precisas. Es decir que suele reducir el número de documentos recuperados.

Ejemplo: para recuperar todos los documentos que traten sobre el turismo en España, la ecuación sería: Turismo AND España, y el resultado sería el conjunto de documentos que tengan ambos términos, marcados con morado en el gráfico.



Consejos de uso: si te fijas, cuantos más términos emplees en la ecuación de búsqueda unidos con este operador, menos documentos recuperarás. De hecho, si pones demasiados términos unidos con este operador llegará un momento en el que no recuperes nada, porque habrás indicado demasiadas condiciones en tu consulta. Por lo tanto, deberías usar este operador con cuidado y no excederte en su uso.



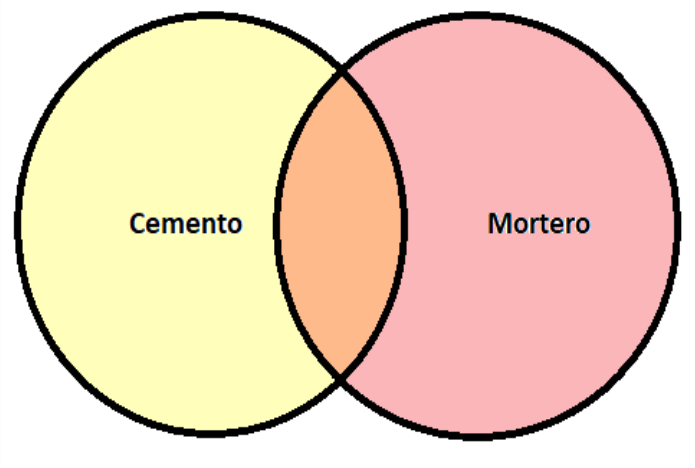
Operador de unión

Función: cuando se usa se logra que el sistema recupere todos los documentos que contengan, al menos, uno de los términos usados. Evidentemente, también recuperará aquellos documentos que contienen ambos.

Símbolo: se suele emplear el símbolo OR, aunque también es posible que haya sistemas que empleen “O”, “|”, “.O”, o cualquier otro. Como en el caso anterior, lo mejor es buscar la ayuda de la base de datos que estés consultando para ver cómo se emplea ahí.

Resultado: su uso recomendado es para hacer búsquedas exhaustivas, es decir, que amplía el número de documentos recuperados.

Ejemplo: para recuperar todos los documentos que traten sobre el cemento o el mortero, la ecuación sería: Cemento OR Mortero, y el sistema mostraría los documentos que tienen o bien uno de ellos, o bien el otro, o bien ambos.



Consejos de uso: este operador se puede emplear para buscar por todos los sinónimos de un mismo concepto (por ejemplo: Fotografía OR Instantánea) o para hacer búsquedas en las que convenga recuperar todo lo que se ha escrito sobre un tema.



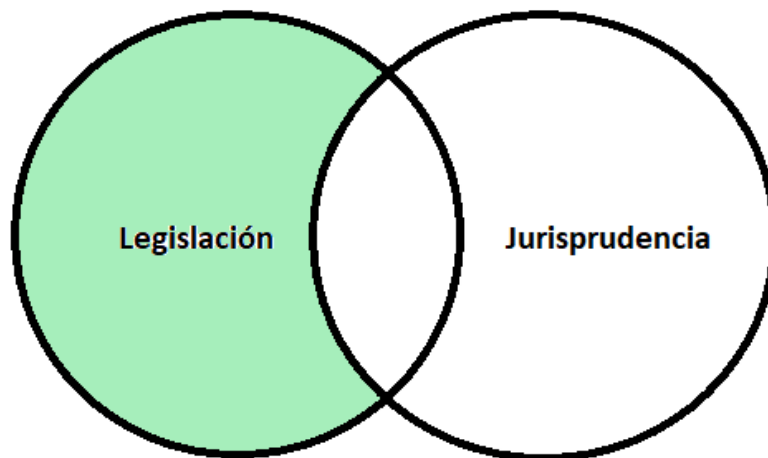
Operador de negación

Función: muestra los resultados que excluyen el término seleccionado.

Símbolo: se suele emplear el símbolo NOT, aunque también es posible que haya sistemas que empleen "NO", "AND NOT", "-", o cualquier otro. Como en el caso anterior, lo mejor es buscar la ayuda de la base de datos que estés consultando para ver cómo se emplea ahí.

Resultado: del conjunto de documentos recuperados eliminará aquellos que contengan el término que quieres excluir.

Ejemplo: Una búsqueda que intente localizar los documentos que traten sobre legislación, pero no los de jurisprudencia tendría como ecuación de búsqueda: Legislación NOT Jurisprudencia, y daría como resultado aquellos en los que solo aparece la palabra Legislación,



Consejos de uso: si estás consultando una base de datos a texto completo, el uso de este operador es muy restrictivo, ya que corres el riesgo de perder documentos que sí hablan de lo que interesa pero que, por la propia forma de escribir de los autores, no recuperarás.



Algo que hay que tener muy en cuenta es que todos los sistemas de recuperación de información suelen emplear un operador booleano por defecto. Esto significa que todos los términos de búsqueda que se escriban se unirán entre sí empleando dicho operador, salvo que se especifique lo contrario.





[Todo](#)
[Noticias](#)
[Shopping](#)
[Imágenes](#)
[Libros](#)
[Más](#)
[Configuración](#)
[He](#)

Aproximadamente 9.290.000 resultados (0,38 segundos)

Diccionario de sinónimos y antónimos - WordReference.com

<https://www.wordreference.com> › [sinonimos](#)

WordReference.com: un **diccionario** de 200.000 **sinónimos** y antónimos online.

Sinónimo

sinónimo - sinónimos de 'sinónimo'
en un diccionario de 200.000 ...

Buscar

buscar - sinónimos de 'buscar' en
un diccionario de 200.000 ...

Diccionario de sinónimos y ...

Más de 30.000 entradas. Más de
200.000 sinónimos y antónimos ...

Inglés

inglés - sinónimos de 'inglés' en un
diccionario de 200.000 ...

En Google el operador booleano que se emplea por defecto es el de intersección. Eso significa que cada vez que se introducen dos palabras sin ningún operador entre medio, el sistema añade este operador. Así que la búsqueda que aparece en la imagen es lo mismo que escribir diccionario AND sinónimos.

3. LOS OPERADORES DE PROXIMIDAD

Estos operadores no establecen relaciones semánticas entre los términos, como sucede con los booleanos. Más bien se centran en especificar cómo de cerca deben estar dos palabras entre sí en el documento que se recupere. Es decir, que en realidad no son más que un operador de intersección (AND) con alguna limitación especificada. No tienen el mismo grado de normalización que los booleanos, lo que implica que no todas las bases de datos los usen y, si lo hacen, es posible que varíe ligeramente la función de la aquí explicada.

A continuación, veremos los más extendidos:



Operador SAME

Cuando lo usamos, en esencia, estamos diciendo: “quiero que recuperes documentos que tengan estas dos palabras en el mismo párrafo o en el mismo campo, pero me da igual cuál aparezca primero”. Por lo tanto, el ejemplo:

Resumen: cáncer SAME tratamiento

ejecutará un AND entre ambas palabras, pero con la condición de que deben aparecer en el mismo párrafo o en el mismo campo, dando igual si primero está cáncer o tratamiento.

En la base de datos Web of Science (que se explicará en el apartado dedicado a fuentes de información para Ciencia y Tecnología) se ha realizado la búsqueda: TÍTULO: (cancer SAME treatment). Si le echamos un vistazo a alguno de los resultados se observa cómo ha recuperado el 9, que incluye los dos términos en ese orden, y el documento 10 que tiene ambos términos, en orden cambiado y con palabras entre ambos:

9. **The iron chelator desferrioxamine synergizes with chemotherapy for cancer treatment.**
- Por: Wang, Lingjuan; Li, Xiaoqing; Mu, Yanxi; et ál.
Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS) Volumen: 56 Páginas: 131-138 Fecha de publicación: 2019-Dec (Epub 2019 Aug 19)
- Texto completo de la editorial Ver abstract ▼
10. **More options for adjuvant treatment of HER2-positive breast cancer: How to choose wisely?**
- Por: Trapani, Dario; Rajasekar, Arun K A; Mathew, Aju
International journal of cancer Volumen: 145 Número: 11 Páginas: 2901-2906 Fecha de publicación: 2019-Dec-01 (Epub 2019 May 31)
- Texto completo de la editorial Ver abstract ▼

Operador W/n y NEAR/n

En ambos casos la n sirve para representar el número máximo de palabras que puede haber entre los dos términos de búsqueda. Por ejemplo:

cepillo W/1 dientes



Servirá para recuperar aquellos documentos que contengan las palabras cepillo y dientes separadas entre sí por hasta un máximo de 1 palabra, con lo que recuperará documentos que hablen de cepillo de dientes. No conviene ser ni demasiado severo ni excesivamente generoso en el número indicado. Así, en nuestro ejemplo hemos sido algo rácanos, ya que poniendo un 1 perdemos documentos relevantes, como “cepillo para los dientes” o “el cepillo y los dientes”, ya que tienen más de una palabra separando ambos términos. Tampoco deberíamos usar un número muy grande, ya que cuantas más palabras separen nuestros términos de búsqueda más opciones tendremos de que los documentos recuperados no traten el tema que nos interesa.

Hay sistemas de recuperación que al operador W (que viene de With) lo denominan NEAR. Independientemente de cómo lo llamen, siempre hará lo mismo. Es importante destacar que con el uso de este operador, además de especificar la cantidad de palabras que debe haber entre las dos seleccionadas, estas pueden aparecer en cualquier orden.

Operador PRE/n

Con este operador se consigue aumentar la precisión, ya que además de especificar el número máximo de palabras que pueden separar los dos términos de búsqueda (sustituyendo “n” por un número, al igual que en el ejemplo anterior) es necesario que se respete el orden en el que se han escrito en la ecuación de búsqueda.

cadenas PRE/1 polímeros

4. OPERADORES DE DELIMITACIÓN O PROXIMIDAD

Se trata de operadores de una gran importancia, ya que nos permiten realizar “búsquedas a la carta”. En realidad, los símbolos que se emplean para estos operadores son los nombres de los campos que se han usado en la base de datos y, en concreto, las etiquetas que ellas emplean para identificar cada campo. Por esa razón también se conocen como “búsqueda por campos”.

Por ejemplo, en algunas bases de datos la etiqueta AB se emplea para referirse al campo que contiene el resumen (abstract en inglés) del artículo. Si se desea localizar aquellos documentos en los que aparezca una palabra concreta solo en el campo resumen, la siguiente ecuación de búsqueda

AB=cefalópodos

Recupera todos los documentos que contengan dicha condición.

Es posible combinar diferentes campos con múltiples operadores, aunque lo más lógico suele ser hacerlo con los booleanos. Así, para localizar documentos que tengan en el resumen una palabra concreta y en el título otra, la ecuación podría ser:

AB=mediterráneo AND TI=cefalópodos



Los campos más empleados para realizar este tipo de consultas suelen ser el de título (TI), resumen (AB), palabra clave (KW) y fecha de publicación (PY).

Con el fin de facilitar el proceso de búsqueda muchas bases de datos incluyen un formulario donde ya aparecen estos campos en su forma desarrollada, y no con la abreviatura, tal y como se muestra en la imagen de abajo, que pertenece a la base de datos Scopus, que se explicará en este MOOC, más adelante.

Scopus Search Sources Alerts Lists Help v SciVal x Create account > Sign in

Document search

Documents Authors Affiliations [Advanced](#)

Search Abstract

E.g., "Cognitive architectures" AND robots

AND

Search Article title

[> Limit](#)

[Reset form](#)

4. SÍMBOLOS RESERVADOS

El nombre se ha tomado como influencia de los lenguajes de programación. En esencia, lo que se hace es reservar el uso de una serie de símbolos para que se empleen solo en una cosa concreta. Son las comillas, los paréntesis y los comodines. Esto significa que se usan en los sistemas de recuperación de información para una labor en concreta, por lo que no podremos emplearlos para otra cosa.

4.1. Comillas

Las palabras que se agrupan dentro de unas comillas serán tratadas como una cadena única de caracteres. Eso quiere decir que es como si todas esas palabras se unieran con un operador booleano AND y, además, se respetara la forma y el orden en el que se han escrito en el casillero de búsqueda.



"juego de tronos"

Recuperará todos los documentos en los que aparezcan esas tres palabras, en ese mismo orden y escritas de la misma forma.

Lo ideal es emplear las comillas en casos muy concretos:

Expresiones literales: "ojo por ojo"

Nombres de personas: "John F. Kennedy"

Nombres de instituciones: "Organización Mundial de la Salud"

Títulos: "El capitán Alatriste"

Expresiones hechas: "frutos secos"

En general, cualquier conjunto de palabras que siempre se escriben de la misma forma.

La gran ventaja que tiene su uso es que permiten restringir mucho la búsqueda, aumentando la precisión. Sin embargo, hay que tener mucho cuidado en su uso, ya que si se escribe mal cualquier palabra dentro de la expresión entrecomillada no se recuperará ningún documento.

4.2. Los comodines

Se trata de un mecanismo muy eficaz para no perder información, ya que permiten buscar por la raíz de un término. Habitualmente se emplean el asterisco (*) para sustituir una cadena indeterminada de caracteres y un símbolo de cierre de interrogación (?) para un único carácter.

Así, con

naranj*

Se recuperarían todos los documentos que tengan cualquiera de las palabras naranja, naranjas, naranjal, naranjo, naranjada, naranjales... Es decir, todos los que comparten la misma raíz y, a partir de ella, las palabras que sean. Se supone que por compartir raíz también comparten significado, por lo que es un método muy sencillo para hacer búsquedas exhaustivas.

Por su parte, el símbolo de interrogación se emplea para sustituir un único carácter. Por ejemplo:

Médic?

Recuperaría documentos en los que aparezcan las médico y médica.



4.2. Los paréntesis

Al igual que sucede en las matemáticas, con los paréntesis se especifica cuál es la operación que queremos que se realice primero. Esto está provocado porque en los sistemas de recuperación de información existe un orden a la hora de ejecutar una ecuación de búsqueda que esté formada por varios operadores diferentes.

En todas las bases de datos, el primer operador que se ejecuta es NEAR (o W), luego SAME, luego PRE y a continuación los operadores booleanos NOT, AND y OR.

Imagina que deseas recuperar documentos sobre la fotografía en documentales o películas, y formulas la ecuación de búsqueda de la siguiente forma:

Fotografía AND documentales OR películas

Teniendo en cuenta lo que hemos comentado anteriormente, el sistema recuperará primero el AND, así que no localizará todos los documentos que traten sobre fotografía y documentales. A continuación, buscará todo lo que trate de películas (independientemente de si es fotografía, historia del cine, movimientos artísticos, directores, etc.). Es decir, que no tiene nada que ver con lo que necesitamos.

Para realizar correctamente esa búsqueda necesitamos romper con el orden de ejecución de los operadores establecido por el sistema. Y justo para eso se emplean los paréntesis, quedando la cosa así:

Fotografía AND (documentales OR películas)



7.- BIBLIOGRAFÍA

- **Codina, Lluís (2018).** “Sistemas de búsqueda y obtención de información: componentes y evolución”. Anuario ThinkEPI, v. 12, pp. 77-82.
- **Cid-Leal, Pilar; Perpinyà i Morera, Remei (2013).** *Cómo y dónde buscar fuentes de información.* Universitat Autònoma de Barcelona, Servei de Publicacions. ISBN 978-84-490- 3896-9
<https://dialnet.unirioja.es/servlet/extlib?codigo=729567>
- **Fuentes Agustí, Marta (2009).** *Aprender a buscar y seleccionar información digital: una experiencia didáctica de enseñanza infusionada.* Aula de innovación educativa, ISSN 1131-995X, Nº 181, 2009, págs. 22-26.
<https://dialnet.unirioja.es/servlet/extart?codigo=2960303>
- **Gómez Díaz, Raquel (2003).** *La evaluación en recuperación de la información* [en línea]. "Hipertext.net", núm. 1 <<http://www.hipertext.net>>
- **Rouet, Jean François; Potocki, Anna (2018).** *De la lectura a la alfabetización documental: aprender a buscar, evaluar e integrar información de diversos textos.* Infancia y Aprendizaje: Journal for the Study of Education and Development, ISSN 0210-3702, ISSN-e 1578-4126, Vol. 41, Nº 3, 2018, págs. 429-446
<https://dialnet.unirioja.es/servlet/extart?codigo=6540965>