

Sintaxis del lenguaje de consulta
LUCENE

Las expresiones de búsqueda pueden ser tan sencillas como simplemente escribir una palabra o bien, estructuras complejas que incluyan paréntesis y operadores lógicos. Por eso es necesario comprender como es la sintaxis a la hora de escribir una consulta.

- **Términos**

Cada consulta ejecutada es dividida en términos, de los cuales hay de 2 tipos: términos simples o frases. Los términos simples son palabras tales como nota, computadora, expediente, etc, etc. Las frases son un grupo de palabras encerradas entre comillas como por ejemplo "jefe departamental". En las frases, el orden en el que se escriben las palabras es importante, ya que se buscarán coincidencias exactas.

Los términos podrán combinarse también con operadores lógicos para formar una consulta más compleja.

- **Modificadores**

El sistema admite la modificación de los términos de la consulta para poder ampliar las opciones de búsqueda. Dichas modificaciones pueden hacerse a través de:

Comodines: es posible reemplazar uno o más caracteres de una palabra (no de una frase) por un comodín genérico. Para realizar un reemplazo de un solo carácter se utiliza el símbolo "?". Para realizar un reemplazo de varios caracteres, se utiliza el símbolo "*".

El "?" busca términos en los cuales coincida la posición del símbolo con el carácter reemplazado. Por ejemplo, para buscar "casa" o "caza" (también quedarían incluidas caja, cama, cala, etc) se escribe la siguiente consulta:

ca?a

También pueden usarse más de un símbolo en el mismo término, como por ejemplo e?ei?a, lo que dará como resultado **ezeiza**.

Por el contrario, el comodín "*" busca cero o más caracteres. La consulta del término gob* traerá como resultado todas las palabras que empiecen con los caracteres gob y terminen de cualquier manera: gobierno, gobernantes, gobernar, etc, etc.

También puede aplicarse en el medio de una palabra: *san**o traerá como resultados a palabras como: santo, santuario, sanitario, santiago, etc.

NO PUEDEN UTILIZARSE “?” NI “*” AL COMIENZO DE LA PALABRA.!!!

- **Búsquedas difusas**

El sistema ofrece la posibilidad de hacer búsquedas que permitan encontrar términos parecidos en cuanto a su deletreo con respecto a la palabra original escrita en la consigna de búsqueda. Para ello simplemente hay que agregar el carácter “~” al final de cada una de las palabras que queremos ampliar. Este símbolo puede escribirse con la combinación de teclas [Alt] + 126.

Por ejemplo, si se está buscando información sobre el aeropuerto de Ezeiza o el Aeroparque pero no sabemos como se escribe correctamente entonces podemos escribir el nombre como nos suene y hacer una búsqueda como esta: *eceisa~* o *niuberi~*. Esto traerá como resultado documentos que incluyan términos parecidos a ese, entre los cuales estará “Ezeiza” o “Newbery”.

También se puede ajustar el grado de similitud que se desea con respecto al término original. Ese grado de similitud es un número entre 0 y 1, siendo 1 la mayor similitud. Si no se especifica ningún número, el valor por defecto es 0,5. Para indicarlo, simplemente se escribe el número luego del símbolo “~”.

Si a la misma búsqueda anterior la hago pero con un valor mayor a 0,5 (que es el valor por defecto) me arrojará otros resultados. Por ejemplo *niuberi~0.7*

- **Búsquedas por proximidad**

También es posible encontrar palabras que estén separadas por una distancia específica. Para hacer una búsqueda de este tipo se utiliza el mismo símbolo que la búsqueda difusa (“~”) pero puesto a continuación de una frase.

Por ejemplo, se puede hacer una búsqueda cómo la siguiente: *"tareas mantenimiento ezeiza"~20*. Esto traerá resultados del tipo:

“...Aer. Ezeiza. Solicita al Concesionario un cronograma para la realización de las tareas de mantenimiento del Cerco perimetral conforme el informe de inspección AVSEC...”

“... Solicita al concesionario para que en el plazo de 10 días remita un cronograma para la realización de tareas de mantenimiento del cerco perimetral, que incluya plazo de ejecución en el Aer. Ezeiza...”

- **Búsquedas por rango**

Las búsquedas por rango permiten encontrar documentos cuyos valores en el campo especificado estén entre los límites inferior y superior del rango expresado. Los límites de estos rangos pueden ser inclusivos o exclusivos. Para que sea inclusivo, se debe expresar el rango entre símbolos de corchete “[“ y ”]”. Para que sea exclusivo, se expresa entre llaves “{“ y }”. Entre ambos valores se escribe la palabra “TO”.

Por ejemplo, una consulta por un rango de números de dictamen que incluyan los límites sería:

[0010 TO 0026]: tanto la nota 0010 como la 0026 están incluidas en el resultado.

También se pueden poner palabras para expresar un rango. Por ejemplo, la consulta por [blanco TO negro] incluiría todas las palabras entre blanco y negro teniendo en cuenta un orden lexicográfico.

- **Operadores lógicos**

Los operadores lógicos permiten combinar a los términos de una búsqueda. Existen 3 operadores disponibles: AND, OR y NOT. Deben escribirse siempre en mayúsculas.

AND

El operador “AND” indica que **todos** los términos de la consulta deben aparecer en la misma página del documento. Por ejemplo, la consulta *bienes AND servicios* traerá documentos que contengan ambas palabras.

OR

El operador “OR” es el operador por defecto del sistema. Esto quiere decir que si no se especifica ningún operador entre dos palabras, entonces el sistema asume un OR. Indica que **cualquiera** de los términos buscados debe aparecer en el documento.

NOT

Este operador **excluye** los documentos que contengan el término precedido por NOT. No puede usarse cuando la consulta posee un solo término. Por ejemplo, la consulta *NOT "jefe departamental"* no es válida ya que solo incluye un término o frase.

- Agrupamiento

Es posible dividir una consulta en subconsultas más específicas usando paréntesis "(" y ")". Esto es muy útil cuando se quiere controlar la lógica de ejecución de una búsqueda. Por ejemplo si quiero encontrar documentos sobre las licitaciones públicas (pero no las privadas) para el servicio de limpieza puedo hacer una consulta como la siguiente:

licitación AND (pública NOT privada) AND limpieza

De esta manera descarto cualquier documento que diga licitación privada y limpieza.

Todas las opciones anteriores pueden combinarse de alguna manera, haciendo que la consulta sea más específica o más amplia. Por ejemplo, en la siguiente consulta se agrupan términos, se usan operadores lógicos y además se hace una búsqueda difusa para el nombre del aeropuerto.

licitación AND (pública NOT privada) AND niuberi~

encontrará todos los documentos que sean de una licitación pública relacionados al aeroparque Jorge Newbery (o algo similar a niuberi).