



Tus dudas son nuestras dudas

Búsqueda de texto avanzada con DocFetcher

Por **Comisión de Recursos Tecnológicos**

¿Qué es DocFetcher?

DocFetcher es un programa de búsqueda avanzada que nos permite buscar palabras, frases o cualquier cadena de texto que esté dentro de los documentos que tenemos en nuestra computadora. Es software libre y gratuito. Además, viene en tres versiones: una versión para Windows, otra para GNU-Linux y una tercera portátil que puede ejecutarse tanto en Windows como en GNU-Linux sin necesidad de instalación. DocFetcher utiliza Java, por lo que se necesita tener instalado Java Runtime Environment (JRE) 1.6 o posterior para que funcione.



¿Qué puede buscar?

DocFetcher puede buscar dentro de los siguientes tipos de archivo: HTML, texto plano (txt), PDF, Microsoft Office (doc, xls, ppt), Microsoft Office 2007 (docx, xlsx, pptx), ODF (odt, ods, odg, odp), texto enriquecido (rtf), AbiWord (abw, abw.gz, zabw), Microsoft Compiled HTML Help (chm), Microsoft Visio (vsd) y Scalable Vector Graphics (svg).

¿En qué se diferencia de otros programas de búsqueda?

Dos cosas diferencian a DocFetcher de otros programas de búsqueda como Google Desktop:

- 1) Nos permite indizar solamente lo que necesitamos y no todo el disco rígido.
- 2) Solamente indiza documentos y omite archivos de imagen, música y video.

¿Por qué es bueno esto?

Cuanto más grande sea el campo de búsqueda, habrá más coincidencias superfluas en los resultados de la misma. Es por esta razón que DocFetcher nos brinda mejores resultados que otros programas similares de su categoría. Además, una cantidad menor de archivos a indizar redundan en una indización más rápida y un menor procesamiento de nuestra computadora, por lo que nuestra máquina no se ralentizará en absoluto, como sí ocurre con otros programas.

El secreto de Google y de otros motores de búsqueda similares radica en la indización de documentos, lo que ofrece una muy buena aproximación del contenido del documento. No obstante, la indización de imágenes o videos no resulta útil, ya que no

podemos acceder al contenido de una imagen o un video mediante la extracción de texto.

Por lo tanto, si quitamos la necesidad de buscar imágenes, música y videos, nos quedan los documentos de texto de todo tipo, y es justamente eso en lo que se destaca DocFetcher y lo que nos compete a los traductores.

Por lo tanto, a todos los traductores que quieran una herramienta poderosa, libre y gratuita que haga solamente lo que se supone que debe hacer, de manera rápida y eficaz, sin violar su privacidad y sin instalar cosas adicionales de dudosa procedencia, les recomendamos este programa.

¿Cómo hago para...?

¿...crear un índice?

Hagan clic con el botón derecho en el panel "Search Scope" de la izquierda de la pantalla, seleccionen "Create Index" y elijan una carpeta a indizar (también pueden arrastrar una carpeta y soltarla en el panel "Search Scope"). Luego de seleccionar la carpeta, aparecerá una ventana de configuración. La configuración que aparece de manera predeterminada será suficiente, así que hagan clic en "Run" y esperen hasta que DocFetcher haya terminado de indizar los documentos. El proceso de indización puede tardar algunos minutos si la carpeta posee muchos archivos, pero normalmente tarda pocos segundos al indizar carpetas pequeñas. Tengan en cuenta la siguiente proporción: Para indizar 200 documentos tarda aproximadamente 1 minuto.

¿...efectuar una búsqueda?

Ingresen la/s palabra/s a buscar en el cuadro de texto que aparece sobre el panel de resultados y presionen la tecla Intro. Los resultados aparecerán en el panel de resultados. DocFetcher listará todos los documentos que se encuentren dentro de la/s carpeta/s previamente seleccionada/s e indizada/s y que contengan esa/s palabra/s. Normalmente tarda 1 segundo en mostrar los resultados.

Podemos navegar a través de las páginas utilizando los botones de flechas que se encuentran a la derecha del cuadro de texto en donde ingresamos la/s palabra/s buscadas.

¿...filtrar los resultados?

En la casilla "Minimum/Maximum Filesize" podemos ingresar las cifras deseadas para excluir de la búsqueda los archivos cuyo tamaño sea inferior al mínimo o superior al máximo allí especificados. También podemos tildar o destildar los elementos del panel "Filetypes" y "Search Scope" para excluir ciertos tipos de archivo o archivos de carpetas específicas.

¿...ordenar los resultados?

Hagan clic en los encabezados de las columnas del panel de resultados para ordenarlos de acuerdo al criterio escogido.

¿...tener una vista previa de los archivos?

Debajo del panel de resultados se encuentra el panel de vista previa. En él podemos observar una vista previa en modo "sólo texto" en donde se resaltan los términos buscados. En el caso de los documentos HTML, se observa un simple navegador web.

¿...abrir un archivo?

Hacemos doble clic en uno de los elementos del panel de resultados. De manera predeterminada, tras abrir un archivo, DocFetcher se minimizará automáticamente en la bandeja del sistema. DocFetcher también se minimizará en la bandeja del sistema si hacemos clic en el botón Minimizar de la ventana del programa. Para volver a maximizarlo, debemos hacer clic en el ícono de la bandeja del sistema o presionar el atajo de teclado Ctrl + F8.

¿...actualizar los índices?

Si agregamos, quitamos o modificamos los documentos de las carpetas ya indizadas, debemos actualizar dichos índices. Esto puede hacerse automáticamente de dos maneras:

- 1) Si DocFetcher está ejecutándose, detecta los cambios y actualiza los índices de manera inmediata.
- 2) Si DocFetcher no está ejecutándose, los cambios son detectados por un pequeño programa que se ejecuta en segundo plano. Luego, los índices afectados son actualizados la próxima vez que iniciemos DocFetcher.

¿...indizar un disco completo?

La respuesta es simple: No debemos indizar discos completos. Al contrario de lo que ocurre con otras aplicaciones de búsqueda integral como Google Desktop, DocFetcher no fue diseñado para indizar discos rígidos completos (por ejemplo C:\). No lo intenten, ya que perderían mucho tiempo en crear los índices y no sería útil en absoluto. DocFetcher fue diseñado para que podamos indizar solamente lo que necesitamos, lo que nos ahorra tiempo y espacio de disco y nos ofrece resultados mucho más precisos.

¿...utilizar sintaxis de búsqueda?

A continuación veremos las principales expresiones que podemos utilizar:

- **Operadores booleanos:** El operador predeterminado es OR. Es decir, si ingresamos perro gato, DocFetcher buscará documentos que contengan una o ambas palabras. Para buscar documentos que contengan ambas palabras, debemos ingresar perro AND gato. Para excluir una palabra, utilizamos el operador NOT o - (signo menos) delante de él.
- **Términos requeridos:** Para indicar que el documento a buscar debe contener un término específico, utilizamos el operador + (signo más), como se ve a continuación: +perro gato.
- **Búsqueda de frase:** Si queremos buscar perro gato en ese orden específico, debemos tipear "perro gato" (entre comillas).
- **Búsqueda en campos específicos:** Podemos buscar en los campos "contents" (contenido), "filename" (nombre de archivo), "title" (título) y "author" (autor).

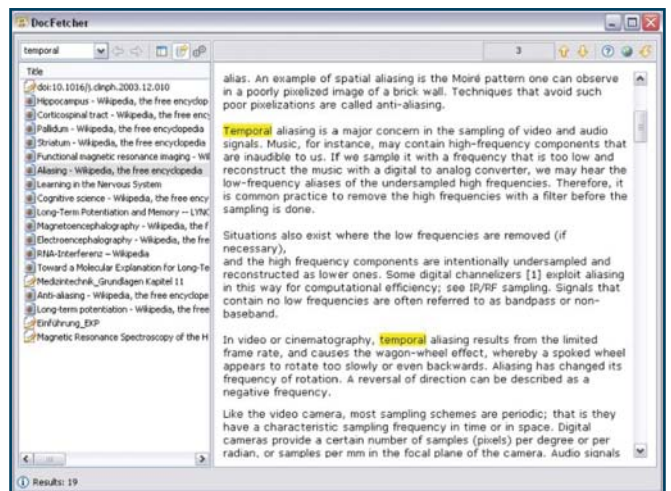
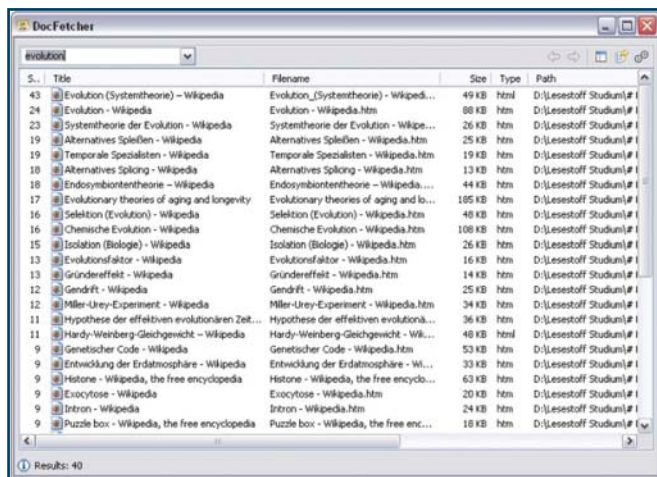
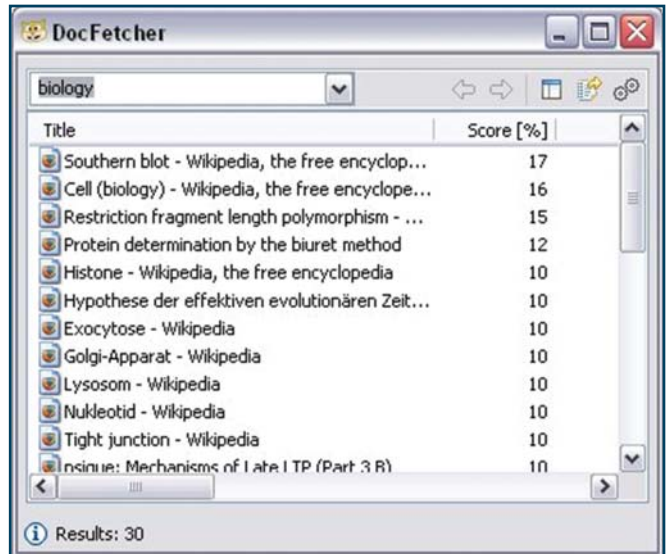
El campo "contents" (contenido) es el campo de búsqueda predeterminado. No obstante, si queremos buscar un documento cuyo título contiene "perro", debemos ingresar title:perro. Para un documento cuyo título tenga "perro" y/o "gato", debemos ingresar title:"perro gato".

- **Comodines:** El signo de pregunta (?) equivale a un solo carácter desconocido. El asterisco (*) equivale a un número desconocido de caracteres desconocidos. Las cadenas de búsqueda no pueden comenzar con ? ni *. Ejemplos: Luc* nos mostrará Luc, Luca, Lucas, Lucrecia, Lucía, Luciano, etcétera. Por otro lado, Luc? solamente nos mostrará Luca.
- **Términos preferenciales:** También podemos asignarles ciertos "pesos" a las palabras. De esta manera, podemos aumentar o reducir el nivel de coincidencia dentro de un documento específico en el que se encuentre/n dicha/s palabra/s. Esto nos permitirá modificar el orden de relevancia del panel de resultados. Ejemplo: perro^4 gato nos mostrará los documentos que contengan "perro" en la parte superior del panel de resultados.

¿...omitir archivos de la indización mediante el uso de expresiones?

DocFetcher nos ofrece la posibilidad de omitir ciertos archivos al momento de la indización mediante el uso de expresiones específicas. Dichas expresiones deben ingresarse en el cuadro "Skip files" del cuadro de diálogo "Indexing Options". Veamos algunos ejemplos:

- **Excluir todos los archivos de MS Excel:** *.xls
El signo "." (punto) equivale a un carácter desconocido. El signo "*" equivale a "el carácter precedente, cero o más veces". Por lo tanto, en este caso ".*" equivale a una serie de cero o más caracteres desconocidos. Como "." ya está utilizado como carácter especial, cuando queremos buscar el punto en un nombre de archivo, debemos hacer una salvedad colocando una "\" (barra invertida) delante de él. Los caracteres A-Z, a-z y 0-9 equivalen a su significado habitual. Esa es la razón por la cual "*.xls" buscará todos los nombres de archivo de cualquier longitud que finalicen con un punto y la secuencia "xls".
- **Excluir archivos que contengan textos de licencias:**
licencia.txt \$ Licencia.txt \$ LICENCIA.txt
El signo "\$" se utiliza para combinar dos o más expresiones. Como resultado, esta orden excluirá todos los archivos con los nombres "licencia.txt", "Licencia.txt" y "LICENCIA.txt". Como ya vimos, debe hacerse la salvedad del punto ya que es un carácter especial.
- **Buscar una secuencia de dígitos:** diario[0-9]*.doc
Supongamos que queremos excluir un conjunto de documentos de MS Word que comienzan con "diario" y terminan con una fecha, como diario2008.doc. Por otro lado, no queremos excluir archivos como "diariodeviaje.doc". Es obvio que una expresión como "diario*.doc" no funcionará. La solución consiste en reemplazar el "." con "[0-9]". Mientras que "." equivale a cualquier carácter, "[0-9]" equivale a un dígito. Por lo tanto, "[0-9]*" equivaldrá a una secuencia de dígitos.



¿Cómo utilizar la función "Buscar y reemplazar" de MS Word de manera avanzada?

Por **Comisión de Recursos Tecnológicos**

Podemos ahorrarnos mucho tiempo al traducir textos legales, financieros y técnicos si utilizamos adecuadamente ciertas fórmulas de búsqueda y reemplazo.

Quizás no todos conozcan las funciones avanzadas de "Buscar y reemplazar" de Microsoft Word. Sin embargo, esta opción puede ser muy útil en el trabajo del traductor. Para acceder a ella, debemos ir a **Edición >> Reemplazar**. Luego aparecerá la ventana de la **fig. 1**

Una vez allí expandimos las opciones mediante el botón **Más** y activamos la casilla **Usar caracteres comodín**, como se observa en la **fig. 2**

Esta característica nos permitirá definir condiciones de búsqueda complejas al utilizar combinaciones de caracteres especiales.

IMPORTANTE: Todas las opciones avanzadas que veremos en esta nota sólo funcionarán si dicha opción se encuentra tildada.

Los símbolos a utilizar están detallados en la ayuda de Word, por lo que no los mencionaremos aquí. No obstante, para demostrar el uso de esta opción utilizaremos cinco ejemplos sencillos.

Nociones preliminares

- Si seleccionamos una porción del texto, la búsqueda se limitará a dicha porción del documento. Word buscará en esa porción y luego nos preguntará si deseamos seguir buscando en el resto del documento. De lo contrario, buscará en todo el documento.
- El reemplazo puede ejecutarse haciendo clic en **Edición >> Reemplazar** o con el atajo de teclado **Ctrl + L**. Lue-