

Metodologías utilizadas en la Búsqueda de Documentos

Nuestro propósito con este estudio es el de explicar los distintos métodos de que disponemos para buscar y recuperar documentos en las Empresas y Organizaciones. Los modernos Sistemas de Gestión Electrónica Documental (SGED) han demostrado que la búsqueda eficiente de la documentación se puede mejorar enormemente respecto a un sistema de archivo manual basado en papel.

La mayoría de los SGED usan uno o varios de los siguientes métodos:

- Búsqueda estructurada, también llamada jerárquica
- Búsqueda por palabras clave
- Búsqueda por metadatos
- Búsqueda por contenido, también conocida como de “texto completo”

Cada uno de los métodos tiene sus ventajas y desventajas respecto a los otros, dependiendo de la aplicación específica. Este estudio explica cada uno de estos métodos y presenta las ventajas e inconvenientes de cada uno.

Analizaremos también algunas de las consideraciones clave que se deben tener en cuenta a la hora de decidirse por el mejor sistema de búsqueda y recuperación para su compañía.

Le propongo un juego: “encontrar la carpeta”

Antes de profundizar en la parte tecnológica de la búsqueda de los documentos, es importante considerar el modo en que se consultan los documentos. La mayoría de las Organizaciones almacenan los documentos en tres o cuatro localizaciones diferentes: archivos físicos centralizados, archivos físicos en las oficinas, carpetas en unidades de disco compartidas y carpetas en unidades locales de disco. Muchas de estas compañías intentan centralizar el archivo en un servidor con espacio suficiente, eliminando la posibilidad de almacenar documentos importantes en los discos locales y en los archivadores de las oficinas. Este es generalmente el primer paso para llevar a cabo la transición a un SGED.

Por ejemplo, el director de Recursos Humanos, Juan, recibe una petición urgente de una empresa de seguros solicitando un documento del historial de un empleado. Este sería el proceso:

- Juan va al archivo de personal.
- Juan tiene la llave del archivo, de modo que lo abre y busca la carpeta del empleado.
- Suponemos que la última persona que necesitó la carpeta la dejó bien colocada, de modo que Juan la encuentra.
- Juan cierra el archivo con llave.

- Juan lleva la carpeta al lugar donde se encuentra la máquina de fax y envía el documento solicitado a la compañía de seguros.
- Juan regresa al archivo.
- Juan abre la puerta y deja de nuevo la carpeta donde la había encontrado.
- Juan cierra el archivo.
- Juan regresa a su despacho y continúa trabajando.

Esta parece la secuencia de una película de cine, de modo que Usted ya habrá imaginado que en la vida real, el proceso de recuperación nunca se desarrolla de forma tan sencilla. Hay incontables obstáculos que aparecerán inevitablemente a lo largo del proceso. Y estos comienzan incluso antes de producirse el primer paso. Juan recibe otra llamada urgente o alguien se presenta en su despacho de forma inesperada, de modo que el viaje al archivo se pospone o se olvida.

¿Cuáles son los otros posibles obstáculos que afectan a la velocidad de la recuperación? Estos son algunos de los más habituales:

- Juan no encuentra la llave.
- Juan dedica algún tiempo a buscar la carpeta, pero esta no aparece. Alguien se la ha llevado antes que él, o ha sido mal archivada.
- Juan se tropieza con Ana y toman un café mientras discuten asuntos importantes o quizá al argumento de la película que vieron el pasado fin de semana.
- Juan deja la carpeta en la sala de fax y se distrae. Un empleado descontento descubre que contiene información privada y se lo lleva. Se produce cierto caos y algunas denuncias contra la empresa.
- Juan devuelve el expediente, pero se equivoca al guardarlo, provocando que la próxima persona que lo necesite no sea capaz de encontrarlo.
- Juan olvida el objetivo del viaje al archivo y decide ir a fumar un cigarro.

¿Es esto es una exageración? Probablemente no. Seguramente Usted sabe que esto ocurre cada día. Como un evento aislado, no parece un caso grave por el que preocuparse (excepto del asunto legal). Pero no es un caso aislado. Este es un proceso que se repite por múltiples personas dentro de su Organización cada día. Incluso en un escenario en el que todo estuviera archivado correctamente, se perdería un tiempo precioso archivando y recuperando documentos.

Esta es la razón por la que existen tantos productos de GED y distribuidores de los mismos dedicados a ayudar que las Empresas y las Organizaciones funcionen de forma más eficiente. Los SGED resuelven la ineficiencia asociada a los procesos de archivo basados en papel. Y el mayor impacto financiero de los entornos que adoptan un SGED es el de reducir los costos laborales. El tiempo empleado en archivar, recuperar y procesar documentos se reduce normalmente en un 50% o más. El mayor ahorro de tiempo se produce en el tiempo dedicado a la búsqueda y la recuperación.

Búsqueda estructurada

Este es el método más cercano a la búsqueda en el archivo tradicional, ya que está basado en una estructura jerárquica de archivo, carpeta y documento. Los mejores sistemas del mercado son capaces de emular el mundo del archivo físico para hacer más sencilla la transición al Sistema de Gestión Electrónica Documental a todas aquellas personas familiarizadas con conceptos como archivadores, carpetas, etiquetas-índice y documentos grapados o unidos por clips. De modo que nuestro protagonista -Juan- puede ahora entrar en el sistema, encontrar y abrir una carpeta, identificar el documento implicado, y enviarlo por email, por fax o imprimirlo mientras su interlocutor está aun al teléfono.

La búsqueda estructurada depende en gran medida del interface que use el SGED. Un buen interface será intuitivo y requerirá apenas formación para buscar y recuperar documentos. Si un usuario puede visualizar la estructura del archivo y navegar por ciertos documentos con un número mínimo de clics del ratón e introducción de datos, entonces podemos asegurar que es un interface “productivo”.

Es importante que una búsqueda estructural proporcione un control similar sobre el acceso a los documentos. En el ejemplo citado anteriormente, solo el personal con la llave del archivo puede acceder a los registros personales. (Si es que no se ha dejado abierto). Un buen SGED debería extender el modelo de seguridad de la “llave del archivo” al nivel de carpetas y documentos. Esto permitiría controlar el acceso de los usuarios solo a aquellos documentos a los que tuvieran permiso.

Otros métodos permiten refinar más la búsqueda, pero el estructural es el más eficiente, pues necesita el menor número de pulsaciones. Si además el sistema nos permite previsualizar los documentos si necesidad de abrirlos, este método será aun más útil. Por ejemplo, si tenemos tres facturas digitalizadas con índices similares, podemos ahorrar mucho tiempo si podemos previsualizarlas para determinar en cuál de ellas estamos interesados.

La principal desventaja de la búsqueda estructurada es que puede llevar más tiempo archivar un documento en una ubicación determinada. La mayoría de los sistemas que ofrecen búsquedas estructuradas disponen de herramientas para archivar los documentos de la forma más eficiente posible reduciendo los tiempos de archivo. Y tenga siempre en cuenta que es más rápido escanear un documento que almacenarlo en un archivador físico.

No todos los sistemas tienen posibilidades de llevar a cabo búsquedas estructuradas. Intente disponer siempre de un interface orientado a carpetas.

Búsqueda por palabras clave

La búsqueda por palabras claves se puede utilizar bien junto a la búsqueda estructural o bien por sí misma. Definiciones de “palabras clave”:

- Entre las palabras que forman un título o forman parte de un documento, las más significativas o informativas sobre su contenido.
- Expresión abreviada de una sentencia.
- Palabra específica cuyo uso es esencial para el significado y la estructura de una oración o un programa.

Fuente: <http://www.wordreference.com>

Un Sistema de Gestión Electrónica Documental permite a los usuarios indexar los documentos con palabras clave. Dichas palabras pueden más adelante usarse en un campo de búsqueda para realizar una consulta y obtener una lista de resultados. Cuantas más palabras clave estén asociadas a un documento, más específica puede hacerse la consulta.

Mientras algunos SGED requieren de una entrada manual de palabras clave, y otros sistemas la automatizan. El modo más eficiente de usar palabras clave es con ambos modelos. Por ejemplo, supongamos que Juan, el atareado responsable de recursos humanos, necesita digitalizar y archivar una aprobación de vacaciones para una empleada llamada Sonia.

Si Juan puede localizar la carpeta de Sonia en el SGED y digitalizar el documento, cierta información sobre Sonia como el nombre completo y número de empleado ya estarán relacionados automáticamente, por lo que no habrá que teclearlos de nuevo. Es obvio que Juan puede añadir algunas otras palabras clave como “aprobado”, “personal”, etc., que ayudarán en el futuro a localizar este documento. Juan puede localizar ahora el documento de dos formas: buscando en la carpeta de Sonia un documento con el título “Aprobación de Vacaciones” o hacer una búsqueda por palabras clave usando “Sonia” y “Vacaciones”.

O mejor aún, Juan puede navegar a la carpeta de Sonia en el SGED y lanzar una consulta por palabras clave de “Vacaciones”. Esta flexibilidad permite a los usuarios encontrar el modo más rápido de llegar a un documento específico.

La mayor ventaja, y a la vez la mayor desventaja, del uso de palabras clave en la búsqueda de documentos es que da al usuario una gran libertad en la indexación y la búsqueda. Un planteamiento de SGED de este tipo puede ser arriesgado, ya que no hay dos personas que utilicen las mismas convenciones para indexar el mismo documento. Este modelo de indexación y búsqueda es más eficiente si se usa en el contexto de una búsqueda estructural. Además, la búsqueda por palabras clave no garantiza que se encuentre el documento que se necesita.

Si el personal que archivó originalmente el documento no añadió palabras clave, este tipo de búsqueda no arrojaría ningún resultado. Por esto es por lo que un SGED debe disponer de múltiples posibilidades de búsqueda.

Búsqueda por metadatos

Además de las palabras clave, muchos SGED permiten el uso de metadatos para realizar búsquedas. Wikipedia define metadatos de la siguiente forma:

Metadatos (del griego $\mu\epsilon\tau\alpha$, meta, «después de» y latín datum, «lo que se da», «dato»), literalmente «sobre datos», son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado recurso. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

Los SGED que soportan el uso de búsqueda por metadatos tienen ciertos metadatos asociados a los documentos. Algunos ejemplos serían: fecha de creación o edición del documento, quien lo creó, quién lo vio por última vez, formato y tamaño del fichero, etc. La verdadera potencia de los metadatos aparece cuando se usan junto a palabras clave. Al añadir esta capacidad, Juan podría ir a la carpeta electrónica de Sonia y lanzar una búsqueda de la palabra clave “Vacaciones” junto a la fecha de creación “21-05-2009” para encontrar exactamente el documento requerido.

La ventaja de usar metadatos es que generalmente estos se aplican de forma automática a los documentos. El SGED aplica ciertos metadatos como “fecha de creación”, “fecha de edición”, etc., sin necesidad de intervención por parte del usuario. Y como los metadatos se asocian de forma sistemática, son plenamente fiables en toda la estructura del SGED, sin importar cuantas personas están indexando y buscando. La desventaja de los metadatos es que están limitados a ciertos tipos de información del documento; por consiguiente, no es muy útil a menos que se combine con otros métodos de búsqueda, como la estructural y por palabras clave.

Búsqueda por contenido

La búsqueda por contenido o “texto completo” (del inglés “Full Text”) es otro modo de buscar un documento. Consiste en encontrarlo basándose en una palabra o frase contenida dentro del texto del documento. Por ejemplo, si Juan quiere encontrar todos los documentos del archivo de Recursos Humanos que contengan la palabra “juicio”, la búsqueda por contenido le sería muy útil.

Lo primero y más importante para poder usar esta búsqueda es que los documentos contengan texto. Un SGED que proporcione búsqueda por contenido será capaz de indexar todos los documentos de texto de sus repositorios. A medida que se archivan e indexan los documentos, estos deben estar disponibles para su consulta. Este es un proceso sencillo para documentos como emails, MS Word, MS Excel y otros documentos basados en texto. Sin embargo, los documentos digitalizados no contienen texto (un documento digitalizado es una imagen) de modo que deben ser convertidos a un formato que contenga texto para poder indexarlo y hacer posible buscar por dicho texto.

Los SGED de hoy proporcionan opciones para llevar a cabo esa conversión. A menudo se conoce como conversión OCR (Optical Character Recognition), que significa Reconocimiento Óptico de Caracteres, y que convierte un formato imagen a formato texto.

El proceso de lanzar una búsqueda por contenido es muy simple. El usuario introduce una palabra o frase y el sistema ejecuta una búsqueda y muestra una lista de resultados. La mayor parte de los SGED que permitan este tipo de búsqueda permitirán también el uso de búsquedas booleanas con “and” y “or”, así como los comodines “*” y “?”, e incluso búsquedas por aproximación, que localizarán palabras parecidas aunque no idénticas a las introducidas por el usuario.

Imaginemos que el OCR nos diera un 99.9% de acierto en el reconocimiento, lo que no estaría nada mal. Si solo nos basáramos en el OCR y la búsqueda por contenido, tendríamos garantizado que perderíamos 1 de cada 1.000 documentos. La mayoría de las Empresas u Organizaciones no se lo pueden permitir.

Existen dos grandes desventajas en este tipo de búsqueda. Primera, el tiempo extra que se consume para convertir imágenes digitalizadas a texto, teniendo en cuenta que el texto no es preciso al 100%. Segunda, la falta de fiabilidad de las búsquedas, como en el caso de las palabras clave, por lo que no debería nunca ser el único método de consulta.

Sin embargo, la potencia de la búsqueda por contenido para localizar documentos tiene muchas aplicaciones. Por consiguiente, cuando se usa junto a métodos de búsqueda estructurales y por palabras clave, este tipo de búsqueda es una magnífica propiedad para redondear un completo SGED.

Archivo vs. Recuperación

Perder unos pocos segundos en indexar y archivar un documento correctamente podría ahorrarle a Usted, y a otras muchas personas, horas en el tiempo de recuperación. Cuanto menos tiempo y esfuerzo emplee en estas operaciones, mas difícil será encontrar los documentos. Por la misma razón, un buen SGED debería proporcionar propuestas razonables de automatización de los procesos de indexación tanto como sea posible. Un procedimiento de archivo e indexación eficiente redundará posteriormente en una recuperación eficiente.

Lamentablemente, no hay un botón mágico en el escáner que se pueda pulsar para archivar de forma automática todos los documentos correctamente. Si lo hubiera, este botón no haría que la recuperación de documentos fuera sencilla. Pero esta circunstancia no debería disuadirle de probar el funcionamiento de un SGED para intentar que el funcionamiento de su Negocio u Organización sea más eficiente. La mejora de la productividad será apreciada por cada una de las personas que archivan, recuperan, crean, editan y procesan documentos.

Conclusión

Como probablemente habrá comprobado, no existe un método ideal que sirva para todos los entornos y todas las aplicaciones. Como ocurre en muchos otros aspectos de la vida, hay muchas consideraciones a tener en cuenta.

Agilizar el proceso de añadir e indexar ficheros en el Sistema de Gestión Electrónica Documental puede suponer que más adelante el proceso de consulta lleve más tiempo. Por otra parte, una indexación detallada de cada documento puede hacer muy sencilla su búsqueda, pero conllevará un aumento considerable en el tiempo destinado a la captura y archivo de los documentos.

La mejor solución es siempre disponer de la posibilidad de ejecutar cualquiera de los tipos de búsqueda analizados en este estudio, dependiendo del proceso de negocio y del enfoque general más eficiente.