

Soluciones de búsqueda para empresas

Incorporación de soluciones de búsqueda en la arquitectura de TI

Las soluciones de búsqueda para empresas pueden ser de gran ayuda para la productividad y el rendimiento de la organización, siempre y cuando se implementen de forma efectiva y eficiente.

INTRODUCCIÓN: IMPORTANCIA DE LA ARQUITECTURA

Las soluciones de búsqueda para empresas están emergiendo como herramientas de grandes prestaciones que ayudan a las organizaciones a conseguir un objetivo universal: optimizar el uso de la información.

Las soluciones de búsqueda de varios niveles que exploran el escritorio, los departamentos, la empresa e Internet para localizar información prometen impulsar la productividad e incrementar la efectividad de la organización. Estas soluciones ofrecen un método para vencer los principales obstáculos con los que se enfrentan las organizaciones hoy en día: depósitos redundantes de datos idénticos, distintos métodos de recuperación y la falta de definiciones semánticas coherentes en la empresa. Estas soluciones también pueden ayudar a cumplir las expectativas de crecimiento de los trabajadores, si los métodos de búsqueda de los sistemas de la empresa son tan sencillos y fáciles de utilizar como el software y los servicios de Internet.

Uno de los aspectos clave para beneficiarse de los sistemas de búsqueda en toda la empresa es integrar de forma efectiva la solución de búsqueda con otros sistemas de tecnología de la organización. BearingPoint ha identificado los principios clave de las soluciones de búsqueda de las organizaciones y, a partir de ello, ha desarrollado una arquitectura recomendada para la correcta implementación de las mismas.

PRINCIPIOS DE LAS FUNCIONES DE BÚSQUEDA EN TODA LA ORGANIZACIÓN

Son varios los principios generales que dictan el desarrollo de la arquitectura de las soluciones de búsqueda:

Simplicidad. El principio operativo del sistema de búsqueda de la organización es disponer de sistemas y soluciones simples. La complejidad inhibe la productividad de los desarrolladores de software, dificulta la planificación, creación y verificación de los productos, genera retos de seguridad y provoca la frustración de usuarios finales y administradores. Las tecnologías informáticas y de comunicación han evolucionado hasta tal punto que es factible disponer de un modelo de servicios. Un software de servicios mejorado y el desarrollo de soluciones de búsqueda más simples dan respuesta a la demanda de los usuarios de soluciones integradas y convincentes que funcionen.

Funcionalidad. Por orden de prioridad, la arquitectura de la solución de búsqueda debe presentar las siguientes características: ser manejable, fiable, segura, ampliable (en términos de usuarios, volúmenes de mensajería y servicios) y poder crecer de forma independiente mediante la adición de nuevos servicios proporcionados por la infraestructura de la solución de búsqueda.

PERSPECTIVA:	
INTRODUCCIÓN: IMPORTANCIA DE LA ARQUITECTURA	1
PRINCIPIOS DE LAS FUNCIONES DE BÚSQUEDA EN TODA LA ORGANIZACIÓN	1
DIRECTRICES DE FUNCIONAMIENTO PARA EL DESARROLLO DE SOLUCIONES	2
ARQUITECTURA DE LA SOLUCIÓN DE BÚSQUEDA	3
Arquitectura conceptual	4
Ejemplo de arquitectura de la solución de búsqueda	4
OPORTUNIDAD DE LA SOLUCIÓN DE BÚSQUEDA	8

Fiabilidad. La plataforma de la solución de búsqueda debe ser una infraestructura sólida y, por consiguiente, debe evitar el uso de tecnologías no consolidadas. Como norma general, en su arquitectura principal utilizará una tecnología que:

- Ofrezca la posibilidad de ser utilizada en cualquier entorno.
- Haya sido probada y sea fiable en otros escenarios de uso comparables.

Exigencia. Una solución de búsqueda para empresas debe considerarse como un componente crucial para la empresa que sea muy seguro, pueda ampliarse y esté disponible las 24 horas del día, 7 días a la semana. La arquitectura de la estructura principal de la solución de búsqueda debe diseñarse teniendo en cuenta estos atributos.

Accesibilidad. A fin de permitir un amplio uso, la arquitectura de la solución de búsqueda debe estar diseñada para un uso intenso. Esta capacidad debe tenerse en cuenta desde que se empieza a diseñar la arquitectura, ya que luego no es factible introducir una funcionalidad de este tipo.

Flexibilidad. El concepto de servicio flexible es la clave del éxito para toda solución de búsqueda. No nos referimos a un servicio de usuario final, sino a servicios que constituyen unidades diferenciadas de funcionalidad disponibles en la plataforma de búsqueda a través de interfaces definidas.

DIRECTRICES DE FUNCIONAMIENTO PARA EL DESARROLLO DE SOLUCIONES

Partiendo de los principios anteriores, la arquitectura de la solución de búsqueda debe ceñirse a las directrices siguientes:

Estándares abiertos, tecnología abierta e interfaces comunes. La solución de búsqueda debe basarse en estándares abiertos, tales como servicios web, J2EE y .NET, que permitan mantener la independencia de la tecnología. Gracias a ello, la solución es más flexible, se evita que deban realizarse cambios en la interfaz al modificar otra interfaz de la aplicación y se oculta la complejidad de la conectividad de la API (interfaz de programación de aplicaciones) entre sistemas y aplicaciones relacionados.

Arquitectura modular. La solución de búsqueda se fundamenta en una arquitectura basada en componentes y de múltiples niveles. Este enfoque permite la abstracción de cada nivel y para cada componente, de modo que los componentes son independientes y la funcionalidad se divide de forma lógica. La arquitectura debe ser flexible y muy granular.

Interoperabilidad. El uso de XML (lenguaje de marcas extensible) y de XSLT (transformación del lenguaje de hojas de estilo extensible) crea interoperabilidad y proporciona una abstracción que separa el dominio de presentación de la lógica subyacente. Dividir la presentación en un subnivel de presentación y un subnivel de interfaz de presentación permite aislar la lógica de presentación. Para conseguir la interoperabilidad, la arquitectura se basa en la independencia de los dispositivos, las plataformas y las tecnologías.

Administración de metadatos. La administración de los metadatos hace posible la creación, el almacenamiento, el acceso, la antigüedad y el mantenimiento de contenido de metadatos, incluido el control y la sincronización de los metadatos compartidos. La administración de los metadatos de la empresa permite a la organización:

- Automatizar la asignación de metadatos a contenido existente. Esto hace posible que las búsquedas sean más efectivas, porque se realizan en los activos de contenido correctos, independientemente de su ubicación.
- Desarrollar una capacidad para asociar fácilmente metadatos con activos de contenido que se encuentren lo más cerca posible a su fecha de creación.
- Establecer una administración organizativa clara para el desarrollo y la gestión continuada de taxonomías y de listas de autoridad.
- Implementar procesos para mantener las taxonomías actualizadas y auditar asociaciones de metadatos.
- Obtener distintas vistas del mismo contenido, por ejemplo, por rol, tiempo y proceso.
- Integrar fuentes de contenido. La solución de búsqueda debe tener una arquitectura abierta que permita la integración de varias fuentes de contenido. Para que una plataforma funcione sin problemas, es preciso disponer de estándares abiertos, especialmente en relación con la interoperabilidad de varias fuentes de contenido.

Las soluciones de búsqueda de varios niveles ofrecen un método para vencer los principales obstáculos con los que se enfrentan las organizaciones hoy en día.

Seguridad. Una solución de búsqueda para empresas debería abarcar y ampliar los principios de seguridad asociados con las distintas fuentes de contenido de la empresa. Es necesario aplicar los principios de autenticación, autorización, auditoría y administración de identidades de modo que únicamente los usuarios apropiados puedan acceder a información privilegiada. La seguridad de la solución de búsqueda debe cumplir la política de seguridad de la empresa y los requisitos reguladores. En algunos casos, para poder hacer un uso efectivo del contenido no estructurado, se precisará seguridad en el nivel de documento.

Infraestructura de la red. La implementación de la solución debe tener en cuenta el impacto en la infraestructura de la red y la organización a la que da soporte dicha infraestructura. En el caso de búsquedas en toda la organización, debe tenerse en cuenta la ubicación de los datos buscados, la posibilidad de que personas externas a la organización utilicen la funcionalidad de búsqueda, la seguridad y el rendimiento general.

A continuación, se indican consideraciones y directrices generales sobre el diseño de la red:

- ¿Cuánto tráfico generará la búsqueda en la red de la empresa y en los segmentos asociados? El diseño estará regido por el número de fuentes de datos que deba explorarse, la ubicación de dichas fuentes de datos y el tipo de infraestructura de red implementada.
- ¿Dónde reside la base de usuarios para la solución de búsqueda: en el mismo edificio, en un campus distinto o en una red de área amplia (WAN)?
- ¿Se ha clasificado el tráfico de la red? Es muy importante tener en cuenta la clasificación del tráfico en un entorno en el que se lleva a cabo la réplica de datos de la red de área de almacenamiento (SAN), como Symmetrix Remote Data Facility (SRDF). La organización debe tener en cuenta el impacto de las búsquedas en la red y determinar políticas adecuadas de segmentación de la red e implementaciones de calidad de servicio (QoS).
- ¿De qué tipo de control de acceso dispone la red? En muchas implementaciones, la solución de búsqueda se utilizará para acceder y servir datos de propiedad y confidenciales, por lo que es crucial que la implementación sea la correcta. Debe tenerse en cuenta una segmentación correcta de la red, el diseño de la VLAN (red de área local virtual), la implementación de cortafuegos y el acceso a la red.
- ¿Qué tipo de resistencia de red se precisa? Teniendo en cuenta que la solución de búsqueda se utiliza en el departamento financiero, en el departamento de recursos humanos y en otros departamentos de gestión, se considera que es una aplicación fundamental que debe estar siempre disponible. Para conseguirlo, la solución de búsqueda se suele implementar a modo de clúster, con opciones de recuperación de errores en uno o varios sitios.

Recuperación tras fallos y continuidad de la empresa.

A fin de poder cumplir con los principios de recuperación tras fallos y de continuidad de la empresa, se requiere redundancia de hardware y de software con el fin de minimizar los puntos de error. La filosofía de diseño general consiste en utilizar componentes redundantes para crear una infraestructura redundante, incrementando así la capacidad de recuperación tras fallos de componentes individuales y mejorando la disponibilidad del sistema para procesos y usuarios tanto internos como externos.

ARQUITECTURA DE LA SOLUCIÓN DE BÚSQUEDA

El concepto de arquitectura para las búsquedas en toda la organización es un sistema robusto para dar respuesta a los objetivos de la empresa. El concepto pasa a ser una realidad en forma de arquitectura de empresa modular y de n niveles que cumple los estándares y los principios indicados anteriormente. Puede dar soporte a una solución de búsqueda que dé respuesta a las necesidades actuales de la empresa, permita el crecimiento, crezca al mismo ritmo que el mercado y reduzca el coste total de propiedad (TCO).

Arquitectura conceptual

En la ilustración 1 se describe la arquitectura conceptual de la solución de búsqueda.

Los componentes de la arquitectura conceptual son los siguientes:

Fuentes de contenido. Las fuentes de contenido incluyen la intranet de la empresa, sitios web, servidores de archivos, sistemas de administración de contenido, como Documentum y FileNet, y aplicaciones empresariales que abarcan todos los sistemas de operaciones, incluyendo la administración de las relaciones de clientes (CRM), la administración de la cadena de suministro (SCM) y la planificación de recursos de la empresa (ERP). Básicamente, este bloque representa los distintos sistemas dentro de la empresa que podrían servir como fuentes de contenido sobre las que los usuarios podrían realizar búsquedas.

Servicios de conectividad. El nivel de servicios de conectividad proporciona los adaptadores básicos para las fuentes de contenido subyacentes. Esta capa incluirá dos tipos de adaptadores: conectores y acceso federado. Los conectores estipulan la creación de adaptadores de distintas tecnologías necesarios para establecer una conexión con las fuentes de contenido. Los conectores de acceso federado proporcionan la conexión al contenido de terceros.

Servicios de metadatos. El nivel de servicios de metadatos proporciona servicios de valor añadido para el contenido. Ofrece dos tipos de servicio: correlación de metadatos y correlación de diccionarios de datos. Los servicios de correlación de metadatos ayudan a definir los datos de la empresa entre distintos sistemas. También proporcionan una única correlación entre elementos comunes a través de la creación del diccionario de datos de la empresa en varios sistemas. El diccionario de datos de la empresa está representado por un conjunto de definiciones de metadatos y de representaciones de elementos de datos.

Acceso federado. El nivel de acceso federado proporciona la creación de intermediarios de consulta que pueden obtener información de sistemas de operaciones subyacentes. Para poder acceder a los datos de la empresa, es preciso crear un índice universal de la semántica relacionada.

Ejemplo de arquitectura de la solución de búsqueda

En la ilustración 2 se describe un ejemplo de arquitectura de la solución de búsqueda derivada de la arquitectura conceptual anterior.

La arquitectura tiene los siguientes niveles:

Interfaz de usuario

Éste es el nivel de interacción del usuario que controla las interacciones de los usuarios a través de un explorador web o incluso de un dispositivo portátil. Este nivel presenta a los usuarios pantallas y datos asociados.

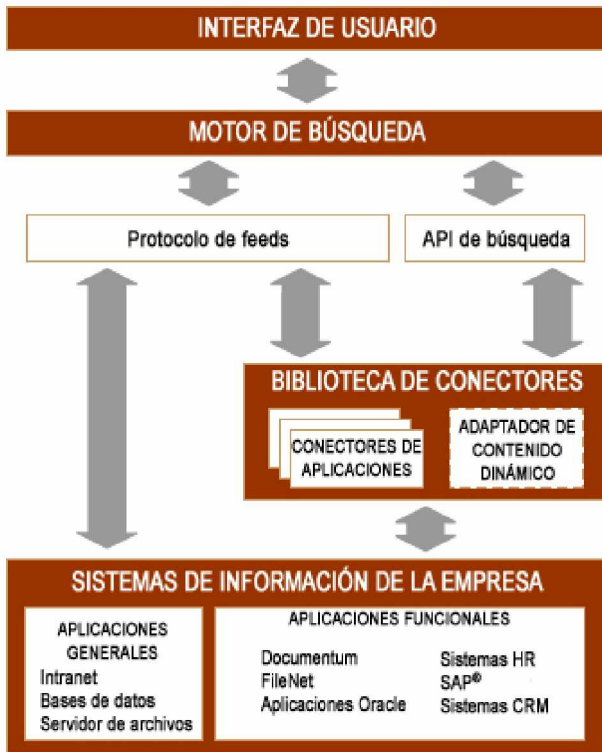
La interfaz de usuario es una interfaz basada en XML que se crea de forma dinámica en el nivel de la aplicación. En función del dispositivo de consulta, se ofrecerá la presentación apropiada mediante archivos XSLT. De este modo, se separa la presentación del contenido y se dividen claramente las responsabilidades. Este nivel también está estrechamente vinculado con los componentes de seguridad de autorización y autenticación, con el fin de que los resultados de búsqueda se muestren a personas que cuenten con la autorización necesaria para verlos.

Los aspectos clave que deben tenerse en cuenta para mejorar la capacidad de uso de la solución de búsqueda incluyen la presentación basada en roles, modos de visualización, de edición y ayuda contextual, y paginación de los resultados de búsqueda.

Ilustración 1. Arquitectura conceptual de la solución de búsqueda



Ilustración 2. Ejemplo de arquitectura de la solución de búsqueda



las características de seguridad necesarias para los agentes de información de seguridad. En lugar de ello, el método de implementación recomendado consiste en tener adaptadores que puedan transferir la consulta al catálogo de contenido básico en nombre de la persona que emite la consulta.

- **Agrupador de resultados.** Este dominio recopila los resultados de distintos catálogos que coinciden con la consulta y los presenta al usuario. Los resultados deben incluir la capacidad de proporcionar comentarios en cualquiera de los resultados mostrados, deducciones para pasar de una consulta de búsqueda a otra y la capacidad de aplicar de nuevo una búsqueda por facetas o categorías al motor de búsqueda.

Motor de búsqueda

El motor de búsqueda es un producto de hardware o software que proporciona, a través de un cuadro, las funcionalidades de búsqueda de la organización. Se espera que pueda realizar búsquedas en cientos de formatos de archivos distintos en cualquier lenguaje y que sea capaz de indexar millones de documentos. Debe disponer de funciones de seguridad con el fin de garantizar que los usuarios sólo ven los documentos que están autorizados a consultar.

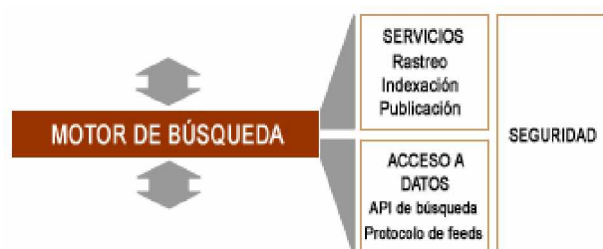
Las funciones principales del motor de búsqueda deben ser las siguientes (véase la ilustración 3):

- **Publicación.** Esta función proporciona una interfaz de búsqueda estándar que, de forma predeterminada, puede presentarse desde el motor de búsqueda. Esta interfaz debe personalizarse para modificar la hoja de estilo XSLT subyacente. Otras funciones, como la concordancia de palabras clave, sinónimos y filtros, pueden ayudar a promocionar páginas web específicas como parte de los resultados de búsqueda. La función de sinónimos sugiere palabras o frases alternativas para las consultas de búsqueda. Otras funciones crean depósitos de información lógica denominados “colecciones”, que ayudan a satisfacer las necesidades de búsqueda de los usuarios.

Un entorno empresarial típico cuenta con varios almacenes de documentos y de contenido web en los confines de la intranet. Dentro de estos almacenes se utilizan varias tecnologías para llevar a cabo el alojamiento real del contenido. El alojamiento incluye mecanismos de seguridad inherentes, catálogos de índices individuales y almacenes de procesamiento del contenido propiamente dicho. La arquitectura de la solución de búsqueda debe proporcionar capacidades dentro de tres dominios:

- **Procesador de consultas.** El procesador de consultas es un servicio central al que puede llamarse desde distintos lugares. Su trabajo consiste en desglosar la consulta, validar la ortografía, traducir a idiomas alternativos, identificar sinónimos, conceptos y el tipo de consulta más apropiado para el almacén de contenido asociado; por ejemplo, palabra clave, concepto, booleano y texto completo.
- **Adaptador de catálogo.** Los adaptadores de catálogo proporcionan el contrato de integración entre la consulta inicial y los catálogos del índice de contenido de procesamiento. Una única capacidad del índice en todos los almacenes de contenido no proporcionará

Ilustración 3. Funcionalidad del motor de búsqueda



- **Rastreo.** Los rastreadores son agentes que solicitan y recuperan documentos de servidores web para la indexación automática. El rastreo de datos puede realizarse según dos métodos: descubrimiento de datos y recuperación de datos. Una interfaz de configuración permite rastrear distintas fuentes de datos. Todo el contenido se agrega para crear un índice maestro, que se actualiza con cada nuevo rastreo.

El rastreo puede considerarse como un proceso de tres pasos: descubrimiento inicial (nuevas URL por rastrear), indexación y rastreo continuado. Cada uno de los documentos nuevos que detecta el rastreador se explora para detectar vínculos. Éstos bien se recopilan de forma inmediata o bien se programa su recuperación posterior. El rastreador trata sin problemas contenido protegido y gestiona las comunicaciones del protocolo de transferencia de hipertexto seguro (S-HTTP, Secure Hypertext Transfer Protocol). El rastreador del motor de búsqueda puede gestionar la autenticación básica, la autenticación NTLM (NT LAN Manager) y el acceso basado en formularios y cookies personalizadas. El motor debería rastrear el contenido de las bases de datos Oracle, SQL Server, MySQL, IBM DB2 y Sybase. En caso de que el rastreador no pueda acceder a un tipo de datos, éste puede pasarse directamente al motor de búsqueda en formato XML.

- **Indexación.** Para indexar los datos, pueden utilizarse varios métodos:
 - Un feed de datos para el sistema de búsqueda mediante adaptadores estándar o personalizados.
 - Acceso directo desde el sistema de búsqueda (el adaptador estándar suministrado con el sistema es el único que puede acceder directamente a los datos).
 - Feeds de datos HTTP o que no sean HTTP.
- **Feeds.** La mejor forma de pasar ciertos tipos de documento al motor de base de datos es mediante feeds, en lugar de encontrarlos a través de vínculos en las páginas web rastreadas. Estos tipos de documento son los siguientes:
 - Documentos que no pueden obtenerse mediante el rastreador. Por ejemplo, registros de una base de datos o archivos de un sistema que no estén habilitados para la web.
 - Documentos que pueden rastrearse, pero que es mejor rastrearlos de nuevo a horas distintas de las establecidas por el programador de rastreo automático que se ejecuta en el motor.

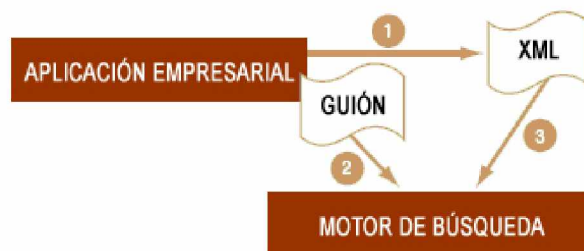
- Documentos que pueden rastrearse, pero que no tienen vínculos en el sitio web y que permiten al rastreador descubrirlos durante un nuevo rastreo.
- Documentos que pueden rastrearse, pero que, debido a problemas del servidor web o de la red, se suben con mayor rapidez si se utilizan feeds.

Los feeds pueden ser de distintos tipos:

- Feeds de contenido: contienen direcciones URL y su contenido. También pueden contener metadatos, como la fecha de la última modificación del registro. En un feed de este tipo puede especificarse cualquier nombre de fuente de datos.
- Feeds de web: en este tipo de feeds, cada uno de los registros contiene direcciones URL, pero no contenido. Estas direcciones URL pueden rastrearse de la forma habitual. En los feeds de este tipo, el término “web” se utiliza como nombre de la fuente de datos. Todas las direcciones URL deben contener un nombre de dominio totalmente calificado (FQDN) en el fragmento de la dirección URL que corresponde al nombre del host.
- Feeds de base de datos: pueden crearse conectores personalizados con el fin de promocionar registros de la base de datos para crear un índice.

El proceso de configuración de los feeds del motor de búsqueda se muestra en la ilustración 4.

Ilustración 4. API de suministro de contenido de la búsqueda



- 1 El guión exporta datos en formato XML para generar el archivo de feed.
- 2 El guión publica el archivo XML en el motor de búsqueda.
- 3 El motor de búsqueda importa el archivo XML e indexa el contenido.

- **Seguridad en el motor de búsqueda.** El motor de búsqueda debería incrementar el sistema de seguridad de la empresa. Debe ejecutar todos los servicios detrás de un cortafuegos, abriendo solamente algunos puertos para permitir la comunicación con el motor a través de éste. El motor de búsqueda rastrea e indexa tanto documentos públicos como confidenciales. Un paquete de seguridad opcional aplica las políticas de seguridad de nivel de documento de la empresa.

La entrega del conjunto de resultados correcto a cualquier usuario se basa en el filtrado de los resultados del índice en el que se ha realizado la búsqueda. El contenido se divide en dos categorías: contenido público y todo el contenido. Los usuarios pueden realizar búsquedas en el contenido público únicamente o en los documentos públicos y los protegidos del índice, según especifique el administrador del índice. De forma predeterminada, el contenido permanece protegido de modo que los usuarios que no tengan acceso no verán ningún contenido para el que no dispongan de autorización.

El motor de búsqueda deberá dar soporte a la autenticación básica y NTML (NT LAN Manager), así como a la autenticación basada en formularios, y gestionará cada uno de estos protocolos de autenticación de forma distinta lo requiere que el administrador lo configure de forma diferente. Cada método puede rastrear e indexar los documentos protegidos de los sitios de intranet y puede realizar una autenticación en las búsquedas realizadas en dichos documentos.

Una API de autorización de búsqueda permite a un servicio web realizar la conversión entre la API de autorización del motor de búsqueda y el servidor de la empresa que proporciona los servicios de control de acceso, conocido como conector de acceso (CA). El CA actúa como una capa entre el motor de búsqueda y el sistema de control de acceso de la organización.

Biblioteca de conectores

Esta capa de software forma el nivel de mensajes o de acceso a distintas fuentes de datos de una empresa, que pueden ser desde fuentes de datos estructurados hasta aplicaciones empresariales. Generalmente, para integrar las fuentes de datos, se crea una plataforma de integración, conectores personalizados o adaptadores.

Los conectores personalizados también pueden crearse siguiendo estos pasos:

- Convertir los datos en XML en el formato especificado en el feed del motor de búsqueda.

- Subir el XML en el motor de búsqueda mediante el protocolo HTTP.

Este nivel forma la estructura básica subyacente para exponer información a través de la empresa independientemente del formato del almacenamiento o de la fuente de datos. Por consiguiente, este nivel permite establecer una correlación entre servicios y adaptadores de integración creados a partir de distintas fuentes de datos.

Los metadatos definidos en este nivel también podrían proporcionar la infraestructura necesaria para consolidar la información entre distintas fuentes de datos de la empresa. Generalmente, esto también permite descubrir y recuperar con rapidez los elementos que coinciden en el nivel semántico. Por consiguiente, un metamodelo unificado establece definiciones de metadatos coherentes y correctas en el nivel contextual para consolidar los distintos tipos de información. También mejora los resultados de la búsqueda con información de vínculos controlados por el contexto a datos estructurados y aplicaciones empresariales.

Fuentes y tipos de información

Actualmente es posible encontrar información en cualquier lugar de una empresa. La información está disponible en formatos estructurados o no estructurados, en depósitos de propiedad o de terceros (véase la tabla 1). Los motores de búsqueda pueden buscar información que puede indexarse para facilitar su recuperación y para que los resultados de la búsqueda sean relevantes. Por consiguiente, es muy importante clasificar las fuentes de datos del mismo modo y luego utilizar los mecanismos de recuperación apropiados.

Tabla 1. Fuentes y tipos de información

	DE PROPIEDAD	NO DE PROPIEDAD
Estructurado	Sistemas de gestión de bases de datos(Oracle, SQL Server, Sybase, Informix) Sistemas de gestión de documentos(Documentum, FileNet, etc.) Aplicaciones empresariales, como ERP y CRM	Sistemas de administración de bases de datos (mySQL)
No estructurado	Sistemas de archivos de red/sistemas de archivos Documentos de Microsoft Office Archivos de imagen, vídeo y audio	Documentos de Open Office Sistemas de archivos / servidor de archivos Archivos de texto Correo electrónico

Estructurado frente a no estructurado. La diferencia principal entre información estructurada y no estructurada es un modelo de información predefinido. La analogía más sencilla para la información estructurada son datos que puedan presentarse en un formato predefinido como HTML o datos que puedan insertarse en una tabla, como ocurre con la tecnología de bases de datos. Otra forma de ver la información estructurada es en términos de metadatos, es decir, información acerca de información. Por otro lado, los datos no estructurados no tienen ningún formato predefinido y, por consiguiente, existen en formularios de flujo libre. Como ejemplos de este tipo de datos cabe citar los archivos de texto y los documentos PDF. En la información no estructurada, el tipo de búsqueda, en términos de las herramientas que se utilizan como solución, supone un problema de recuperación de la información.

Depósitos de propiedad frente a depósitos de terceros. Actualmente, la información de la empresa está disponible en sistemas ERP, CRM y en sistemas de administración de conocimientos, así como en portales estáticos y dinámicos. Para poder recuperar información, algunos de estos sistemas proporcionan formatos de propiedad para publicar información. Algunos utilizan formatos que no son de propiedad, como XML, que cada vez va ganando más terreno. Asimismo, si bien el texto sigue dominando en muchas aplicaciones, la importancia de otros tipos de medio, como imagen, audio y vídeo, aumenta a un ritmo constante. En este caso, se requiere una combinación eficiente de recuperación automática de texto, recuperación de metadatos (generalmente creados de forma manual) y recuperación basada en contenido de datos multimedia.

OPORTUNIDAD DE LA SOLUCIÓN DE BÚSQUEDA

Una solución de búsqueda para empresas puede ser el catalizador para optimizar la productividad de los empleados y para poder contar con mayor información en el proceso de toma de decisiones. No obstante, cualquier solución de búsqueda que se utilice en la empresa debe introducirse conociendo bien sus requisitos de arquitectura y sus capacidades. Para optimizar el uso de las características que ofrece una solución de búsqueda, es muy importante planificar minuciosamente la arquitectura y su implementación.

Si desea obtener más información sobre la forma en que nuestras soluciones pueden aportar valor añadido a su empresa, [hablemos](#).

ADMINISTRACIÓN GLOBAL Y CONSULTORÍA DE TECNOLOGÍA PARA EL CONTEXTO EMPRESARIAL ACTUAL

BearingPoint es un proveedor líder en servicios de consultoría tecnológica y gestión global para las compañías Global 2000 y muchas de las compañías de servicios públicos más importantes del mundo. Nuestros expertos ayudan a las empresas de todo el mundo a prepararse para conseguir sus objetivos y crear valor empresarial. Al adaptar sus procesos empresariales a los sistemas de información, les ayudamos a adquirir una ventaja competitiva y a obtener resultados rápidamente. Para obtener más información, póngase en contacto con nosotros en el número 1.866.661.FIND (+1.603.589.4089 si llama desde fuera de Estados Unidos y Canadá) o visite nuestro sitio web en www.bearingpoint.com.

BearingPoint proporciona servicios de consultoría estratégica, servicios de aplicaciones, soluciones tecnológicas y servicios de gestión a las empresas Global 2000 y a entidades gubernamentales.



BearingPoint
1676 International Drive
McLean, VA 22102
www.bearingpoint.com