

Recuperación de información y sistemas para la gestión del conocimiento.

Ricardo Eito Brun

Dpto. Biblioteconomía y Documentación
Universidad Carlos III de Madrid

Resumen:

La gestión del conocimiento ha adquirido una gran importancia en los últimos dos años. Uno de los factores claves en el desarrollo de políticas para la gestión del conocimiento es la infraestructura tecnológica que las soporta.

Dentro de esta infraestructura debemos incluir herramientas software de nueva generación, diseñadas exclusivamente con el propósito de gestionar conocimiento. Sin embargo, las aplicaciones disponibles actualmente y que se proclaman válidas para la gestión del conocimiento presentan perfiles diferenciados.

La comunicación analiza la arquitectura lógica de estas aplicaciones y presenta los rasgos que tienen en común entre ellas.

Se describe con detalle el enfoque que proponen para el subsistema de recuperación de información y se presenta un cuadro clasificatorio para estas aplicaciones.

Introducción

La gestión del conocimiento es uno de los aspectos que ha acaparado una mayor atención en los últimos dos años. En este tiempo, ha proliferado un gran número de contribuciones bibliográficas, cursos, seminarios, etc., dedicados a este tema.

Esta situación ha llevado a algunas personas a considerar la gestión del conocimiento como un término de moda capaz de captar inversiones en proyectos de consultoría y tecnologías de la información. Nada más lejos de la realidad. La gestión del conocimiento, de la misma forma que ha sucedido con otros movimientos de gestión empresarial (por ejemplo el aseguramiento y control de la calidad, reingeniería, etc.), puede otorgarnos importantes ventajas competitivas sostenibles a largo plazo si somos capaces de implantar y mantener un sistema adecuado para su administración.

Por otra parte, aunque la gestión del conocimiento se relaciona normalmente con el mundo empresarial, debemos indicar la posibilidad - y, por qué no, la conveniencia - de aplicar las técnicas y metodologías que propugna en otros sectores: administraciones públicas, universidades, sector sanitario, etc. Pero, ¿a qué nos referimos cuando hablamos de gestión del conocimiento?

Gestión del conocimiento

La gestión del conocimiento es una práctica organizativa y de gestión cuyo propósito es lograr la reutilización de los conocimientos y habilidades individuales de sus miembros, para así aumentar la productividad de la organización y potenciar su capacidad de innovar.

de esta definición tenemos que destacar tres ideas:

- a) el alcance corporativo de las estrategias destinadas a la gestión del conocimiento (en un proyecto GC deben estar implicados - sin excepciones - todas las personas que forman la organización),
- b) la necesidad de reutilizar y compartir los conocimientos de los individuos que conforman una organización y
- c) la productividad y la innovación como objetivos últimos.

Una estrategia destinada a gestionar el conocimiento nos permitirá ser más productivos, ya que la experiencia adquirida por cualquier miembro de la organización en la realización de su trabajo podrá ser reutilizada por otras personas en la ejecución de tareas similares.

Por otra parte, una estrategia para la gestión del conocimiento debe potenciar el intercambio de ideas como catalizador de la innovación y origen de posibles mejoras.

La táctica y las herramientas para la GC

Una vez señalados los principales objetivos que podemos lograr si adoptamos un compromiso con la gestión del conocimiento, debemos plantearnos la siguiente pregunta: ¿qué debemos hacer para implantar un sistema para la gestión del conocimiento?

Un proyecto de gestión del conocimiento debe plantearse la creación de una infraestructura capaz de optimizar la distribución de la información y de la experiencia adquirida en la actividad profesional de todos los miembros de una organización. Para ello, se utilizarán las siguientes herramientas:

- a) repositorios que recojan el conocimiento explícito de la organización,
- b) directorios de expertos que permitan dirigirnos con la menor brevedad posible a las personas que puedan proporcionarnos información sobre un tema específico,
- c) árboles de competencias y
- d) información procedente de fuentes externas que facilitan el seguimiento del mercado, competidores, clientes reales y potenciales, etc.

Las primeras herramientas - conocidas generalmente como *memorias corporativas*, consisten en bases de datos documentales que reúnen las experiencias documentadas de los trabajadores de la organización.

Los directorios de expertos - con los que están familiarizados los profesionales de la documentación que conozcan la iniciativa *IRM (Información Resources Management)* promovida por **Forest W. Horton** -, resulta ser una de las herramientas más fáciles de implementar y es, probablemente, la que ofrece un mayor retorno de la inversión en el menor tiempo.

Un árbol de competencias detalla el conjunto de conocimientos necesarios para desempeñar un trabajo u ocupar un puesto en el organigrama de la empresa.

Finalmente, la información externa continúa siendo un factor clave en las estrategias para la gestión del conocimiento, de la misma forma que lo ha sido en la gestión de información tradicional.

El objetivo final de una implantación para la gestión del conocimiento es conseguir la interacción de estas cuatro herramientas. Así, las contribuciones aportadas a las memorias corporativas permiten identificar expertos en áreas específicas así como grupos de personas con intereses comunes que establecen relaciones de colaboración (a pesar de que esa relación no sea explícita ni en el organigrama organizativo ni pueda deducirse a partir de la participación en proyectos comunes).

A su vez, los expertos en áreas de conocimiento deben participar en la gestión de las memorias corporativas para validar contribuciones y fomentar un proceso de diálogo y mejora continua en torno a ellas.

Por otra parte, el sistema para la gestión del conocimiento debe ofrecer a cualquier persona responsabilizada de completar una tarea toda la información - generada internamente o procedente de fuentes externas -, que le permita realizar su trabajo de manera óptima.

La tecnología en la GC

La tecnología es una pieza clave en cualquier estrategia para la gestión del conocimiento. La gestión de documentos, y especialmente las tecnologías que hacen posible la colaboración y la teleformación, son imprescindibles en la concepción de una estrategia válida.

De todas ellas, las tecnologías que hacen posible la colaboración garantizan el mayor retorno de la inversión en el menor intervalo de tiempo. Identificar expertos, acceder a ellos y reutilizar su experiencia ha demostrado ser, en palabras de **Davenport**, el beneficio más inmediato e importante en los proyectos desarrollados para la gestión del conocimiento.

Sin obviar el papel de la tecnología como capacitador para la GC, debemos señalar que existen factores más importantes - y de naturaleza más crítica -, en una estrategia GC. Nos referimos a la cultura organizativa y a la actitud que adopten las personas ante esta nueva práctica empresarial. La gestión del conocimiento es un proyecto con alcance corporativo, y por lo tanto debe hacerse un gran esfuerzo para dar a conocer los beneficios que se pueden obtener a través de ella e implicar a todos los miembros de la organización.

Por otra parte, para que la participación y la implicación de las personas que conforman la organización se haga efectiva, es necesario alinear la estrategia de Recursos Humanos de la organización (planes de compensación, remuneraciones, identificación de necesidades formativas, etc.), con los procesos de la gestión del conocimiento.

GC y recuperación de información

En la literatura es posible encontrar con bastante frecuencia una idea errónea que plantea la gestión del conocimiento como una solución al gran volumen de información disponible en la actualidad y a la diversidad de su origen. Esta visión de la gestión del conocimiento está muy vinculada al problema de la sobrecarga de información, y su origen se debe a la estrategia de posicionamiento que han adoptado los fabricantes de motores de búsqueda e indexación ante la evidencia de que la gestión del conocimiento iba a captar un amplio margen de la inversión en tecnologías de la información en los próximos años.

Aunque la recuperación de información es una de las tecnologías imprescindibles en una estrategia GC, la identificación de la gestión del conocimiento con la recuperación de información - e incluso con la integración de fuentes de información heterogéneas -, resulta tremendamente inapropiada y simplista.

En el apartado anterior hemos enumerado cuatro herramientas de valor crítico en la implantación de un sistema para la gestión del conocimiento: memorias corporativas, directorios de expertos, árboles de competencias e información externa.

Es importante recordar la interacción que se debe establecer entre estas cuatro herramientas, y que nos obliga a definir procedimientos para acceder a todos ellos mediante un único interface común (entendemos por interface tanto a las aplicaciones informáticas como a los mecanismos de recuperación de información y navegación por el sistema).

Para poder acceder a la información disponible en estas fuentes, es necesario disponer de un sistema de recuperación de información. De hecho, tanto los sistemas comerciales para la gestión del conocimiento como los definidos ad hoc, hacen un uso intensivo de las tecnologías relacionadas con la recuperación de información textual: recuperación booleana, ordenación por relevancia, query by example, etc.

La utilización de estas tecnologías está ampliamente difundida. Los sistemas de gestión documental tradicionales y los motores de indexación popularizados a raíz del auge de las intranets corporativas han contribuido enormemente a su generalización.

Sin embargo, a medida que acumulamos una mayor cantidad de información, los problemas para acceder a ella se hacen más asfixiantes y los paradigmas tradicionales para la recuperación y categorización han cedido su paso a otros criterios para la recuperación de información: la autoridad y en la experiencia compartida. Tanto una como otra se consideran factores determinantes en la transmisión del conocimiento, y el concepto de recuperación de información que se maneja en un proyecto GC está más próxima a un proceso de aprendizaje y experimentación compartida que al modelo tradicional de recuperación de información.

Auge para los lenguajes controlados

La necesidad de disponer de mejores mecanismos de acceso a la información nos ha obligado a fijar nuestra atención en los lenguajes controlados. De hecho, estos se han revelado como la única solución para el problema que afrontamos cuando queremos contextualizar contenidos con procesos de trabajo y grados de experiencia.

La cadena para la creación del conocimiento se define a menudo mediante la siguiente ecuación:

Datos+Contexto = Información + Experiencia = Conocimiento

Es decir, los datos contextualizados constituyen información, y ésta a su vez, al ser relacionada con la experiencia y los procesos de trabajo, constituye el origen del conocimiento.

Para ser efectivo, el sistema de indexación debe permitirnos contextualizar la información que recibe el sistema con los procesos de trabajo que se desarrollan en la práctica diaria de la organización y con la experiencia adquirida en su ejecución.

Para lograr esto, hay que crear unos sistemas de indexación que reúnan las siguientes características:

- a) orientados hacia la resolución de problemas en procesos de negocio.
- b) términos y unidades relativas a áreas de experiencia y tareas que integren los procesos de trabajo propios de la organización,
- c) estructura jerárquica - taxonomías - y con relaciones explícitas entre sus unidades,
- d) el repositorio de términos debe contener información sobre todas las entidades que intervienen en el proceso de creación/utilización del conocimiento: usuarios, niveles de competencia, necesidades de formación, tareas, procesos de negocio, etc).

Partiendo de esta base se hace manifiesta la necesidad de disponer de un sistema de indexación multidimensional (o facetado), que permita no sólo describir el contenido informativo tratado en las aportaciones a las memorias corporativas, sino también contextualizarlas en las prácticas de la organización (en respuesta a la pregunta ¿en qué contexto es útil esta información?), y en el nivel de competencia de sus posibles receptores (en respuesta a preguntas como ¿para quién puede ser útil esta contribución? o ¿cual es el nivel de conocimiento apropiado para poder comprender y aplicar la información consignada en esta contribución?)

Como resultado de la interacción entre memorias corporativas, autores de contribuciones, directorios de expertos y árboles de competencias, el sistema de indexación cuenta con la suficiente información sobre sí mismo para ser capaz de deducir relaciones entre las entidades utilizadas como criterios de indexación y así promover un proceso de recuperación activo basado en la interacción de distintas dimensiones.

Difusión selectiva, sí. Pero... ¿agentes?

En los últimos años la difusión selectiva de información basada en el uso de agentes automáticos se ha presentado como uno de las mejores soluciones - si no la definitiva -, para el problema del exceso de información.

Sin ánimo de entrar en polémica, debemos indicar la utilidad de estas herramientas. Pero los sistemas en los que se basan para deducir la pertinencia de una contribución para un perfil de usuario (basados generalmente en la presencia o ausencia de una serie de términos en el contenido textual de la contribución), hace que difícilmente puedan sustituir la labor de intermediarios humanos que canalicen la información a las personas o centros de interés apropiados.

Como alternativa a los sistemas push y a los mecanismos de difusión selectiva basados en agentes software y en el análisis automático de textos, en el área de la gestión del conocimiento se otorga una importancia creciente al llamado *collaborative filtering*. Con este término de difícil traducción, se hace referencia a la actividad desarrollada por un grupo de expertos para seleccionar información y dirigirla a otras personas a las que puede ser útil para sus necesidades de formación y para el desarrollo de su actividad profesional.

La característica que diferencia el *collaborative filtering* de la difusión selectiva tradicional basada en perfiles definidos mediante palabras claves consiste en que los editores son expertos en los temas tratados en las contribuciones, y en que se tienen en cuenta el nivel de conocimiento de los usuarios a los que se hace llegar la información, así como sus necesidades formativas.

Otra forma de aplicar el *collaborative filtering* consiste en asignar una prioridad a las contribuciones, de forma que los usuarios del sistema únicamente tienen que dedicar tiempo al estudio de aquellas contribuciones que se consideran más importantes en un momento dado. Los factores que determinarán la prioridad de una contribución frente a otra serán tan diversos como la calidad de su contenido, el prestigio o el reconocimiento que tenga su autor en la organización, o el proyecto en el que esté implicado el receptor de la contribución.

Podemos decir que el *collaborative filtering* parte de una definición más precisa del contexto en el que se va a aplicar la información, con el fin de garantizar que ésta sea procedente y pueda aplicarse inmediatamente en la resolución de problemas de negocio.

Del documento a las personas

El subsistema de recuperación que incorpore nuestro sistema de gestión de conocimiento no sólo debe ser capaz de ofrecer acceso a los documentos disponibles en las memorias corporativas. También debe ser capaz de unir a todas las personas que han participado en la creación de una contribución con todos sus lectores potenciales. De esta forma, en respuesta a una

ecuación de búsqueda el sistema no sólo nos ofrecerá una lista de documentos, sino también la forma que debemos utilizar para contactar con sus autores.

Esto último es especialmente importante en el caso de los documentos que contienen experiencias de trabajo documentadas. En este caso, el saber los motivos por los que se ha tomado una decisión es - en muchas ocasiones - más importante que la decisión en sí misma, y el tiempo para documentar es, habitualmente, escaso.

Las memorias corporativas deben actuar como foros de discusión que promuevan el intercambio de ideas y sean la base para un proceso de mejora continua para los procesos de trabajo de la organización.

Herramientas

En los apartados anteriores hemos enumerado las principales características que debe reunir el subsistema de recuperación de información de un sistema para la gestión del conocimiento.

En respuesta a estos marcos conceptuales, distintos proveedores de aplicaciones informáticas han lanzado al mercado productos que se presentan como soluciones idóneas para implantar una estrategia para la gestión del conocimiento.

En líneas generales, las respuestas de los fabricantes de software pueden agruparse en las siguientes líneas:

- a) un nuevo posicionamiento de los fabricantes de aplicaciones para la gestión electrónica de documentos y motores de indexación y recuperación textual,
- b) los proveedores de soluciones groupware (aplicaciones para el trabajo en grupo) han reclamado como suyos los modelos conceptuales planteados por la gestión del conocimiento y
- c) aparición de pequeñas empresas que ofrecen programas para la gestión del conocimiento, en ocasiones basadas en tecnologías novedosas.

La estrategia del cambio de posicionamiento

Ciertamente, la gestión de grandes volúmenes de documentación es una práctica necesaria en una estrategia para la gestión del conocimiento. Pero éste no es, ni mucho menos, el factor más importante, especialmente si pensamos que - en una estrategia GC -, es necesario anteponer la calidad de los contribuciones a su cantidad (de allí la importancia de los procesos de valoración y la adopción del *collaborative filtering* como principal medio de difusión selectiva en lugar de los agentes software basados en perfiles de palabras clave).

Un sistema para la gestión electrónica de documentos tradicional puede ser una pieza clave en la implantación de un sistema para la gestión del conocimiento, ya que es necesario disponer de un repositorio controlado de documentación con el que puedan interactuar el mayor número posible de usuarios. Sin embargo, no debemos caer en el error de identificar gestión del conocimiento con gestión documental.

Con los fabricantes de sistemas de indexación y recuperación textual sucede algo similar. En este caso, la estrategia seguida por estas empresas se basa en hacer equivalente la gestión del conocimiento con el acceso a fuentes de información creadas y gestionadas por sistemas informáticos heterogéneos (bases de datos relacionales, documentos en distintos formatos, mensajes de correo electrónico, etc.)

Aunque los sistemas de indexación y recuperación son un componente imprescindible en los sistemas de gestión electrónica de documentos, resulta extremadamente simple igualar gestión del conocimiento con acceso a documentos. La accesibilidad al documento es únicamente una parte en la cadena de gestión del conocimiento, en la que tenemos que considerar otros eslabones más importantes: los que facilitan su creación y son capaces de crear entorno de discusión.

¿Groupware es igual a gestión del conocimiento?

Las aplicaciones de trabajo en grupo (en inglés groupware), se aproximan en gran medida a las ideas que están detrás del modelo teórico de gestión del conocimiento: gestión de bases documentales, foros de discusión, servicios de directorios, etc.

Pero la existencia de foros de discusión o bases documentales en una empresa no significa que esa organización esté realizando realmente gestión de conocimiento. ¿Por qué no podemos igualar estos dos conceptos?

En este área se corre un importante peligro: invertir en tecnologías para la colaboración y en aplicaciones informáticas para comprobar que, una vez pasado el entusiasmo inicial, únicamente hemos logrado tener bases de datos de cuyos contenidos no podemos garantizar ni la calidad, ni su pertinencia con los procesos de trabajo desarrollados por la empresa.

Para evitar que esto sea así, debemos considerar la gestión del conocimiento como un proceso en el que la actitud y la participación de las personas es el factor realmente crítico para alcanzar o no el éxito. Por lo tanto, para que una estrategia GC permita alcanzar los beneficios esperados, es necesario alinear el sistema de GC con los modelos y las aplicaciones utilizadas para la gestión de los Recursos Humanos en la organización.

De esta forma, el sistema GC debe estar vinculado al sistema de competencias de la organización, nos puede servir para detectar necesidades formativas, o como base para diseñar los sistemas de compensación y promoción profesional.

En resumen, si bien el modelo propuesto por las aplicaciones para el trabajo en grupo es reutilizable en una estrategia GC, éste no es el componente más importante y debemos volver nuestra atención hacia otros modelos centrados en la gestión de personas. Tenemos que recordar que la gestión del conocimiento es un compromiso organizativo que requiere la participación de todos sus miembros, y que debe ser un proceso continuo, y no una inversión puntual en tecnologías. Para gestionar conocimiento debemos ante todo gestionar personas y ser capaces de motivarlas para trabajar y contribuir en esta dirección.

Nuevos entrantes

Inevitablemente, con la popularización de la gestión del conocimiento han aparecido nuevas empresas cuyos desarrollos se centran exclusivamente en este área.

En líneas generales, podemos decir que estas empresas ofrecen cuatro tipos de aplicaciones:

- a) aplicaciones de colaboración (*groupware*), que incorporan la posibilidad de utilizar taxonomías y lenguajes controlados como principal mecanismo de acceso a la información,
- b) aplicaciones para ver la información de forma más sencilla, generalmente en forma de mapas conceptuales
- c) programas para el *information mapping* y
- d) programas de indexación y recuperación textual avanzados.

Aplicaciones para la colaboración

En el primer grupo entrarían aplicaciones como *Wincite*, *Grapevine* u *Intraspect*. Se trata de aplicaciones que permiten crear foros de discusión, incorporan listas predefinidas de términos estructurados en forma de jerarquías para indizar las contribuciones (estas listas podrán ser personalizadas para cada organización), y afrontan el *collaborative filtering* de distinta forma.

Por ejemplo, con *Intraspect* los usuarios pueden asignar un valor a los documentos que reciben, de forma que los documentos que alcancen un mayor valor se copiarán automáticamente a unas carpetas públicas y serán indexados para estar disponibles para todos los miembros de la organización.

Estos programas tienen como principal inconveniente el mismo que las aplicaciones *groupware* tradicionales: no vinculan la gestión del conocimiento con el resto de procesos de gestión de Recursos Humanos (promoción, formación, desarrollo de competencias, etc.)

Mapas conceptuales

El segundo grupo de aplicaciones reúne a una serie de aplicaciones cuyo origen suele encontrarse en las necesidades de departamentos de márketing e inteligencia competitiva. Estos programas suelen consistir en módulos que nos permiten indexar manualmente los contenidos de los documentos indicando de qué conceptos (o personas u organizaciones) se está tratando en ellos, y qué relaciones se dan entre estos conceptos.

A partir de esta información, el subsistema de recuperación será capaz de dibujar mapas conceptuales en los que los distintos conceptos u objetos se representan mediante iconos y las relaciones mediante líneas que los unen.

Un ejemplo de este tipo de aplicaciones lo encontramos en el programa *KnowledgeX*, desarrollada por la empresa con el mismo nombre que fue adquirida a comienzos del pasado año por *IBM*. En la misma línea se encontraría la aplicación *ThemeMedia*.

Information Mapping

Dentro de este grupo encontramos aplicaciones que tratan de relacionar la información (o el conocimiento), con su ubicación física.

Un ejemplo citado con mucha frecuencia en la bibliografía es el programa *RepTools*, desarrollado por *Xerox* y que permitía movernos por una oficina y ver qué tipo de información residía en los distintos lugares.

Con independencia de la riqueza visual que ofrezcan estos programas, la idea no es nueva para los conocedores del modelo *IRM (Information Resource Management)* propuesto por **Forest W. Horton** a comienzos de los ochenta. Pero son una buena muestra de que, en ocasiones, el acceso a las personas que pueden facilitarnos información es más importante que el acceso a la documentación propiamente dicha.

Recuperación de información avanzada

El cuarto grupo de aplicaciones que se presentan a sí mismas como alternativas válidas para la gestión del conocimiento está formado por sistemas de indexación y recuperación textual que incorporan características avanzadas.

El mejor ejemplo dentro de este grupo lo constituye el programa *Sovereign-Hill*, adquirido a finales del pasado año por otro clásico de la recuperación textual: *Dataware Technologies*.

Pero al igual que sucedía con los fabricantes de programas de indexación, estos programas resuelven tan sólo una pequeña parte de las necesidades planteadas por una estrategia GC.

Conclusiones

La gestión del conocimiento está formada por el conjunto de procesos que permiten crear, diseminar y utilizar el conocimiento, entendiendo por conocimiento la experiencia colectiva adquirida por todos los miembros de una organización en la realización de su trabajo.

Una estrategia de gestión del conocimiento debe estar alineada con la estrategia de la empresa para gestionar sus Recursos Humanos, y debe constituir la base que nos permita identificar necesidades formativas y diseñar los planes de compensación y desarrollo profesional de la empresa.

Una infraestructura para la gestión del conocimiento debe incluir un subsistema para la recuperación de la información.

El subsistema de recuperación debe premiar la calidad frente a la cantidad, permitir el acceso a información interna y externa compaginar el acceso en texto libre con la utilización de taxonomías y lenguajes controlados. Para garantizar la calidad es importante establecer filtros en los procesos de captura de datos.

Respecto a la utilización de lenguajes controlados, si bien los sistemas de recuperación basados en el texto libre son un componente imprescindible, poco a poco están siendo sustituidos por el lenguaje controlado como principal mecanismo de acceso al conocimiento.

En lo referente a la indexación, el sistema tiene que permitirnos poner en relación el conocimiento y la experiencia con los procesos de trabajo, las áreas de responsabilidad y el nivel de conocimiento requerido para poder aplicar los conocimientos consignados en el documento en la resolución de problemas de negocio.

En el área de la difusión selectiva, el *collaborative filtering* está resultando ser la opción preferida frente a la difusión basada en agentes software y perfiles de búsqueda formados por palabras clave.

Al hablar de sistemas de recuperación del conocimiento, quizás deberíamos sustituir el nombre "subsistema de recuperación" por el de "subsistema de acceso", ya que no sólo se pretende obtener un documento disponible en el sistema, sino también poner en contacto a todas

las personas que hayan participado en la creación de ese documento con sus usuario potenciales para así establecer un proceso de interacción y mejora continua.

Bibliografía

DAVENPORT, Thomas; PRUSAK, Laurence. *Working Knowledge: how organizations manage what they know*. Harvard Business School, dic. 1997.