

Función de los lenguajes documentales en el tratamiento de la información en las organizaciones

Blanca GIL URDICIÁN
Profesora de la EUBD Complutense

La información existente en una empresa es heterogénea, proviene tanto de dentro como de fuera de su ámbito y, presentada en los más diversos soportes, puede ser textual, gráfica, o puede tratarse de datos o de comunicaciones verbales. Los medios que la canalizan son igualmente variados: teléfono, correspondencia, fax, etc. Teniendo en cuenta estas características, podría agruparse toda la información recibida en una organización en tres categorías:

1. Publicaciones de todo tipo: libros, publicaciones periódicas, normas, patentes, informes, publicaciones de régimen interno.
2. Datos puramente administrativos y comerciales
3. Documentación interna, que comprendería: correspondencia, expedientes, normas de funcionamiento de la empresa,...

La primera categoría ha sido gestionada tradicionalmente por el bibliotecario-documentalista. En la segunda los medios informáticos son de extrema necesidad. La tercera categoría, un tipo de documentación diversificada, apenas estructurada y formalizada, ha venido siendo manejada por personal de los distintos departamentos empresariales, con independencia entre ellos. La desconexión entre esos departamentos impide la accesibilidad a los documentos que llegan a cualquier otra dependencia de la misma empresa, de forma que se archivan y recuperan en exclusiva para aquel al que en principio iban destinados, con independencia de los demás departamentos. Estos documentos se almacenan y recuperan, por lo general, sin ningún tipo de codificación, lo que impide en muchos casos su recuperación.

El amplio campo informativo que abarca esta tercera categoría de documentos exige, sin embargo, su normalización y su tratamiento de forma integrada con el resto de la información recibida en la organización.

Esta idea de tratamiento uniforme de la documentación independientemente de sus características no es comúnmente aceptada, pero los beneficios que de ella derivan merecen que se tenga en cuenta. El análisis integrado de la información implica la comunicación interdepartamental sin la cual la gestión informativa o no es posible o se lleva a cabo con inexactitudes que son uno de los más graves inconvenientes en la profesión del gestor de la información.

La elaboración de un lenguaje documental apropiado es clave para la consecución de ese flujo informativo, que lleva implícita la necesidad de utilizar un código común.

OBJETIVOS DE UN SISTEMA DE GESTION DE LA INFORMACION

Donald W. King, teórico de la gestión de la información en las organizaciones, en su obra «Obstáculos para futuros sistemas de información»¹ manifiesta que un sistema de gestión debe perseguir, entre otros, estos cuatro objetivos:

- «1. Transferir siempre la dosis exacta de información necesaria.
2. Presentar la información en el lenguaje del usuario y a niveles de comprensión apropiados para los diferentes consultantes.
3. Proporcionar información en el momento exacto en que se necesita.
4. Llevarlo a cabo en el más corto espacio de tiempo posible.»

Se tratará de poner de relieve aquí la efectividad de los lenguajes documentales para conseguir esos fines, no sin antes apuntar que la satisfacción de las necesidades de información de los usuarios o grupos de usuarios requiere el conocimiento de sus necesidades de información, que se pueden prever a través de autoanálisis solicitados a los propios consultantes, encuestas y entrevistas y demandas de información.

En este punto procedería hacer mención a un tema que merece capítulo aparte, como es la formación de usuarios. Se habla siempre de la formación del documentalista como único responsable de un proceso en el que es el usuario el personaje principal. Tanto el investigador como el simple consultante necesitan conocer los servicios de documentación, primero para saber adónde tienen que

¹ KING, Donald W.: «Obstáculos para futuros sistemas de información». En VARIOS AUTORES: *Telecomunicación y bibliotecas*. pp. 288-289.

acudir en busca de información y, en segundo lugar, deben aprender a utilizar esos servicios. El olvido de este aspecto origina la precariedad en el uso y beneficio de los sistemas de información y documentación. Podría establecerse todo un reglamento del usuario que comprendiera las directrices para su formación teórico práctica en el uso de la documentación, pero, quizá lo más eficaz sería la introducción, en los distintos planes de estudios, de una disciplina encaminada hacia este objetivo. Ello permitiría conocer, entre otras cosas, la estructura de los lenguajes documentales utilizados, los cuales, a su vez, por estar diseñados ad hoc, darían conocimiento simultáneo de la totalidad de parcelas informativas presentes en una organización determinada.

Un lenguaje científico debe cumplir las siguientes condiciones: dará acceso rápido, seguro y multidimensional a la información y ante una búsqueda, será capaz de encontrar la documentación pertinente dentro del sistema, por compleja y detallada que sea. Pero no todo lenguaje es capaz de dar respuesta al perfil descrito por King: quizá para la consecución de los planes previstos por una organización en materia documental sería el tesoro el instrumento idóneo, ya que sigue los principios de la indización coordinada, que permiten memorizar y encontrar cualquier tipo de documento, dando además acceso a esa información a partir de varios elementos del contenido, es decir, de forma multidimensional.

La utilización de un tesoro diseñado a propósito para una organización, —con temática específica— permitiría la transferencia de la dosis de información necesaria en un momento *dado*. Un lenguaje documental, respetando el principio de relevancia, estaría formado por descriptores con una gran carga de significación, lo que evitaría cualquier género de ambigüedad en el momento de la búsqueda.

La existencia de una adecuada codificación permite, en efecto, la recuperación de la documentación pertinente y propicia además la conexión con aspectos informativos que quizá el usuario no se planteó en principio y que localiza en el momento de la búsqueda gracias a la estructura de ciertos lenguajes que permiten conectar cada concepto con aquellos que tienen relación con el que buscamos. Esta riqueza de matices es posible por la previsión con que se elaboran; las relaciones paradigmáticas que existen entre los descriptores se utilizan para adaptar todavía mejor los resultados de la búsqueda documental a las necesidades del usuario. Esto y lo expuesto anteriormente apuntan al tesoro como el instrumento por excelencia para la consecución de estos objetivos. Su capacidad de control de temáticas especializadas como las que se tratan, por lo general, en las organizaciones, le convierte en el hilo conductor de la información. Los tesoros, antes de incluir un término en su estructura, analizan previamente cada aspecto concreto con profundo estudio de probabilidades, de modo que pueden abarcar las materias de la empresa en cuestión bajo todos los posibles puntos de

vista sin velar significados, algo que ocurre a menudo en la jerga de las disciplinas.

El segundo objetivo que se plantea un sistema de gestión de la información es, como señalé anteriormente, el de *presentar la información en el lenguaje del usuario y a niveles de comprensión apropiados para los diferentes consultantes*. El procedimiento ideal para conseguir este propósito sería la utilización del lenguaje natural, que permite el diálogo abierto con el sistema y una gran exhaustividad, eso sí, en detrimento de la precisión. La ambigüedad que genera la búsqueda documental a través de este tipo de lenguaje exige, por lo tanto, una reflexión previa a su adopción para el manejo de la información. En un primer análisis habría que considerar las características de heterogeneidad documental con que contamos. En ese examen se tendría en cuenta el material disponible y la riqueza conceptual de la documentación a manejar, ya que, dependiendo de ello, en ocasiones se hace necesaria la combinación de los procedimientos de búsqueda en texto libre y mediante los descriptores de un tesaurus; se debe llegar, en definitiva, a conjugar el nivel de especificidad del lenguaje con las necesidades reales de forma que el tiempo de búsqueda y recuperación sea mínimo.

La propuesta que aquí se hace es la de aplicar un tesaurus a las tareas del análisis documental por su capacidad para la descripción de cualquier temática especializada sin rechazar el posible uso del lenguaje natural como alternativa, en determinados casos, o proceder a la combinación de ambos. Es bien conocida la oposición: lenguaje libre versus lenguaje controlado, sin embargo, la adopción de ambos procedimientos propicia, en casos tipificados, una estrategia de carácter complementario muy favorable para la recuperación. Un caso ejemplar lo constituye la documentación manejada en centros de documentación de prensa, caracterizada por el uso de un lenguaje (vocabulario) sectorial perteneciente al mismo tiempo a diversas disciplinas. Esa universalidad conlleva la necesidad de tratamiento compartido; si se estuviera únicamente frente a una diversidad de materias, la adopción de un lenguaje de estructura jerárquica sería adecuada, pero el sistema de polijerarquías y relaciones que caracterizan al vocabulario periódico hace pensar en otro género de herramientas para su análisis.

Aparte las excepciones, si se decide la conveniencia de indizar por medio de un tesaurus, lo aconsejable es proceder a la adopción de alguno ya publicado sobre la temática que interesa o, en su defecto, de uno de un nivel conceptual más genérico. El siguiente paso sería la adaptación de ese instrumento a las necesidades informativas de la organización incorporando términos de la documentación que se maneja en la empresa, que habremos recogido de la indización de una muestra de varios cientos de documentos representativos de cada uno de los grandes campos a cubrir. Por último excluiríamos del tesaurus de referencia aquellos términos que no se identifican con la materia de interés.

Van Slype² describe el proceso de adaptación de un tesoro de la siguiente forma:

- «1. Se adquieren los tesauros disponibles sobre la temática en cuestión.
 2. Se prepara una muestra de documentos para indizar (unos mil).
 3. Se indizan esos documentos en lenguaje natural.
 4. Se intenta traducir los descriptores libres de la indización en descriptores controlados extraídos de cada uno de los tesauros examinados, uno tras otro.
 5. El tesoro que haya proporcionado la mayor cantidad de descriptores que respondan a las necesidades del sistema documental será el candidato para la elección.
- * Si proporciona más del 95 por ciento de los descriptores necesarios, podrá ser adoptado como tesoro.
- * Si proporciona del 80 al 95 por ciento de los descriptores necesarios, podrá ser completado por los descriptores que falten y llegar a ser adecuado.
- * Si proporciona del 50 al 80 por ciento de los descriptores necesarios, sólo podrá ser utilizado como una de las fuentes de terminología incorporable al tesoro del sistema documental, el cual deberá ser construido.
- * Si proporciona menos del 50 por ciento de los descriptores necesarios, sólo podrá servir de fuente secundaria de terminología del tesoro del sistema, el cual deberá ser construido.»

INDIZACION MANUAL VERSUS INDIZACION AUTOMATICA

Los dos últimos objetivos señalados por Donald King como prioritarios de un sistema de gestión de información: proporcionar información en el momento exacto en que se necesita y *llevarlo a cabo en el menor tiempo posible*, están intrínsecamente relacionados y llevan implícita la necesidad de automatización del sistema de recuperación de la información. Actualmente la mayoría de las organizaciones disponen de procedimientos automáticos; algunas operan en texto libre a la hora de realizar las búsquedas, de modo que acceden al documento a través de todas las palabras significativas de que está compuesto. De esta particularidad deriva su susceptibilidad al ruido documental, inevitable a pesar de los recursos con que cuentan los programas informáticos documentales. La velocidad conseguida con el uso de procedimientos automáticos en el proceso de indización se ve afectada por esta razón. La necesidad de establecer un control lingüístico se hace evidente ya que la falta de precisión generada por el tratamiento mediante lenguaje libre no favorece a los objetivos de la empresa. La indización manual permite la profundidad en el análisis y evita problemas de

² VAN SLYPE, Georges: *Los lenguajes de indización*. Fundación Germán Sánchez Ruipérez, Madrid, 1991, pp. 88-89.

polisemia, ya que el indizador es quien elige los términos más convenientes para definir los conceptos contenidos en un texto. La operación manual permite, por tanto, una mayor racionalización del proceso y posibilita la extracción de conceptos no expresados de forma explícita por el autor del texto. En cualquier caso, se trate de indización inteligente o automática, la propuesta que aquí se hace es la del control del vocabulario por las ventajas que aporta en el momento de la recuperación.

La indización ayudada por ordenador se sirve de métodos sintácticos y por asignación, entre otros, que evitan la dispersión y facilitan la formulación de las ecuaciones de búsqueda. El procedimiento sintáctico se basa en la combinación sintáctica de los términos por medio de enlaces o «links», que permiten el análisis contextualizado de los conceptos y garantizan la coherencia entre documento solicitado y documento recuperado. El método de indización por asignación, por su parte, realiza la operación en base a los descriptores registrados en un tesoro, evitando cualquier accidente lingüístico. El procedimiento por asignación consta de dos etapas: en la primera se localizan los descriptores del tesoro correspondientes a los conceptos que figuren en el texto; en la segunda, se traducen esas entradas por los descriptores del tesoro.

Además de estos métodos, Chaumier³ hace referencia al sistema documental del INSPEC, basado también en la utilización de un tesoro para la indización automática. El procedimiento seguido por este sistema es el siguiente: «Los documentos son indizados manualmente en lenguaje libre por unos indizadores. El conjunto de las indizaciones realizadas se somete al sistema... que opera en combinación con uno de indización ayudada por ordenador».

En España la mayoría de las bases de datos trabajan con lenguaje controlado, aunque no en todos los casos se puede decir que utilizan tesauros. La precariedad de su elaboración afecta no sólo a los aspectos formales, sino que trasciende al plano de los contenidos. Algunas de estas herramientas que se presentan bajo la denominación de tesauros, no son más que simples listas alfabéticas provistas de relaciones de equivalencia, otras presentan índices sistemáticos sin desarrollar, acompañados, en algunos casos, de índices permutados; pero son muy pocas las bases de datos que manejan un tesoro propiamente dicho.

Tanto operando de forma manual como automática, se plantea la cuestión del control lingüístico, como factor determinante para el funcionamiento de los servicios de información de las organizaciones. El conocimiento del procedimiento más adecuado para cada caso concreto requiere el estudio detenido del tipo de información con que se cuenta, de las características del sistema y de los recursos humanos y tecnológicos de que se dispone.

³ CHAUMIER, Jacques: *Análisis y lenguajes documentales*. Mitre, Barcelona, 1986, pp. 57-58.