

El archivo inteligente

INFORMATIZAR LA DOCUMENTACIÓN PERIODÍSTICA NO ES UN PROBLEMA TRIVIAL, COMO ALGUNOS VENEDORES DE SOFTWARE LO HACEN PENSAR. UN ARCHIVO REALMENTE EFICIENTE, ES DECIR, UN VERDADERO SISTEMA DOCUMENTAL, DEBE SER ANALÍTICO, Y NO SIMPLEMENTE UN DEPÓSITO AMORFO DE DATOS INCONEXOS. AL CONSTRUIRLO HAY QUE TENER EN CUENTA UN PRINCIPIO ESENCIAL: EL TIEMPO INVERTIDO EN LA CONFECCIÓN DE LOS REGISTROS SE RECUPERA CON CRECES AL UTILIZAR LA INFORMACIÓN.

Raymond Colle*

* Raymond Colle es Licenciado en Teología, profesor de la Escuela de Periodismo de la Universidad Católica.

Los Sistemas Documentales cumplen una función cada día más importante y vital en todo de tipo de organización. Máxime en una empresa cuya razón de ser consiste en informar. Mucha información que era hasta hace poco (no más de 10 años) del dominio «del lápiz y del papel» se ciñe hoy a formatos estándares e ingresa a sistemas computacionales que facilitan no sólo su utilización posterior sino también la obtención de «subproductos» informativos de alta utilidad para la toma de decisiones.

Por otra parte, cierto tipo de información —como la bibliográfica, científica o tecnológica—, que era siempre requerida pero de difícil o lenta difusión, ha ganado enormemente en posibilidades de acceso gracias a los sistemas informáticos de gestión de «bancos de datos» y a las redes informáticas nacionales e internacionales.

Nacidos en los Estados Unidos a principios de los 60 —junto a la tercera generación de computadores— los sistemas documentales informatizados sólo fueron cobrando mayor difusión en los 70, con el nacimiento del llamado «modelo relacional», alcanzando cierta madurez en los 80. Esto explica que el mercado mundial, en el que circulaban unos 500 de estos bancos de información en 1975, haya pasado a 2.500 diez años después. Así, han empezado a transformar los hábitos de trabajo de los docentes, investigadores, profesionales y ejecutivos, usuarios que hoy suman más de 50 millones de personas en el mundo.

Si bien en Estados Unidos la informatización de los centros de documentación de los medios de prensa se inició en los 70, en Europa empezó a mediados de los 80. Y el costo de infraestructura (especialmente para conservar grandes cantidades de datos en medios magnéticos) llevó a muchos medios a constituir sistemas que estuvieran al servicio de conjuntos de empresas (como la Asociación de Editores de Periódicos en Bélgica, o el para-estatal «Banco de Información Política de Actualidad» de Francia). Gracias a la posterior evolución de los

«AL INVERTIR MUY POCO TIEMPO PARA CONFECCIONAR LOS REGISTROS, SE PIERDE GRAN CANTIDAD DE TIEMPO EN LA RECUPERACIÓN Y SE DESALIENTA A LOS USUARIOS DEBIDO AL ENORME RUIDO INFORMATIVO QUE SE GENERA.»

sistemas de compresión de datos y el abaratamiento de los soportes magnéticos, hoy muchos cuentan al menos con Bases de Datos donde se conserva información sobre toda publicación propia.

SUPUESTOS BÁSICOS

Abordar adecuadamente la informatización de la documentación periodística no es un problema trivial, como lo hacen pensar algunas veces los vendedores de software. Lo más común es encontrarse con ofertas de un tipo de software que retoma el texto publicado y lo archiva *full text* en forma de fichas (registros), ofreciendo a la vez facilidades de búsqueda/recuperación. Esta forma de trabajar parece bastante común en Norteamérica pero no permite responder a preguntas precisas, como «¿Qué dijo Fulano acerca de tal tema?», típica de un periodista del sector político. Tampoco lleva a la correcta implantación de un verdadero *sistema* documental, el cual se compone necesariamente de múltiples archivos que se complementen mutuamente. Típico error sistémico sería, por ejemplo, la ausencia de un Archivo Biográfico, obligando a leer numerosos registros de un archivo *full text* de crónicas para reconstruir el currículum de algún personaje público.

Esto lleva a señalar dos condiciones básicas a tener presentes antes de informatizar la documentación periodística:

1. Mientras más analítico es el archivo diseñado, más fácil y más precisamente se podrá recuperar la información buscada a posteriori.
2. Es indispensable configurar un verdadero *sistema documental*, y no solamente construir diversos tipos de archivos, de acuerdo a la intuición o las necesidades del momento.

De la primera condición se deriva otro principio esencial:

El tiempo que se invierte en la confección de los registros (ingreso de información) se recupera con creces en la recuperación de la información.

Este es justamente el principio que infringen los modelos no-analíticos: al invertir muy poco tiempo para confeccionar los registros (puede hacerse automáticamente), se pierde gran cantidad de tiempo en la recuperación y se desalienta a los usuarios debido al enorme *ruido* informativo que se genera (registros que no responden a las preguntas que guían la búsqueda).

Pero los modelos analíticos tienen además otras ventajas, muy importantes hoy. Permiten ligar la entrada de información con nuevas formas de salida de la misma. En efecto, cuando se cuenta con múltiples atributos (como, por ejemplo, el nombre del «actor» de un hecho noticioso, el nombre de la institución a la cual pertenece, el tipo de acción y el tema tratado —si es una declaración— etc.) se pueden generar estadísticas tales como «quiénes han hecho declaración sobre tal tema, con qué frecuencia, durante tal o cual período». El trabajo interpretativo del periodista se ve grandemente facilitado. Y la misma tabla de resultado constituye un nuevo producto también publicable.

Paralelamente, es posible instalar un sistema de *diseminación* de información (como se explica más adelante), aproximándose así a un sistema de «diario a la carta», en que los suscriptores reciban boletines personalizados, de acuerdo a sus intereses particulares.

SISTEMA DOCUMENTAL

Sólo se puede hablar de *sistema documental* si el conjunto de archivos se estructura en forma de «espacio de información», permitiendo diversas maneras de lectura de la información que contiene y contemplando la existencia de relaciones entre los distintos archivos que lo conforman.

El diseño de sistemas documentales debe partir del análisis de las características propias e intrínsecas de los referentes (objetos y eventos) representados, además de tener en cuenta las necesidades y los hábitos de los usuarios, es decir la forma en que los destinatarios de la misma tenderán a requerirla. Pero los hábitos de los usuarios no pueden ser los únicos que han de guiar el diseño, por cuanto ello podría impedir un mejor aprovechamiento de la información para nuevas tareas o productos que se pueden diseñar hoy o en el futuro.

Sólo un sistema documental construido de acuerdo a estos principios y administrado con herramientas in-

formáticas que permitan mantener relaciones entre los datos de varios archivos constituye un verdadero y provechoso *banco de datos*.

BANCO DE DATOS

El término *base de datos* se utiliza generalmente como sinónimo de «banco de datos». Sin embargo, se hace algunas veces una distinción entre estos apelativos: el «banco» hace referencia más bien al contenido (significación), mientras la «base» se relaciona con la estructura lógico-matemática y al tipo de software utilizado, poniendo el énfasis en el carácter de «fundamento» y punto de partida sobre el cual se construye. De ahí que los programas informáticos (software) se llamen habitualmente «Sistemas de Gestión de Bases de Datos» (SGBD) y no «sistemas de gestión de *bancos* de datos».

Todo SGBD debe cumplir como mínimo las siguientes condiciones:

1. Las estructuras de datos (espacio informativo) son simples e independientes del programa que genera los datos.
2. Varios espacios informativos –si los hay– se asocian mediante la presencia de al menos un atributo común¹.
3. Un conjunto de operadores permite la definición, búsqueda y actualización de los datos.
4. Un conjunto de requisitos de integridad define el estado coherente de la base de datos.

SISTEMA DOCUMENTAL PERIODÍSTICO

Al desarrollar un banco de datos con fines **periodísticos**, se ha de tomar en cuenta una multiplicidad de factores y, en particular, la necesidad de desagregar información para poder recomponer la información. ¿Qué entendemos por esto? La médula de la información periodística es la actualidad. Pero la actualidad está compuesta de acciones que se suceden unas a otras. Los acontecimientos son tan efímeros como los días y las horas, mientras las personas involucradas y –más aún– los escenarios en que ocurren pertenecen a un orden de permanencia muy superior.

Por otra parte, en los medios de prensa que conocemos existe habitualmente una brecha entre la bibliote-

ca (atlas, libros de historia, «quién es quién», enciclopedia, etc.) y el archivo de crónica: vacío de información y ausencia de nexos que conduzcan de un tipo de registro a otro. Para conocer los antecedentes de un personaje público, se requiere a veces revisar noticias (banco de datos noticioso o carpetas de recortes); para conocer la historia reciente de un país debe hacerse lo mismo pero con una gran cantidad de noticias separadas, lo cual puede resultar engorroso e ineficiente.

Consecuentemente, una adecuada economía (de espacio y de tiempo de búsqueda) en un sistema de archivos implica separar las informaciones más permanentes de las más efímeras. Así, datos relativos a personas, instituciones, lugares, instrumentos técnicos o conceptos científicos deberán ser separados de los registros de acontecimientos (o «Archivo de Crónica», como lo llamamos), reduciendo de este modo la longitud de éstos y facilitando las consultas necesarias para posteriores trabajos de redacción.

Los archivos que componen un Sistema Documental Periodístico son típicamente los siguientes:

- Crónica
- Biografías
- Instituciones
- Países (Sistema político, estructuras ejecutiva y legislativa, etc.)
- Fotografías
- Enciclopedia (Definiciones, explicaciones de objetos, técnica, etc. que no estén en enciclopedias publicadas)
- Bibliografía
- Estadísticas

«SÓLO SE PUEDE HABLAR DE SISTEMA DOCUMENTAL SI EL CONJUNTO DE ARCHIVOS SE ESTRUCTURA EN FORMA DE «ESPACIO DE INFORMACIÓN», PERMITIENDO DIVERSAS MANERAS DE LECTURA DE LA INFORMACIÓN QUE CONTIENE Y CONTEMPLANDO LA EXISTENCIA DE RELACIONES ENTRE LOS DISTINTOS ARCHIVOS QUE LO CONFORMAN.»

¹ Un «atributo» es una variable de descripción de un referente (evento u objeto que se documenta).

«UNA ADECUADA ECONOMÍA (DE ESPACIO Y DE TIEMPO DE BÚSQUEDA) EN UN SISTEMA DE ARCHIVOS IMPLICA SEPARAR LAS INFORMACIONES MÁS PERMANENTES DE LAS MÁS EFÍMERAS. ASÍ, DATOS RELATIVOS A PERSONAS, INSTITUCIONES, LUGARES, INSTRUMENTOS TÉCNICOS O CONCEPTOS CIENTÍFICOS DEBERÁN SER SEPARADOS DE LOS REGISTROS DE ACONTECIMIENTOS.»

En el futuro estos archivos podrán ser integrados gracias a los nuevos sistemas de *hipertexto* e *hiperarchivos*, de modo que facilite «navegar» de un archivo a otro sin siquiera tener a la vista las separaciones entre uno y otro, tal como lo hemos demostrado en un sistema experimental y demostrativo creado en nuestra Escuela de Periodismo, con el *HyperCard* de Macintosh².

ATRIBUTOS MÍNIMOS DEL ARCHIVO DE CRÓNICA

La selección de las variables mediante las cuales se describirán los referentes (e.d. lo observado que se va a «documentar») ha de partir del análisis de la naturaleza de éstos. Una primera distinción consiste en separar eventos (acontecimientos, objetos de la «crónica») de objetos (personajes, instituciones, etc.), tal como lo hicimos para definir los archivos mencionados³.

Este análisis intrínseco ha de complementarse con el análisis de la «conversación-tipo» que un usuario tendría con un documentalista con el fin de encontrar lo que busca. En el caso de hechos noticiosos, hemos de preguntarnos, en consecuencia, cuáles serían las preguntas más comunes que harían los periodistas. Encontramos por ejemplo:

- ¿Qué hizo Fulano en tal momento o período de tiempo?
- ¿Cuándo hizo tal cosa Fulano?
- ¿Cuántos hechos de tal tipo ocurrieron en tal período?
- ¿Qué pasó en tal país desde tal fecha?
- ¿Quiénes fueron los más afectados por tal tipo de hecho?

De estas preguntas y del modo de trabajar del analista de información noticiosa, se deduce una serie de puntos de acceso o «entradas» necesarias en el banco de datos relativo a acontecimientos:

1. Fecha
2. Lugar (país, ciudad)
3. Actor (nombre propio)
4. Objeto (tipo de hecho/referente; representado por descriptores, según *Tesaurus*)
5. Implicados (nombres propios de pacientes, destinatarios u otras personas citadas, como cuando un político se refiere a las declaraciones de otro)
6. Reseña del acontecimiento

PRODUCTO

Con los antecedentes acumulados podemos –a posteriori– realizar consultas, como buscar todas las noticias relativas a un determinado tema (p.ej.: el plebiscito de 1988, que tendría la clave temática «41») o las intervenciones de una determinada persona (p.ej.: «Büchi, H.»), o generar «informes» copiando la totalidad o parte de los registros que respondan a estos criterios de búsqueda o produciendo tablas ordenadas con datos correspondientes a algunos de los atributos.

El procesamiento en tablas nos conduce a otra posibilidad: la confección de estadísticas, generando de este modo nueva información o *meta-información* (e.d. información sobre la información). Todo ello podrá ser de sumo interés para confeccionar artículos que sintetizen la actualidad durante cierto período, expliquen antecedentes o factores contextuales, demuestren el rol preponderante de ciertas personalidades, sugieran proyecciones futuras, etc. Se estima que la *meta-información* será un componente de importancia creciente en el próximo siglo. Un adecuado sistema documental permitiría por lo tanto «ganar puntos» en la carrera del futuro.

DISEMINACIÓN

Se considera que un sistema documental cumple realmente su propósito si permite hacer llegar eficientemente la información a los posibles usuarios. Estos se clasifican en dos categorías: los «ocasionales» y los «ha-

² COLLE, RAYMOND: «El Periodista», en *Cuadernos de Información*, N° 6, 1990, pág. 119.

³ Por más detalles, ver COLLE, RAYMOND: *Documentación periodística: principios y aplicaciones*, Ed. Pontificia Universidad Católica, Santiago, 1992.

bituales». Para los ocasionales, la posibilidad de interrogación a través de un documentalista que maneje el sistema es suficiente: encontrará respuesta –más rápida– a sus consultas cuando se produzcan (esto puede llevar a sistemas de información por vía telefónica).

Pero el mayor poder de un sistema documental computarizado consiste en hacer llegar regularmente a los usuarios habituales una selección de informaciones que responda a intereses específicos. Esto es lo que se entiende por proceso de *diseminación*. Implica la existencia de un archivo con datos acerca de los usuarios habituales (nombre, dirección) y especialmente la *descripción* de lo que les interesa, en el mismo lenguaje documental utilizado para describir los referentes. Las variables de «descripción», por lo tanto, han de ser atributos comunes a los archivos de datos sobre referentes y sobre usuarios. En este caso, con un SGBD relacional, es factible automatizar la comparación (referentes que pueden interesar ciertas personas) y producir listas de antecedentes que se envíen a los interesados. Esta es la *diseminación automática*.

¿QUÉ PASA CON LOS RECORTES?

El archivo de prensa tradicional se componía esencialmente de recortes. La función de los recortes es doble: en primer término asegura la conservación de la información histórica acerca de los eventos y objetos (función cubierta por un Sistema Documental como el que hemos descrito), en segundo término consiste en conservar diferentes versiones -procedentes de diferentes publicaciones- de los relatos y comentarios, especialmente para estudiar tendencias interpretativas. Esta última función es secundaria en relación al núcleo de la información periodística y, por lo tanto, la empresa ha de ponderar la importancia de su conservación para ese exclusivo fin.

Obviamente, si se cuenta con un sistema de conservación digital de texto completo, este nuevo archivo reemplazará el de recortes de la propia publicación. Pero se puede perfectamente optar por no conservar el texto completo en forma digital, ya que no juega ningún papel en la producción automática de «meta-información» (la cual, como lo hemos visto, requiere archivos analíticos). En este caso, basta que los registros del Archivo de Crónica contengan la referencia física del texto (fecha de publicación y página), con lo cual se remite al original en

«EL PROCESAMIENTO EN TABLAS NOS CONDUCE A OTRA POSIBILIDAD: LA CONFECCIÓN DE ESTADÍSTICAS, GENERANDO DE ESTE MODO NUEVA INFORMACIÓN O META-INFORMACIÓN (E.D. INFORMACIÓN SOBRE LA INFORMACIÓN). TODO ELLO PODRÁ SER DE SUMO INTERÉS PARA CONFECCIONAR ARTÍCULOS QUE SINTETICEN LA ACTUALIDAD DURANTE CIERTO PERÍODO, EXPLIQUEN ANTECEDENTES O FACTORES CONTEXTUALES, DEMUESTREN EL ROL PREPONDERANTE DE CIERTAS PERSONALIDADES, SUGIERAN PROYECCIONES FUTURAS, ETC.»

papel (de la colección que se conserva siempre) o microfilmado.

El microfilm es también la solución adoptada por muchos medios europeos para la sustitución de los recortes de publicaciones ajenas. En este caso lo que se asegura es el acceso a los microfilmes mediante un archivo digital de referencias bibliográficas, en el que se incluyen algunos campos de datos que aseguren la relación lógico-semántica con el archivo de crónica si corresponde (con atributos comunes como los descriptores, la fecha del acontecimiento y los actores).

CONCLUSIÓN

El acercamiento de las metodologías del trabajo científico y del quehacer periodístico es hoy una exigencia cada vez más importante, especialmente porque el público requiere mayor rigor en el trabajo informativo. Sin este rigor, el hombre con responsabilidades sociales (político, empresario, profesional, etc.) no se encuentra preparado para interpretar la realidad y ajustar sus decisiones al contexto en que debe actuar. Esto hace indispensable que la prensa cuente con una base documental más sólida, mejor estructurada y de más fácil acceso y con un «procesamiento» más fino y al alcance de los nuevos usuarios. ■

