

Digitalización

*y acceso en línea a la **prensa canaria***



*Digitalizar el patrimonio periodístico de Canarias y ponerlo al alcance de la comunidad universitaria es ya una realidad consolidada en la Biblioteca de la Universidad de las Palmas de Gran Canaria. Con ello no sólo se ha logrado preservar sino también **difundir un material de interés social**, haciendo verdadero honor a lo que hoy conocemos como Sociedad de la Información y la Comunicación.*

EL COMIENZO DE LA AVENTURA

A mediados de 1996, la Biblioteca de la Universidad de Las Palmas de Gran Canaria auspició el interés, entre diversas entidades afines de Gran Canaria y Tenerife, por acometer la digitalización de todas las cabeceras de la prensa corriente publicada en las dos provincias canarias.

La idea, bastante simple, consistía en contratar —habida cuenta de la inexistencia de medios propios con los que acometer un objetivo de esta envergadura— a una empresa especializada que procediera a digitalizar la totalidad de las ediciones diarias de la prensa publicada y remitiera posteriormente, a cada uno de los centros participantes, las imágenes escaneadas almacenadas en CD-Rom para su consulta en monopuesto por los usuarios correspondientes.

Para ello contactamos con una empresa de Madrid especializada en digitalización, Centro de Tratamiento de la Documentación (CTD), y se concertó el envío periódico de lotes de los ejemplares recibidos por suscripción. La escasa dotación presupuestaria con que contaban algunas bibliotecas interesadas en participar, hizo desistir en un principio a algunas de ellas, mientras que las restantes, tampoco en una situación económica boyante, se decantaban por unos mínimos que aseguraran la reproducción digital periódica, aunque sin incluir una digitalización que incluyera los tonos grises necesarios para incluir toda la gama cromática de las fotos incluidas o el color de determinadas páginas y suplementos.

Bajo ese panorama nació esta iniciativa en la que se incluyeron las ocho cabeceras que se publicaban a diario en la región (*El Día*, *Jornada Deportiva*, *La Gaceta de Canarias*, *Diario de Avisos*, *La Provincia*, *Canarias7*, *Diario de Las Palmas* y *La Gaceta de Las Palmas*). Unos meses después se recibían los CD-R conteniendo un visor propio en cada disco conjuntamente con las imágenes monocromas en formato TIFF de 1 bit con compresión CCITT G-4/ITU-T T.6, dispuestos bajo una estructura de subdirectorios con ordenación cronológica.

De este “consorcio” se beneficiaban entidades como las bibliotecas universitarias de Las Palmas de Gran Canaria y La Laguna, la Biblioteca municipal de Santa Cruz de Tenerife, la Sociedad Científica El Museo Canario, la Biblioteca Insular de Gran Canaria, la Biblioteca Pública del Estado de la provincia de Las Palmas y otra serie de organismos que han ido sumándose o renunciando en determinados momentos y que, lógicamente, ha afectado a los costes que ha de pagar cada uno de los participantes, por cuanto los mismos se reparten entre los entes interesados, siendo por supuesto más barato cuantos más participen.

Ulteriormente, problemas en los tiempos de recepción de los discos y la detección de una significativa cantidad de errores nos hicieron replantearnos nuestra relación con CTD, pasando a contratar a Tecnologías de la Documentación (Tecnodoc), empresa con la que seguimos desde entonces. Pese a dicho cambio, en que se mejoraron los plazos de entrega y se corrigieron los errores, el procedimiento de almacenamiento y estructura de los datos no cambió sustancialmente, ya que los discos seguían incorporando una estructura similar y su contenido se mostraba también mediante un visor muy sencillo desarrollado por la propia empresa.

Mas adelante, Tecnodoc ofreció un producto complementario a la simple visualización de imágenes, consistente en la posibilidad de búsqueda dentro del texto de las noticias, al haber implementado una herramienta de reconocimiento óptico de caracteres. El visor ofrecido incorporó entonces la posibilidad de hacer dicha búsqueda y además poder enlazar con la imagen correspondiente, a fin de poder leer la información dentro de su contexto. Por desgracia, la inexistencia de un sistema de índice inverso con punteros que estructurase la información como una base de datos, lejos de dar una respuesta casi inmediata a las búsquedas, hacía muy lenta la búsqueda del texto deseado.

Pero una nueva estructuración de dichos datos textuales y la evolución del visor proporcionado por Tecnodoc mejoraron sustancialmente dicha búsqueda. Los



usuarios de la Biblioteca Universitaria accedían, desde diversos ordenadores establecidos al efecto en la Hemeroteca de nuestro Edificio Central, a la prensa digitalizada que, transcurridos los años, superaba ya los dos millares de CD-ROM, lo cual suponía ya un considerable problema en su gestión, acceso y almacenamiento físico.

LA MIRADA TAMBIÉN HACIA EL PASADO

Pasados algunos años desde su comienzo, las propias editoras de los periódicos publicados en Gran Canaria, bajo los auspicios de la empresa encargada de su digitalización, accedieron a la propuesta para digitalizar todo su fondo retrospectivo desde el mismo año de su fundación. Fue así como se acometió la tarea de recopilar la totalidad de los números aparecidos colaborando con otras entidades, como la Hemeroteca de El Museo Canario y la Casa-Museo León y Castillo de la localidad grancanaria de Telde, al no conservar todos los ejemplares la propia empresa editora, Editorial Prensa Canaria (integrada en Editorial Prensa Ibérica). De esta forma fue posible digitalizar el *Diario de Las Palmas*, cabecera aparecida en 1893 hasta su ulterior desaparición en 1999, y *La Provincia*, desde su creación en 1911 hasta la actualidad. Este corpus se encuentra actualmente a la venta en formato CD-ROM. Aún así, existen determinados periodos cuyos ejemplares continúan en paradero desconocido, razón por la cual continúan las investigaciones para conseguir completar la totalidad de dichos títulos.

Igual proceder siguió a posteriori la empresa Inforcasa, editora de la cabecera *Canarias7* y que asumió la completa digitalización de su periódico, mucho más reciente, desde 1982 hasta hoy. La Biblioteca de la Universidad de Las Palmas de Gran Canaria conseguía así contar con la copia electrónica de todos los periódicos publicados en la actualidad en nuestra provincia desde el mismo momento de su creación, lo que añadía un valor considerable para las investigaciones que tuvieran que hacer uso de la prensa.

Sin embargo, y pese a estos considerables avances de recopilación de información, constatábamos la escasa utilización

de dicho corpus documental debido a dos razones: el poco conocimiento de este recurso entre nuestros usuarios potenciales y el hecho de existir un único punto de acceso localizado en la Hemeroteca del Edificio Central de la BULPGC con un horario, aunque amplio, limitado a jornadas laborales estándar. Además de ello, el hecho de que su localización se encuentre a considerable distancia física de los investigadores más interesados en hacer uso de ello —los de las facultades de humanidades— agravaba aún más dicha situación.

Por ello y gracias a la incorporación de nuevo personal informático asignado a la BULPGC, y al hecho de haberse aumentado en gran medida la capacidad —aumentada a cuatro terabytes— del servidor de nuestra Biblioteca Digital, nos vimos motivados para acometer **una planificación tendente a poner toda la digitalización efectuada hasta entonces en línea, para posibilitar el acceso desde cualquier punto de la Universidad a todos nuestros usuarios potenciales.**

La planificación se dividió en dos frentes. Por un lado consistió en establecer unos nuevos parámetros, acordes a lo que se pretendía, para la digitalización efectuada a partir de enero de 2004, eliminando un formato como el TIFF, poco favorable a su visualización en línea, y también eliminar el sistema estructurado en infinitos ficheros monopágina. Así pues, se encargó, sólo para la BULPGC —pero con posibilidad de que los restantes participantes en este proyecto cooperativo se unieran posteriormente, como así fue—, que la digitalización entregada en el futuro tuviera los siguientes parámetros:

- Suministro en soporte DVD-R estándar de 4,7 Gb.
- Formato PDF con opciones de compresión ajustadas a compresión de texto y gráficos lineales e imágenes monocromas con JBIG-2 multipágina, OCR incorporado en formato de “imagen original con texto oculto”, no encriptado, tamaño de página ajustada a A4, visualización de miniaturas y calidad para visualización de gráficos a 300 ppp.
- Un solo fichero por periódico y día.
- Nombre del fichero según una sintaxis determinada que permitiera identificar por sí mismo y sin necesidad de abrirlo

- el periódico del que se trate así como la fecha a la que corresponda.
- Ubicación en el directorio raíz sin ninguna estructura de directorios.

Por lógica, y en aras de la deseable homogeneidad, pedimos también presupuesto para la conversión, con los mismos nuevos parámetros descritos, de todos los ficheros de los periódicos digitalizados hasta la fecha. Sin embargo, el excesivo coste de la operación nos hizo desistir de dicho propósito.

TRANSPARENCIA PARA EL USUARIO

El otro ámbito de planificación consistía en el estudio de la manera más adecuada de poner en línea toda esta documentación, teniendo en cuenta que debíamos integrar dos sistemas muy distintos (TIFF monopágina con estructura de directorios conjuntamente con PDF multipágina sin estructura de directorios) en un único sistema de forma transparente al usuario. La totalidad de la plataforma en la que se basa la Biblioteca Digital de la ULPGC se encuentra en entorno Linux Red Hat, con programación PHP y base de datos MySQL y que, por consiguiente, las soluciones adoptadas debían adaptarse a ello.

Para los ficheros en formato TIFF, que incluía toda la prensa aparecida hasta 2003, descartamos desde el comienzo tener un visor específico para dicho formato, por lo que se hacía necesario contar con una programación, software o librería que permitiera convertir “al vuelo” la página seleccionada por el usuario y la mostrara en formato PDF en el mismo momento de la búsqueda. Una vez encontrada dicha solución, programamos en PHP un sistema que mostrase al usuario en primer lugar la cabecera de su elección y luego una plantilla de calendario en la cual, en primer lugar se da la posibilidad de elegir la década y el año, y después el sistema construye una página-calendario en el que aparecen, de forma resaltada, todas las fechas disponibles sobre las que solo hay que pulsar para que se abra el Adobe Reader y muestre el periódico seleccionado. La única diferencia apreciable para el usuario es que en los periódicos de antes de 2004 debe ir pulsando dentro de cada una

de las páginas de su elección, a fin de que el sistema se la vaya convirtiendo “al vuelo” y poder mostrársela en PDF, mientras que a partir de dicha fecha se navega por las páginas dentro del propio fichero al ser PDF nativo. Asimismo, en el segundo caso, el usuario tiene la posibilidad de copiar y pegar texto y gráficos, ventaja que no es posible para las fechas anteriores.

Mientras se fue desarrollando esta programación, desde la propia Hemeroteca se estuvieron cargando en el servidor durante varios meses, vía FTP seguro, el contenido completo de los más de 2300 CD-ROM y DVD-ROM.

A partir de la inauguración del acceso en línea vía web –disponible en pruebas desde finales de 2004 para todos los miembros de la Comunidad Universitaria de la ULPGC a través de la Intranet de la Biblioteca Digital–, constatamos, según era nuestro propósito, un prodigioso aumento del número de consultas. A dicho incremento colaboró el relieve dado a la noticia por la propia prensa y la difusión realizada por el Gabinete de Prensa de la ULPGC y, por otro lado, el hecho de haber entrado en funcionamiento el Servidor de Autenticación PAPI que permitía que un usuario registrado pudiera acceder a su consulta en cualquier momento y desde cualquier lugar. También las peticiones llegadas desde fuera de la Universidad por parte de particulares y entidades para acceder a dicho recurso y que –tanto por razones de propiedad intelectual como de capacidad de respuesta del servidor– resultaba imposible atender, atestiguan la gran demanda que un servicio de este género está teniendo, no sólo dentro de la propia Universidad, sino también en el conjunto de la sociedad canaria. En la actualidad,

este corpus documental reúne un conjunto de más de tres millones de páginas, convirtiéndose así en uno de los mayores a nivel nacional de entre los que se encuentran accesibles en línea vía web.

En cualquier caso, esto sólo supone la primera fase del plan, ya que nuestro propósito –apoyado ya por la realización de un proyecto fin de carrera en Informática– es aumentar su accesibilidad incorporando la capacidad de búsqueda en texto plano (tanto sobre TIFF nativo como sobre los propios PDF) que permita al usuario encontrar cualquier cadena de caracteres –circunscrita o no a un periodo cronológico determinado– y mostrar acto seguido la página dónde se refleja dicho contenido.

OTROS TÍTULOS DE PUBLICACIONES SERIADAS

No quiero finalizar este artículo sin reseñar el especial interés de la Biblioteca Digital de la ULPGC en el rescate y divulgación de la hemerografía canaria, por cuanto, además de la prensa, estamos acometiendo a un tiempo, tanto con medios externos como propios, la digitalización de significativas publicaciones seriadas vigentes de especial interés para la investigación sobre Canarias y América, como el *Anuario de Estudios Atlánticos* (desde 1955) o los *Coloquios de Historia Canario Americana* (desde 1976), y que suman hasta la fecha más de 40.000 páginas. También publicaciones ya desaparecidas, como el periódico *La Opinión*, publicado en Las Palmas de Gran Canaria entre 1872 y 1873, y revistas literarias o de análisis social, como *Sansofé* (1969-1972), *Aguayro* (1970-1996) o *Millares* (1964-1967), por citar algunos títulos. ■

AUTOR: Macías Alemán, Víctor M.

ILUSTRACIONES: Universidad de Las Palmas de Gran Canaria.

FOTOGRAFÍAS: Revista *Mi Biblioteca*.

TÍTULO: *Digitalización y acceso en línea a la prensa canaria*.

RESUMEN: Con gran cantidad de contratiempos y gracias a la colaboración y cooperación de varias instituciones, la Biblioteca de la Universidad de Las Palmas de Gran Canaria ha podido digitalizar la prensa canaria. Se describe aquí un largo camino recorrido que comenzó en 1996 y que hoy tiene su recompensa: facilitar el acceso de forma sencilla y rápida a todo ese material en formato de publicación periódica.

MATERIAS: Bibliotecas Universitarias / Nuevas Tecnologías / Internet / Prensa / Acceso a la Información / Bibliotecas Digitales / Digitalización.