

HEMEROTECA DIGITAL

El reto de la digitalización de prensa: Hemeroteca Digital de la BNE

LOLA RODRÍGUEZ FUENTES

*Jefe del Servicio de Gestión de Colecciones de Publicaciones Seriadadas
Biblioteca Nacional de España*

RESUMEN: La digitalización de prensa histórica en las grandes colecciones patrimoniales plantea algunas dificultades propias a las que la Biblioteca Nacional de España ha tenido que enfrentarse, fundamentalmente podríamos centrarlas en la cantidad de títulos que las integran, la preocupación por el estado de conservación y la necesidad de dar acceso público a un fondo indispensable para los estudiosos de la historia y la cultura española. Las políticas de preservación que en el siglo pasado se basaban en la microfilmación se han rendido a los beneficios que la digitalización aporta, tanto en materia de preservación como de acceso a la información. En esta línea se ha diseñado la Hemeroteca digital que con 5 millones de páginas y 1.065 títulos accesibles en Internet, recibe un elevado número de consultas, cumpliendo así el objetivo para el que se creó.

PALABRAS CLAVE: OCR, Digitalización de prensa, Servicios en línea, Hemerotecas

ABSTRACT: Digitization of historic press in huge patrimonial collections raises several issues. The National library of Spain had to face the great number of titles which integrate the collection, the concern for the conservation conditions and the need to give public access to this essential collection for those who are interested in the Spanish history and culture. In the last century, the preservation policies were based on microfilm, now they have surrendered to the benefits shown by digitization, the Hemeroteca Digital was designed keeping this in mind and it has now up to 5 million pages and 1.065 titles which are

available on Internet. This website receives a high number of visits, carrying out the goal it was created for.

KEY WORDS: OCR, Digitization newspapers, On-line services, Newspapers libraries

INTRODUCCIÓN

La Hemeroteca Digital de la BNE nace en el año 2007 para dar respuesta a las numerosas demandas de consulta de prensa histórica y marcándose como objetivo convertirse en un referente para la investigación de la prensa española.

Como pionera en el uso de las nuevas tecnologías de la información, la BNE ya había puesto en marcha en 1999 un programa de digitalización de prensa actual, que permitía atender las demandas de los investigadores de una forma más rápida, cómoda y eficaz que la realizada en los propios periódicos o en sus reproducciones en microfilm. Al mismo tiempo, al retirar los originales de la consulta y de la continua manipulación que ésta conlleva y almacenarlos adecuadamente en depósitos de conservación, garantizaba la preservación de un material tan frágil.

A lo largo de estos años ha ido incrementándose el número de títulos y páginas digitalizadas hasta alcanzar los 1.397 títulos con más de 50 millones de páginas de prensa y revistas históricas y actuales.

Las herramientas de consulta también han ido evolucionando en paralelo a los avances tecnológicos, desde la consulta en monopuesto mediante discos ópticos al acceso a las imágenes a través de Internet, e incorporando nuevas funciones como la consulta integrada con BDH (Biblioteca Digital Hispánica), funcionalidades 2.0, o la adopción de estándares internacionales que nos permiten formar parte de los grandes portales digitales como EUROPEANA.

LA PRENSA COMO FUENTE DE INVESTIGACIÓN HISTÓRICA

Históricamente la prensa ha sido un material relegado en las bibliotecas españolas. Aunque los profesionales reconocen el interés que tiene como información de actualidad, la mayoría de los centros nunca han tenido los recursos necesarios para conservar colecciones históricas con unos requisitos de conservación muy costosos y con una necesidad de espacio tan significativa, por tanto, se deshacían de los periódicos al año de su publicación. Se pueden encontrar colecciones importantes en algunas hemerotecas y algunos archivos, aunque pocos centros tienen estos fondos catalogados (en el Catálogo Colectivo de Publicaciones Periódicas¹ solamente aparece recogidos

¹ *Catálogo Colectivo de Publicaciones Periódicas* [en línea]. [Fecha de acceso 1 octubre 2012]. Disponible en: http://catalogo.bne.es/uhtbin/cgisirsi/x/0/0/57/49?user_id=CCPPWEB

los fondos de prensa de la BNE), esto dificulta enormemente el conocimiento de las colecciones de prensa existentes en nuestro país y, por tanto, el acceso a ellas.

Por el contrario, desde hace algún tiempo, los investigadores valoran muy positivamente la prensa como fuente indispensable para la investigación de la historia reciente, lo que se traduce en un aumento considerable de las consultas de prensa en todos sus ámbitos (nacional, regional e incluso local). Hasta la digitalización esto acarrearba viajes, innumerables visitas a las instituciones con fondos de prensa, solicitudes de información telefónicas, por carta, consultas de catálogos, repertorios, etc. Y mucha frustración por la complicación para localizar las colecciones. Además, la carencia de herramientas de búsqueda de contenidos en prensa (índices, sumarios, bases de datos, etc.) hacía necesaria la consulta de gran cantidad de fondos para localizar alguna información, e incluso, para terminar no localizándola.

Con la digitalización se ha realizado una reproducción totalmente fiel de la publicación, no del aspecto con el que salió de la imprenta, sino del que presentan los ejemplares custodiados en la BNE, con sus faltas, mutilaciones, sellos, anotaciones, etc.; es decir, se ha plasmado de forma fidedigna la vida de las publicaciones, para que el investigador disponga de todos los elementos de estudio sin tener que consultar el original. Además, poder aplicar técnicas de OCR a las imágenes digitales ha permitido que la consulta se agilice de forma asombrosa, hasta el punto de que no siendo en absoluto una herramienta precisa, como puede ser un índice o una base de datos de vaciado de artículos, es la forma más empleada en la consulta de prensa y ha incrementado considerablemente el uso de las publicaciones seriadas como valiosas fuentes de información en numerosos aspectos de la investigación histórica, como son:

- La crónica política.
- Acontecimientos históricos.
- Datos biográficos de personajes.
- Investigación socioeconómica.
- Evolución científica y técnica.
- Memoria de actividades culturales, deportivas, etc.
- Medio de transmisión de textos literarios (folletines, poesía, novelas por entregas).
- Estudio de la historia de la prensa.

Otro aspecto importante asociado a la facilidad de acceso a los fondos de prensa es el incremento de la actividad investigadora, que ya no se limita a entornos docentes o eruditos, sino que llega a documentar películas, literatura de ficción, publicaciones divulgativas y se ha democratizado hasta llegar a cualquier persona curiosa o interesada en cualquier tema.

LA COLECCIÓN DE LA BNE

La colección de publicaciones seriadas de la BNE está formada por los fondos que albergaba la antigua Sección de Revistas, cuya procedencia eran los depósitos recibidos por la institución en virtud de las distintas leyes de Propiedad Intelectual y Depósito Legal que han existido en España, a los que se suman la compra, el canje, los donativos y las incautaciones de bibliotecas nobiliarias y de órdenes religiosas. A esta colección se une la de la antigua Hemeroteca Nacional, creada en 1943 en apoyo de la recién creada Escuela Oficial de Periodismo, cuyos fondos procedían, en buena parte, de la «censura previa», es decir, de los ejemplares que los periódicos tenían que enviar a la censura gubernativa derivada de las leyes de Prensa de 1938² y de 1966³. La Hemeroteca también ingresó, por compra, interesantes colecciones históricas, como la de periódicos españoles de la Guerra de la Independencia, propiedad de D. Emilio de Urarte y llamada González Echegaray por el director de la institución que la estudio y que hizo su catálogo⁴, indispensable para conocer una gran cantidad de folletos, hojas volantes, pasquines y toda clase de publicaciones periódicas y no periódicas que vieron la luz en esa convulsa época. Además de la compra, el donativo fue un método de ingreso muy significativo; por este modo ingresaron todas las cabeceras de prensa a partir de 1984⁵ lo que permitió crear una colección de prensa actual muy completa. A partir de 1985⁶ se fusionan estas dos instituciones y la recepción del D.L. de la prensa pasa a gestionarse en la Hemeroteca, que se ocupa del mantenimiento de la colección hasta que en el año 1996 desaparece como institución al integrarse físicamente en la Biblioteca. La gran colección que albergaba el edificio de la calle Magdalena se dividió entre los dos edificios actuales de la BNE: Recoletos y Alcalá de Henares.

Merced a todos estos ingresos, en la BNE podemos encontrar la colección más importante de prensa impresa española, desde las primeras Gazetas⁷ a las cabeceras actuales; todas las publicaciones oficiales españolas (boletines oficiales estatales, autonómicos, provinciales, municipales, etc.); una colección de prensa extranjera con los títulos más representativos de Europa y América, destacando la prensa iberoamericana del siglo xx, y de las antiguas colonias, de Cuba, Filipinas y Puerto Rico. En el conjunto total de las revistas, además

² Ley de Prensa. *B.O.E.* 23 abril 1938, n. 549, pp. 6.915-6.917.

³ Ley 14/1966, de 18 de marzo, de Prensa e Imprenta. *B.O.E.* 19 marzo 1966, n. 67, pp. 3.310-3.315.

⁴ GONZÁLEZ ECHEGARAY, Carlos. *Periódicos y revistas de la Guerra de la Independencia y reinado de Fernando VII (1808-1833) que existen en la Hemeroteca Nacional.*

⁵ R.D. 2089/1984, de 14 de noviembre, para Ayudas a empresas periodísticas y agencias informativas.

⁶ R.D. 565/1985, de 24 de abril, por el que se establece la estructura orgánica básica del Ministerio de Cultura y de sus Organismos Autónomos, y R.D. 848/1986, de 25 de abril, por el que se determina las funciones y la estructura orgánica básica de la B.N., desarrollado por la Orden del MCU de 10 de junio del mismo año. *B.O.E.* 21 junio 1986, n. 148.

⁷ *Relación o Gazeta de algunos casos particulares, así políticos como militares sucedidos en la mayor parte del mundo...* Madrid: Julián de Paredes, 1661-1663. ISSN 0212-1212.

de la práctica totalidad de las revistas españolas de todas las épocas, cabe destacar la colección de las principales revistas científicas europeas de los siglos XVII al XIX y la colección de revistas extranjeras actuales que ingresan por compra, canje o donativo.

Actualmente la colección de publicaciones seriadas de la BNE supera los 160.000 títulos, con un crecimiento en torno a los 3.000 títulos anuales, y en esta colección están representados todos los formatos que hasta hoy han empleado las publicaciones seriadas en su publicación, desde el papel hasta las memorias USB.

POLÍTICAS DE PRESERVACIÓN Y DIFUSIÓN

La encuadernación ha sido la primera medida de preservación que se ha aplicado a los fondos de prensa y revistas, gracias a ella las colecciones del siglo XIX, de grandes formatos y pésima calidad de papel, han pervivido hasta ahora, aunque no en muy buen estado. La degradación del papel, la mala colocación en estanterías poco adecuadas (pensadas para formatos más pequeños) y la constante manipulación (traslados, consultas y reproducciones) que el uso comporta, han operado en contra de su pervivencia. A pesar de que algunas encuadernaciones, pésimamente realizadas, han causado estragos en muchos volúmenes, no hay que dudar de su eficacia en la conservación de la prensa que ha llegado a nuestros días.

La colección de prensa custodiada por la BNE presentaba un estado que reclamaba medidas urgentes que garantizaran su conservación a medio y largo plazo. Ante la imposibilidad de acometer un plan de conservación integral (de soportes y de contenidos) en una colección de este tamaño y teniendo en cuenta el compromiso de la BNE en la conservación y pervivencia de las colecciones patrimoniales, las políticas de preservación, respecto a la prensa y a las publicaciones seriadas, se han centrado en abordar la preservación de los contenidos mediante su reproducción en otros soportes.

En 1991 comenzó el Programa de Microfilmación de Prensa y hasta el año 2009 se han microfilmado la práctica totalidad de las cabeceras españolas hasta 1998 y las colecciones de revistas en peor estado y más solicitadas por los usuarios. Las cifras totales correspondientes a este programa son las siguientes: unos 3.000 títulos que suman alrededor de 79.000 volúmenes y que han dado lugar a 33 millones de fotogramas.

En 1997 se realizó la primera tentativa en materia de digitalización de prensa, digitalizando desde el microfilm 187 títulos del siglo XIX. Los resultados no fueron satisfactorios: las imágenes de originales ácidos tenían un fondo muy oscuro que apenas contrastaba con el texto, que en muchos casos era ilegible, las de grandes formatos con una tipografía muy pequeña tenían muy mala calidad y las ilustraciones (grabados) no tenían nitidez.

En 1998 se había dado una solución a la consulta de los fondos históricos mediante el programa de microfilmación, que se encontraba ya en fase muy

avanzada, pero con motivo de las obras de remodelación del edificio de Recoletos hubo que cerrar la «Sala de Consulta de Prensa Diaria» y las publicaciones se trasladaron a los depósitos de Alcalá de Henares, por tanto, había que buscar una solución para la consulta de la prensa actual que no supusiese su continuo traslado de un edificio a otro. Comenzaban entonces a aparecer, en España, los primeros proyectos de digitalización de fondos de prensa por los propios editores, como el de ABC, que digitalizó su colección histórica y la editó en CD-ROM. También en la Biblioteca se pensó en la digitalización como solución a la consulta de la prensa actual.

En 1999 se inició el Programa de Digitalización de Prensa como continuación al de microfilmación. De este modo la consulta se realizaba en soportes de gran capacidad (micros y cedés) que por su reducido tamaño podían almacenarse en la misma sala de consulta, lo que facilitaba enormemente el servicio de peticiones y la consulta de los investigadores.

PROYECTO DE DIGITALIZACIÓN DE PRENSA ACTUAL

Un importante factor a tener en cuenta para defender la digitalización de periódicos contemporáneos, a pesar de que la práctica totalidad de los mismos tienen una versión en Internet, es la gran diferencia existente entre las ediciones en línea y las ediciones impresas. Con Internet, el periodismo electrónico se caracteriza por sus servicios de valor añadido, fundamentalmente por su contenido multimedia, interactividad, actualización constante de contenidos y acceso universal. Por lo tanto, muy diferente a la edición impresa tradicional, no solo la forma es diferente, sino que también difieren los contenidos (diferencia en editoriales, artículos de opinión, anuncios, publicidad, esquelas, etc.), esto hace necesario tener que contar con las dos versiones, si queremos conservar toda la información. En la edición impresa, el lugar y la maquetación de cada noticia está ya dándonos importante información sobre ella y su contexto, información que se pierde en la versión en línea. Es también frecuente que está última excluya artículos firmados (por cuestión de derechos de autor) y que cambien anuncios y publicidad, en función de los distintos recursos que ofrecen dos productos tan dispares. Aunque algunos medios ofrecen su versión impresa en su página digital, no todos los medios lo hacen y no de todas sus ediciones (se pierde así la información de las ediciones regionales o locales), por lo que disminuye su interés como fuente de investigación, frente a la tradicional edición impresa.

El principal objetivo de este proyecto era facilitar la consulta de los fondos de prensa, proporcionando un acceso rápido al mayor número de cabeceras posible. En 1999 se comenzó digitalizando 100 diarios nacionales de información general de los 132 recogidos en la Guía de Medios⁸, lo que nos da una cobertura del 76 %. Además, se incluían 5 diarios deportivos de los 7

⁸ G. M.: *Guía de los medios de comunicación de España*, diciembre 1998, n. 136.

que citaba la Guía de Medios y 3 financieros de los 4 que recogía el citado directorio. A este alto grado de exhaustividad hay que añadirle una representatividad regional absoluta, ya que incluía títulos de todas las comunidades autónomas, incluyendo Ceuta y Melilla. El último indicador, pero no el menos importante, que se tuvo en cuenta para realizar la selección de cabeceras que formaron parte del proyecto, fue el uso, obtenido a partir de las estadísticas de consulta de prensa.

Desde el comienzo se decidió evitar, en la medida de lo posible, duplicar trabajos que ya hubiesen realizado en otros organismos con la intención de ahorrar costes y esfuerzos a la comunidad bibliotecaria. Por ello, además de estar al día de los proyectos de digitalización de prensa existentes para tratar de llegar a acuerdos de cooperación, se diseñó una herramienta de gestión de imágenes capaz de gestionar distintos tipos de ficheros (tiff, jpeg, pdf) y que permitiese integrar en el mismo sistema de consulta imágenes de procedencias variadas.

El número inicial de cabeceras ha ido incrementándose hasta las 332 actuales con las siguientes incorporaciones:

- Las imágenes de las primeras experiencias en digitalización incluyen títulos históricos, algunos con derechos y otros de dominio público: son las imágenes más antiguas y de peor calidad.
- Imágenes del proyecto de digitalización de prensa de la Biblioteca que comenzó en 1999 con 108 títulos y que se ha ido reduciendo a medida que algunos de ellos se han incluido en otros proyectos ajenos a la BNE.
- Colecciones digitalizadas ingresadas en la Biblioteca en cumplimiento de la ley de D.L., fundamentalmente el diario ABC que digitalizó su archivo histórico y toda la prensa balear.
- Imágenes adquiridas a los propios editores de diarios, que han decidido realizar una copia digital de sus colecciones históricas, con fines de archivo y consulta.
- Imágenes procedentes de convenios con otras instituciones.

Además, a todas las imágenes digitalizadas a partir del año 2004 se les aplicó un programa de tratamiento de textos con lectura óptica de caracteres (OCR), que permite realizar búsquedas a texto libre. Esta herramienta supuso un servicio de valor añadido de gran importancia que potenció el uso de la prensa digitalizada de forma significativa.

Todas las imágenes se han almacenado en discos ópticos. Se conservan dos copias, una de ellas con la función de master de seguridad desde el que se pueda migrar la información a cualquier otro soporte que garantice su pervivencia.

Junto a la herramienta de gestión, se diseñó una aplicación adecuada para la consulta de estas imágenes, con tres utilidades que cubren las necesidades específicas de este material:

- Herramientas de búsqueda: permiten la recuperación por los tres elementos básicos en la consulta de prensa: título, fecha y lugar de difusión. Posteriormente se incluyó la opción de búsqueda a texto completo.
- Herramientas de visualización: incluyen diferentes posibilidades de navegación a partir de los resultados de la búsqueda y también de visualización (miniaturas, doble página, zoom, lupa, giros, etc.).
- Obtención de documentos: permite imprimir ejemplares, páginas o selección de artículo o fotografía. También se pueden guardar las imágenes seleccionadas en un soporte externo o exportarlas a un fichero para crear un dossier de prensa.

En el año 2010 hubo que cortar este proyecto por falta de recursos. Se habían digitalizado ya 11 años (1999-2009) de la mayoría de títulos de la prensa española. Obligados a priorizar, se consideró más importante destinar los recursos existentes a la digitalización de la prensa histórica, que al estar libre de derechos puede difundirse en Internet y así, posibilitar el acceso a nuestra colección desde cualquier parte del mundo.

Actualmente, se ha comenzado la migración de los más de 45 millones de páginas a servidores. Este proceso, que supone una ardua tarea, tiene dos beneficios indiscutibles: sirve de «refresco» a las imágenes pasándolas a un soporte más seguro y además optimiza los recursos permitiendo una consulta, mediante red local, más ágil, con más posibilidades y con varios usuarios concurrentes. Ya hay 6.382.137 páginas (13 títulos) disponibles en la red interna⁹ (Intranet) de la BNE y esperamos poder continuar a este ritmo hasta completar la migración.

PROYECTO DE DIGITALIZACIÓN DE PRENSA HISTÓRICA: HEMEROTECA DIGITAL

En 2006, tras el éxito obtenido en la consulta de prensa digitalizada y atendiendo a numerosas sugerencias de los investigadores, se comenzó la digitalización de prensa histórica. En marzo de 2007 se disponía ya de 143 títulos y se lanza la Hemeroteca Digital¹⁰ que como se dijo en su presentación: «se convierte en la primera aportación a la Biblioteca Digital Hispánica, en la que se integra» (14 de marzo de 2007)¹¹.

Nace con el objetivo de incrementar el acceso público a la colección de prensa histórica de la Biblioteca Nacional, convertirse en el principal recurso para la investigación y la consulta de la prensa española y abrir cauces de cooperación con otros centros para localizar y completar colecciones y para compartir los recursos disponibles.

⁹ <http://prensaactual.bne.es/>: accesible solamente desde el edificio de la BNE.

¹⁰ *Hemeroteca Digital*. Disponible en: <http://hemerotecadigital.bne.es/index.vm>

¹¹ *Presentación de la Hemeroteca digital* [en línea]. [Fecha de acceso 2 octubre 2012]. Disponible en: <http://www.bne.es/es/Actividades/ActosCulturales/CicloCitaBN/CitaBN2007/HemerotecaDigital.html?pagina=1>

Selección

La magnitud de los fondos de prensa histórica con que cuenta la Biblioteca, imposibilita su reproducción total a corto y medio plazo, por lo que se impone un proceso de selección de títulos. Este proceso se complica según aumenta el número de criterios a tener en cuenta, pero al mismo tiempo, solamente un proceso selectivo, minucioso y exhaustivo, en el que se tengan en cuenta todos los factores que operan en un proyecto de tal envergadura, puede asegurar el cumplimiento de los objetivos marcados. En nuestro caso los criterios a evaluar fueron:

- La representatividad de las publicaciones, tanto de su época, como de la colección existente en la Biblioteca, recogiendo todas las materias y tipos de publicaciones, de forma que quedase reflejada la riqueza temática de la edición hemerográfica hispana.
- Las características de las colecciones custodiadas por la Biblioteca, primando los títulos con menos lagunas y los ejemplares únicos. Se han analizado todas las colecciones de cada título, en el caso de existir más de una, para elegir los volúmenes en mejor estado de conservación y viendo la posibilidad de que se puedan completar unos con otros.
- Las preferencias de investigadores y usuarios, para lo que se han tenido muy en cuenta las estadísticas de consulta en sala y las sugerencias y demandas recibidas durante estos años, tanto de investigadores, como de otras unidades de la Biblioteca.
- Respaldo a otros proyectos culturales llevados a cabo por la propia Biblioteca, como pueden ser exposiciones o eventos importantes, que han supuesto la digitalización de algún título; por ejemplo, se han digitalizado las Memorias de la BNE¹² con motivo del tricentenario de la institución. También se ha digitalizado algún título en colaboración con otras instituciones para completar la digitalización de sus publicaciones históricas.

Como resultado de esta selección, la Hemeroteca Digital cuenta actualmente con 1.065 títulos y unos 5 millones de páginas. De esta primera selección aún unos 300 títulos están en proceso y faltan por cargar en la aplicación, esperamos poder hacerlo lo antes posible. Aunque nos gustaría continuar digitalizando el rico patrimonio que custodiamos, los recursos necesarios para llevarlo a cabo son costosos, por lo que no está asegurada la continuación de este proyecto.

Metodología

- Revisión: una vez realizada la selección, se consultaron los distintos catálogos existentes en la BNE para localizar todas las posibles colecciones

¹² BIBLIOTECA NACIONAL (ESPAÑA). *Memoria remitida al Ministro de Fomento... por el Director de la Biblioteca Nacional*. Madrid: [Biblioteca Nacional], 1865-. ISSN 2174-9450.

de cada título y sus distintas ubicaciones y se revisaron todos los ejemplares para elegir los más idóneos. De algunos títulos importantes se pueden tener hasta cuatro colecciones incompletas de diferentes procedencias. Esta ha sido la tarea más laboriosa del proyecto y ha consumido mucho tiempo y gran parte de los recursos humanos del Departamento de Seriadas.

- Gestión y control del proyecto: se realizó una base de datos que recoge toda la información de las distintas etapas del proceso, desde los motivos que se hayan podido dar en la selección de cada título hasta las posibles incidencias que se puedan plantear en su consulta. En el caso de no haber podido completar ejemplares o títulos se hace constar en las imágenes, con los mensajes de «FALTA PÁGINA» o «FALTA EJEMPLAR», y se ha guardado toda esa información para intentar ir completando colecciones mediante adquisiciones, donativos o cooperación con otros centros. En algunos casos se ha tenido que proceder a pequeñas intervenciones de restauración para consolidar el soporte antes de la digitalización, pero en otras ocasiones la pérdida de texto ya era irreparable.
- Procesos de digitalización: mediante concurso público bianual se han adjudicado a una empresa externa las tareas de digitalización, al principio en sus propias instalaciones y posteriormente se habilitó una zona en la Sede de Alcalá, próxima a los depósitos de prensa para que los traslados fueran los mínimos. Las imágenes resultantes del proceso empezaron volcándose en cedes, igual que la prensa actual, posteriormente se volcaban en discos duros con más capacidad y al trasladarse la digitalización a la Biblioteca se cargaban directamente en dos servidores (uno para los masters y otro para los derivados), desde los que se indexaban en la aplicación de consulta y se editaban en Internet.
- Características técnicas: la digitalización se ha realizado con una resolución de 300 puntos, al principio en color, blanco y negro o escala de grises, dependiendo de las características del original, pero dada la gran demanda de imágenes de calidad y en color para todo tipo de usos (exposiciones, catálogos, publicaciones multimedia, etc.) se optó por realizar todas las imágenes a color.
- Los formatos empleados han sido tiff para los ficheros master y de ellos se extraen los derivados en fichero pdf para la consulta. Los ficheros de consulta se han comprimido a 150 puntos para facilitar la descarga de imágenes en Internet.
- Todas las imágenes que forman parte de Hemeroteca Digital cuentan con lectura OCR.
- Siguiendo fielmente las publicaciones originales, las imágenes están en ficheros pdf monopágina (para agilizar las descargas) pero agrupadas por ejemplares, de forma que se pueda descargar o imprimir el ejemplar completo.

Bienvenido | Benvingut | Ongi etorri | Benvido | Welcome | Bienvenue

Biblioteca Digital Hispánica

Consulta | Resultados | Detalles | Listado de títulos | Novedades | Próximamente | Otras hemerotecas digitales | Contacto | Ayuda

Documentos que contengan las palabras en

contengan las palabras en

contengan las palabras en

Título

- La Abeja del Turia
- Abeja española
- La Abeja manresana
- La Abeja (Barcelona)
- AC. Documentos de actividad contemporánea
- Academia heráldica
- Acción española (Madrid)
- La Acción social (Zaragoza)
- La Acción (Madrid, 1916)
- Actualidades (Madrid, 1893)

Ámbito geográfico

- Álava
- Albacete
- Alicante
- Argentina
- Asturias
- Badajoz
- Baleares
- Barcelona
- Brasil
- Burgos

Año

- 1683
- 1685
- 1689
- 1690
- 1691
- 1692
- 1693
- 1694
- 1696
- 1744

Fecha entre el 01 / 01 / 1683 >> y el 31 / 12 / 1993

Buscar Limpiar Recuperar Guardar Cargar

Documentos ordenados:

Mostrar: Títulos Ejemplares Páginas

CrasZandora v4.17

Pantalla de consulta de Hemeroteca digital.

Herramienta de consulta

Para diseñar la aplicación de consulta de publicaciones seriadas digitalizadas, se ha tenido muy en cuenta la experiencia obtenida en los muchos años de atención a los usuarios en las salas de consulta de prensa. Desde las primeras versiones de esta herramienta se ha querido responder a las búsquedas más comunes, que son:

- Consulta de una publicación determinada: para lo que necesitan, o conocer el título exacto o herramientas que les permitan localizarlo por el comienzo o por alguna palabra que contenga.
- Consulta de los títulos publicados en un lugar o área geográfica determinada: cuando se quiere localizar noticias o referencias muy locales, es interesante poder delimitar nuestra consulta a la prensa local o regional.
- Consulta de los títulos publicados en una fecha concreta o en un rango de fechas: esta consulta nos puede dar una amplia visión de como se trató en las distintas publicaciones existentes un hecho deter-

minado (por ejemplo, el 23F), o qué tipo de publicaciones había en un periodo más o menos amplio (por ejemplo, en la Primera República Española).

Hasta la digitalización, la búsqueda de un tema, de un personaje o de un evento no era posible con las herramientas de búsqueda tradicionales (catálogos y listados).

En las grandes bibliotecas, en las que se están llevando a cabo importantes proyectos de digitalización (véase TELplus project)¹³, se está analizando la conveniencia de aplicar técnicas de OCR a periódicos antiguos y la exactitud de los resultados obtenidos. No hay que olvidar que el OCR siempre será una búsqueda textual no controlada, por lo que aunque su transcripción fuese exacta seguiría dándonos resultados erróneos (y mucho «ruido» en nombres comunes y fechas).

En nuestro caso concreto se comenzó a aplicar OCR a algunos de los títulos de prensa moderna y a pesar de que los resultados no eran totalmente satisfactorios, sí pudimos observar el impacto que tuvo en la investigación y consulta de la prensa. Los títulos más utilizados fueron los que contaban con esta herramienta, e incluso se usaba la búsqueda en OCR para averiguar las fechas de los eventos que se trataban de localizar y poder buscar después esa fecha en otros títulos más interesantes para el investigador. Además, cuando los usuarios comenzaron a adiestrarse en el uso de la herramienta fueron capaces de diseñar perfiles de búsqueda que respondían suficientemente a sus consultas minimizando los fallos del reconocimiento de caracteres. A continuación comenzaron a demandar que se digitalizase y se aplicase OCR a toda la prensa histórica, que estaba microfilmada para facilitar su consulta.

Con este escenario, al acometer el proyecto de digitalización de prensa histórica en la BNE se optó por dotar a Hemeroteca Digital con la única herramienta de búsqueda de contenido que actualmente podemos ofrecer a nuestros investigadores y, a pesar de su inexactitud, ellos nos lo han agradecido con una elevada cantidad de consultas y felicitaciones por nuestro trabajo.

A pesar de no haber realizado un estudio pormenorizado de los resultados obtenidos en la búsqueda mediante OCR, lo que sí hemos hecho ha sido un análisis automático sobre una muestra de 268.696 páginas, pertenecientes a 23 títulos diferentes. Los resultados del análisis afirman haber reconocido 4.000.867.107 caracteres, consideran caracteres correctos o con un muy bajo umbral de duda a 3.459.095.760 caracteres, por tanto el porcentaje medio correcto de todos ellos es 86,46 %. Los porcentajes por título están entre un 98% el de mejor resultado de lectura y un 82% el de peor. Estos datos indican una relativa buena lectura para tratarse de prensa histórica.

También es posible combinar todos los criterios de búsqueda, lo que permite afinar los resultados de forma significativa. Por ejemplo, si en el

¹³ Proyecto TELplus [en línea]. [Fecha de acceso 1 octubre 2012]. Disponible en: <http://www.theeuropeanlibrary.org/confluence/display/wiki/TELplus+project>

texto buscamos «Jacinto Benavente» el número de resultados es de: 8.107; si queremos localizar noticias de estrenos de sus obras, centramos la búsqueda entre los años 1894 y 1922 (año en que fue Premio Nobel) y los resultados son: 4.727; si además delimitamos la búsqueda a un título concreto, *El arte del teatro*¹⁴, el número de resultados es 15 y todos sobre sus estrenos teatrales.

Entre las reacciones de los usuarios recogemos algunos ejemplos que nos han llegado a la dirección de correo: gracias a la búsqueda textual un investigador, que estaba realizando un estudio sobre la historia de la edición en España, localizó la primera prensa mecánica que llegó a España y con la que se imprimía en 1837 el *Semanario pintoresco español*¹⁵. Una imagen de Baldomera Larra (encontrada en Hemeroteca Digital) ilustra un artículo sobre ella en *El País*¹⁶. Numerosas exposiciones y/o publicaciones sobre personajes regionales ilustres se han enriquecido con retratos, noticias y datos curiosos, como del diputado por la provincia de Huesca D. Agustín Viñuales (localizado en la revista *Por esos mundos*¹⁷). También algunos municipios han recabado información interesante sobre su historia.

Servicios de valor añadido

Para lograr el objetivo de ser el recurso de referencia para la investigación en la prensa española, Hemeroteca Digital cuenta con una serie de herramientas de gran ayuda para el investigador, como son:

– *Catálogo Bibliográfico de la BNE*¹⁸:

Los títulos digitalizados tienen un enlace al catálogo de la Biblioteca, donde se puede consultar su registro bibliográfico normalizado y todas las colecciones, tanto originales como reproducidas, existentes en la Biblioteca de ese título. También, a la inversa, se puede llegar a las imágenes digitalizadas desde el registro bibliográfico del catálogo.

– *Descripción de publicaciones*:

Las publicaciones digitalizadas tienen una pequeña presentación, donde se explica la importancia de la publicación, su historia, los autores que publicaron en ella y el papel que representó en su época. Esto puede ser muy útil para localizar obras de autores conocidos, no firmadas, y para contrastar el posible sesgo ideológico al tratar cualquier tema. Además, también se puede realizar búsquedas por texto libre en este campo.

¹⁴ *El arte del teatro: revista quincenal ilustrada*, Madrid, 1906-1908.

¹⁵ *Semanario pintoresco español*, 1 junio 1838, n. 118, p. 5.

¹⁶ TORRES, Rosana. El arte de la estafa. *El País*, 21 marzo 2009.

¹⁷ *Por esos mundos*, 1 diciembre 1907, p. 37.

¹⁸ *Catálogo BNE*. Disponible en: <http://catalogo.bne.es/uhtbin/cgiirsi/vCx3O8Lt5N/BNMADRID/46540073/2/11>

– *Otras hemerotecas digitales*¹⁹:

Hemeroteca Digital cuenta con un directorio donde se recogen los enlaces a las principales hemerotecas digitales españolas con una breve presentación de las mismas y de sus recursos

– *Biblioteca Digital Hispánica*²⁰:

En 2011 Hemeroteca Digital se integró en BDH, de esta manera es posible realizar una consulta conjunta de los registros de Biblioteca Digital Hispánica y de Hemeroteca Digital desde la interfaz de BDH.

– *Servicios de la página Web de la BNE*²¹:

La Web de la Biblioteca presta una serie de servicios en línea de gran utilidad, en apoyo de la investigación, como son: un directorio con los enlaces a todos los títulos de prensa española actuales en línea, el gestor de recursos electrónicos SFX que nos permite localizar los periódicos y revistas electrónicas actuales disponibles en las numerosas bases de datos accesibles desde la propia Web, servicio de reproducciones, funcionalidades 2.0, información bibliográfica y sobre horarios y funcionamiento de las salas de consulta.

– *Proyectos internacionales*:

En 2012 se ha puesto en producción una nueva versión de la aplicación de Hemeroteca Digital, que cumple con los estándares internacionales OAI (Open Archives Initiative)²² y EUROPEANA. Ello permitirá formar parte de dicho proyecto y de otros muchos que se están llevando a cabo en el ámbito de The European Library²³, para compartir sus recursos con las principales bibliotecas digitales europeas y realizar búsquedas sobre sus colecciones digitales de forma simultánea.

BIBLIOGRAFÍA

- FERNÁNDEZ POUSA, Ramón. *Catálogo de los diarios y revistas existentes en la Hemeroteca Nacional*. Madrid: Hemeroteca Nacional, 1949.
- GARCÍA-PUENTE, Elena y RODRÍGUEZ, Lola. The Hemeroteca Digital of the National Library of Spain. En MESTROVIC DEYRUP, Marta (editor). *Digital scholarship*. New York: Routledge, 2009, pp. 49-71.
- GONZÁLEZ ECHEGARAY, Carlos. *Periódicos y revistas de la Guerra de la Independencia y reinado de Fernando VII (1808-1833) que existen en la Hemeroteca Nacional*. Madrid: Instituto Bibliográfico Hispánico, 1981.

¹⁹ *Otras Hemerotecas Digitales* [en línea]. [Fecha de acceso 2 octubre 2012]. Disponible en : <http://www.bne.es/es/Catalogos/HemerotecaDigital/OtrasHemerotecas/>

²⁰ BDH. Disponible en: <http://bdh.bne.es/bnearch/>

²¹ BNE. Disponible en: <http://www.bne.es/es/Inicio/index.html>

²² *Standards for Web Content Interoperability* [en línea]. [Fecha de acceso 2 octubre 2012]. Disponible en: Open Archives Initiative: <http://www.openarchives.org/>

²³ *The European Library*. Disponible en: <http://www.theeuropeanlibrary.org/tel4/>

- GONZÁLEZ GÓMEZ, Joaquín. *Publicaciones periódicas de la Guerra Civil (1936-39) en zona republicana, existentes en la Hemeroteca Nacional*. Madrid: Dirección General del Libro y Bibliotecas, 1986. (Publicaciones de la Hemeroteca Nacional, n. 21).
- GÓMEZ IMAZ, Manuel. *Los periódicos durante la Guerra de la Independencia (1808-1814)*. Madrid: Tip. de la Revista de Archivos, Bibliotecas y Museos, 1910.
- MARTÍN, Flory, PAZ, Adela y VINATEA, Pilar. *Revistas de arte en la Biblioteca Nacional*. Madrid: Biblioteca Nacional, 1989.
- RODRÍGUEZ FUENTES, Lola. Hemeroteca digital de la BNE: tres claves para su éxito. *MEI: Métodos de información* [en línea]. 2011, vol. 2, n. 2. [Fecha de acceso 2 octubre 2012]. Disponible en el documento *web*: <http://www.metodosdeinformacion.es/mei/index.php/mei/issue/view/49/showToc>
- ZAMORA LUCAS, Florentino y CASADO JORGE, María. *Publicaciones periódicas existentes en la Biblioteca Nacional*. Madrid: Dirección General de Archivos y Bibliotecas, 1952.

