

Gazeta: el fondo histórico del BOE en base de datos

Por María Jaudenes Casaubón

A FINALES DE 1997 comenzaron los trabajos de valoración de la colección histórica de la *Gaceta de Madrid*, de la que existe un ejemplar íntegro microfilmado que cubre el período desde 1711 a nuestros días. Para poder preservar debidamente la versión en papel (completa desde 1832), así como también para ofrecer un mejor acceso a la información y al documento en su formato original, se plantea el proyecto de digitalización e incorporación a las bases de datos del BOE de esta colección de textos antiguos.

Por su complejidad el proyecto se contempla como un plan a varios años con dos objetivos:

— Obtener una colección de imágenes digitalizadas del diario, que se integrará en el archivo electrónico existente, que abarca desde 1968 hasta hoy, permitiendo la recuperación por fecha de publicación y por número de BOE.

— Conseguir una información textual estructurada en campos que facilite el acceso a una base de datos histórica de la gaceta oficial. Los registros así obtenidos quedarán vinculados a las imágenes correspondientes de la gaceta de modo que se pueda leer el texto completo de cada disposición en su formato original.

En 1998 se realizó un estudio previo que ha permitido conocer el volumen de datos a tratar (alrededor de 1.500.000 páginas) y las características formales de la gaceta a lo largo de su historia: contenido, estructura, formatos y estado de conservación del fondo. Así mismo, se han realizado pruebas de digitalización comparadas a partir de microfilm y de papel para poder evaluar la calidad de los resulta-

dos, y se han hecho pruebas de OCR (reconocimiento óptico de caracteres) sobre páginas de diferentes épocas que dieron unos resultados muy poco convincentes: para obtener un grado de calidad mínimo era necesario verificar y retocar cada texto manualmente casi en su totalidad (unos 30 minutos/página).

En primer lugar se llevó a cabo una “prueba piloto” que permitió comprobar el planteamiento del proyecto y subsanar posibles errores. Este primer trabajo incorporó a las bases de datos actuales los diarios publicados entre 1961 y 1967. Son en total 166.270 disposiciones y casi 136.000 páginas en formato *tiff*. En el mes de octubre de 1999 se lanzó la primera fase, cuya ejecución terminará a mediados de este año 2000. Tras ella se incluirán en *Gazeta* las disposiciones e imágenes de los diarios oficiales hasta 1931. En total se trata de casi 400.000 páginas, con casi la misma cantidad de disposiciones publicadas en el *Boletín Oficial del Estado*, la *Gaceta de Madrid*, la *Gaceta de la República* y el *Boletín Oficial de la Junta de Defensa Nacional de España*.

Gazeta es, en primer lugar, una herramienta de acceso a las páginas del *Boletín Oficial del Estado* anterior a 1968. Permite la búsqueda por palabras del título, rango, fecha de publicación, de aprobación, número oficial y departamento emisor. Pero, para que no sea solo un índice de imágenes sino una verdadera publicación, se acompañará de un trabajo documental exhaustivo de revisión, depuración y selección de las disposiciones a analizar, establecimiento de criterios de análisis jurídico y documental y ejecución del análisis propiamente dicho.

María Jaudenes Casaubón.
maria.jaudenes@docu.boe.es

SUMARIO

Artículos

Bibliotecas digitales: situación actual y problemas
Por Pedro Hipola, Benjamín Vargas-Quesada
y José A. Senso4

Mercado

Vista desde la galería
Por Steve Borley14

Los bibliobuses en España15

El Parlamento Europeo escoge Loris16

Recursos informativos en las redes

Gazeta: el fondo histórico del BOE en base de datos
Por María Jaudenes Casaubón3

Corc: hacia la biblioteca digital
Por Cristina García Testal17

Sistemas de información

La holografía en el mundo de la documentación
Por Lola García-Santiago20

Fundamentos para la organización de un banco de fotografías
Por Lluís Codina31

La gestión de la información
en las organizaciones35

Aprender tendiendo puentes mentales36

¿Transferencia o transacción de información?37

Utilidad y usabilidad en la búsqueda
de información38

Diseño de información: buscar la eficiencia en la
comunicación informacional40

Resenas

Políticas locales de información
Por Assumpta Bailac40

Número especial de
Cuadernos de documentación multimedia43

Primer congreso nacional de periodismo digital44

Agenda

.....46

Información para los autores

.....48

Los trabajos de la sección **Artículos** son aprobados según el sistema tradicional "peer review": al menos dos expertos en el tema, del *Consejo Asesor* la revista y/o externos, deben dar el visto bueno antes de su publicación.