

# Servicio de digitalización con OCR y generación de ficheros ebook

2018

## OBJETIVOS Y RETOS INICIALES QUE SE PLANTEARON

El cliente de este caso de éxito necesitaba generar varios ficheros eBook a partir de 74 libros con una media de 200 páginas.

Para la realización del servicio el cliente facilitó 36 libros en su soporte original papel, que necesitaban ser digitalizados, y 38 en formato electrónico, con una suma total de 14.800 páginas.

Este proyecto se realizó utilizando la tecnología OCR (Optical Character Recognition).

El Reconocimiento Óptico de Caracteres permite convertir diferentes tipos de documentos, tales como documentos en papel escaneados, archivos de PDF o imágenes captadas por una cámara digital en los datos editables y con opción de búsqueda.



El servicio de digitalización y generación de ebook prestado por GEDSA consistió en las siguientes etapas:

### **Recogida y recepción del fondo a procesar:**

Los 36 libros en formato original papel deberán estar agrupados y listos para su recogida en el lugar acordado.

Previamente, el cliente se informó a GEDSA de la relación de libros a recoger, identificándolos unívocamente y como mínimo por su título y autor. De este modo fue posible cotejar que los fondos recogidos corresponden con los previstos y detectar cualquier incidencia que pudiera producirse.

Así mismo, se puso a disposición del clientes un repositorio FTP en el que depositar los ficheros PDF correspondientes a los 38 libros restantes.

### **Escaneo:**

Tras la recepción de los fondos en la sede de GEDSA, se procedió a los trabajos de conversión digital.

El primer paso, fue la adecuación de la documentación física para que su procesamiento mediante escáneres de alta producción.

Una vez separadas las hojas, se digitalizaron a una resolución de 400dpi en blanco y negro. De esta forma, se mejoró la calidad del resultado del proceso de OCR.

El proceso de digitalización en cadena de GEDSA contempla una etapa de verificación y corrección de la calidad de la imagen, garantizando que el proceso de conversión no ha supuesto la pérdida de contenido o calidad.



### **Reconocimiento óptico de caracteres:**

La tarea de procesado de OCR se realizó sobre los libros físicos y sobre los libros que se remitieron en PDF que no tenían información textual embebida.

El resultado del OCR permitió obtener el texto que formó las distintas líneas del documento final.

### **Creación de índices interactivos:**

Una de las grandes ventajas que ofrecen los libros electrónicos es la opción de poder crear índices y tablas de contenido con llamadas entre los elementos, de forma que pulsando sobre el enlace nos dirigirá exactamente al punto buscado.

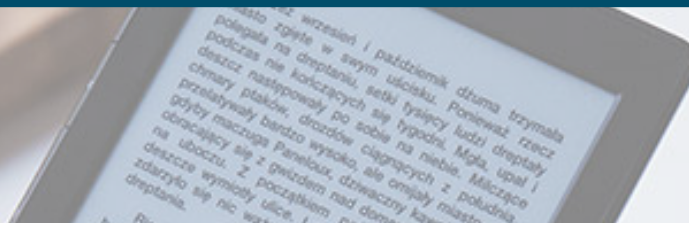
### **Entrega de ficheros al cliente:**

El resultado del proceso de conversión digital se pone a disposición del cliente como enlace de descarga desde el servidor de GEDSA.

Para cada libro se crea un fichero ZIP que contiene las tres versiones de libro digital (MOBI, EPUB, PDF) así como las imágenes digitalizadas en formato TIFF de los libros que hayan sido escaneados. Esta descarga está disponible hasta 30 días después de la notificación de la finalización del servicio y solo podrá ser accedida por el cliente.

Junto con los originales, se hizo entrega de una memoria USB con capacidad suficiente para almacenar todos los ficheros digitales generados. En la memoria se grabó una copia de cada libro con las 3 versiones de libro electrónico y los ficheros de imagen generados.





### **Devolución de los libros en formato papel:**

Los libros guillotizados fueron introducidos en subcarpetas de cartulina y sujetando las hojas con pinzas abatibles para garantizar su agrupación a su devolución.

- Eficiencia y seguridad de los libros en formato físico.

-Rapidez en la entrega y cumplimiento de los plazo previstos.

---

### **RESULTADOS Y MEJORAS EXPERIMENTADAS**

Como resultado de los trabajos de digitalización y generación de ebook se obtuvieron los siguientes resultados, seleccionados específicamente por su importancia para el cliente:

- Una corrección de la calidad de la imagen que ha mantenido el contenido y la calidad tras el proceso de conversión.

