

Aplicación práctica de las NTI: archivo electrónico con disco óptico

María BARCELÓ LLAUGER

I. INTRODUCCION

La implantación de las Nuevas Tecnologías de la Información es ya un requisito indispensable en el desarrollo de las empresas del sector privado. El aumento de la plantilla, el incremento del volumen de negocio y otras consideraciones asociadas implican el crecimiento de los fondos documentales de todo tipo —con los consiguientes problemas básicos de gestión de papel—, hasta un punto en el que es necesario plantearse la estrecha relación de los Sistemas de Información con los objetivos y requerimientos de una empresa/organización.

Tras esta reflexión y después de realizar un análisis del entorno del profesional de la División de Asesoría Legal y Fiscal de Price Waterhouse nació el Proyecto SICRA —Sistema Informático de Consulta y Recuperación de Antecedentes. Con una gran visión de futuro en la fase inicial del proyecto y tras valorar las distintas posibilidades que se ofrecían en el mercado, se optó por un sistema electrónico de archivo documental con soporte óptico. La novedad y, a su vez, el reto suponía integrar una base de datos documental, que implica una gran variedad en las posibilidades de recuperación de información, con un sistema de almacenamiento óptico de imágenes.

El objetivo básico del proyecto era y es el de proveer al profesional de la División con una herramienta de trabajo —en cuanto a la recuperación de antecedentes— de amplias posibilidades.

Este enunciado engloba otros objetivos de carácter más específico y práctico: agilizar la recuperación de información, ampliar los criterios de búsqueda y dar

una gran fiabilidad en el sentido de proporcionar una información precisa y exhaustiva del tema requerido.

Por otra parte, destacar que, desde el inicio del proyecto y dado que la División de Asesoría Legal y Fiscal está implantada en distintas ciudades de nuestra geografía, se planteó que, en lo referente a la recuperación de información, no iba a ser un sistema centralizado en Madrid.

II. METODOLOGIA

La metodología de trabajo ha pasado por una serie de fases que explican de forma estructurada el proceso seguido. Por una parte, nos centraremos en los aspectos documentales y, posteriormente, en el diseño informático: con una breve descripción de la red, servidores, estaciones, etcétera.

II.1. Aspectos documentales

Selección de Información

En esta fase se realizó una valoración de la información que se manejaba en la División —en cuanto a contenidos y tipología— y una selección del material que se iba a introducir en el sistema.

Los tipos de documentos son los siguientes: monografías, documentación interna, informes, cartas, etc. Las obras de referencia —diccionarios, repertorios— los boletines y la información de prensa quedó fuera del proyecto desde el principio.

En lo que se refiere a los contenidos, dada la especialización del área de trabajo y, por tanto, de sus fondos documentales, estos incluyen legislación, jurisprudencia, consultas y trabajos internos.

Otro considerando que se realizó en esta fase previa es el del período temporal que abarcaría la carga inicial de documentos: cinco años. A medida que avanzaba la realización del proyecto, se decidió incluir el año en curso en una puesta al día de la información. De este modo, la base de datos contiene información relevante de los años 1986 a 1991.

Descripción de las bases de datos

Una vez seleccionada la información, se diseñaron cinco bases de datos que, a efectos del software de gestión, constituyen una única base de datos con todos los campos contenidos en cada una de ellas.

Se trabajó en el diseño de las bases, en los campos, en su caracterización y recuperabilidad.

Cada base consta de un modelo de registro o ficha con una serie de campos que abarcan contenidos muy extensos —que, entre otros, incluyen resumen y descriptores— a los que se asocia un número de imágenes con los textos completos correspondientes. No es, pues, un sistema de información referencial sino que el registro en sí contiene el documento completo.

La complejidad del diseño estructural de las bases de datos contribuye a permitir criterios de búsqueda diferentes y a establecer combinaciones entre los mismos y entre las distintas bases de datos.

Los campos son bien de texto completo (full-text), bien de fecha, bien de descriptores, en función del tipo de información que contienen.

Se diseñó la Hoja de Recogida de Información (HRI), que se corresponde con los distintos modelos de registros de cada una de las bases y se crearon unas instrucciones de cumplimentación de las mismas.

Captura masiva de la información

El trabajo se repartió entre el personal de todas las oficinas que han participado en el proyecto. Ellos fueron quienes rellenaron las HRI: cumplimentando datos, elaborando resúmenes y adjuntando las imágenes de texto asociadas a cada ficha o registro según las instrucciones dadas previamente. Esto supuso un buen método de trabajo en el sentido de que constituía la carga inicial de las bases y el volumen de información era considerable; de este modo, no hubo que contratar personal adicional. También fue interesante porque implicaba a toda la División en el proyecto. Como contrapartida, era personal técnico y especializado en el campo jurídico y fiscal pero ajeno a temas documentales y ello originó algunos problemas en cuanto a falta de homogeneidad en el vocabulario e inconsistencia en los descriptores.

Esto último trajo consigo un gran esfuerzo por parte del personal de documentación en cuanto a verificación y corrección del trabajo realizado. También fue el servicio de Documentación el que se ha ido encargando de la base de datos de Bibliografía.

Una vez depurada la información, se procedió a la contratación del grupo de trabajo de grabación, compuesto por tres personas: dos para los datos y una en el escáner. La introducción de datos se hizo de forma lenta pero con un control muy elevado del proceso con el fin de evitar errores, verificar toda la información dos veces y mantener un alto grado de calidad en la introducción de datos e imágenes. Este proceso, por cuestiones funcionales, se dividió a su vez en dos fases intercambiando al personal implicado.

En lo que se refiere al archivo electrónico de las imágenes en disco óptico surgieron inicialmente problemas de tipo técnico y fue preciso hacer algunos

ajustes tanto en el software como en el hardware. Por todo ello, la comprobación de las imágenes ha sido exhaustiva hasta el final, verificando tanto los originales como las copias de discos ópticos que iban a ser destinadas a otras oficinas.

Terminología y Optimización

Inicialmente, una vez recopilada la información, se realizó una prueba con documentación real —previamente se había hecho una demo para presentación de las prestaciones del producto— sobre los documentos.

Se probaron las bases de datos, la entrada de información y la recuperación de la misma. En ese momento se pasó por una fase de optimización de las bases de datos en la que trabajaron conjuntamente el equipo informático y el documental asociados al proyecto.

Paralelamente a la optimización de las bases de datos se detectó, desde el punto de vista documental, un gran problema en la falta de control de la terminología, especialmente en lo que afectaba al campo de materias —descriptores— existente en cada una de las bases de datos. Se había trabajado directamente sobre el lenguaje natural, no controlado.

Como consecuencia de este último aspecto, se tuvo que establecer un vocabulario controlado a posteriori, a partir de la información recibida. Se elaboraron unos índices para el campo de descriptores y, como se requería cierta homogeneidad en el tratamiento de todos los datos, se hizo un control terminológico de toda la información, mediatizado por la propia idiosincrasia del lenguaje propio de los temas fiscales y legales. En un futuro se pretende realizar un tesoro basado en los índices ya establecidos.

II.2. Diseño informático

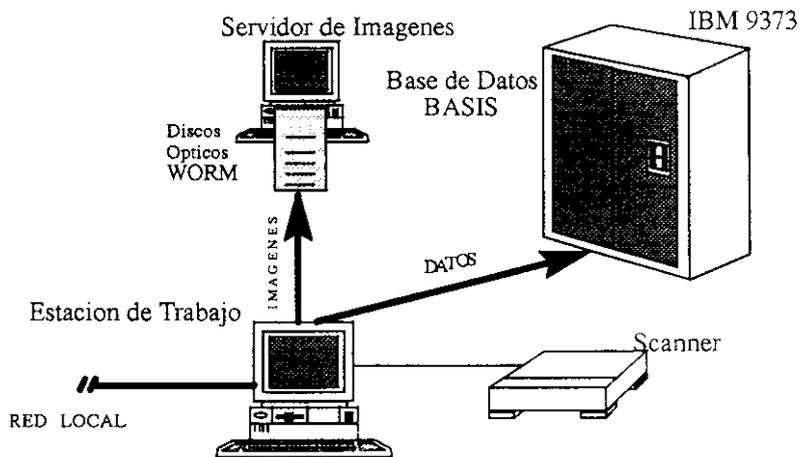
Hasta aquí se ha hecho una exposición de los aspectos documentales del desarrollo metodológico del proyecto pero una cuestión fundamental que queda por abordar es el diseño del sistema informático utilizado.

A una red de área local Ethernet (red NOVELL) se conectó un IBM 9373-025 en el que reside la base de datos BASIS con una aplicación hecha a medida.

Consta de un servidor de red, un servidor de imágenes y un servidor de impresora. El servidor de imágenes tiene asociada una disquetera de cuatro unidades de discos ópticos WORM (no regrabables). Los discos ópticos empleados son LM-D501W, de 470 MB por cara sin formatear.

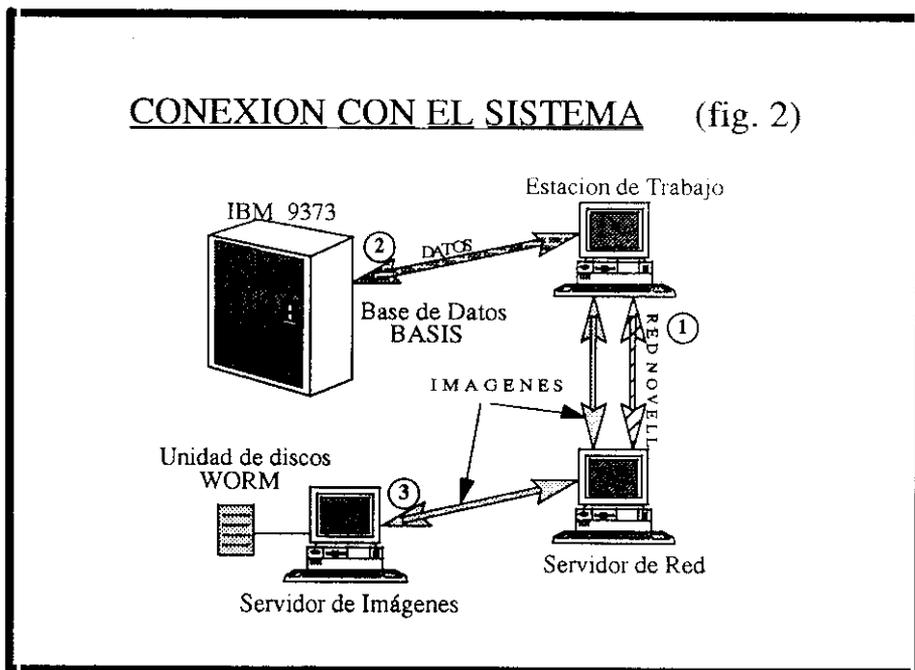
Además, en Madrid, hay cinco estaciones de trabajo con monitores de alta resolución para las imágenes. De estas, una está conectada con el escáner y tiene

ENTRADA Y ALMACENAMIENTO DE LA INFORMACION (fig. 1)



un gestor de imágenes, es la estación de introducción de imágenes (ver fig. 1); el resto son estaciones de consulta y actualización y están accesibles a todo el personal de la División. Todas las estaciones están conectadas al IBM y a la red.

La conexión con el sistema se establece desde las estaciones de trabajo a través de la red—mediante el servidor de red. Una vez establecidas las comunicaciones se accede a la base de datos residente en el IBM para formalizar la petición de información. Se realiza la consulta requerida y como resultado de la misma, aparece en pantalla una selección de registros a la búsqueda efectuada. Cada registro contiene datos de referencia y una o varias imágenes asociadas. Desde cualquiera de los registros se llama a las imágenes correspondientes, que están almacenadas en discos ópticos, a las que se accede mediante el servidor de imágenes, que contiene un sistema de gestión de las mismas. Finalmente, aparecen en pantalla las imágenes con la información completa del registro previamente seleccionado (ver fig. 2).



- ① Conexión a la Red y establecimiento de comunicaciones se almacena en C: de la estación de trabajo el software necesario para comunicarse con otros elementos del sistema (soft. de manejo de imágenes, etc...)
- ② Conexión al IBM para manejo de bases de datos (SICRA)
- ③ Conexión al servidor para capturar las imágenes (SICRA)

Respecto a las otras oficinas, distantes de Madrid geográficamente, se conectan a la base de datos del IBM vía modem por la RTC. Debido a los problemas que suelen surgir en la transmisión de imágenes se optó por proveer a cada oficina implicada en el proyecto de un servidor de imágenes y una disquetera de discos ópticos (WORM). De este modo, aunque la carga y actualización de las bases de datos está centralizada en Madrid, el acceso en modo de consulta es igual para toda las oficinas, que disponen de una red de área local propia (ver fig. 3). La recuperación de las imágenes se realiza siempre en redes de area local evitando así los problemas de transmisión remota. Desde Madrid se mandan las copias de los discos ópticos correspondientes y las sucesivas actualizaciones.

III. CONCLUSIONES

En la actualidad el SICRA es un sistema de gran utilidad en cuanto a la recuperación de información. Ya que se encuentra en una fase inicial de su implantación todavía no se ha podido recoger una evaluación de rendimientos basada en datos estadísticos; sin embargo, su uso se va extendiendo a medida que el personal comprueba su rentabilidad en el propio entorno de trabajo. Hay que destacar que en todo momento se ha implicado al usuario final en el proceso y que a medida que el proyecto avanzaba se han impartido unos cursos muy básicos de uso de la aplicación.

Todavía estamos trabajando en la periodicidad de las actualizaciones, en la posibilidad de exportar las imágenes y otras posibilidades de optimización y mejora del sistema. El disco óptico nos ha permitido tener una base de datos a texto completo de gran capacidad y fiabilidad. En la actualidad, el SICRA contiene unos 3.500 registros con más de 17.000 imágenes asociadas.

En cuanto a los objetivos marcados para el proyecto, cabe destacar una serie de cuestiones unidas a un eje central que podría resumirse en una recuperación de la información efectiva y fiable:

- ahorro considerable de espacio
- rapidez en las búsquedas
- posibilidad de consulta simultánea
- facilidad de acceso a un documento por criterios de búsqueda diferentes
- fiabilidad en el almacenamiento: WORM

Frente a otras tecnologías en uso y ya implantadas, el archivo electrónico de documentos con disco óptico ofrece claras ventajas en lo que se refiere a la duración (tiempo de vida) del disco óptico, al tiempo de acceso a la información y a la gran fiabilidad en la grabación digital, de disco a disco.

ESQUEMA GENERAL DEL SISTEMA SICRA EN MADRID (fig. 3)

