

## **PUBLICACION ELECTRONICA DE PATENTES Y SU DIFUSION EN INTERNET**

Carlos Turmo, Jefe de la Dependencia Informática de la OEPM

### **1.-LA PUBLICACION DE PATENTES**

Dentro de las actividades definidas por la Ley de Patentes 11/86, está la de publicación de la especificación de la patente en su Artículo 32.:

- 1. Transcurridos dieciocho meses desde la fecha de presentación de la solicitud o desde la fecha de prioridad que se hubiera reivindicado, una vez superado el examen de oficio y hecha por el solicitante la petición del informe sobre el estado de la técnica a que se refiere el Artículo 33, el Registro procederá a poner a disposición del público la solicitud de patente, haciendo la correspondiente publicación en el "Boletín Oficial de la Propiedad Industrial" de los elementos de la misma que se determinen reglamentariamente.*
- 2. Al mismo tiempo se publicará un folleto de la solicitud de patente que contendrá la descripción, las reivindicaciones y, en su caso, los dibujos y los demás elementos que se determinen reglamentariamente.*
- 3. A petición del solicitante podrá publicarse la solicitud de patente, en los términos establecidos en el presente artículo, aun cuando no hubiera transcurrido el plazo de dieciocho meses mencionado en el apartado 1.*

El objeto de la publicidad en los procedimientos de concesión de patentes es difundir la información contenida en esa documentación a efectos de enriquecer los conocimientos de la sociedad innovadora. En la legislación anterior (Estatuto de la Propiedad Industrial de 1929) se limitaba la actividad a la mera publicidad administrativa, es decir a la disponibilidad al público de las memorias presentadas por los solicitantes en los archivos del entonces RPI (Registro de la Propiedad Industrial) y definidos como públicos en el BOPI. La publicación activa del contenido de las solicitudes de patentes mediante la edición de un panfleto distribuido por la OEPM requirió la puesta en marcha en 1986 de un nuevo servicio: SPEP, Sistema de Publicación Electrónica de Patentes. Este servicio elabora por medios electrónicos un producto que finalmente permite la publicación de los contenidos de las patentes.

### **2.-DISEMINACION DE LA INFORMACION PATENTE EN LA OEPM**

La política de información de la oficina cuenta con dos importantes herramientas para la diseminación de la información patente CIBEPAT y CD-CIBEPAT. Siendo esta la versión on-line y en CD-ROM de la base de datos documental de patentes. Su distribución con coste para los suscriptores ha

venido siendo la principal herramienta de la diseminación de la información patente en castellano.

En conjunción con la Oficina Europea de Patentes se elaboran productos en CD-ROM (ESPACE-ES) que sirven para distribuir en formato facsímil las patentes españolas que no reivindican prioridad, es decir que no tienen equivalentes en otros idiomas.

La Oficina Española de Patentes y Marcas publica desde 1998 el Boletín de la Propiedad Industrial en Internet, contando por tanto el público especializado con las notas de publicación de las solicitudes en la red.

Espacenet, un proyecto de cooperación dentro de la Organización Europea de Patentes permite distribuir de forma coordinada con todas la oficinas pertenecientes a la Organización la mayor fuente de información patente en la red. Unos treinta millones de documentos en facsímil en diferentes idiomas están disponibles a los usuarios.

La disponibilidad en formato electrónico de un fondo documental en castellano de gran interés tecnológico, con características técnicas que favorecen su diseminación en Internet y el apoyo decidido a difundir de forma activa y eficaz la documentación patente en castellano, determinaron la creación de un servicio que pone a disposición de cualquier usuario de la red un importante repertorio de tecnología y que enriquece de forma notable los contenidos de la red en ese idioma.

### **3.-EL SISTEMA DE DOCUMENTACION ELECTRÓNICA DE PATENTES (SPEP)**

Como ya hemos mencionado desde 1986 la OEPM viene utilizando medios electrónicos para la publicación de las especificaciones de patentes. El sistema cuenta con las secciones siguientes:

- A. Captura de Datos
- B. Tratamiento de textos.
- C. Edición.
- D. Almacenamiento temporal.
- E. Aprobación e impresión.
- F. Almacenamiento permanente.
- G. Administración del Sistema.

#### A) Captura de Datos

Su misión es incorporar los documentos al sistema. Esto se realiza mediante lectoras ópticas de caracteres (OCR) en las cuales está integrada la función de aprendizaje en la de lectura mediante el software *ReadStar*. *ReadStar 6* es un

sistema que puede tratar tanto impresos creados por un periférico de microordenador como cualquier otro mecanografiado, transformándolos en ficheros de caracteres ASCII. Se adapta a todos los paquetes de tratamiento de texto. El sistema permite la lectura de documentos de naturaleza muy diversa. Este sistema está compuesto por dos elementos: un lector óptico conectado al ordenador por un cable paralelo que transmite una imagen "blanco y negro" de la página leída y un paquete de explotación que reagrupa dos sistemas diferentes: El sistema de reconocimiento de caracteres y el sistema de aprendizaje, que permite la determinación de características que aseguran el reconocimiento de nuevos juegos de caracteres y que permite un porcentaje de caracteres reconocidos en torno al 99%.

El ingreso de los documentos en soporte magnético es una operación que se reduce a convertir y copiar ficheros ASCII desde el soporte en que llegan a los discos de las unidades de entrada. Los dibujos y planos son capturados utilizando las mismas lectoras mediante el software *EyeStar*.

#### B) Tratamiento de textos.

Realiza varias tareas:

a) Los documentos introducidos en el sistema mediante digitalización son ficheros ASCII en los cuales algunos caracteres no han sido reconocidos y el sistema automáticamente los transforma en asteriscos. Una de las tareas de la unidad es la corrección de dichos caracteres para lo cual se usa el programa *PACSTAR*.

b) Mediante un software específico se codifican automáticamente los acentos, la ñ, se suprimen los guiones de las palabras cortadas y ciertos caracteres ocultos.

c) Se efectúa la conversión, el tratamiento y copia de las memorias presentadas en soporte magnético.

#### C) Edición.

Los documentos introducidos en el sistema son ficheros ASCII. Esto implica que a cada letra del alfabeto o símbolo de puntuación corresponde un único código, lo que no permite distinguir ni tipos de letras, ni tamaños, ni modalidades tipográficas, ni reconocimiento de fórmulas matemáticas. A partir del anterior fichero y empleando el editor de textos *Aurora* se realiza la composición tipográfica insertando en dicho fichero comandos *LaTeX* (programa de composición tipográfica), lo que permite componer el texto, crear espacios para la inserción de imágenes, escribir fórmulas matemáticas, componer tablas, etc.

Una vez realizado este paso se compila el fichero, generando un fichero que, mediante el adecuado driver de impresión, compone las páginas con calidad tipográfica en impresoras láser.

Los ficheros generados se pueden visualizar en pantallas gráficas de alta resolución mediante el programa *MAXVIEW* para pruebas de resultados, disminuyendo el coste de originales de prueba en las impresoras láser.

#### D) Almacenamiento Temporal.

Aprobado en su conjunto el trabajo realizado se efectúa su almacenamiento.

Existen tres aspectos distintos en las necesidades de almacenamiento temporal:

1. La base de datos que controla el sistema.
2. Almacenamiento temporal de ficheros de textos y de imágenes en curso de elaboración.
3. Almacenamiento temporal de solicitudes y concesiones publicadas de patentes españolas y europeas y de modelos de utilidad. Estos documentos una vez impresos pasan a almacenarse en discos ZIP de 100 Mb y posteriormente en CD-R.

#### E) Aprobación e impresión

Los ficheros, a través de la red, se reciben en los puestos de impresión donde, agrupándose los ficheros que componen el folleto de patente, se insertan las imágenes y se imprime de modo automático mediante impresoras láser un ejemplar del mismo efectuándose un minucioso control de calidad.

#### F) Almacenamiento permanente.

El almacenamiento definitivo se realiza en CD-R, en los cuales quedan permanentemente los ficheros de forma que, en cualquier momento, se pueda facilitar el folleto, o la parte del mismo, que se requiera mediante la búsqueda en la base de datos, proporcionándolo bien impreso o mediante visualización en pantalla. Al poder ser el soporte leído por cualquier lector de CD-ROM, se pueden facilitar copias a cualquier unidad del Organismo que precise imprimir o visualizar los folletos, como ya se hace en la Biblioteca Técnica.

#### G) Administración del sistema.

El administrador del sistema controla los diferentes subsistemas, la base de datos, el almacenamiento y, en general, la operatividad del mismo. El sistema dispone de una unidad de análisis y programación que asume las tareas de

mantenimiento de las aplicaciones, de la programación y de la base de datos, así como del equipo físico.

El sistema ha venido funcionando satisfactoriamente y, desde Febrero de 1988, ha venido imprimiendo más de 250.000 páginas/año conteniendo muchas de ellas texto e imágenes.

#### **4.-EDICION EN PDF**

El formato PDF (Portable Data Format de Adobe) resulta de especial interés para la diseminación de información en Internet. En el caso de nuestro sistema de publicación electrónica, la organización de la información ya existente permite preparar de forma rápida una versión PDF de nuestra documentación con virtudes muy apreciables en relación a los antecedentes. De esta forma adquiere un gran valor añadido la labor realizada por un sistema que cuenta con mas de una década de actividad. La información preparada para los productos Espace-ES y Espacenet está basada en documentos facsímil de grandes dimensiones (800KB de media por documento) y baja información por su naturaleza facsímil. Las ventajas de los documento PDF en español se basan pues en su mejor distribución mediante soportes físicos (un CD puede contener unos cuatro mil documentos) y por Internet ya que de media un documento de patentes ocupa 150 KB y puede ser redistribuido con facilidad por otros medios, como el correo electrónico etc.

Además de iniciar la publicación en paralelo a su publicación legal de los fascículos en CD, se preparó una colección retrospectiva a partir de los documentos electrónicos preparados desde el inicio de la actividad del sistema (1988). En total unos 175.000 documentos que forman la base de eficaces servicios documentales electrónicos.

#### **5.-DISTRIBUCION EN INTERNET**

La recuperación de esta documentación está basada en la base de datos documental MistralX que corre en un sistema Sun Solaris (SUN Enterprise 5000). Esta base recoge referencias de mas de un millón de documentos. Los documentos PDF se encuentran actualmente en un servidor NT en la que ocupan 30 GB, estando en fase de proyecto la dotación del servidor Sun de un sistema de almacenamiento suficiente para las previsiones de este servicio, esto es unos 200 GB netos.

Las bases de datos documentales existentes en la OEPM se basan en el sistema de recuperación MISTRAL, que es un software para el diseño, creación, actualización, gestión y tratamiento de

la información de bases de datos, desarrollado para búsquedas documentales rápidas mediante la recuperación tanto on-line como off-line. Gestiona la carga, mantenimiento y recuperación de múltiples bases de datos, estando estructuradas en dominios y éstos en campos de longitud variable.

Las listas de documentos recuperados cuentan entre sus campos, además de con los datos bibliográficos e imágenes del documento referenciado, con la url de su publicación si esta existe. La activación de esta imagen desde un navegador con el plug-in de Acroread permite acceder directamente a la consulta del documento con todos los servicios que presta el plug-in, incluida la búsqueda de palabras dentro del documento, servicio de gran utilidad en el estudio de las patentes.

Los documentos pueden cargarse on-line mediante un comando o por un proceso batch, con la posibilidad de utilizar para la grabación de los datos, cualquier procesador de textos. Mediante el programa de carga se generan diccionarios para cada campo, lo que permite que la recuperación de la información sea inmediata, con tiempos de respuesta muy cortos.

Existe un programa adicional que en tiempo real facilita la depuración continuada de las Bases de Datos, en cualquier campo, además de la incorporación puntual por este proceso de los documentos y las imágenes.

Después de un proceso de migración realizado durante el primer trimestre de 1999, el software de diseño y recuperación, es ahora MISTRAL/X, arquitectura cliente/servidor, siendo el cliente Windows/95 o superior. Este nuevo software permite al usuario realizar sus propios diseños de consulta, lo que hace que el producto sea una herramienta apropiada para todo tipo de usuarios. Para otro tipo de consultas más especializadas, el software permite la utilización de una ventana experta con todas las posibilidades para usar operadores booleanos, y mezclas entre campos. En la migración se ha prestado especial atención a la incorporación al nuevo interfaz de usuario de todas las funciones ya existentes en el formato anterior, así como de nuevas posibilidades cuyo desarrollo resultaba imposible con el mismo y que se habían visto necesarias a lo largo de su tiempo de uso.

A las actuales bases de datos OEPMPAT Y MODINDU se les han incorporado las imágenes publicadas en el BOPI desde 1998. Además se está realizando una carga retrospectiva con imágenes procedentes del SPEP que cubrirá desde el año 1986 en Patentes y Modelos de la Nueva Ley. Se ha previsto un campo que permitirá la consulta del documento completo en formato PDF utilizando el visor de Acrobat. Este campo permitirá la consulta de los documentos de patentes referenciados como se

indica en extenso en otro punto. En la Base de Datos LATIPAT se está haciendo una retrospectiva con las imágenes desde 1991.

Además del desarrollo cliente-servidor, la migración a MISTRAL/X ha incluido la implementación de un servicio de consulta vía páginas HTML, que se ha elaborado con un doble diseño intranet - internet. El acceso vía intranet permite la interrogación por todos los campos de la base, y aunque sus posibilidades de búsqueda son limitadas en comparación con las que permite el programa cliente de MISTRAL/X, resultan suficientes para muchas aplicaciones. El acceso vía internet se ha diseñado para ofrecer una consulta no profesional al público en general.

## **6.-CONTENIDOS Y ALCANCE**

El servicio de publicación de patentes en Internet permite explotar una información de alta calidad como son los documentos de patentes en modo mixto, indexados mediante la bases de datos documental Mistral.

La interrogación de la base en conjunción con el uso de imágenes relevantes de la primera página y los datos bibliográficos permiten centrar las búsquedas en los documentos relevantes de forma ágil. La herramienta de búsqueda da acceso a una importante colección de documentación en castellano que supone un servicio de singular interés para la comunidad innovadora hispanoparlante.

El fondo en castellano formado por los documentos PDF referenciados actualmente en el sistema, contiene 175.000 documentos libremente consultables y reproducibles. De esta documentación 31.570 documentos corresponden a modelos de utilidad, 70 documentos a publicaciones de patentes bajo el antiguo estatuto, 30.000 publicaciones de patentes de la denominada nueva ley, 85.000 documentos de patentes europeas con efecto en España y 27.000 patentes PCT. Mensualmente se incorporan los documentos publicados en los BOPI del mes, dándole a la información un alcance muy significativo. Actualmente esta documentación ocupa 30 GB de un servidor NT.

## **7.-PERSPECTIVAS Y CONCLUSIONES**

Extensión de la documentación española. A través de la colaboración con la Oficina Europea de Patentes se está realizando un proyecto que permitirá convertir de forma automática a PDF los documentos españoles que en formato ST35 de OMPI (estándar de propiedad industrial combinación de SGML y TIFF Grupo IV para los documentos de patentes) se encuentran en el sistema BNS de la Oficina Europea. Esta documentación es de menor calidad que la publicada por la OEPM ya que está en

facsimilar, no tiene texto buscable, ocupa una media de 800 KB frente a los 150 KB de los documentos publicados por SPEP. Este proyecto incorporará cerca de doscientos mil documentos mas al servicio, publicados entre 1968 y 1986, enriqueciendo notablemente el fondo documental en español. Están previstas unas necesidades de almacenamiento netas de 160 GB para esta documentación.

Documentación Iberoamericana. Dentro de los proyectos de la OEPM encuentra gran prioridad la creación de servicios en castellano mas allá de las propias actividades de publicación. En este sentido existe el proyecto de publicar en Internet las patentes de los países iberoamericanos, cuyos datos se recogen ya en nuestra base de datos bibliográfica Cibepat.

Este proyecto considera en muchos casos la incentivación de la actividad de publicación electrónica en las oficinas iberoamericanas de propiedad industrial. La OEPM proyecta actividades de cooperación internacional en conjunción con otros organismos (EPO, OMPI) a fin de promover el desarrollo de la actividad electrónica de dichas oficinas.

Integración con otras bases. La integración de la información ofrecida en Internet permitirá conocer directamente información relacionada con las publicaciones de patentes consideradas y contenida en otras bases de datos de nuestra oficina y de otros organismos, como la situación jurídica de esas solicitudes, el documento equivalente europeo cuando exista o las publicaciones equivalentes de la "familia" de patentes. Esta actividad requiere la integración con otras bases de datos disponibles en Internet, como Espacenet, etc. con cuyos administradores colaboramos para estos fines.

La integración de procedimientos electrónicos de publicación de patentes y el uso de Internet permiten además de dar la publicidad eficaz a la solicitud que el procedimiento exige, dar un gran alcance a la tecnología disponible en la documentación patente que debe beneficiar de forma singular a los agentes de la innovación en el entorno hispanoparlante y de forma particular a la pequeña y mediana empresa.

Figura 1  
 FORMULARIO DE INTERROGACION DE CIBEPATNET

SearchWay WEB - csepapat - Microsoft Internet Explorer

MINISTERIO DE CIENCIA Y TECNOLOGÍA Oficina Española de Patentes y Marcas csepapat

### Consulta de la base de datos Inventiones españolas

Introduzca los valores de búsqueda deseados y haga clic en "Ejecutar". Para verificar la existencia de términos en el diccionario, haga clic en el botón "Diccionario" del campo correspondiente. Para más detalles, haga clic en "Ayuda". Para obtener un formulario en blanco, pulse en el botón "Anular"

F. Pub. Conces.	1999+	Consultar el Diccionario
Provincia Española	48	Consultar el Diccionario
*Números Solicitud		Consultar el Diccionario
*Números Publicac.		Consultar el Diccionario
*Clasificaciones		Consultar el Diccionario
*Titulo-Resumen		Consultar el Diccionario

Dominios utilizados para la búsqueda: MCOEPI / MCOONLP

Ejecutar Ayuda Anular

http://www.csepai.es/publico/csepapat\_es?ACTION=HELP

Figura 2  
PRESENTACION DE LA LISTA DE RESULTADOS

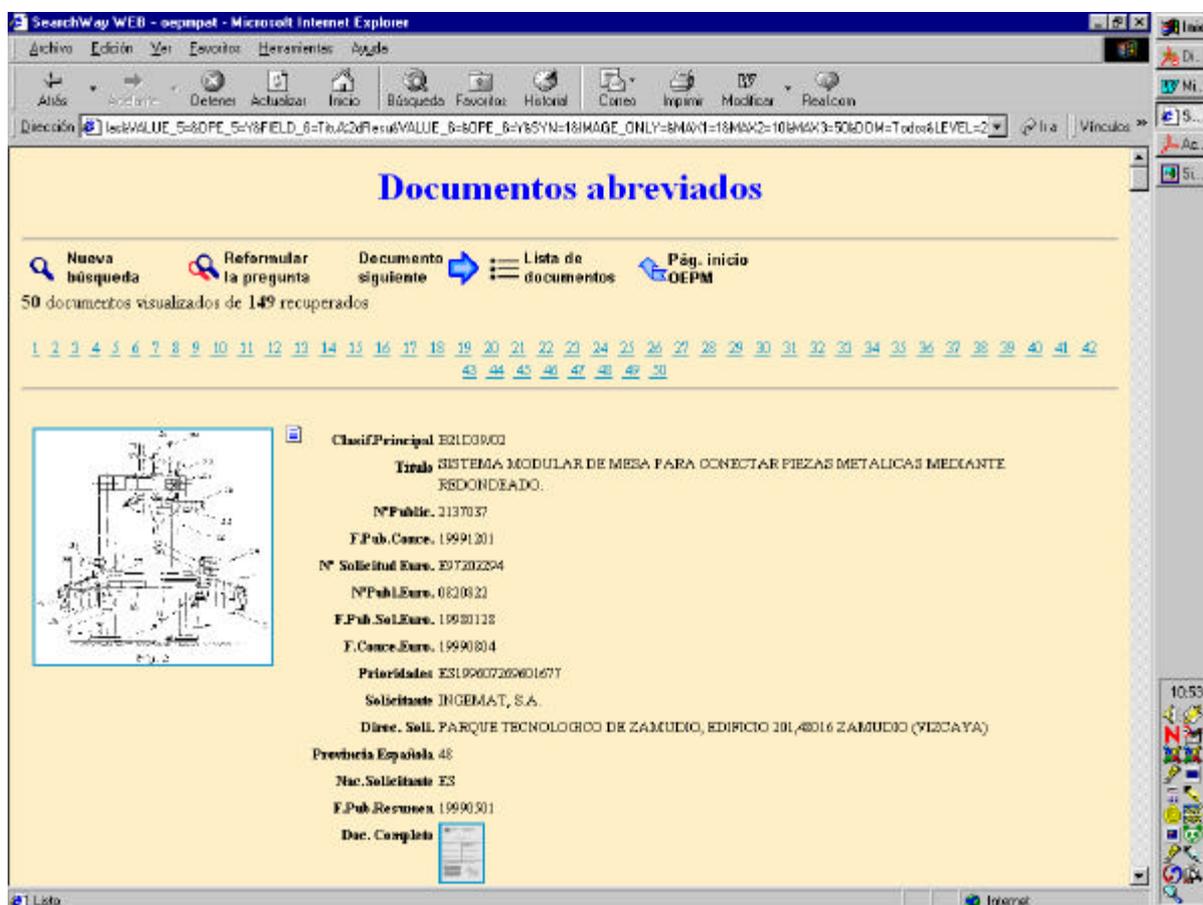


Figura 3  
PRESENTACION DE UN DOCUMENTO DE PATENTES EN PDF

