

B.5. Europeana, bibliotecas digitales y repositorios. ¿Pueden competir con Google como herramientas de búsqueda documental?

Por Luis Rodríguez-Yunta

23 diciembre 2008

Rodríguez-Yunta, Luis. "Europeana, bibliotecas digitales y repositorios. ¿Pueden competir con Google como herramientas de búsqueda documental?". *Anuario ThinkEPI 2009*, EPI SCP, pp. 52-54.



Resumen: En torno a la puesta en marcha de Europeana, se plantea la necesidad de reflexionar sobre los sistemas de recuperación que se están implementando en este tipo de recursos. Con frecuencia repositorios, bibliotecas digitales y agregadores no disponen de sistemas de recuperación de información que puedan competir con un buscador como Google. Tampoco se están aplicando las herramientas de calidad existentes en los recursos documentales tradicionales.

Palabras clave: Bibliotecas digitales, Sistemas de recuperación de información, Buscadores, Google, Internet.

Title: Europeana, digital libraries and repositories. Can they compete with Google as document-based search tools?

Abstract: In relation to the beginning of Europeana, the need to reflect on the information retrieval systems that are being implemented in this type of resources is considered. Frequently, digital repositories, libraries and harvesters are not applying retrieval systems that can compete with a search engine like Google. The existing quality tools in the traditional documental resources are not being applied either.

Keywords: Digital libraries, Information retrieval systems, Search engines, Google, Internet.

SI HAY QUE ELEGIR una noticia de 2008 de especial relevancia e impacto social en el ámbito de los recursos documentales, hay que destacar sin duda el lanzamiento de Europeana.

<http://www.europeana.eu/>

Aún en su versión beta, este proyecto ha levantado grandes expectativas y recabado la atención de los medios. Su importancia radica en su carácter agregador de los recursos digitales europeos, por lo cual se presenta como futuro competidor directo de Google para focalizar las búsquedas de documentos dispersos en diferentes archivos y bibliotecas digitales en Europa.

Europeana representa la culminación de un esfuerzo colectivo en la apuesta por los contenidos libres en internet. Los poderes públicos y numerosas instituciones se han implicado en la digitalización de obras de interés general libres de derechos.

Se está poniendo a prueba el potencial democratizador de la Red, especialmente en dos ámbitos: el del patrimonio cultural, con las obras depositadas en bibliotecas, archivos o fototecas, y el de la documentación académica y científica, que universidades, centros de investigación o los

propios investigadores, están depositando en archivos abiertos.

Repositorios institucionales, plataformas de revistas electrónicas de acceso gratuito, archivos y bibliotecas digitales aportan un creciente número de documentos a texto completo en internet.

"La importancia de Europeana radica en su carácter agregador de los recursos digitales europeos, por lo cual se presenta como futuro competidor directo de Google"

A ello se suman imágenes, vídeos y registros sonoros. Un beneficio indudable para bibliotecarios y documentalistas, implicados tanto en el diseño como en la consulta y difusión de estas nuevas herramientas de trabajo. Sin embargo, desde el punto de vista profesional hay que cuestionarse si el diseño de estos recursos es el adecuado, si



los usuarios llegarán a través de las interfaces de usuario de estos sistemas, o aterrizarán directamente en los documentos a través de buscadores como *Google*.

¿Qué ventajas aporta al usuario la búsqueda a través de *Google*? Sin duda, son fáciles de encontrar:

- La interrogación directa del texto completo de los propios documentos, que le permite localizar datos concretos, frases, citas, numerosos elementos que son indispensables para un amplio tipo de búsquedas.
- La navegación entre documentos, a través de citas y enlaces.
- La presentación de resultados en función de criterios que pretenden adaptarse a los usos sociales.

Por el contrario, repositorios institucionales, bibliotecas digitales y agregadores, con frecuencia no disponen de sistemas de búsqueda que puedan competir con estas opciones. Están limitados a la información contenida en los metadatos, excluyen tanto el texto completo como la explotación de las citas y emplean sistemas básicos para la ordenación de resultados.

Estas limitaciones son las mismas que pueden achacarse a los *Opac* o a las bases de datos referenciales. Sin embargo, estas herramientas tradicionales han cumplido una importante función en la tradicional búsqueda de documentación impresa y para los profesionales no cabe duda sobre las características que avalan la calidad de estos recursos:

- El control de autoridades y la opción de búsqueda a través de índices, que permiten delimitar

la forma más apropiada para realizar una búsqueda tan exhaustiva como pertinente.

- La existencia de una amplia batería de campos que permiten un análisis sistemático de la documentación.

- El empleo de lenguajes documentales para la búsqueda por materias: clasificaciones, tesauros...

Estos elementos de la tradición profesional están quedando al margen en los proyectos de acceso a fuentes digitalizadas, cuando podrían ser el punto diferenciador con un buscador global.

Entre los propios profesionales es frecuente escuchar críticas a cualquier diseño de una interfaz de usuario que no se asemeje a la pantalla simple de *Google*. Se piensa que los usuarios se alejan de los sistemas complejos por su falta de amigabilidad, pero la realidad es que hacerlos más sencillos no atrae un mayor número de consultas.

Un sistema aislado no puede competir con el alcance de *Google*, ya que difícilmente puede mejorar la exhaustividad de sus resultados. ¿Por qué no diseñar por tanto sistemas complejos, sabiendo que van destinados a un uso profesional, para resolver las necesidades de búsqueda que *Google* no alcanza?

“¿Por qué no diseñar sistemas complejos, sabiendo que van destinados a un uso profesional, para resolver las necesidades de búsqueda que *Google* no alcanza?”

Si tomamos el diseño de *Europeana*, en su página principal, se aprecia con claridad este fenómeno de imitación de la estética de *Google*, un formulario de búsqueda con un único recuadro.

La búsqueda avanzada ofrece poco más: limitaciones por título, creador, fecha y materia, en suma, los metadatos más básicos.

A ello se añaden las opciones de refinamiento por idioma, país, fecha, proveedor o tipo de documento. Las posibilidades de búsqueda en este contexto son limitadas.

Por ejemplo, un mapa histórico del siglo XVIII figura en el apartado de imágenes, igual que cualquier fotografía contemporánea. ¿Cómo distinguirlos dentro del mismo tipo de fuente? No por fecha, pues este campo parece referirse al año de la digitalización, no de la imagen original.

El mismo análisis puede realizarse sobre los repositorios institucionales o archivos de *e-prints*.

Será imposible competir con *Google* si se apuesta por interfaces de usuario que limitan la búsqueda a los metadatos sin el contrapunto de herramientas avanzadas (control de autoridades, tesauros multilingües, clasificaciones...).

Los esfuerzos invertidos en la construcción de *Europeana*, o en cualquiera de los archivos y bibliotecas digitales, no serán nunca baldíos.

En cualquier caso se está contribuyendo a potenciar la accesibilidad del conocimiento en una dimensión desconocida hasta ahora. Pero el debate profesional debe dirigirse también hacia el diseño de herramientas de recuperación, para ofrecer un servicio de mayor calidad y dirigido también al uso profesional.

* * *

La importancia del control de autoridades

Por **Ricardo Santos-Muñoz**

Me he alegrado bastante al leer esta nota, que expresa gran parte de mis opiniones sobre hacia dónde van los *Opacs*. Creo que la *googleización* de los interfaces de búsqueda de las bibliotecas (la barra única de búsqueda, monda y lironda), es un error que desaprovecha clamorosamente buena parte del trabajo técnico de la catalogación, esto es, el control de puntos de acceso, su normalización y sus relaciones.

Precisamente, cuando medio mundo profesional se machaca el cráneo inventando la manera de poner un orden en las búsquedas por internet, tanto en normalización como en asistencia a la obtención de resultados por medio de enlaces, etc., las bibliotecas, que ya tenemos ese trabajo hecho, utilizamos unos interfaces que ignoran o infrutilizan el trabajo de autoridades.

La búsqueda por texto libre de un registro bibliográfico es una excelente posibilidad, pero que no tiene en cuenta las cosas que no están en los registros, esto es, los numerosos reenvíos y entradas alternativas que proveen los registros de autoridad, que abren las puertas hacia el camino hacia el multi-lingüismo, por ejemplo, cosa que debería ser fundamental en *Europeana*.

Tampoco le vayamos a echar toda la culpa hacia la influencia involuntaria de *Google*; los *Opacs* de la era pre-*Google* también fallaban en este aspecto, de modo que sólo en la búsqueda por índice se aprovechaba el potencial de la normalización de los puntos de acceso, pero actualmente, los nuevos interfaces se dirigen, desde mi punto

de vista, hacia el lado contrario de donde se debería de ir.

La actual tendencia en muchos programas de catalogación en cuanto al enriquecimiento de los registros bibliográficos es estupenda, pero empecemos la casa por aprovechar el enorme trabajo que ya tenemos.

El modelo *Frbr* parece mostrar el camino al que deben ir los *Opacs*, es decir, responder a las consultas desde el nivel de obra-expresión y no desde la manifestación, como se ha venido haciendo.

Para decirlo de otra manera, en lenguaje más asequible, que cuando alguien busca en esas barras de búsqueda anteriormente mencionadas, se responda con vocabulario controlado, con puntos de acceso controlados, desde autoridades y, desde ahí, ir bajando.

O si no, haced este experimento: probad en *Europeana* a poner **Durero** y **Dürer**, y comparad los resultados. Un simple paso por autoridades antes de empezar a mostrar contenido podría solventar este pequeño gran problema.

El control en la recuperación de documentos en *Europeana*

Por **Francisco-José Diago**

Quizás deberíamos comparar *Europeana* no con el buscador *Google*, sino con *Google Books*, que tiene las ventajas de *Google* y añade la posibilidad de buscar por autor, *Isbn*, título, idioma, fecha de publicación, etc., al margen de otra serie de opciones y funcionalidades.

De todas maneras, el sistema elegido por *Europeana* de redirigir a la institución que contiene el documento para poder acceder al mismo, me parece poco acertado. Se pierde el control de los documentos recuperados y hay que utilizar sistemas de visión de documentos de lo más peregrinos.

El sistema de registro de usuarios quizás mejore estas cuestiones y ofrezca otras posibilidades, pero "*Registration is not enabled yet*".

¿Puede competir con *Google*?

Si buscamos "*en un lugar de la mancha de cuyo nombre no quiero acordarme*", obtenemos los siguientes resultados

http://books.google.es/books?as_brr=1&q=en+un+lugar+de+la+mancha+de+cuyo+nombre+no+quiero+acordarme&btnG=Buscar+libros

<http://www.europeana.eu/portal/brief-doc.html?start=1&view=table&query=en+un+lugar+de+la+mancha+de+cuyo+nombre+no+quiero+acordarme>

Para este viaje no hacían falta alforjas.