

El ámbito tecnológico

En sus inicios, la BVMC asumió los compromisos más avanzados del acervo tecnológico disponible (etiquetado XML, metadatos, principios de la web semántica), al tiempo que una tecnología (servidores, hardware, software, etc.) estandarizada y muy contenida en costes, en contraste con las inversiones al uso de la época. En la actualidad la conversión de las imágenes digitalizadas a texto, su corrección manual y su conversión a páginas HTML supone una costosa y lenta operativa que restringe el alcance y velocidad de la digitalización de obras. Los actuales modelos de digitalización empleados por Google, Microsoft, y presumiblemente, por la Biblioteca Digital de la Unión Europea (BDUE) concilian una doble exigencia de rigor y operatividad. Por un lado, el usuario accede a la imagen de la obra original digitalizada (documento, manuscrito, libro, etc.) lo que le garantiza la mayor fiabilidad; por otro, los procesos de búsqueda se realizan a través del texto reconocido –no corregido– por el software especializado existente, llegando en muchos casos a niveles de exactitud superiores al 99%.

La disponibilidad de esta tecnología abre un enorme potencial para la BVMC. Potencial que debe evaluarse y especificar dentro de la operativa actual. En sí misma, la digitalización de un impresionante corpus literario de unas 20.000 obras representa ya un inmenso tesoro, ya sea para la aplicación de técnicas de concordancias o para la investigación en general. Además, la consideración de un método alternativo de digitalización más operativo, no excluye en modo alguno adicionar nuevos documentos u obras que por su interés deban sumarse al corpus existente.

Así mismo, la tecnología del futuro se proyecta en el marco de la Web 2.0 con propuestas y potenciales realmente interesantes (folksonomies, para la catalogación popular de documentos, wikis, API públicas, geoposicionamientos...) para su desarrollo y aplicación en la BVMC.

Los autores encargados son Rafael Carrasco, Catedrático de Lenguajes y Sistemas Informáticos. Director Adjunto de la BVMC, Responsable actual del desarrollo tecnológico de la Biblioteca Virtual en la actualidad. **Manuel Marco** Director de Tecnología del Grupo Prensa Ibérica. Responsable del proyecto de digitalización de la hemeroteca histórica del Grupo. Ex Vicerrector de Nuevas Tecnologías de la Universidad de Alicante y responsable de tecnología de la BVMC durante su periodo como vicerrector (1997-2000). **Pedro Pernías**, Director del proyecto Contenidos Abiertos. Profesor de las asignaturas de “Nuevas Tecnologías aplicadas a la educación” y “técnicas multimedia”. Director del Secretariado de Nuevas Tecnologías de la UA (1996-2001). Director de Contenidos de Universia, de su Biblioteca de Recursos de Aprendizaje y del área de formación y cursos en línea (2002- 2005).

El ámbito tecnológico

Manuel Marco, Rafael Carrasco, Pedro Pernías.

1.- Contexto tecnológico: De la Web 2.0 a la BVMC 2.0

- Introducción
- La Web es la plataforma.
- Nueva concepción del contenido.
 - Programación abierta.
- Foco en el usuario.
 - Web participativa: relaciones con los usuarios.
 - Gestión del conocimiento colectivo: sabiduría de la multitud.
- Beta perpetua.
- Desarrollo científico.

2.- Descripción funcional de la BVMC 2.0.

- Proporciona contenido para las alianzas estratégicas; (internacionalización).
- Explorar nuevos modelos de negocio (sostenibilidad).
- Liderazgo tecnológico y cultural (factor académico-científico).

3.- Modelo tecnológico.

4.- Arquitectura Tecnológica.

1. Contexto tecnológico. De la Web 2.0 a la BVMC 2.0

Introducción

La Biblioteca Virtual Miguel de Cervantes (en adelante, BVMC) abrió en 1999 una puerta a la relación entre la tecnología y la cultura tratando de poner en un mismo sitio lo mejor de ambos mundos en el entorno iberoamericano.

La BVMC es algo más que un catálogo bibliográfico más o menos especializado: también es un portal-entorno al que los usuarios comenzaron a desarrollar cierta actividad.

Pero la Red es cambiante y, en la actualidad, fruto de la reflexión sobre la actividad de millones de usuarios en todo el mundo, ha emergido la idea de que la Web se reinventa a sí misma y que la nueva Web posee una serie de características que suelen agruparse bajo el paraguas conceptual de la llamada Web 2.0¹ (en oposición a la Web tradicional o Web 1.0, aunque nadie la bautizara de aquella manera).

Precisamente por el carácter “virtual” que tienen las bibliotecas digitales y a que su ecosistema es el de la Red, las bibliotecas de este tipo no podían permanecer ajenas a estos cambios y a las propuestas que de él se han derivado. Esta reflexión es la que, paralelamente ha venido a denominarse “Biblioteca 2.0”²

¹ Tim O’reilly anuncia el concepto “Web 2.0” tras unas reflexiones que tuvieron lugar a finales del 2005 sobre el nuevo rumbo que estaba tomando la Web y que reflejó en el artículo “Qué es la Web 2.0. Patrones de diseño y modelos de negocio para la siguiente generación de software”
<http://sociedaddelainformacion.telefonica.es/jsp/articulos/detalle.jsp?elem=2146>

² Michael Casey es el primero que en su *blog*
http://www.librarycrunch.com/2005/10/working_towards_a_definition_o.html anuncia el término y escribe unas interesantes reflexiones sobre el mismo

Para presentar el modelo tecnológico, primero hemos de presentar muy condensadamente los elementos fundamentales de la Web 2.0 y la Biblioteca 2.0:

- La Web como plataforma.
- Nueva concepción del contenido.
- La Web colaborativa.
- Programación abierta.

Trataremos de extraer de la reflexión consiguiente conclusiones útiles para enfocar el modelo tecnológico que debe poseer la BVMC para dar el salto a la que, de manera mimética, también denominaremos BVMC 2.0.

A continuación, presentaremos las funcionalidades que proponemos debe aspirar a poseer la BVMC y por último describiremos el modelo informático que lo hará posible.

A este respecto es muy importante señalar una cuestión clave: la Web no se puede improvisar. Una construcción como la que se ambiciona exige una cuidadosa y sofisticada arquitectura informática para hacer posible los proyectos propuestos sin que los costes de desarrollo, en tiempo y en esfuerzo económico y humano, hagan inviable su ejecución o la demoren perdiendo la oportunidad que en un momento dado se pueda presentar.

Por ello, la primera y segunda parte de este estudio debe considerarse un desiderátum razonado y la tercera la propuesta que lo hace factible.

La Web es la plataforma

Hasta ahora, se asumía que el ordenador era el sistema que recogía, procesaba, almacenaba y mostraba la información. La Web solo era un canal para hacer llegar más lejos el teclado y la pantalla de un servidor a fin de permitir el trabajo remoto.

La arquitectura cliente/servidor tradicional de las Redes de ordenadores ha inspirado la Web tradicional. Un servidor emite información que es recogida por clientes distribuidos con poca o limitada capacidad de interacción con el origen de la información

Hoy, la conquista por parte del usuario de la Web hace pensar que más bien es la propia Web, formada por sistemas informáticos, datos y usuarios, el sistema que permite recoger, procesar y distribuir la información. Entendiendo la biblioteca como una parte de la Web se comprende la necesidad de interactuar con ella. Se trata de abandonar la idea de servicio aislado en competencia con otros similares para tratar de encontrar la forma de conectar con ellos y buscar la oportunidad de sacar partido de la relación del usuario con la Red.

La aparición de aplicaciones *online* que vienen a sustituir –o cuando menos a complementar- a las aplicaciones tradicionales de ofimática, tratamiento de imagen, sonido, vídeo y otros, constituyen la prueba de que el ordenador del usuario extiende su capacidad de proceso a la propia capacidad de proceso de la Red.

Algunos ejemplos:

Programa	Función	URL
----------	---------	-----

Google docs Numbler Calc5 Zoho	Servicios ofimáticos que permiten trabajo colaborativo (textos, hojas de cálculo, presentaciones, calculadoras).	http://docs.Google.com http://numbler.com http://www.calc5.com http://notebook.zoho.com
Pixer.us Mojito FD's flickr toys Picresize Eyespot Shinkpictures Bublesnaps	Tratamiento de fotos, vídeos e imágenes.	http://pixer.us http://mojiti.com http://bighugelabs.com http://www.picresize.com http://eyespot.com http://www.shrinkpictures.com http://www.bublesnaps.com
Gliffy, Cumulate Dra. Mindomo	Dibujo de esquemas, gráficos, etc usando la capacidad de proceso de un ordenador remoto y sus posibilidades.	http://gliffy.com http://www.cumulatelabs.com http://www.mindomo.com
Clip2net	Clipboard en la Red para copiar y pegar entre ordenadores conectados.	http://clip.net
Zamzar Media-convert Html2pdf Convert VX2	Servicio Web para cambiar el formato de un fichero.	http://zammzar.com http://media-convert.com http://html2pdf.seven49.net http://convert.vx2.com/
LIM	Autoría de documentos específicos (educativos, ocio, etc).	www.educalim.com
Net2ftp	Servicio web que actúa como un cliente FTP.	www.net2ftp.com
Incompetech	Proveedor de documentos pautados específicos.	www.incompetech.com

Thinkkature	Pizarra para trabajo colaborativo.	http://thinkature.com
Mind42	Servicio para realizar mapas conceptuales.	http://www.mind42.com/
Bibsonomy	Sistema para mantener listas bibliográficas.	http://www.bibsonomy.org
CiteUlike		http://www.citeulike.org
Connotea		http://www.connotea.org

Nueva concepción del contenido

El dato, el elemento básico entorno al cual se construye un servicio o actividad es también el centro de la Web 2.0. Pero el dato se concibe de manera distinta en la actualidad. La propiedad del dato o la exclusividad sobre el mismo ya no es el motivo para desarrollar un modelo de negocio. Hay muy pocas oportunidades de que esto ocurra y son muy frágiles desde el punto de vista comercial. Pero el dato sigue siendo importante porque es la excusa para provocar la actividad del usuario. Mediante su cesión, el poseedor del dato conversa con el usuario y le propone usos específicos, a la vez que se abre a propuestas de uso más imaginativas procedentes del propio interlocutor.

El ejemplo más sobresaliente lo han dado las empresas que proporcionan mapas o imágenes satelitales de la superficie terrestre de gran resolución. Sin que los proveedores principales de este tipo de imágenes (como por ejemplo, Teletlas o Navteq) vean amenazado su negocio, empresas como Google, Microsoft o Yahoo permiten a los usuarios que vean, usen, descarguen y posicionen en sus propios lugares web, *weblogs* o donde deseen, los mapas o imágenes que les ofrecen. La posesión del dato no es crítica: lo es la actividad que el usuario es capaz de desarrollar sobre él.

Otro ejemplo lo tenemos en servicios como Craigslist, (anuncios clasificados). En esta ocasión se trata de una enorme base de datos de anuncios clasificados accesible por terceros directamente, sin pasar por la página principal o usar el buscador web proporcionado por craigslist.org. Un caso especial de utilización de esta base de datos se produce cuando un tercero (*housingmaps*) funde la geolocalización que proporciona Google con la base de datos de Craigslist, proporcionando un servicio de búsqueda de pisos por ubicación geográfica.

Una biblioteca digital posee fondos propios sobre los que orbita la actividad de sus usuarios. Normalmente estos fondos se consultan mediante unos protocolos previstos por los gestores de la biblioteca. Los usos previstos lo son por parte de los diseñadores de la interacción de la Web de la biblioteca digital, si es que existe ese planteamiento. Pero el uso que otros usuarios podrían dar al fondo documental que la biblioteca posee puede ser muy diferente al previsto por los organizadores de la misma. La oferta del documento almacenado en el formato más sencillo y manipulable posible y la observación de estos casos es una fuente de información esencial para conocer los deseos de los usuarios y atraerlos a la biblioteca.

Un ejemplo de utilización de un fondo documental bibliotecario para desarrollar actividad del usuario es LibraryThing (<http://www.librarything.com/>) Esta biblioteca digital permite al usuario ponga en interacción su propia colección bibliográfica con la de librarything para ayudarle a catalogar sus libros, buscar versiones, comentarlos con otros usuarios, recomendarlos y, si es el caso, comprar volúmenes similares a los que ya posee.

La BVMC puede extender sus fondos actuales con los catalográficos de otras iniciativas importantes como **Boogle Books** la **Biblioteca Digital Europea**, o recolectar documentos universitarios procedentes de las principales fuentes de este tipo de información (bibliotecas universitarias, la Biblioteca de Recursos de Aprendizaje de Universia, etc.) a fin de dialogar con sus propios usuarios, y hacerles llegar la información precisa relacionada con sus necesidades como lectores.

Surge así el concepto de “sindicación³ del contenido”. Hasta ahora, syndicar un conjunto de datos bajo unas condiciones económicas era el modelo de negocio de algunos servicios web pero por las razones expuestas anteriormente de fragilidad de la exclusividad del mismo y competencia con iniciativas libres y gratuitas, cada vez es menos frecuente encontrar este tipo de servicio. En su lugar, la sindicación ha pasado a ser simplemente la posibilidad de que el dato circule entre servicios sin cargar con componentes gráficos o de presentación. La sindicación de información está hoy tan extendida que es rara la web donde no encontramos el reconocible icono de XML o RSS que indica cuál es la puerta de acceso para la sindicación del contenido puro de la misma. La mayoría de los *blogs* que se escriben en el mundo poseen esta características; los periódicos y cualquier otra fuente de información como microportales o portales temáticos también.

La sindicación no implica el establecimiento de una relación contractual o ni siquiera formal entre lugares Web. Pero se busca como factor de éxito y el conocimiento de la audiencia de esa sindicación habla de la importancia y uso de una Web. Servicios como feedburner (www.feedburner.com) o technorati (<http://www.technorati.com>) se dedican a ello.

³ El término correcto para utilizar es el de “Redifusión” en vez de “sindicación” que es considerado un barbarismo. Sin embargo, hemos preferido mantener el concepto de “sindicación” ya que este es el utilizado en la mayor parte de la literatura tecnológica y científica sobre este tema.

La tecnología implicada en la “liberación” del dato tiene nombre: AJAX, arquitecturas REST, modelos de programación “ligeros”... En definitiva, son metodologías de programación y arquitectura de sistemas que permiten que el usuario interactúe en diferentes niveles con el proveedor de información, desde una manera simple y unidireccional como es el tradicional uso de la Web hasta una manera más interactiva e intensa como la desarrollada por servicios de ofimática *online* que hemos mencionado anteriormente. En cierta medida, el usuario se convierte en un elemento más de la cadena del desarrollo de un servicio. En este sentido, la idea de proyecto finalizado desaparece siendo desplazada por la de “beta perpetua”. El usuario de la Web ayuda a mejorar el servicio que se presta y sus cambios de preferencias y gustos mantienen a la Web en continuo perfeccionamiento. En otras ocasiones, se proporcionan a los usuarios muchos grados de libertad para acceder al dato y sacarle el mayor partido posible. Estamos hablando de la naturaleza *remezclable* del dato como motivación para muchos usuarios en la búsqueda de nuevos usos para el mismo. Más adelante, en el capítulo dedicado al modelo informático propuesto se verán varias de estas tecnologías puestas en juego.

Foco en el usuario: la Web Social

La Web participativa

La idea de centrarse en el usuario no es nueva ni es patrimonio de la Web2.0, lo que ha cambiado es la forma de llevar a cabo esta relación.

“Internet son conversaciones” y las conversaciones deben poseer interés. No solo hay que escuchar al usuario para darle lo que quiere: también hay que escucharlo porque lo que tenga que decir puede ser de interés para otros.

El usuario añade valor a la Web, al depositar en ella parte de su conocimiento, aunque éste sea privado. Pero el valor real utilizable de la Web se consigue cuando son muchos los usuarios que hacen esto. El valor total del conocimiento de los usuarios es mucho más que la suma de las partes. Existen varios servicios en línea que tienen como ventaja que el usuario no necesita albergar nada en su ordenador. Esto le proporciona ubicuidad –puede usar su información desde cualquier lugar de la Red -y le abre la posibilidad de compartir la información con otros para, por ejemplo, realizar una tarea de manera colaborativa –por ejemplo Google Docs o Zoho- o simplemente conocer las opiniones de otros usuarios.

Permitir y registrar la participación del usuario exige una arquitectura informática que permita extender el dato propio con el valor que el usuario le proporciona. Esta es la clave para recoger el conocimiento del usuario y construir algo más valioso que el propio núcleo de datos.

Los problemas relativos a la propiedad intelectual de los documentos públicos – los privados quedan protegidos adecuadamente- se suele solucionar echando mano de las licencias *abiertas* como Creative Commons (<http://www.creativecommons.org>) que aplican al mundo de los contenidos ciertas filosofías procedentes del mundo del *Open Source* o Software Abierto.

Gestión del conocimiento colectivo

Los usuarios suman conocimiento que puede llegar a ser anónimo y se aprovechan del que ya existe. La gestión del conocimiento adquiere una dimensión nueva cuando hablamos de la Red. Hace unos años, antes de que nadie hablase de la Web 2.0, ya existían aplicaciones que conseguían sumar el conocimiento de un numeroso grupo de usuarios y gestionarlo adecuadamente para dar un servicio a esos mismos usuarios.

Por ejemplo, el sistema de control de SPAM para clientes de correo electrónico CLOUDMARK (<http://cloudmark.com>). Cloudmark se instalaba en el cliente de correo electrónico del usuario y analizaba cada uno de los mensajes que recibía comparándolos con una base de datos para determinar si era correo basura o no. La virtud del sistema estribaba en cómo se construía la base de datos consultada. Si un usuario encontraba un correo electrónico no deseado en su buzón, lo marca como “basura” y el sistema añadía ese dato a la base de datos centralizada. Independientemente de que el sistema recordase que para ese usuario concreto, próximos correos similares se deberían considerar también como “basura”, si el sistema central recibía notificaciones de varios usuarios al respecto, lo añadía a la base de datos principal. Así, otros usuarios se beneficiarían de la acción individual de “limpiar” el correo electrónico propio.

Este ejemplo, intencionadamente extraído de la era pre Web 2.0, anunciaba lo que vendría a continuación: miles de sedes web que tratan de concentrar la actividad de usuarios determinados para extraer de ellos un “conocimiento colectivo” y administrarlo adecuadamente. La excusa para hacerlo bien podría ser el uso de un determinado dato o servicio, como ya se ha dicho anteriormente. Pero en el fondo, lo que se busca es el la gestión de la interacción del usuario con la Web y, a través de esa gestión, encontrar la oportunidad de un modelo de negocio no necesariamente definido en un primer momento.

La conversación que el usuario mantiene con la propia Web añade valor a la misma si se fomenta, se cuida y se gratifica. Esta gratificación no necesariamente económica constituye una auténtica “moneda social” que es enormemente valorada por los usuarios actuales de la Web.

Hoy son muchos los lugares web de éxito que utilizan esta fórmula. Podemos presentar varios ejemplos en diferentes estadios de desarrollo: desde aquellos que ya han formalizado un modelo de negocio que comienza a rendir beneficios y otros que aun no lo han hecho.

Algunos ejemplos

Nombre	Objeto	URL
Del.icio.us tagsAhoy	Almacén público/privado de los hallazgos Web de los usuarios.	http://del.icio.us http://tagsahoy.com
Meneame	Servicio donde los usuarios referencias noticias y éstas son valoradas por los usuarios.	http://meneame.net
Digg	Ídem en Ingles. Original.	http://digg.com
Flickr	Servicio Web para depositar fotografías públicas o privadas. En la actualidad, dispone de de más de 12 millones de fotografías reutilizables con licencias abiertas tipo CC.	http://www.flickr.com
Youtube Metacafé	Servicio Web para albergar vídeos de los usuarios.	http://www.youtube.com http://www.metacafe.com
11870	Tarjetero virtual privado/publico con comentarios y	http://www.11870.com

Salir	opiniones.	http://www.salir.com
Twitter	<i>Blog-rápido para contar públicamente ideas rápidas.</i>	http://www.twitter.com
Booking	Reserva de hoteles y búsquedas en base a criterios comentados por el resto de los usuarios.	http://www.booking.com

La BVMC, debido a su naturaleza, no aspira a establecer una relación con todo tipo de público. Aún considerando ampliar el espectro de usuarios, estos representan un segmento reducido de la población en general. Sin embargo, dos factores juegan a favor: el medio –Internet- permite que la actividad de la biblioteca se extienda en el mundo y aunque el grupo de usuarios buscados sea en términos porcentuales, una fracción de la población, en términos absolutos pueden ser multitud. El otro factor es que el poder atractivo de la literatura permite estrechar una relación intensa que trasciende a la de la mera utilización de un servicio y es de suponer que el nivel de *fidelización* del usuario pueda mantenerse alto.

Sin embargo, esto no garantiza que el usuario de la BVMC vaya a quedarse si no encuentra interés en la relación que establece con el sistema y, sobre todo, si no recibe algo a cambio de su participación.

Programación abierta: la versión Beta perpetua

La beta perpetua

En el mundo de los programadores de ordenadores “beta” es la denominación que se le da a una aplicación informática que se considera suficientemente terminada como para exponerla en público y comenzar a usarla pero que aún es susceptible de cambios que se determinarán tras un periodo de prueba. Los sistemas tradicionales basados en Web esperaban hasta que estuviesen finalizados para ponerse en servicio al público, suponiendo que los defectos y, sobre todo, las limitaciones que pudiesen encontrarse en el sistema arruinarían la relación con los usuarios. Esta característica está heredada de la concepción del servicio de web como una aplicación cerrada que debía funcionar en una computadora con un grado de perfección alto para satisfacción del cliente.

A su vez, esto inducía la necesidad de que el cliente especificase todas y cada una de las características y funciones que la aplicación debía poseer ANTES de que esta fuese construida. Durante la fase de construcción, si el cliente modificaba el criterio, a menudo suponía una enorme fuente de problemas que retrasaban el lanzamiento y producían una relación frustrante entre el programador y el usuario. Pero para las aplicaciones que circulan a través de la Red y en base a la idea de que la plataforma es la Web, esto ya no es más de esta manera.

Ahora el cliente no especifica previamente por la sencilla razón de que “no existe tal cliente” como persona a quien se le puede consultar. Por ello, la dinámica de construcción de una aplicación Web 2.0 comienza en base a una idea que se esboza y que se resuelve informáticamente lo antes posible y de la manera más simple posible. A continuación se lanza a la Red con un grado razonable de terminación pero con numerosas limitaciones y se espera a recibir la retroalimentación de los usuarios para dar prioridad a las tareas siguientes y mejorar el servicio. Y esta dinámica de mejora se mantiene durante la vida útil del servicio.

Recordemos el título de “beta” que Google mantuvo durante años en su buscador de imágenes o el también status de “beta” que Writely mantuvo hasta su compra y fusión dentro de la iniciativa Google Docs. Aún hoy día, este servicio, que ha comenzado a desarrollar su modelo de negocio, está claramente inacabado, lo que mantiene la atención de los usuarios puesta sobre las novedades mientras usan los elementos ya finalizados.

Es de suponer que el mosaico de servicios con que la BVMC debe contar también poseerá esta característica. Manteniendo un perfil de bajo coste de producción de cada uno de los elementos pero explorando todas las posibilidades, se tratará de provocar la posible apertura de nuevas maneras de relacionarse con los usuarios, que permitan a su vez explorar nuevos modelos de negocio.

La arquitectura informática por capas de la BVMC ha de permitir este tipo de modularización de servicios, haciendo independiente el dato y la gestión del mismo del proceso de flujo de trabajo que finalmente el usuario podría desarrollar con él.

La evolución de las bibliotecas digitales

Aunque la BVMC ha sido considerada en numerosos aspectos un ejemplo a seguir por sus buenas prácticas por la Comisión Europea (<http://www.minervaeurope.org/bestpractices/listgoodpract.htm>), el rápido avance de la investigación y el desarrollo en este campo hace imprescindible un

esfuerzo de actualización e innovación constante para mantener esta calificación. Para ello conviene tener en cuenta que las líneas de desarrollo más probable durante los próximos años en el ámbito de las bibliotecas digitales son las siguientes:

Mejora de los sistemas de transcripción de texto

La Comisión Europea ha establecido como objetivo prioritario del Séptimo Programa Marco el desarrollo de nuevos sistemas y procedimientos que permitan la creación de bibliotecas digitales con mayor eficiencia , lo han denominado como *cost effective digitization*. Por ello, financiará varios proyectos de mejora de los sistemas de digitalización y, en particular, del reconocimiento automático de texto OCR (Optical Character Recognition).

Para los proyectos de este tipo, la BVMC constituye un banco de prueba único en sus características, ya que posee la mayor colección de textos digitalizados en español transcritos y supervisados por expertos. Por ello, es recomendable la creación inmediata de alianzas estratégicas con los consorcios que desarrollan estos proyectos.

Creación de herramientas de explotación cooperativas (Web 2.0)

A medio plazo, la explotación de los textos con transcripciones de gran calidad que se irán generando, requerirá el desarrollo de nuevos servicios y herramientas de explotación que todavía no existen.

Conviene por tanto, iniciar un programa de adopción de este tipo de herramientas, desarrollando código abierto para permitir que otros desarrolladores colaboren en su creación y para facilitar la cooperación con otras bibliotecas digitales.

- Creación de nuevos servicios para simplificar el uso de las bibliotecas digitales.
- Técnicas como la aplicación de métodos automáticos de traducción o de resumen de textos permitirán a los usuarios una experiencia más rica en su interacción con la biblioteca.
- Personalización de las bibliotecas digitales.
- Mayor importancia de contenidos no textuales (audio, vídeo)

Dado el carácter eminentemente textual de la BVMC, actualmente debe estudiarse la asociación con otras bibliotecas digitales o instituciones con contenidos audiovisuales.

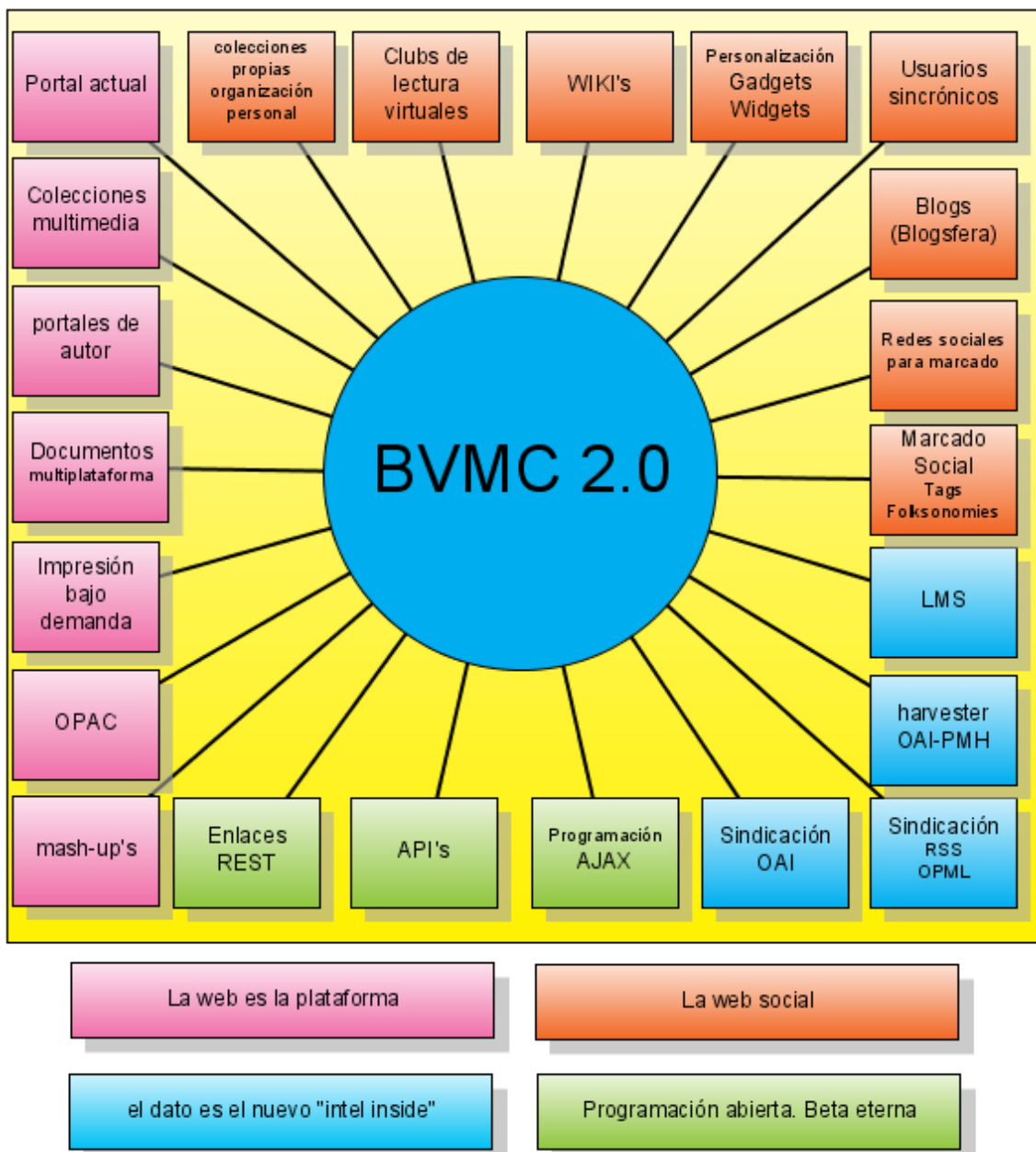
- La interlocución con propietarios de derechos (como ha hecho Google Book Search) para ofrecer modelos de negocio.

- Mejoras de eficiencia en los sistemas de producción.
- Recomendaciones de la Comisión Europea a tener en cuenta:
<http://www.minervaeurope.org/publications/costReduction.htm>

2. Descripción funcional de la BVMC 2.0

Los servicios de la BVMC 2.0

En relación con la idea de que la “Web es la plataforma” surgen unos interesantes elementos algunos de ellos ya presentes en la actual Web de la BVMC.



El portal actual

Con la vista de catálogo por diferentes criterios, accesos por áreas temáticas y secciones de material multimedia, así como las noticias o los accesos a los portales de autor.

Sin embargo, en la propuesta de BVMC 2.0, la página principal no es la única vía de acceso a los distintos componentes de la biblioteca. La indexación efectuada por motores de búsqueda externos hace que una fracción creciente de los usuarios accedan directamente a la página con el contenido o servicio que buscan sin pasar necesariamente por la página principal. En cierto sentido se puede afirmar que todas las paginas de una Web pueden ser una página principal. Ello conduce a una primera idea fundamental: **todas las paginas de una web han de ser analizadas con la minuciosidad que se suele aplicar a la pagina principal**. Pero también **reduce la presión de la página principal** como única oportunidad para capturar la atención del usuario y derivarlo hacia otras áreas del portal.

La web y sus contenidos habrán de estar diseñados pensando en que no solo va a ser consultada por seres humanos a los que fácilmente se les puede dar una explicación sobre el contenido propuesto, sino que en muchas ocasiones, serán robots de búsqueda automatizados los que lean la página correspondiente tratando de determinar cuál es el contenido de la misma, para clasificarla y poder ofrecer resultados de búsqueda mas exactos a los usuarios que acuden a ellos.

Por ello, el contenido de una página web de la biblioteca ha de poseer esta doble orientación: ser útil a los usuarios de la Web para que su uso –explícitamente se menciona *el uso* en vez de *la consulta*, para enfatizar la naturaleza interactiva de

la Web- y a la vez, ofrecer información específica a los buscadores para que entiendan el contenido concreto de esa página.

El diseño gráfico también pierde protagonismo, en el sentido de que ha de estar enfocado hacia el servicio y hacia el usuario y este suele preferir páginas simples, claras a las simplemente bonitas o con gran alarde gráfico; páginas donde sea posible averiguar fácilmente el medio de utilizar el servicio propuesto y donde las explicaciones convengan por su sencillez. Ello exige de un esfuerzo de planificación y diseño gráfico muy poco frecuente.

Sugerencias

- Orientar el diseño hacia los robots de búsqueda tanto como para el usuario.
- Simplificar la comprensión de las páginas mediante estructuras gráficas sencillas.
- Estudiar las páginas internas de manera minuciosa y pormenorizada.

Las colecciones multimedia

La Web actual es multimedia. Los usuarios han asaltado la Web con sus imágenes, sus vídeos, sus grabaciones... Producir material audiovisual es tan simple como instalar una barata *webcam* y apretar el botón de grabación. La exigencia de calidad técnica y narrativa no es alta y esta sustituida por la oportunidad o novedad del material y por la brevedad del mismo. Los medios de comunicación se están haciendo eco de este tipo de actividad y algunos de los

mejores periódicos digitales existentes ofrecen a sus usuarios la posibilidad de compartir las grabaciones que han realizado como parte de los contenidos mediáticos que éstos ofrecen.

La BVMC posee una colección audiovisual de gran valor y de excelente calidad técnica. En torno a estas colecciones podrían ubicarse las aportaciones de los usuarios, que podrían ser sus propias lecturas en voz alta de obras expuestas o producciones propias relacionadas de alguna manera con los elementos culturales presentados en la página. Por otro lado, el uso del material audiovisual debería facilitarse en los términos que otros servidores audiovisuales lo hacen. Una de las claves de Youtube (<http://www.youtube.com>) es la facilidad con la que cualquier usuario puede incrustar los vídeos que allí se encuentran en otras páginas web, como pueden ser sus propios *weblogs*. Esto permitiría la difusión de los contenidos más allá de las propias páginas del portal. No se pretende cerrar con esta breve enumeración las posibilidades pues serán los propios usuarios quienes las descubran y orienten.

Sugerencias:

- Hacer participar a los usuarios con sus propias colecciones.
- Sindicar (RSS, OPML, etc) los materiales audiovisuales de manera simple.
- Elaborar medios transversales de hiperenlazar el material audiovisual con los demás elementos de contenido y servicio del portal de manera parecida al formato de youtube o similares.
- Asistir a los usuarios en el manejo de información audiovisual.

Los portales de autor

Uno de los elementos de más acierto de la BVMC ha sido, sin duda, la presencia de los llamados “portales de autor” o “portales temáticos”. Éstos presentan una vista adornada del catálogo interno de la biblioteca siguiendo un criterio concreto de tema o autor. Frente a la aridez de consulta de un catálogo ordenado alfabéticamente o la frialdad de un buscador, un portal de autor permite tratar un tema con el esmero y cuidado.

La aplicación de los criterios de la Biblioteca 2.0 implicará hacer partícipe al usuario del crecimiento del portal. Su opinión, su colaboración, su contraste informativo es contenido muy valioso para que un portal de autor no se quede sólo en la publicación ordenada de lo que la BVMC sabe sobre un autor o tema. Las colaboraciones de los usuarios mantendrían viva la actividad del portal y una observación estrecha de esa actividad daría como resultado una medida del éxito de la iniciativa cara a invertir más esfuerzos en consolidarla.

Sugerencias:

- Permitir la colaboración de los usuarios más allá de la inclusión de un foro temático: crear una wiki específica por cada tema.
- Permitir la inclusión de material de los usuarios que pueda ser contrastado y valorado por los mismos para ampliar el contenido informativo del portal.

Documentos multiplataforma

La Web 2.0 también es multiplataforma. Son ya numerosos los dispositivos que permiten acceder a contenidos especialmente formateados para ellos. Casos específicos son los relacionados con reproductores de mp3 y mp4 como la iPod de Apple, la consola portátil con navegador PlayStation Portable (PSP) o la consola de salón de Nintendo Wii, que se conecta inalámbricamente a la Red de casa y ofrece canales informativos vía Internet así como uso de la Web mediante Opera, un navegador de dominio público.

También conviene citar la existencia de numerosos dispositivos de telefonía móvil con capacidad multimedia que permiten reproducir vídeos y para los cuales algunas productoras ya han comenzado a elaborar material específico. Mobuzz.tv (<http://mobuzz.tv>) ofrece como un canal de televisión por Internet multiplataforma o KYTE (<http://www.kyte.tv>) permite al usuario crear un canal de televisión aprovechando su Webcam o su teléfono móvil.

Si bien parece que ninguno de los dispositivos mencionados es muy apropiado para la lectura de un documento extenso, no debemos olvidar que uno de los principios que hemos desarrollado en líneas anteriores suponía la reelaboración de materiales existentes y la no restricción a documentos escritos.

Sugerencias:

- Separar el contenido de la presentación a fin de aumentar el rango de posibles formatos de presentación del contenido –cualquiera que sea éste– que se encuentre en la biblioteca.

Impresión bajo demanda

Si bien está relacionado con el punto anterior, presenta unas características específicas que lo hacen singular, tenemos el caso de un formato de publicación de documentos que es a la vez tradicional y novedoso: la impresión de documentos bajo demanda.

No escapa a nadie que muchos los documentos albergados hasta la fecha por la BVMC y los que se espera que pueda incorporar en un futuro tienen su medio natural de consulta en el formato impreso y, hasta que la tecnología encuentre un sustituto ergonómico para éste, el libro sigue siendo un vehículo de transmisión de la cultura cómodo, barato y simple de usar.

La modularización en capas con funciones diferentes del modelo informático de la BVMC permitirá el acceso directo al contenido para una reproducción fiel al original. En algunas situaciones, si éste posee las correspondientes marcas de formato, la reproducción puede variarse cara a una edición comentada o similar.

El problema fundamental reside en que los servicios de impresión de documentación para tiradas cortas o para ejemplares sueltos precisan en la actualidad de formatos específicos para realizar su tarea. Por ejemplo, este es el caso de QOOP (<http://qoop.com>), constituido en un estándar de facto para la realización de impresiones de todo tipo con bajo coste. QOOP posee servicios “Business to Business” (B2B) para que aquellas webs interesadas en proporcionar servicios de publicación a sus clientes puedan establecer una

pasarela directa al servicio correspondiente de QOOP y lanzar una impresión de un determinado documento. Este es el caso del proyecto de la Rice University (<http://www.rice.edu>) denominado “CONEXIONS” (<http://www.cnx.org>) . Este proyecto consiste en un repositorio de recursos educativos de libre utilización (contenidos abiertos) que proporciona herramientas para que los docentes pueden editar sus materiales de aprendizaje y que utiliza técnicas de marcado en XML para poder separar claramente la presentación del contenido de cada documento. Así, es posible disponer de varias instancias en diferentes idiomas para un contenido determinado y asignarles el mismo aspecto gráfico, lo que incluye grabados, formulaciones matemáticas, imágenes, etc. El resultado final es la posibilidad de disponer de ediciones diferentes de un mismo documento en varios idiomas. Como servicio de extrema utilidad para los docentes, Connexions ofrece una pasarela directa con QOOP para que a petición del usuario y, asumiendo este el coste de hacerlo, le sea transmitido a QOOP el documento en las condiciones tecnológicas que ha marcado este último servicio y pueda generarse una impresión cuya tirada puede ir desde un ejemplar a varios cientos. Existen otros servicios similares como Lulu (<http://www.lulu.com>), Llumina (<http://www.llumina.com>) o Exlibris (<http://www2.xlibris.com>).

La BVMC tendrá la capacidad de concentrar la atención de los usuarios del mundo universitario que desean realizar una impresión bajo demanda de tirada corta de alguno de los documentos que albergue su catálogo aunque también actuará como servicio atractivo para aquellos interesados en realizar una impresión de un documento elaborado artesanalmente y que podría pasar a pertenecer a los fondos del propio catálogo de la BVMC.

Sugerencias:

- Crear servicios B2B para conectar los fondos de la BVMC con servicios de impresión bajo demanda y ofrecer a los usuarios la posibilidad de realizar impresiones de manera sencilla.

OPAC

OPAC (Online Public Access Catalog) son aquellos sistemas que proporcionan acceso público a un catálogo y que con la llegada de Internet suelen utilizar la arquitectura Web para llegar a sus usuarios. Aparentemente, un acceso a un catálogo o OPAC posee una estructura tan simple que es difícil imaginar variantes del mismo que añadan algo a lo que lleva haciéndose desde el principio de la digitalización de las bibliotecas. Sin embargo, las reflexiones producto de la aparición de la Web 2.0, centradas en el usuario más que en la aplicación, atacan un aspecto fundamental de cualquier OPAC: el interfaz de usuario.

Un OPAC tradicional se concibe como una manera de proporcionar al usuario acceso al dato que contiene la biblioteca. Para ello, si se dispone de una buena estructura catalográfica, se le pide al usuario que elija un criterio de búsqueda sobre cualquiera de las categorías que la componen y se realiza una búsqueda tratando de igualar criterio y resultado. Sin embargo si lo enfocamos bajo la filosofía 2.0 necesitamos reflexionar más acerca de las auténticas necesidades de los usuarios que están buscando algo y mejorar el interfaz de búsqueda para que pueda definir mejor lo que necesita. Esta expresión se traduce –sin que el usuario necesite intervenir- en las correspondientes órdenes de búsqueda que producirán los correspondientes resultados.

Las técnicas de procesamiento de lenguaje natural son tecnologías que vienen a ayudar a los usuarios a expresar sus deseos de búsqueda en términos de lenguaje humano y que un interfaz desarrollado con tal fin se ocupa de traducir en criterios de búsqueda.

Pero no sólo existe esa aproximación: pensando en usuarios concretos, han surgido nuevas maneras de concebir los interfaces de un OPAC para que esos

usuarios concretos encuentren el entorno y la ayuda necesaria para que su búsqueda sea exitosa.

Por ejemplo en ChildrenLibrary (<http://www.childrenlibrary.org/icdl/SimpleSearchCategory>) o Yakpack (<http://test.indexdata.com/yakpac/>) encontramos casos de OPAC orientado específicamente para niños.

Una mirada a estos ejemplos nos da una idea del aspecto que una biblioteca para niños podría tener. Veremos que se articulan entorno a un sistema de búsquedas que en el caso de ChildrenLibrary pueden llegar a ser muy pintorescas –poco ortodoxas desde el punto de vista académico- tales como “libros grandes” o “personajes de animales verdaderos” que, sin embargo, parecen aproximarse más a la habitual *serendipia* que un escolar puede realizar sobre una biblioteca.

Sugerencias:

- Construir vistas de catálogo para públicos objetivos –empezando por el público escolar y de enseñanzas medias- que hagan amigables las exploraciones sobre la biblioteca.

Los sitios *mush up*

Un *mush up*, es un sitio web o aplicación web que usa contenidos de varias fuentes para crear un nuevo servicio. Los *mush up* han cambiado el panorama de Internet al permitir que la creatividad sobre el uso de ciertos conjuntos de

contenidos no se limite solo a los poseedores del mismo. Como hemos enunciado antes, el dato entregado a la colectividad permite que otros usuarios descubran nuevos usos para el mismo. La manera de organizar esos datos disponibles y combinarlos con otros para construir nuevos servicios es lo se denomina *mush up*.

Existen dos orientaciones fundamentales entorno a los *mush up*: usar el concepto para combinar datos de otros o estimularlos produciendo datos para que otros hagan sus *mush up*.

Los mashups están teniendo mucho éxito en Internet debido a que la programación necesaria para realizarlos no suele ser de alto nivel. El proveedor del dato proporciona un interfaz interno simple para que solicitar un elemento y recogerlo no implique modificaciones en la programación del proveedor y que sea esencialmente simple (arquitecturas REST y SOAP). Ello redundando directamente en que los ciclos de desarrollo del *software* sean cortos y baratos, por lo que cuesta poco esfuerzo poner una idea de *mush up* en el mundo de Internet y ver el resultado.

Un caso cercano y que está teniendo cierto eco mediático es el de Panoramio. Panoramio es una Web que permite ubicar una fotografía en un mapa de Google. (<http://www.panoramio.com>) Esta aplicación hace un buen uso de las API de Google Maps para permitir que el usuario encuentre la localización de una toma fotográfica y la posición en el mapa públicamente y que cualquiera escriba comentarios sobre ella.⁴

Otro caso, más cercano temáticamente, es el de la “geografías literarias” de la Biblioteca Municipal de Muszik.

⁴ Panoramio ha sido comprado por Google Inc, por lo que la integración de este mashup en Google Earth y Google Maps es ahora total.

(<http://www.muskiz.com/b2/mapalit.html>) En este caso, pasajes de obras literarias que referencian algún lugar geográfico real, se enlazan con un mapa procedente de Google Maps y muestran el punto exacto señalado. Inversamente, se puede navegar por un mapa buscando lugares referenciados en alguna obra literaria y ojear el libro y pasaje donde se hace.

La arquitectura informática de la biblioteca nos permite desarrollar ciertos servicios internos a modo de *mush up*, con lo que el tiempo de desarrollo para poner a prueba nuevas ideas debería ser corto.

La otra propuesta de uso de los *mush up* es la de permitir que otros utilicen los datos propios para realizar estas re-mezclas. El que cualquier otro portal realice un trabajo sobre un dato procedente de la BVMC no implicaría de ningún modo el establecimiento de una relación formal entre ambos portales o instituciones. A lo más que se podría llegar es a tratar de censarlas para darles correcta difusión, pero no es descartable mantener una absoluta neutralidad como la que mantiene Google con sus “remezcladores” (¡Claro está, hasta que llega el momento de comprarlos!).

Sugerencias

- Aprovechar la arquitectura de capas y apis propuestas para la BVMC a fin de contruir mashups propios que incluso integren datos externos.
- Permitir el “remezclado” de los datos de la BVMC por parte de terceros y mantener un alto nivel de observación de esas iniciativas. Para ello, construir las API apropiadas y documentarlas y hacerlas públicas de la manera más eficaz posible.

Colecciones de usuarios

Para los académicos, resulta de extremada utilidad la utilización de herramientas que simplifiquen la gestión de información bibliográfica y permitan citar adecuadamente las referencias científicas en sus propios trabajos.

Existen herramientas⁵ comerciales de gran calidad que se pueden instalar en los ordenadores propios, como EndNote (<http://www.endnote.com>), Procite (<http://www.procite.com/>), Referente Manager (<http://www.refman.com/>), o incluso alternativas dentro del ámbito del Software Libre, como Bibus (<http://bibus-biblio.sourceforge.net/>) o Jabref (<http://jabref.sourceforge.net/>). Estas aplicaciones se instalan en el ordenador propio y en general, son capaces de importar y exportar referencias bibliográficas en múltiples formatos.

Pero gracias a Internet, han surgido algunas aplicaciones *online* que son capaces de realizar esta tarea sin precisar de la instalación de ningún *software* en el ordenador propio, e incluso, dada la naturaleza de la aplicación web, permiten aprovechar la citación realizada por otros para completar la propia. Así se consigue el aprovechamiento de la inteligencia colectiva.

De las aplicaciones *online*, destacaríamos la característica principal de hacer públicas las propias referencias para que otros puedan consultarlas e incluso añadirlas a las propias. Algunas de estas aplicaciones son masivamente utilizadas, como CiteUlike (<http://www.citeulike.org>), que permite acceder desde

⁵ En http://en.wikipedia.org/wiki/Comparison_of_reference_management_software se puede encontrar una interesante comparativa de este tipo de herramientas, con una descripción bastante completa de cada una de ellas elaborada por personas independientes de los autores o fabricantes.

el primer momento al conocimiento generado por una masa crítica de usuarios que hace muy útil su uso. Además, también existen herramientas de software libre que permiten ofrecer este tipo de servicio a los propios usuarios, como Wikindx (<http://wikindx.sourceforge.net>).

Además de los usos más académicos, la BVMC podría proporcionar a los usuarios no especializados una manera de organizar sus propias colecciones privadas de libros y documentos. Se trataría de asistir a los usuarios en el proceso de catalogación, completando la información que el usuario necesita con la que puede residir en la base de datos catalográfica de la BVMC. Además de ayudarle a mantener su colección de libros ordenada, el sistema proporcionaría una excusa perfecta para hacer llegar al usuario novedades relacionadas con sus gustos literarios o culturales o para ponerlo en relación con otros integrantes de una comunidad de intereses parecidos.

Sugerencias:

- Explorar la posibilidad de instalar herramientas de citación bibliográfica en la BVMC y adaptarlas para aprovechar la base de datos catalográfica que se posee a fin de tomar como partida una gran base de datos que sustituya a la masa crítica de usuarios precisa.
- Crear un área de colecciones privadas donde los usuarios pudiesen aprovechar el potencial tecnológico de la BVMC para mantener sus propias colecciones de libros, ya sean digitales o no, respetando la privacidad sobre aquellos elementos que lo precisen (digitalizaciones no publicables) compartiendo la información catalográfica y comentarios que puedan ser susceptibles de interés para otros y relacionando sus colecciones privadas con otros focos de interés, como pueden ser novedades editoriales, culturales, servicios de impresión bajo demanda o servicios de venta de entradas.

Clubs de lectura

Una costumbre que existe en ciertos ámbitos culturales es la de pertenecer a un club de lectura.

Un club de lectura es un grupo de personas que leen al mismo tiempo un determinado libro, cada uno a su ritmo, pero con el suficiente paralelismo como para poder comentar simultáneamente las páginas o capítulos leídos. Los clubs de lectura son elementos dinamizadores de muchas bibliotecas físicas e incluso el Ministerio de Cultura (<http://travesia.mcu.es/receta.asp>) recomienda y asiste en el uso de este recurso para la creación de una comunidad cultural entorno a las bibliotecas.

Una biblioteca digital también necesita dinamizar a sus usuarios y puede aprovechar que los usuarios pueden estar ubicados en cualquier sitio del mundo para extender las redes culturales a lugares remotos, encontrar suficientes lectores como para formar el grupo necesario y crear microportales temáticos sobre contenidos del catálogo propio.

Un club de lectura “virtual” tendría acceso a los documentos digitalizados que se proponen como lectura –aunque podría no restringirse exclusivamente a material digitalizado -lo que podría ser visto también como una oportunidad comercial en relación con “primera vista”- y proporcionaría herramientas de comunicación y trabajo colaborativo como:

- Salas de chat con posibilidades multimedia.

- Foros.
- Podcasting.

Sugerencias:

- Crear una zona de Club de Lectura con herramientas colaborativas. Dinamizar la lectura de algún libro o documento de la biblioteca digital con voluntarios.
- Permitir que otras bibliotecas utilicen el área de “club de lectura virtual” para conducir sus propios clubs.

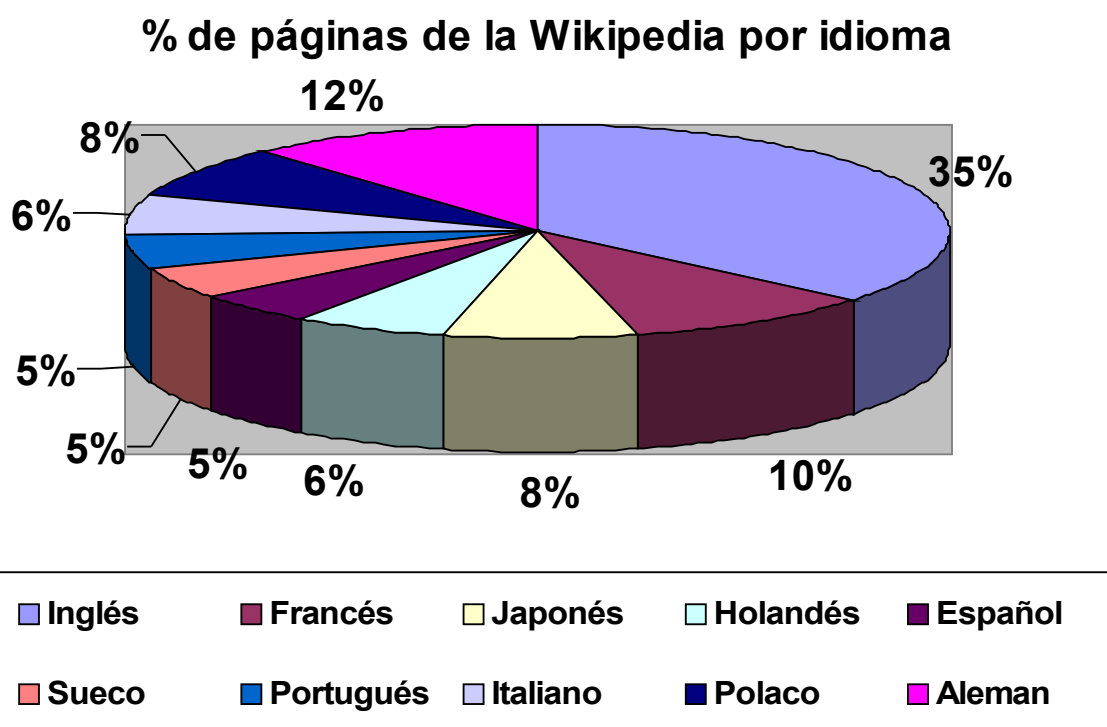
Wiki

Como define la Wikipedia, el sitio más popular que utiliza este sistema, “Un (o una) wiki (del hawaiano wiki wiki, «rápido») es un sitio web colaborativo que puede ser editado por varios usuarios. Los usuarios de una wiki pueden así crear, editar, borrar o modificar el contenido de una página web, de una forma interactiva, fácil y rápida; dichas facilidades hacen de una wiki una herramienta efectiva para la escritura colaborativa”.

Dicha Wikipedia alberga ya más de 4.090.000 términos sólo en los diez idiomas más utilizados. Cada uno de esos términos ha sido desarrollado por un autor de manera no retribuida, cuyo trabajo ha sido revisado –de hecho lo sigue siendo siempre- por cualquiera de los lectores de la wikipedia que se considere apto para corregir un contenido determinado.

En la aparente anarquía de creación de contenidos de la Wikipedia, sin embargo reina una etiqueta exquisita que impide que los autores usen el medio para

beneficio propio, auto citándose o depositando contenidos copiados. La Red y la comunidad de usuarios de la wikipedia se autorregula y corrige rápidamente los errores.



La wikipedia ya incluye miles de páginas de autores de literatura clásica iberoamericana. Se hace necesaria una comparación de los contenidos actuales de la biblioteca con los de la propia Wikipedia y parte del posible trabajo de los redactores del portal podría ser la revisión de los contenidos importantes de la Wikipedia a fin de que a la BVMC se le reconozca el valor académico que posee. Por otro lado, los contenidos de la Wikipedia son de libre utilización y también

podrían formar parte, a modo de extensión, de los contenidos propios que la BVMC alberga.

Por otro lado, el formato wiki es muy interesante para que los usuarios realicen sus propias aportaciones. La *buscabilidad* de las páginas de una wiki es extraordinaria por la simpleza del diseño y la estructuración del contenido. Por tanto, permitir una vista de los fondos de la biblioteca bajo este formato y permitir a los usuarios que extiendan los contenidos propios con sus colaboraciones Redundaría en beneficio de la biblioteca.

Sugerencias:

- Revisar los contenidos de la wikipedia con la óptica de la BVMC a fin de realizar aportaciones significativas que ubiquen a la BVMC como proveedor de contenidos de rigor académico y de excelencia.
- Seleccionar y volcar parte del fondo de contenidos de la BVMC (incluyendo multimedia) en formato wiki para:
 - aumentar la buscabilidad de ciertos términos
 - extender los contenidos propios con colaboraciones

Personalización

La Web 2.0 también se define como una web personalizable. El usuario ya no es un espectador pasivo de contenidos que van desfilando de página en página sino que proactivamente busca y selecciona lo que quiere consultar y utilizar.

Los grandes portales se han dado cuenta de ello y han propuesto mecanismos originales que vienen a sustituir a la antigua “página de inicio”. La propuesta de Netvibes (<http://www.netvibes.com>) consiste en proporcionar una página de inicio totalmente configurable a base de añadir secciones (llamadas canales) que enlazan con fuentes de sindicación de contenidos de otras páginas, pequeños dispositivos para mostrar informaciones varias, como el tiempo o calendarios o con los servicios más utilizados por los usuarios, como sus cuentas de correo, búsquedas, etc.

Google y Yahoo (<http://www.Google.com> y <http://www.Yahoo.com>) también han seguido esta tendencia y proponen sendas páginas de inicio “i-Google” o “my Yahoo!”. En estos casos, las mencionadas firmas también proporcionan información técnica para que cualquier proveedor de contenidos pueda crear canales específicos o dispositivos (llamados *gadgets* o *widjets*). Con ellos, el usuario puede configurar su página web de iGoogle o My Yahoo para tener información directa de estas y otras fuentes cada vez que se abre el navegador.

Sugerencias:

- Seleccionar aquellas fuentes de interés que sean susceptibles de figurar como canales en las páginas Web iniciales de los usuarios y construir los dispositivos apropiados para que figuren en los portales de mayor presencia.

Blogosfera

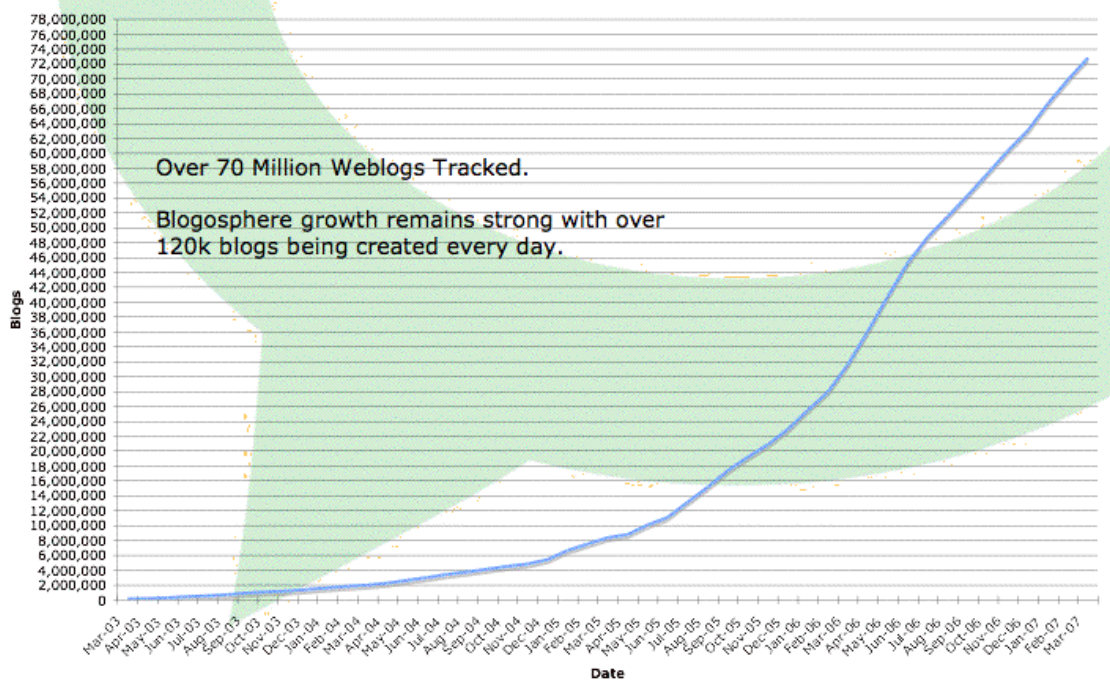
Uno de los fenómenos de la nueva Web que más impacto ha causado es la aparición del fenómeno de *weblog*. El *weblog*, más popular en su diminutivo *blog*, es una página de muy fácil composición que los usuarios usan a modo de diario personal.

La composición de este tipo de páginas web es muy sencilla y no requiere de ningún conocimiento técnico para que los usuarios realicen sus aportaciones y vayan construyendo sus relatos de descubrimientos, impresiones, opiniones, etc.

Desde su aparición, los *blogs* han ido creciendo de forma exponencial y hoy son más de 72 millones (controlados por Technorati). Al día, se crean 120.000 *blogs* y debido a su naturaleza (usuarios sin control y organización mediática) se han convertido en una de las principales fuentes de contenidos de todo tipo de la Web a la que los buscadores prestan especial atención. Es habitual que el resultado de una búsqueda realizada en cualquiera de los servicios habituales (Google, Yahoo, MSN...) incluya entre los primeros resultados a varios *blogs*.

Los *blogs* han supuesto el acceso a la Web por parte de los usuarios de manera masiva. Hay *blogs* sobre cualquier tema imaginable y existe un tejido de relaciones entre ellos que ha desarrollado vida propia y hace la competencia a los medios tradicionales de comunicación (gráfico 2).

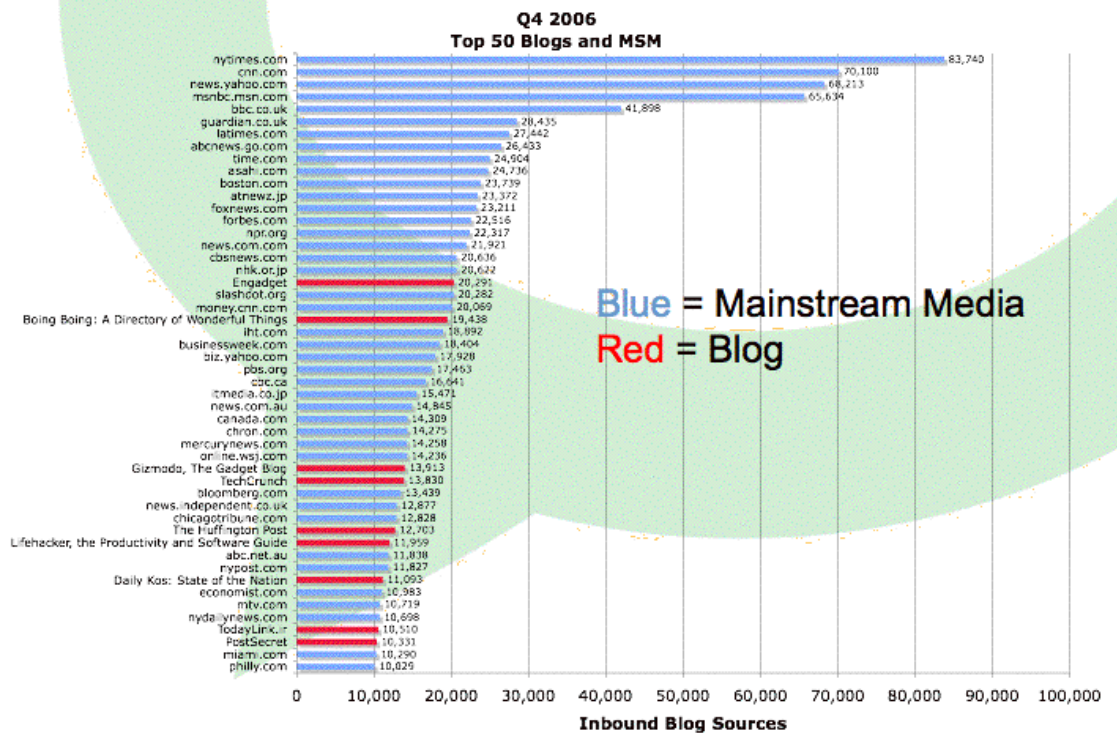
Weblogs Cumulative: March 2003 - March 2007



Este sistema de relaciones entre *blogs* ha recibido el nombre de *blogsfera*, y constituye el campo de batalla mediático en el que muchos medios tradicionales se están jugando parte de su futuro.

Es frecuente que, entre los acreditados a los grandes eventos políticos ya se incluya a varios *bloggers*, puesto que las agencias de imagen de los candidatos ya perciben el poder de creación de opinión que tienen estos medios⁶.

⁶ El Mundo (2/08/2004) da cuenta de la primera vez que esto ocurre.
<http://elmundo.es/navegante/2004/07/29/esociedad/1091088812.html>



Sin embargo, la posición del español en la *blogsfera* no es importante. El español es el quinto idioma en % por detrás del inglés, el japonés, el chino y el italiano.

Las razones para ello pueden ser varias, pero podríamos mencionar:

- El grado de madurez tecnológica de la sociedad donde reside el internauta

- El desconocimiento del impacto de las propias opiniones
- Llegada tardía al ámbito cultural iberoamericano de las herramientas tecnológicas necesarias

Sin embargo, desde nuestro punto de vista, esto se transforma en una oportunidad para que los contenidos de la biblioteca y su plataforma tecnológica sirva de polo de atracción para un sector de la *blogsfera* y que la BVMC reclame parte del espacio que le podría corresponder

Sugerencias:

- Instalar una plataforma para crear y albergar *blogs*. Estos pueden ser de los agentes dinamizadores de la biblioteca (personas de cierto prestigio cultural a los que se les anima a colaborar con los contenidos de la BVMC) o público en general que desea abrir una cuenta bajo el paras de la BVMC.
- Mantener uno o varios *blogs* como sección dinámica de cada una de las áreas del portal. Esto permite establecer relaciones con el resto de la *blogsfera*
- Usar los contenidos de los blogs de los usuarios como extensiones firmadas de terceros sobre los contenidos propios.

Redes sociales y *mercado social*

Muchas webs han desarrollado servicios que permiten que los usuarios clasifiquen sus contenidos utilizando métodos diversos:

- Un sistema de valoración del interés que un determinado contenido que tiene para ellos.
- Un sistema de registro de palabras clave que el usuario asigna al contenido y que le permite encontrar lo que busca con posterioridad. (*tag*). Este registro de claves, se puede transformar en una auténtica clasificación o *folksonomy* de los contenidos de un portal.

Ambos métodos tienen una vertiente pública: conocer la valoración que el resto de los usuarios hace de un determinado contenido puede ayudar a un usuario concreto a encontrar lo que necesita. Por otro lado, a la hora de elegir las palabras clave para definir un contenido, también puede ser útil conocer las que otros usuarios han escogido.

De cualquier forma, la aportación individual, ya sea en forma de valoración o de palabra clave, se suma para clasificar los contenidos sin la intervención de agentes externos: los contenidos adquieren así una clasificación de orden práctica que en algunas ocasiones resulta más útil que las realizadas por expertos.

La valoración de los usuarios permite establecer relaciones entre los contenidos de un portal al poder mostrar elementos similares siempre según criterios de los propios usuarios.

Una manera de mostrar este tipo de contenidos es la llamada “nube de tags” que no es otra cosa que una manera de visualizar una clasificación haciendo que los términos más frecuentes aparezcan más remarcados – por color o tamaño- que los otros.



El mercado social permite agregar la aportación individual y crear un conocimiento colectivo que es posible gestionar y administrar y convertir en servicio de utilidad para los usuarios.

Sugerencias:

- Añadir un sistema de marcado a base de *tags* de los contenidos del portal.
- Añadir a los contenidos del portal la posibilidad de enlazar directamente con contenidos marcados de manera semejante por otros usuarios.

LMS

El valor educativo que posee la BVMC solo podrá desarrollarse si es posible para docentes y posibles alumnos encontrar, acceder y reutilizar los materiales que la biblioteca proporciona. Los recursos que la biblioteca provee son, a menudo, paquetes de contenido cerrado que versan sobre un tema y que pueden incorporar o no una intencionalidad educativa. La producción de los recursos más directamente aplicables a la educación supone una inversión en tiempo y esfuerzo importante que limita su desarrollo en grandes cantidades. Sin embargo, es posible automatizar ciertos elementos de la producción para hacer posible la reutilización de los recursos de la biblioteca bajo ciertas condiciones.

En concreto, la mayoría de los LMS o campus virtuales de los centros educativos –tanto universitarios como de otros niveles de la enseñanza- suelen incorporar la característica de importar materiales educativos en formato estándar. Por tal, se entiende uno de los formatos que se utilizan en el mundo de la educación para etiquetar y empaquetar recursos educativos, como son el IMS o el SCORM.

Sin entrar en detalles acerca de los estándares educativos, es de señalar que la biblioteca podría proporcionar una buena parte de su catálogo en este tipo de formatos, simultáneamente a la presentación actual del mismo como páginas web o como archivos PDF.

En el mundo ya existen varios repositorios de recursos de aprendizaje que incluyen grandes volúmenes de recursos educativos que pueden ser utilizados por cualquier docente⁷, pero uno de estos repositorios tiene especial interés por cuanto los proyectos guardan algunas similitudes en cuanto al objetivo cultural y

⁷ CAREO (<http://careo.netera.ca/>) , MERLOT (<http://careo.netera.ca/>)

geográfico pretendido: nos referimos a la Biblioteca de Recursos de Aprendizaje de Universia⁸, que en estos momentos alberga ya más de 3.000.000 de referencias a recursos utilizables educativamente y procedentes de varias decenas de colecciones de materiales científicos y docentes de todo el mundo.

Los fondos de la BVMC podrían formar parte de esta colección, ocupando un lugar especial en ella, ya que son especialmente referidos al ámbito iberoamericano y su utilidad educativa como materiales docentes para todos los niveles de enseñanza de nuestro ámbito cultural es incuestionable.

Sugerencias:

- Proporcionar una vista del catálogo de los fondos de la biblioteca como recursos empaquetados bajo estándares educativos tipo IMS o SCORM.
- Desarrollar un área educativa de la biblioteca donde sean presentados los recursos educativos y se hagan propuestas concretas sobre su utilización.
- Abrir una puerta de sindicación OAI-PMH que permita la recolección del catálogo por parte de la Biblioteca Virtual Miguel de Cervantes.

⁸ <http://biblioteca.universia.net>

Sindicación

La *sindicación* de contenidos es una de las más importantes claves para entender la dinámica de la Web actual.

La *sindicación* de contenidos hace referencia a los mecanismos tecnológicos que se establecen para que los contenidos de un portal web puedan ser transmitidos a otros portales o a otros usuarios sin que los aspectos de su presentación gráfica interfieran: se trata de la transmisión del contenido “puro” y de información relacionada con él (metainformación) que cuando es recibida por el usuario final o por otro servicio informático, es procesado, reformateado y preparado para su visualización.

La sindicación de contenidos ha permitido, por ejemplo, que la información que los *blogger* escriben en los millones de *blogs* existentes se disemine rápidamente ya que cuando alguien encuentra un *blog* de su interés, es posible suscribirse a recibir sus contenidos directamente cada vez que haya novedades directamente en el escritorio del propio ordenador o en una aplicación propia que permite coleccionar todas las suscripciones (*agregador*) Así no es necesario visitar un rosario de páginas para ver si hay novedades en ellas: los contenidos viajan rápida y oportunamente hasta el ordenador del usuario.

También es posible establecer otro nivel de sindicación de contenidos entre servicios similares, como por ejemplo bibliotecas digitales de ámbitos diferentes pero que desean compartir parte de su catálogo. En esta ocasión, la sindicación de contenidos permite que la información catalográfica de una biblioteca pueda ser utilizada por otra y viceversa. Los usuarios de la Web actual acostumbran a encontrar los elementos sindicables de un portal mediante la búsqueda de unos

iconos específicos –aunque no haya normas estilísticas para ellos- que les indican que allí se encuentra una fuente de información o *feed*.



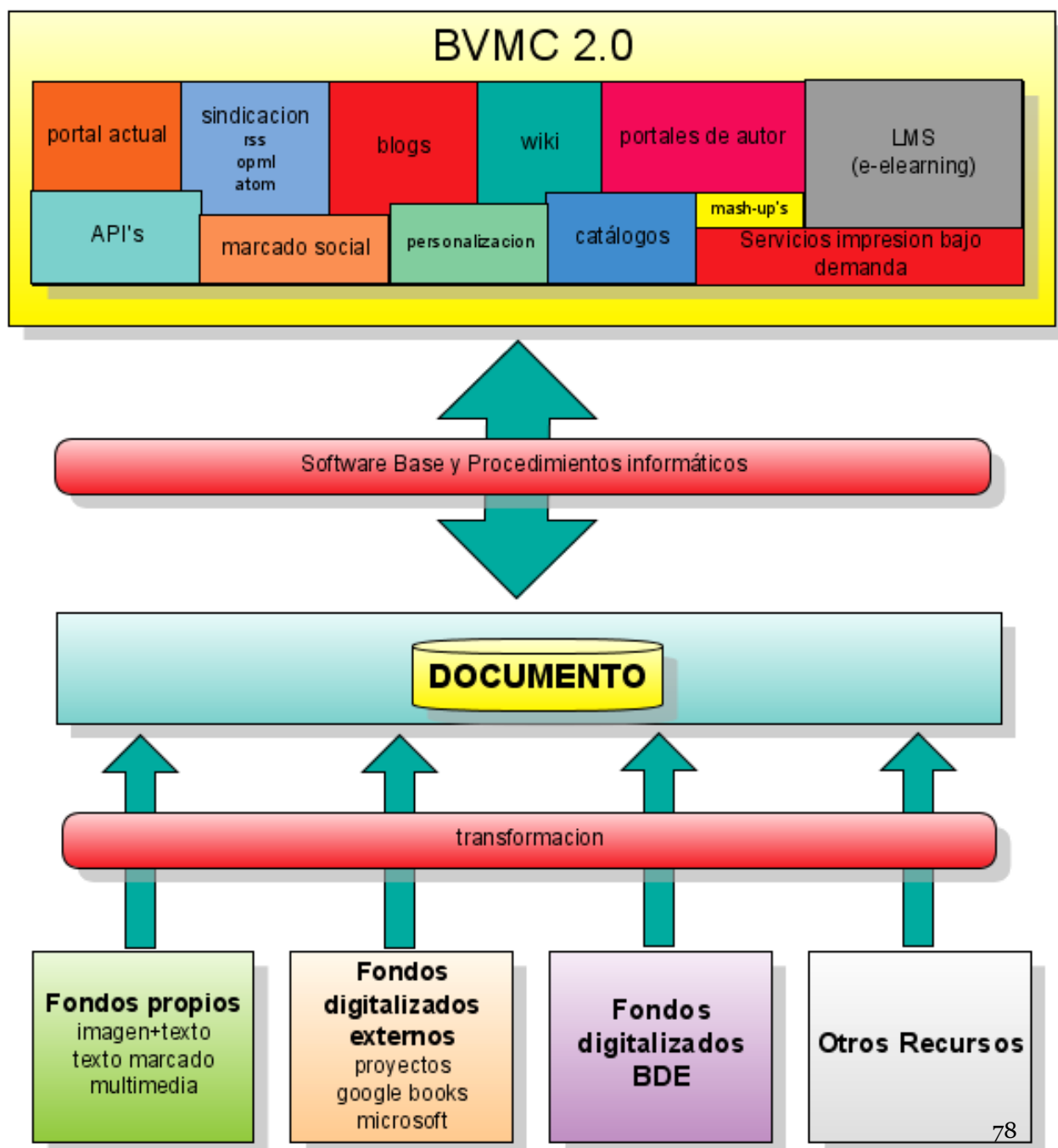
En la Web actual, y debido a la proliferación de herramientas para agregar las distintas fuentes de sindicación de contenidos que los usuarios van encontrando, ya podemos encontrar una auténtica colección de iconos que permiten e incluso incitan al usuario a que se suscriba dependiendo de sus preferencias de herramientas:



En cualquier caso, la sindicación de contenidos es una de las maneras más eficaces de mantener a los usuarios fidelizados a un servicio de información dinámico que es en lo que se han convertidos los portales hoy en día.

Sugerencias:

- Identificar las fuentes de contenidos sindicables (blogs, noticias, etc) y ofrecer a los usuarios los medios para que estos se suscriban a estas fuentes
- Desarrollar un buscador interno orientado a la sindicación del resultado de búsqueda que permita suscribirse a cualquier novedad temática elegida bajo criterios determinados por los usuarios en sus búsquedas.



Modelo conceptual

Para que las funcionalidades expuestas puedan llevarse a cabo, es necesario que el modelo informático de la BVMC posea una serie de características.

Estas, esencialmente serian las derivadas de una arquitectura en capas, separando las relativas a los contenidos y a la adquisición de los mismos de las dedicadas a gestionar internamente la información o a presentarla.

El modelo conceptual siguiente pretende esbozar los elementos de la arquitectura informática que será presentada más adelante.



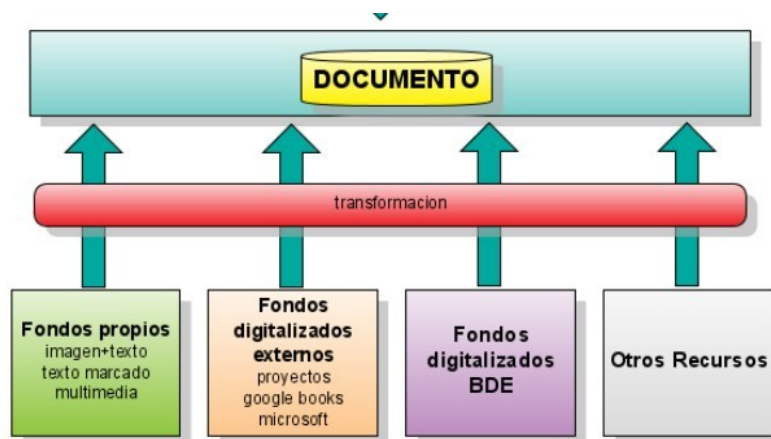
Análisis del modelo

Las entradas de contenido de la biblioteca son:

- Sus propios fondos.
- Fondos procedentes de otros proyectos de digitalización existentes y que posean API's para poder manejarlos.
- Fondos procedentes de los proyectos de digitalización entorno a la Biblioteca Digital Europea.

- fondos digitalizados externamente pero recolectados mediante protocolos como el OAI-PMH o mediante z3250.

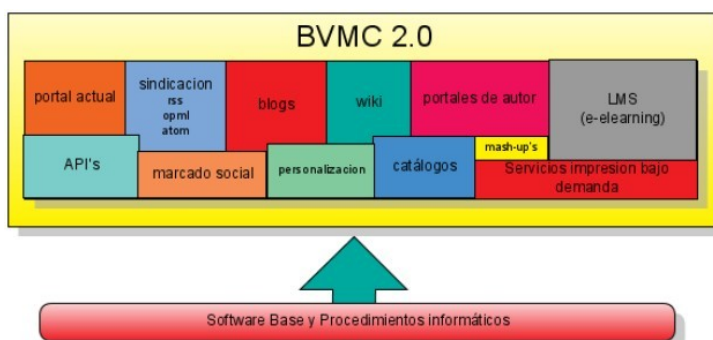
Estos fondos son procesados para integrarse de manera homogénea en la biblioteca, procesando tanto los posibles formatos en los que residen como los datos catalográficos y metadatos que los componen. De esta manera, todas las aportaciones pueden ser tratadas de igual manera por la mecánica de la biblioteca, respetando, claro está, las consideraciones relativas al grado de elaboración del contenido, asuntos de propiedad intelectual o de disponibilidad.



La BVMC poseería una operativa de bajo nivel que le permitiría disponer de un gran número de operaciones sobre los datos disponibles a petición de los programadores. Esta característica se orienta hacia la idea de que los tiempos de desarrollo de las funcionalidades externas de la biblioteca puedan ser más reducidos.



Gracias a los procedimientos informáticos pre-establecidos, la biblioteca puede contar con un mosaico de opciones que desarrollan las diferentes funciones expuestas.

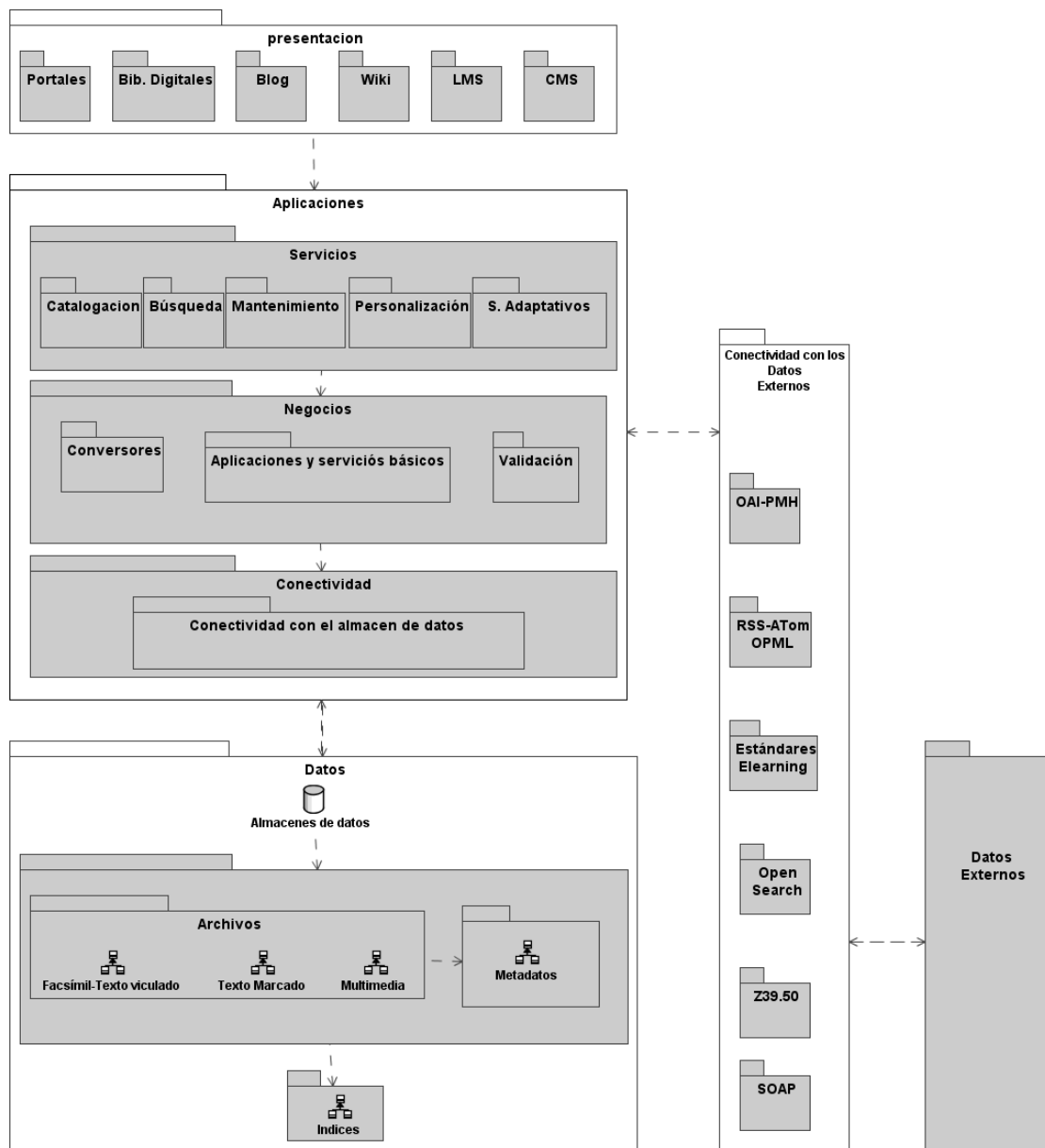


Propuesta de arquitectura tecnológica

Para la realización del sistema descrito funcionalmente la propuesta más viable es seguir una arquitectura en capas que permita separar claramente los componentes y que permita reutilizar el mayor número posible de elementos y que simplifique la creación de las distintas vistas que hemos propuesto para la

biblioteca y para la conexión con los otros sistemas, como también se ha propuesto.

El modelo que se propone se puede plantear en las 3 capas clásicas: **Presentación, Aplicación y Datos**, aunque como veremos la capa intermedia, a su vez se puede dividir en otras 3.



La idea fundamental es que los componentes informáticos que se creen sean reutilizables, que sirvan para distintos sistemas. Por ejemplo, el componente de

catalogación, servirá tanto a los bibliotecarios de la biblioteca desde el gestor de contenidos que se programe para el mantenimiento del portal, como para los usuarios que estén catalogando sus recursos, igual podríamos decir del buscador.

Otra ventaja de una arquitectura de este tipo es que permite evolucionar continuamente al sistema, ya que se puede mejorar o incrementar la capa de presentación puesto que se mantienen las otras dos, o mejorar la capa de aplicación con nuevos servicios, ya que la otra se mantiene. La orientación del sistema hacia los servicios web, permitirá poder conectarse con el mayor número de sistemas que se encuentren en la Web.

Descripción de las capas

Capa de presentación

En esta capa se encuentran los sistemas que se pueden ejecutar directamente desde el navegador: Explorer, Firefox... Están descritos exhaustivamente a lo largo del informe serían por ejemplos:

El portal de la biblioteca

Bibliotecas digitales

Blogs

Wiki

Gestor de Contenidos, CMS.

Sistema de gestión de aprendizaje (LMS), campus virtuales.

Capa de aplicaciones

Contiene un conjunto de servicios y aplicaciones que son utilizados por las diferentes herramientas de la capa de presentación para acceder y gestionar la información que se almacena en la biblioteca. Contiene además funciones básicas que aseguran la validación, conversión y conectividad con la capa de datos.

Para garantizar la conectividad con datos existentes en bibliotecas digitales, portales y otros repositorios de información institucionales, se ha diseñado una capa de conectividad que contiene diferentes componentes orientados al intercambio de información con aplicaciones y datos externos a la biblioteca utilizado para ello protocolos e iniciativas que permiten obtener tanto información básica como acceso a los servicios y datos mediante SOAP, para aquellas aplicaciones que lo permitan.

Conectividad

Para la conexión con aplicaciones externas hemos valorado múltiples protocolos e iniciativas como son las siguientes:

OAI-PMH: está destinado mejorar el acceso a archivos de publicaciones electrónicas e incrementar la disponibilidad de las publicaciones científicas pero puede ser usado para cualquier material digital. El protocolo se basa en la comunicación mediante metadatos.

RSS-Atom-OPML: son formatos sencillos de datos utilizados para syndicate (redifundir) contenidos a suscriptores de un sitio web, permitiendo distribuir contenido sin necesidad de un navegador.

Estándares e-learning: los conforman un conjunto de iniciativas, protocolos y procedimientos que permiten la clasificación estandarizada de materiales educativos, así como la distribución y empaquetamiento de los contenidos para usos docente. Tendremos en cuenta en el diseño la mayoría de las alternativas existentes.

Open Search: conjunto de [tecnologías](#) que permiten publicar los resultados de una búsqueda en un formato adecuado para la sindicación y agregación. Es una forma para que las páginas [web](#) y los [motores de búsqueda](#) publiquen sus resultados de forma accesible. OpenSearch 1.0 fue publicado en [marzo](#) de [2005](#).

Z39.50: es un estándar internacional para la comunicación entre sistemas informáticos. Se ha extendido su uso en el desarrollo de sistemas bibliotecarios interrelacionados.

SOAP (Simple Object Access Protocol): es un [protocolo estándar](#) por Microsoft, IBM y otros, está actualmente bajo el auspicio de la [W3C](#) que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos [XML](#). SOAP es uno de los protocolos utilizados en los [servicios web](#).

Capa de Datos

Almacén de datos

Este sistema permite la mayor explotación de los datos que contiene. Se superponen sobre las bases de datos y se puede obtener una información más integrada de la que ofrecen estas.

Además de los textos, multimedia contenidos en la Biblioteca, tendremos los datos de los usuarios la utilización que estos hacen de los recursos. Aplicando las técnicas propias de la minería de datos podremos conocer mejor el comportamiento de los usuarios y poder aplicar técnicas adaptativas que permitan sugerir usos nuevos a los lectores.

Imagen-texto vinculado.

Sistema utilizado en Google Libros y en Live Books Search de Microsoft. Se reproducen las páginas del libro original con una imagen de las mismas. Y se le vincula el texto creado con un OCR (sistema óptico de reconocimiento de caracteres), de tal forma que en los procesos de búsqueda por palabras el sistema encuentra la imagen de la palabra en la página correspondiente.

Los sistemas robotizados utilizados por Google y por Microsoft son 4DIGITALBOOKS y KIRTAS respectivamente, que son capaces de digitalizar 3.000 páginas por minuto, lo que permitiría digitalizar unos 50.000 libros anuales. El coste de estos sistemas está por encima de los 150.000€.

Mientras el libro es digitalizado, el libro con la sola intervención humana de colocarlo en el sistema, se inicia un flujo de trabajo consistente en reparación de la imagen, reconocimiento de caracteres, vinculación del texto a la imagen, indexación del texto, creación del volumen, incorporación al sistema.

Varios formatos pueden presentar el libro PDF de Adobe, LIT de Microsoft, eBook. En la Open content Alliance tienen un grupo para definir estos formatos.

Texto marcado

El flujo de trabajo es mas complejo, ya que se trata de trabajar con el texto resultante del reconocimiento de caracteres, con lo que se necesita pasar una fase de corrección, y posterior marcado del texto para de dotarle de estructura: títulos, índices, párrafos, ...

Este proceso ha de realizarse por personal especializado y supone un coste por libro elevado. Evidentemente el valor añadido es mayor que en el caso de la imagen con texto vinculado.

También entrarían en esta categoría los textos creados para la biblioteca que no tengan origen en un libro, pero que deben ser considerados como objetos del repositorio general y por tanto ser catalogados para poderse reutilizar en otros contextos.

Multimedias

Otros documentos que incorpora la biblioteca no tienen texto asociado, imágenes, fotos, ficheros de sonidos, vídeos. Todos ellos si deben de estar convenientemente catalogados para permitir su búsqueda y reutilización.