

La metáfora de la biblioteca tradicional para el diseño de una biblioteca digital

Guadalupe Muñoz¹, Ignacio Aedo², Paloma Díaz²

¹Escuela Superior de Ciencias Experimentales y Tecnología.
Universidad Rey Juan Carlos.
C/ Tulipán s/n – Móstoles
Madrid, 28933
gmunoz@escet.urjc.es

²Universidad Carlos III de Madrid
Avda de la Universidad, 30 – Leganés
Madrid, 28911

Resumen

Aunque actualmente, y gracias al avance de la tecnología y de Internet, a los usuarios les resulta relativamente fácil obtener cualquier tipo de información, ya no resulta tan fácil que consigan la información que realmente les interesa. Esto es debido a que la información a la que acceden no está organizada. Este problema de la organización de la información ha sido considerado a lo largo de la historia por las bibliotecas. Internet podría convertirse en una biblioteca digital si organizara la información como lo hacen las bibliotecas tradicionales. En este trabajo se presenta VILMA (VIRtual Library with a Multi-layer Architecture), una biblioteca electrónica que por un lado utiliza la metáfora de la biblioteca tradicional para modelar las funcionalidades de las bibliotecas digitales, y por otro, hace uso de la tecnología virtual en tres dimensiones para facilitar el acceso a dicha biblioteca y para poder llevar la metáfora al ámbito de la interacción del usuario.

Palabras clave: Organización de la información, Internet, Bibliotecas Digitales, Interfaces de usuario 3D.

1. Introducción

Actualmente, y debido al avance de la tecnología y de Internet, la información es mucho más accesible que hace unos pocos años. Esta accesibilidad es engañosa porque aunque es más fácil que los usuarios obtengan cualquier tipo de información, ya no resulta tan fácil que consigan la información que realmente les interesa. Esto es debido fundamentalmente a que la información no está organizada, por lo que los usuarios tienen muchos problemas para obtener lo que realmente quieren.

El problema de organizar información ha sido considerado a lo largo de la historia por las bibliotecas. Las bibliotecas son las entidades en las que se han ido almacenando documentos que se consideraban importantes, y que además se han encargado de gestionar el acceso más adecuado en cada momento a dichos documentos.

Las bibliotecas han ido evolucionando para adaptarse a los medios que iban apareciendo en cada momento y al material del que estaban hechos los documentos que gestionaba, de forma que podemos hablar de tres tipos de bibliotecas: las tradicionales, que gestionan documentos físicos como libros, las

automatizadas, que gestionan documentos físicos usando algunos mecanismos electrónicos, y las digitales, que gestionan tanto documentos físicos como electrónicos utilizando mecanismos electrónicos.

Las bibliotecas digitales deben proporcionar los mismos mecanismos que las tradicionales, pero también proporcionar otros nuevos que sean posibles gracias a la tecnología [Gladney et al. 94]. En opinión de los autores, la mayoría de las bibliotecas electrónicas actuales todavía no proporcionan un acceso adecuado a la información que gestionan porque ni proporcionan los mismos servicios que las tradicionales, ni se aprovechan de las posibilidades de la tecnología.

Internet podría convertirse en una biblioteca digital si se organizara la información como lo está en las bibliotecas tradicionales. De hecho, se ha intentado utilizar algunos de sus mecanismos para intentar solucionar el problema de falta de organización de la información que se encuentra en Internet. Así, aparecen los motores de búsqueda que simulan alguno de dichos mecanismos. También aparecen las llamadas bibliotecas electrónicas o digitales que intentan atajar el problema de Internet utilizando más mecanismos análogos a los de las bibliotecas tradicionales que los mecanismos de búsqueda.

En este trabajo se presenta VILMA (VIRtual Library with a Multi-layer Architecture), una biblioteca electrónica que por un lado utiliza la metáfora de la biblioteca tradicional para modelar las funcionalidades de las bibliotecas digitales, y por otro, hace uso de la tecnología virtual en tres dimensiones para facilitar el acceso a dicha biblioteca y para poder llevar la metáfora al ámbito de la interacción del usuario. A continuación se presenta cómo se han desarrollado cada una de estas características en VILMA.

2. La metáfora de la biblioteca tradicional en VILMA

Se ha descrito una biblioteca electrónica como una estructura que proporciona un acceso tanto intelectual como físico a la inmensa y creciente red de información codificada en formatos electrónicos multimedia [Birmingham et al. 94]. La cuestión es cómo hacer llegar a los usuarios, de entre toda esa información que ofrece la biblioteca, aquella que les interesa [García-Molina et al. 95], utilizando el único medio disponible que es la interfaz, ya que cada usuario llegará a la computadora con ideas preconcebidas basadas en sus experiencias personales, estudios y cultura [Bikson 96]. El hecho

de que los usuarios estén acostumbrados a usar bibliotecas físicas para obtener información ha llevado a varios autores ([Nielsen 90], [Bogaerts et al. 92], [Pejtersen 92]) a proponer el uso de la metáfora de la biblioteca tradicional para las bibliotecas electrónicas, ya que una interfaz que use como metáfora dicha biblioteca será familiar, facilitando su manipulación.

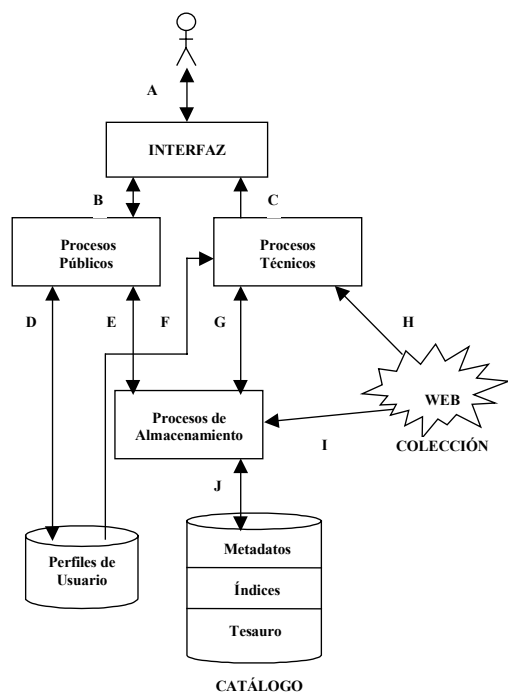
De acuerdo con [Nürnberg et al. 95], hay tres elementos principales en una biblioteca tradicional: entidades de información, metadatos y procesos. Una entidad de información es cada uno de los objetos que almacena la biblioteca (individual o múltiple siempre y cuando sea una unidad de información) que proporciona información completa sobre algo. Los metadatos son información sobre las entidades de información de la biblioteca y cómo pueden ser accedidos a través de índices, y deben identificar unívocamente a cada una de las entidades de información. Además, estos metadatos se organizan para formar el tesauro, de forma que se facilite su acceso. El conjunto de los índices, metadatos y tesauro forma el catálogo de la biblioteca. Los procesos son todas las acciones que se realizan sobre las entidades de información y sobre los metadatos de la biblioteca.

2.1. VILMA

VILMA intenta integrar estos tres elementos, entidades de información, metadatos y procesos, para bibliotecas electrónicas sin perder modularidad, flexibilidad y escalabilidad. La figura 1 muestra una vista general de VILMA, que se explica brevemente a continuación. Consta de tres módulos principales que se llaman "Procesos Públicos" (PP), compuesto de aquellos procesos que interactúan directamente con el usuario, como suscripción, identificación, consulta casual, búsqueda, personalización, consulta experta y movimiento, "Procesos Técnicos" (PT), compuesto de aquellos procesos que gestionan la colección, como la selección de documentos, adquisición, clasificación y catalogación, indización y mantenimiento, y "Procesos de Almacenamiento" (PA), a cargo de la comunicación entre los módulos PP y PT y el catálogo. Estos módulos se comunican entre sí para proporcionar la funcionalidad de la biblioteca como sigue.

La colección de VILMA, es decir, las entidades de información, se obtienen de web, y permanecen en web. El módulo PT selecciona documentos de web basándose en la clasificación de la biblioteca y en las preferencias de los usuarios (flecha F) y los adquirirá para incluirlos en la biblioteca (flecha H). Cuando se adquiere un documento, se clasifica y se cataloga, es decir, se obtienen sus metadatos, y es

indizado en el catálogo a través del módulo PA (flechas G y J). Este módulo también está a cargo del mantenimiento del catálogo, es decir, de mantener los metadatos de la colección actualizados, y de notificar nuevas adquisiciones a los usuarios (flecha C). El módulo PP se encarga de todos los procesos relacionados con el usuario. Obtendrá toda la información sobre cada usuario (flechas A y B) para guardarla en los perfiles de usuario, y para permitirles personalizar su propia vista de la biblioteca a través de las flechas D, B y A. También tomará las consultas de búsqueda de los usuarios y se las pasará (flecha E) al módulo PA para que encuentre la información requerida del catálogo (flecha J). A través de la flecha I, el módulo PA podrá mostrar un documento completo cuando un usuario lo requiera. A continuación se explica brevemente de qué se encarga cada módulo.



- A: Toma las entradas del usuario y le devuelve las salidas.
- B: Comunica las entradas del usuario a los otros módulos y le devuelve las salidas a la interfaz.
- C: Notifica al usuario nuevas adquisiciones.
- D: Guarda datos acerca de los usuarios para construir los perfiles de usuario y devuelve las vistas personales.
- E: Comunica las consultas del usuario al módulo PA.
- F: Toma las preferencias del usuario para buscar en web.
- G: Pasa los metadatos de los documentos seleccionados al módulo PA.
- H: Toma un documento completo de web.
- I: Toma los documentos seleccionados de web.
- J: Guarda los metadatos de los documentos seleccionados y los indiza en el catálogo. Construye el tesoro.

Figura 1. Vista general de VILMA

Procesos Públicos

Este módulo se encarga de aquellos procesos que interactúan con el usuario: suscripción, identificación, consulta (analítica, experta y casual), y personalización. Cuando un usuario llega a la biblioteca, puede suscribirse dando información personal y preferencias. Una vez que el usuario pertenece a la biblioteca bastará con que se identifique para poder utilizar el resto de los servicios, como consultar los documentos en las estanterías, leerlos, realizar búsquedas en el catálogo, consultar al bibliotecario, o personalizar su propia vista de la biblioteca.

Procesos Técnicos

Este módulo se encarga de todos los procesos técnicos, que básicamente son: selección de documentos, adquisición, clasificación y catalogación, indización, mantenimiento y notificación:

Selección y Adquisición: Todos los documentos que pertenecen al prototipo VILMA deben cumplir tres requisitos basados en la naturaleza de la propia biblioteca: 1) El "mercado" de esta biblioteca es World Wide Web, de forma que deben estar localizados en algún sitio web; 2) El sistema de metadatos que se emplea es el conjunto Dublin Core porque es el más usado para documentos web y por su facilidad de uso e interpretabilidad [Miller 96], por lo tanto, los documentos deben contener este tipo de información; y 3) VILMA gestiona documentos sobre informática, y el sistema de clasificación que utiliza es el de 1998 de ACM (con permiso de uso de ACM). Los documentos deberán pertenecer a alguno de los temas o encabezados de la clasificación de la biblioteca. Por lo tanto, los documentos deben sufrir un proceso de pre-clasificación. Un robot escrito en Java rastrea la web para encontrar documentos que, al cumplir los requisitos anteriores, puedan ser incorporados a la biblioteca.

Clasificación y Catalogación: Un módulo de clasificación de documentos escrito como paquete Java, utiliza una versión simplificada del modelo de Boole extendido para la recuperación de la información. En el modelo de vectores espaciales para la recuperación de la información [Salton 71], los documentos se modelan como vectores en un espacio dimensional de muchos cientos de términos. Los términos se obtienen de las palabras y frases del documento, y son "pesados" según su importancia con respecto al documento y el corpus del resto de documentos. El vector que representa a cada documento permite realizar comparaciones en el espacio de vectores de forma que puede utilizarse como herramienta para clasificar el documento con

respecto al sistema de clasificación de la biblioteca, o resolver consultas del usuario. En VILMA se utiliza un conjunto de "documentos de entrenamiento", que son documentos clasificados manualmente y asignados al sistema como ejemplos de documentos que pertenecen a cada clasificación. A partir de ellos, el sistema construye vectores de términos representantes de cada una de las clasificaciones. Cuando el sistema encuentra un nuevo documento que ha de ser clasificado, mide la similaridad entre el vector del documento y cada uno de los representantes de las clasificaciones. Este modelo se ha usado con éxito en algoritmos para categorización y agrupación de documentos entre otros [Jones et al. 97], [Baeza-Yates et al. 99]. Después que se clasifica un documento, un módulo de extracción de metadatos obtiene todos los metadatos que se necesitan para cumplir el conjunto Dublin Core, versión 1.0, que son: Título, Autor, Materia, Descripción, Editor, Colaborador, Fecha, Tipo de Recurso, Formato, Identificador del recurso, Fuente, Idioma, Documentos relacionados, Alcance, Gestión de Derechos; y lo cataloga.

Indización: después de la clasificación y extracción de metadatos del documento, se indizan los metadatos extraídos en los almacenes de datos de la biblioteca para actualizar el tesoro y facilitar el acceso a estos metadatos cuando un usuario realice alguna consulta.

Notificación: cuando se incluye un documento en la biblioteca, un programa de notificación comprueba en los perfiles de usuario qué usuarios estaban interesados en documentos que pertenezcan a la misma clasificación que el incluido para mandarles un aviso de su incorporación.

Mantenimiento de la información: dado que los documentos en VILMA se obtienen de web, un robot también escrito en Java se encarga de comprobar que estos documentos siguen existiendo. En caso negativo se eliminan de la biblioteca, y en caso afirmativo, comprueba que sus metadatos siguen siendo los mismos. Si no lo son, se encarga de actualizarlos. Dado que el período de vida de los documentos web no es fijo, se está realizando un estudio sobre la duración de este período para documentos que utilizan el conjunto de metadatos Dublin Core para proporcionar a este robot el período de comprobación de metadatos más adecuado.

Procesos de Almacenamiento

Este módulo se compone de los almacenamientos de datos para el catálogo de la biblioteca, que guarda los metadatos de todos los documentos que han sido indizados. También se compone de los mecanismos

de acceso que permiten realizar consultas en el catálogo. Este catálogo es una base de datos PostgreSQL en un servidor con sistema operativo Linux Debian accediendo a los módulos anteriores por medio de JDBC.

Actualmente, los registros del catálogo almacenan el conjunto de metadatos específico Dublin Core, pero están preparados para almacenar cualquier ampliación de dicho conjunto, o bien algún otro esquema de metadatos que en un futuro pudiera considerarse más adecuado.

2.2. Realidad virtual como interacción en VILMA

Además, y para incrementar la transparencia de significado hacia el usuario se incluye como elemento básico una interfaz con una representación espacial de los elementos tradicionales de la biblioteca.

Se propone como implementación de esta representación espacial el uso de la tecnología virtual en tres dimensiones. Esta decisión está basada en que algunos experimentos muestran que los usuarios prefieren gráficos tridimensionales en vez de bidimensionales cuando interactúan con la computadora [Levy et al. 96].

En la interfaz de VILMA, cada sala de la biblioteca se representa gráficamente como un modelo virtual en tres dimensiones. Se ha modelado usando el estándar internacional VRML97 ISO/IEC 14772-1:1997, y "External Authoring Interface (EAI)" para añadir comportamiento externo producido por los procesos escritos en el lenguaje Java de la biblioteca.

La biblioteca se representa como un edificio con tres plantas:

- La planta baja proporciona los servicios de suscripción e identificación.
- La primera planta está dedicada a la recuperación de la información en la vista general de la colección, es decir, a las consultas de los documentos de la biblioteca, y a procesos de personalización sobre estos documentos. Estas consultas son de tres tipos: la consulta casual, que consiste en que el usuario pueda ojear los documentos en las estanterías, la consulta analítica, que permite una búsqueda en el catálogo usando mecanismos automatizados, y la consulta experta a través de la cual, el usuario puede realizar consultas al bibliotecario.
- La segunda planta proporciona la vista personal de la biblioteca al usuario y los

mecanismos de consulta analítica y casual, y de personalización sobre estos documentos.

Las tres plantas se comunican mediante un ascensor como muestra la figura 2.

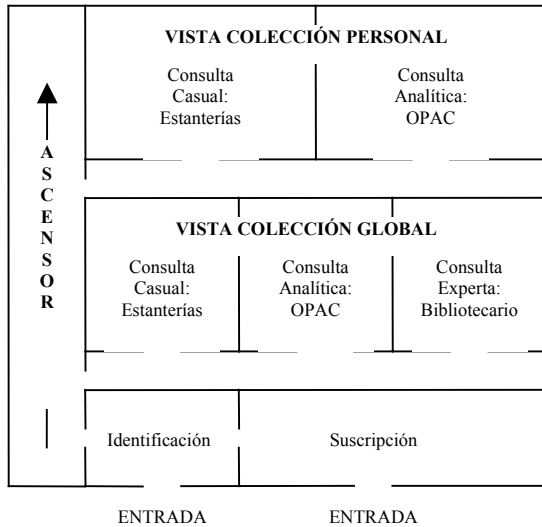


Figura 2. Diagrama general del edificio de la biblioteca

Cuando un usuario llega a la biblioteca, se encuentra con el edificio y su entrada como se muestra en la figura 3. En la planta baja podrá suscribirse si es que todavía no pertenece a la biblioteca, o bien identificarse una vez suscrito para poder acceder al resto de la biblioteca. Una vista desde el vestíbulo muestra la entrada a la sala de suscripción-identificación y al ascensor en la figura 4.



Figura 3. Fachada de VILMA



Figura 4. Planta baja de VILMA

Una vez identificado, el usuario podrá acceder a la primera planta a través del ascensor. Esta primera, es la planta dedicada a la recuperación de la información en la vista general de la biblioteca, es decir, al conjunto global de documentos, o bien a la segunda planta, que es la que se utiliza para la vista personal del usuario. La figura 5 muestra las estanterías que el usuario puede encontrarse en cualquiera de estas plantas. En estas estanterías, cada libro se representa como un libro real, de forma que se organizan espacialmente agrupándolos por temas de la biblioteca. La figura 6 muestra el OPAC que el usuario puede utilizar para realizar consultas analíticas.



Figura 5. Planta baja de VILMA

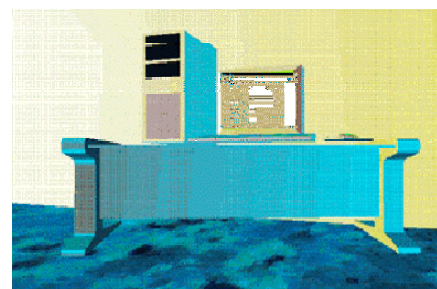


Figura 6. OPAC

3. Conclusiones

Se ha presentado una biblioteca digital que usa la metáfora de la biblioteca tradicional de dos maneras: implementa los servicios de las bibliotecas tradicionales adaptándolos al medio electrónico en la creencia de la generalidad de dichos servicios, y utiliza una metáfora espacial para los objetos de la interfaz. También utiliza la realidad virtual para permitir una interacción más realista con esta metáfora espacial.

En opinión de los autores, las características anteriores mejoran la gestión de documentos electrónicos en una red de dos maneras. Por un lado, porque proporciona todos los servicios que son necesarios en la gestión de documentos, y por otro, porque se aprovecha de las ventajas que proporciona el medio electrónico facilitando el acceso a los documentos, permitiendo al usuario, por ejemplo, personalizar sus preferencias y su vista personal de la biblioteca. Además de facilitar la interacción del usuario mediante el uso de la realidad virtual. Dicha afirmación se basa en evaluaciones informales realizadas con estudiantes universitarios.

El valor de esta biblioteca se comprobará de manera formal cuando se termine la construcción del prototipo VILMA. Entonces se llevará a cabo una evaluación con expertos (bibliotecarios, expertos en Interacción Persona-Ordenador, etc.), y otra con usuarios reales y potenciales. En la evaluación experta, podrán obtenerse valoraciones tanto de la funcionalidad del prototipo VILMA, es decir, de su calidad como sistema informático, como de su facilidad de uso a través de su interfaz. Pero en la evaluación que se lleve a cabo por usuarios reales, las valoraciones que se obtengan serán con referencia a la facilidad y rapidez de la interfaz.

Referencias

- [Baeza-Yates et al. 99] Baeza-Yates, R. y Ribeiro-Neto, B.: *Modern Information Retrieval*. Addison-Wesley/ACM Press (1999)
- [Bikson 96] Bikson, T.: *Identifying Real Information Needs and Developing Digital Libraries to Meet those Needs*. Actas de Conference on Social Aspects of Digital Libraries. (16-17 Febrero 1996).
- [Birmingham et al. 94] Birmingham, W.; Drabenstott, K.; Frost, C.; Warner, A. y Wills, K.: *The University of Michigan Digital Library: This is Not Your Father's Library*. Actas de Digital Library 94 Conference (DL'94) Schnase, J. L.,

- Leggett, J. J., Furuta, R. K., Metcalfe, T. (eds.) Texas, EE.UU. (19-21 Junio 1994). 53-59.
- [Bogaerts et al. 92] Bogaerts, W.F. y Agema, K. S.: *Active library on corrosion*. Elsevier Science Publisher BV con National Association of Corrosion Engineers. Amsterdam (1992)
- [Garcia-Molina et al. 95] Garcia-Molina, H.: *A Strawman Report*. Actas de IITF Digital Libraries Workshop. Virginia, EE.UU. (18-19 Mayo 1995)
- [Gladney et al. 94] Gladney, H.; Fox, E.; Ahmed, Z.; Ashany, R.; Belkin, N. y Zemankova, M.: *Digital Libraries: Gross Structure & Requirements*. Actas de Digital Library 94 Conference (DL'94). Schnase, J. L.; Leggett, J. J.; Furuta, R. K.; Metcalfe, T. (eds.) Texas, EE.UU. (19-21 Junio 1994).
- [Jones et al. 97] Jones, K. S. and Willet, P. eds. *Readings in Information Retrieval*. Morgan Kaufmann, (1997)
- [Levy et al. 96] Levy, E.; Zacks, J.; Tversky, B. y Schiano, D.: *Gratuitous Graphics? Putting preferences in perspective*. Actas de ACM SIGCHI Conference on Human Factors in Computing Systems (CHI96). Bilger, R.; Guest, S. y Tauber, M. J. (eds). Vancouver, British Columbia, Canada. (13-18 Abril 1996).
- [Miller 96] Miller, M.: *Metadata for the masses, Ariadne, the web version. 5*, (Septiembre 1996)
- [Nielsen 90] Nielsen, J.: *Through hypertext*. Communications of the ACM. 33: 3. (Marzo 1990). 297-310.
- [Nürnberg et al. 95] Nürnberg, P. J., Furuta, R., Leggett, J. J., Marshall, C. C., Shipman III, F. M.: *Digital libraries: issues and architectures*. Actas de "The second annual conference on the theory and practice of digital libraries, Digital Libraries' 95"- Austin, Texas, USA. (11-13 Junio 1995)
- [Pejtersen 92] Pejtersen, A. M.: *New model for multimedia interfaces to online public access catalogues*. The electronic library. 10 : 6 (Diciembre 1992). 359-66.
- [Salton 71] Salton, G.: *The SMART retrieval system - experiments in automatic document processing*, Prentice-Hall, Englewood Cliffs N.J. (1971)