

Capítulo 2. Bibliotecas Digitales

Hoy en día se han producido y coleccionado miles de libros, artículos, estudios formales e informales, reportes, revistas, periódicos y videos. La capacidad que tiene el ser humano para absorber esta información e interpretarla para usarla de manera inteligente es limitada no sólo por la cantidad y variedad de datos disponibles, sino también por las complejas relaciones que existen entre los tipos de información y las dificultades que resultan para difundirlos.

Para poder almacenar y organizar la información se cuenta con bibliotecas, las cuales tienen la misión de proporcionar acceso intelectual y físico al acervo cultural así como la preservación del mismo. Una biblioteca es fundamentalmente un conjunto organizado de recursos, que incluyen elementos físicos (espacio, equipo y medios de almacenamiento), componentes intelectuales (políticas de colección que determinan materiales que serán incluidos y esquemas organizacionales que determinan como se va a acceder a esta colección) y personal que maneja los componentes físicos e intelectuales que interactúan con los usuarios para resolver problemas de información. Las funciones de las bibliotecas convencionales incluyen la colección, organización, representación, acceso, recuperación, análisis, síntesis y disseminación de la información.

La "explosión de la información" de los últimos años ha motivado que surja un nuevo concepto conocido como biblioteca digital.

2.1 DEFINICIÓN DE BIBLIOTECA DIGITAL

En el capítulo 1 se mencionó la importancia y el auge que ha cobrado la investigación de bibliotecas digitales en los últimos años. En este capítulo, además de presentar algunas de las definiciones que existen sobre lo que es una biblioteca digital, se plantea el reto que conlleva la construcción de una biblioteca digital, así los aspectos que implica esta nueva tecnología.

También se describe en este capítulo algunas arquitecturas propuestas por diversos investigadores del área. Se presentan algunos ejemplos de universidades y centros de investigación que se han enfrentado al desafío de crear una biblioteca digital. Se hace énfasis en aquellos trabajos que

incorporan la tecnología de videoconferencia en la creación de bibliotecas digitales, puesto que es en este contexto que se desarrolla la presente tesis. Finalmente, en la última sección se describe como se usa esta tecnología en el área de bibliotecas digitales.

Por su novedad, el término "biblioteca digital" provoca aún una impresión diferente en cada lector. Para algunos es simplemente la automatización de bibliotecas convencionales. Para otros es una colección de manuscritos digitalizados. Birmingham [1995] usa el término de biblioteca digital como un nombre genérico para estructuras dinámicas federadas que proporcionan acceso intelectual y físico a redes crecientes de información codificada en formatos digitales multimedios.

Existen muchas definiciones de biblioteca digital, y términos como biblioteca electrónica y biblioteca virtual son a veces utilizados como sinónimos.

En particular [Gladney et al. 1994] proponen una definición precisa al término y es ampliamente aceptada en el ambiente de bibliotecas digitales: *"Una biblioteca digital, es un ensamblaje de cómputo digital, mecanismos de comunicación y software necesario para reproducir, emular y extender los servicios provistos por bibliotecas convencionales como son la colección, catalogación, búsqueda y diseminación de la información"*. Un servicio completo de biblioteca digital, debe estar acompañado de los servicios esenciales de las bibliotecas convencionales además de explotar las ventajas de almacenamiento, búsqueda y comunicación digital.

Otras definiciones incluyen las de [Fox et al. 1993], [Birmingham 1994], [Marchionini y Maurer 1995], [Nürnberg et al. 1995] y [Graham 1995].

La misión de las bibliotecas digitales es adquirir información, organizarla, hacerla disponible y preservarla, además de diseminar el conocimiento, estrechar la comunicación y colaboración entre los usuarios de la misma; una biblioteca digital es vista popularmente como una versión electrónica de una biblioteca tradicional, pero la sustitución del papel por almacenamientos electrónicos, trae consigo *grandes diferencias* como son:

El almacenamiento de libros, revistas, periódicos, discos y videos en forma digital.

La interacción directa por parte del usuario con la información, con el fin de obtener el material deseado.

La copia del material en versión original.

El uso y diseminación de la información por parte del usuario.

El intercambio de puntos de vista, experiencias e información entre

usuarios en forma electrónica.

[Birmingham 1994] menciona algunas características que distinguen a las bibliotecas digitales de las convencionales:

1. Proporcionan información en cualquier tiempo y en cualquier lugar.
2. Proporcionan acceso a colecciones de información multimedia, construidas mediante la integración de texto, imágenes, gráficas, audio y video.
3. Permiten el ingreso y la representación personalizada de la información, incluyendo soporte para recuperar sólo información relevante.
4. Mejoran las actividades colaborativas tales como la investigación, el aprendizaje y el diseño, al reducir las barreras del tiempo y la distancia.

Para [Gladney et al. 1994] las ventajas de una biblioteca digital sobre una biblioteca física, son similares a aquellas que cualquier base de datos digital tiene sobre su contraparte en papel:

1. Rápido acceso a la colección de datos.
2. Mejoramiento en la funcionalidad de revisión y búsqueda.
3. Rápida distribución desde puntos de creación y los puntos de almacenamiento al punto de uso.

Un tema importante en el desarrollo de bibliotecas digitales es el acceso y localización de colecciones de información, además de conferencias de gran importancia disponibles en video y elementos multimedia.

En el futuro, y través de bibliotecas digitales, la gente colaborará con colegas por medio de videoconferencias sin importar las distancias geográficas o temporales; se usarán esas bibliotecas para organizar espacios de información personal y para leer, escribir, enseñar, aprender y crear. El trabajo intelectual será compartido con otros por medio de la biblioteca, y sus contribuciones e interacciones serán elementos de una biblioteca accesible, global y universal que podrá ser usada por mucha gente en diferentes comunidades [Schnase et al. 1994].

2.2 RETOS EN LA CONSTRUCCIÓN DE BIBLIOTECAS DIGITALES

Las bibliotecas digitales no residen en una construcción física, ni manejan objetos físicos como libros, catálogos o archiveros. De hecho las

bibliotecas digitales, deben trasladar los elementos físicos y conceptuales de la biblioteca física y adaptarlos al ambiente digital e incluso se deben añadir nuevos elementos digitales que no aparecen en la biblioteca física [Garret 1993]. Esto nos conduce a una serie de problemas que deben ser resueltos para lograr crear una verdadera biblioteca digital.

En particular, es necesario cuestionarse en cuanto a: ¿cómo se comunican entre bibliotecas?, ¿cómo sabrá el usuario donde buscar?, ¿cómo compartir experiencias y comentarios?, ¿cómo deben ser ordenados, seleccionados, unidos y transmitidos los resultados al usuario?, ¿cómo asegurar la integridad y seguridad de los datos?, ¿cómo digitalizar los objetos y ponerlos en línea?, ¿cómo encontrar fallas y corregirlas?, ¿cómo actualizar la información?, ¿cuánto costará la infraestructura de comunicaciones?, ¿cuánto costará la infraestructura de cómputo? , ¿cuánto costará la información?.

Para preservar la información es necesario que las bibliotecas digitales migren continuamente sus datos. Pero el cambio continuo en los métodos para grabar información, en los formatos de almacenamiento y en las tecnologías que se usan para mantener la información, hace que el costo de la migración se elevado e incluso incalculable [Kuny 1996].

Para [Griffiths y Kertis 1990], una representación apropiada de los recursos de información conduce a una identificación y selección exitosa de los mismos. Esto es importante debido a que el ambiente de red se cuenta con un gran número de recursos digitales que contienen muchas formas de información (texto, imágenes, sonido, video, software, esculturas de datos complejos entre otros), así que es difícil para los usuarios conocer que recursos están disponibles.

2.3 COMPONENTES DE UNA BIBLIOTECA DIGITAL

El problema central con la traducción de datos físicos es decidir que aspectos deben tomarse en consideración durante el proceso de traducción. Esto no es sencillo, ya que las características físicas pueden representar algo para unas personas y otra cosa muy diferente para otras. El problema con la traducción de metadatos de la biblioteca física es que a veces los metadatos o su aplicación están influenciados por el dato físico. El problema clave con la traducción de procesos de la biblioteca física es que son ejecutados por humanos y muchas veces los procesos de ejecución son informales.

Otras arquitecturas se encuentran detalladas en [Shen et al. 1994] y en [Srihari et al. 1994].

Finalmente y a pesar de las diversas arquitecturas propuestas, el objetivo

de toda biblioteca digital es servir de mediadora entre los usuarios (lectores, autores, bibliotecarios y editores) y la información almacenada.

Tabla 2.1 Taxonomía de elementos en una biblioteca digital (Adaptada [Nürnberg et al. 1995])

	Traducción de entradas de una biblioteca física	Nuevas entidades de una biblioteca digital
DATOS	Libro Revista Video	Hipernovela Visualización científica Programa
METADATOS	Indice estático Clasificaciones Ordenamiento espacial	Indice dinámico Estructura personalizada Anotaciones
PROCESOS	Adquisición de datos Sugerir fuentes Ayudar a localizar fuente	Búsqueda de textos. Presentación personalizada Recuperación por agentes

2.4 ALGUNAS BIBLIOTECAS DIGITALES

Desde el artículo "As we may think" de Vannevar Bush [1945] hasta la fecha, mucho se ha dicho, escrito y hecho en materia de bibliotecas digitales. En la actualidad se cuenta con diversos proyectos de bibliotecas digitales con muy variados objetivos, entre los que podemos mencionar:

Un proyecto de la Universidad de Cornell auspiciado por la Comisión de Preservación que intenta rescatar el contenido de materiales frágiles [Kenney y Personius 1993; Webster 1993].

Un proyecto de Sequoia que pretende crear una herramienta específica para el manejo de datos que componen una investigación científica masiva [Kochevar et al. 1993; Stonebraker y Dozier 1991].

En el proyecto Intermedia de la Universidad Brown, el objetivo es concebir acervos para materiales interactivos de instrucción [Yankelovich et al.1988].

En un proyecto IBM/Universidad Reserve Western, la biblioteca tiene la intención de ser un componente de un "ambiente de aprendizaje electrónico", como parte de un sistema de información [Backer y Kingman 1993].

El proyecto de la Biblioteca de Memoria del Congreso Americano, intenta ampliar al acceso publico al material histórico [Culshaw 1992; Polly y Lyon 1992].

El proyecto Envision del Instituto Politecnico de Virginia se enfoca a la recuperación de información e interfaces de objetos estructurados y reusables, así como a tecnologías para aprendizaje acerca de la ciencia computacional con material de ACM, IEEE-CS y otros [Fox et al. 1993].

Un proyecto conjunto de la Biblioteca del Vaticano, la Universidad Pontificia Católica de Río de Janeiro e IBM pretenden dar acceso desde el WWW a los materiales históricamente significativos así como a hermosas ilustraciones copiadas de la Divina Comedia de Dante y a los cuatro viejos manuscritos de los poemas de Virgilio [Gladney et al.1994].

El proyecto Nacional de Bibliotecas Digitales (NDLP) de la Biblioteca del Congreso pretende construir un sistema experimental para organizar y almacenar colecciones históricas. Se van a convertir las colecciones históricas a forma digital para hacerlas disponibles en Internet, esto incluye texto, fotos e ilustraciones [Arms et al. 1997].

La Biblioteca Digital de imágenes de Astronomía (ADIL) va a coleccionar imágenes astronómicas para hacerlas disponibles a la comunidad astronómica y al público en general. Los clientes ingresan a través de Internet usando cualquier programa cliente o navegador multimedios para buscar imágenes. Una vez localizadas dichas imágenes, los usuarios pueden salvarlas en sus computadoras locales para posteriormente analizarlas [Ye 1997].

El proyecto CoLib en un esfuerzo multi - institucional, inter e intradisciplinario que pretende establecer a gran escala una biblioteca digital botánica distribuida. El proyecto engloba a cinco proyectos

particulares que son: el proyecto de la Flora de Norte América (FNA), el proyecto de Flora de Perú, el proyecto de Flora de Madagascar, el proyecto de La Flora Mesoamericana y el proyecto de Flora de China. los mayores participantes en el proyecto CoLib son la Universidad de Washington, Texas A&M University, el Jardín Botánico de Missouri. A raíz del proyecto se han creado diversas líneas de investigación, en las siguientes áreas: a) sistemas de bibliotecas colaborativas y hipermedios, b) cómputo multimedia interactivo y sistema de comunicaciones, y c) relaciones bibliotecarias en la biblioteca digital botánica [Schnase et al.1994].

Estos ejemplos muestran claramente la importancia que ha cobrado el desarrollo de bibliotecas digitales.

En particular hay una iniciativa para bibliotecas (DLI) que está conjuntamente financiada por la Fundación Nacional de Ciencias (NSF), la Agencia de Proyectos de Investigación Avanzada del Departamento de Defensa (DARPA) y la Administración Nacional de Aeronáutica y del Espacio (NASA). Este proyecto de investigación se originó a raíz de la creciente demanda por sistemas y servicios, capaces de proveer a las diversas comunidades de usuarios, con un acceso coherente a bases de datos extensas y geográficamente dispersas.

El objetivo de la iniciativa es incrementar dramáticamente los medios de obtener, almacenar y organizar información en forma digital y hacerla disponible para la búsqueda, recolección y procesamiento a través de redes de comunicación. Un aspecto importante de la iniciativa es el de establecer mejores conexiones entre ciencia básica e innovación tecnológica.

La investigación sobre bibliotecas digitales explora temas variados que son relevantes para el desarrollo y uso de colecciones bibliográficas distribuidas electrónicamente. Las colecciones de interés pueden contener texto, imágenes, mapas, registros de audio, video y películas, además de combinaciones de estos medios (multimedios). La investigación sobre bibliotecas digitales se concentra también en temas sociales, económicos y culturales asociados con el desarrollo y uso de estos recursos.

Seis proyectos multidisciplinarios están siendo financiados a lo largo de un periodo de cuatro años. Cada proyecto tiene equipos de investigadores de una universidad líder con investigadores de otras organizaciones. Más de 75 organizaciones diferentes han establecido relaciones de asociación con los proyectos. Las organizaciones asociadas representan diversos intereses que incluyen empresas líderes de los Estados Unidos en los rubros de computación y comunicaciones, instituciones académicas en todos los niveles, bibliotecas, editoriales, gobierno, agencias estatales, asociados profesionales otras organizaciones interesadas en bases de datos y manejo de información a gran escala.

En la Universidad de California Berkeley existe un planeamiento ambiental y sistemas de información geográfica. Este proyecto producirá un prototipo de biblioteca digital enfocado a la información ambiental. La meta del proyecto es desarrollar tecnologías para el acceso inteligente a colecciones distribuidas de fotografías, imágenes de satélite, mapas, documentos completos y documentos multivalentes. El trabajo está concentrado en la recuperación de información geográfica (GIR) [Larson 1996].

Coleccionará información diversa sobre el ambiente para la preparación y evaluación de datos ambientales, reportes de impacto ambiental y material relacionado. La investigación pretende crear sistemas de producción CERES, para crear el prototipo los investigadores necesitarán producir tecnologías que permitan que los usuarios no entrenados contribuyan para hallar información relevante en otros sistemas de biblioteca digital en el WWW. Las áreas de investigación incluyen indexado automático, recuperación inteligente, procesos de búsqueda, tecnología de bases de datos para soportar aplicación de bibliotecas digitales. Los participantes y colaboradores en el proyecto incluyen a Xerox Corporation, Agencia de Recursos de California, La Biblioteca Estatal de California, La Biblioteca del condado de Sonoma, La Asociación de Estados de San Diego, La Oficina de Educación del Condado de Shasta, y Hewlett Packard.

La Universidad de Carnegie Mellon trabaja en una biblioteca de video digital llamada Informedia. Esta habilitará a los usuarios para ingresar, explorar y recuperar material de ciencia y matemáticas desde archivos de video [Christel 1995]. Informedia trabaja con integración de pláticas electrónicas, imágenes y tecnologías para el entendimiento del lenguaje natural. La biblioteca inicialmente contendrá 1000 horas de video. Una biblioteca digital de video debe ser eficiente para dar a los usuarios precisamente el material que ellos necesitan, para hacer recuperación rápida, la biblioteca digital de video necesitará soportar el video en pequeños clips y representaciones alternativas del video. Aspectos como interacción humano-computadora, precio y cobro por uso de video digital, y seguridad y privacidad, son las líneas a seguir como parte del programa de investigación. Participantes industriales incluyen a Microsoft Corporation, Digital Equipment Corporation, Bell Atlantic Network Services, Comunicaciones QED, Universidad Abierta, y Escuelas Públicas del condado de Fairfax Virginia.

La Universidad de Stanford investiga mecanismos interoperacionales entre servicios heterogéneos. El proyecto de la Biblioteca Digital Integrada de Stanford desarrollará tecnologías para integrar una biblioteca "virtual" que proporcione acceso uniforme a colecciones de información digital. La biblioteca digital integrada creará un ambiente compartido que ligará desde colecciones personales de información, las colecciones halladas hoy en bibliotecas convencionales, hasta la gran cantidad de colecciones

de datos compartidos por científicos. Se pretende crear tecnología que permita a los usuarios ingresar a los recursos digitales desde documentos estáticos para la exploración dinámica de la información y el manejo de servicios, sin tener que conocer los detalles del formato de cada documento y los mecanismos de servicio. Esta tecnología es llamada InfoBus, la cual proporciona un marco de trabajo unificado que reúne los servicios proporcionados por el WWW, así como los servicios de recuperación de información y nuevos tipos de servicios que están siendo desarrollados. La investigación incluye: modelos de comunicación, interfaces de información para clientes, y servicios para hallar información. Las organizaciones participantes incluyen a ACM, Bellcore, Dialog, EIT, Laboratorios Hewlett Packard, ITC, O'Reilly y Asociados, NASA Ames, WAIS Inc., y PARC.

El protocolo "InfoBus", proporciona una manera uniforme para ingresar a una variedad de servicios y fuentes de información través de programas escritos para servicios de Internet específicos, como http, ftp y telnet, que se ejecutan en un servidor con dos conexiones a red, y actúan como un servidor hacia la aplicación cliente y como un cliente hacia el servidor de la aplicación ("proxies"), donde actúan como intérpretes entre el protocolo InfoBus y el protocolo nativo [Paepcke 1996].

InfoBus está acompañado por una interfaz que proporciona un modelo consistente para los usuarios. Se trabaja además en una interfaz de usuario llamada DLITE [Cousins 1997], la cuál permite personalizar las tareas del usuario, DLITE esta diseñada para facilitar a los usuarios a interactuar con los diferentes servicios mientras se enfoca en una tarea específica. DLITE introduce el concepto de centros de trabajo (workcenter) que es un lugar donde las herramientas para una tarea están disponibles para usarse. Las componentes de DLITE caen en 5 categorías básicas: documentos, colecciones, queries, servicios y representaciones de la gente. Usa la metáfora de workcenters conteniendo componentes y mecanismos de retroalimentación como animación, en el contexto de l prototipo de una biblioteca digital.

Además de las interfaces de usuario, una de las áreas de investigación con mayor empuje en el proyecto de la Biblioteca Digital de Stanford es ayudar a los usuarios a hallar información. Se han iniciado una serie de proyectos en este sistema. Se han buscado maneras para que las herramientas de búsqueda puedan ser usadas a través de múltiples fuentes que usan diferentes sintaxis o lenguajes. También se han desarrollado herramientas que proporcionen filtrado estadístico o colaborativo para localizar artículos relevantes. Entre los proyectos que se tienen, se pueden mencionar:

GLOSS, el proyecto Servidores de Glosario, está diseñado para localizar fuentes de información relevantes.

SenseMaker, herramienta que le ayuda a los usuarios a reformular iterativamente su información a través de organizar y activar grupos multidimensionales de búsqueda de resultados.

GRASSROOTS, groupware para hallar información, combinar correo electrónico, noticias y web en un sencillo ambiente con distribución de listas.

STARTS, es un conjunto de estándares informales para negociar entre los vendedores de búsqueda y los usuarios para facilitar interoperaciones.

ComMentor, es una plataforma de anotaciones en el web que proporciona maneras de compartir información, y mantener notas.

InterOp, es el corazón de InfoBus, este protocolo describe métodos de acceso para buscar colecciones, adquirir resultados y encontrar fuentes.

InterBib, es una herramienta para mantener información bibliográfica. Capaz de leer y escribir en muchos diferentes formatos, actúa como un repositorio unificado de registros bibliográficos.

La cantidad y diversidad de los proyectos mencionados, es un indicador del auge que está tomando el área de bibliotecas digitales. Como se ha visto es un campo multi - disciplinario que cubre aspectos de áreas como recuperación de información, interacción humano - computadora, protocolos de comunicación e inteligencia artificial.

2.5 USO DE VIDEO EN BIBLIOTECAS DIGITALES

Un tema importante en el desarrollo de bibliotecas digitales es la comunicación entre usuarios, el acceso a colecciones de información y de conferencias de gran importancia disponibles en video y elementos multimedia. El poder intercambiar con otros participantes puntos de vista, documentos, elementos importantes en el desarrollo de una investigación, el dar referencia de documentos que se encuentran en la biblioteca digital, permite realizar una adecuada diseminación de la información.

Con la videoconferencia es posible compartir información, intercambiar puntos de vista, mostrar y ver todo tipo de documentos, dibujos, gráficas, acetatos, fotografías, imágenes de computadora y videos, en el mismo momento, sin tener que trasladarse al lugar donde se encuentra la otra persona [Draoli 1996]. Este es uno de los puntos cruciales para una biblioteca digital.

Morales Salcedo, R. 1999. **Aplicaciones de la Videoconferencia en Bibliotecas Digitales**. Tesis Maestría. Ciencias con Especialidad en Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas Puebla. Mayo. Derechos Reservados © 1999.