
Herramientas para las Bibliotecas Digitales del Nuevo Milenio



Rodrigo Torr ns Heeren

Centro de C culo Cient fico

Unidad de Teleinformaci n

Universidad de Los Andes

torrens@cecalc.ula.ve

M rida Octubre 2001

Introducción

- Existe mucho interés hoy en día en las llamadas “Bibliotecas Digitales” ...

(por eso estamos aquí hoy por ejemplo)

- Una búsqueda en “Altavista” con las palabras...

“Digital Libraries”

...produce mas de 120.000 respuestas...

- Existen cientos de instituciones estudiando y generando proyectos e información relacionada con el tema....

Porque?

(pregunta abierta por ahora ...)

¿Qué es una Biblioteca Digital?

El término “Biblioteca Digital” es el más reciente de una serie de nombres para un concepto que ha sido escrito casi desde la época del desarrollo de la primera computadora:

“Una biblioteca
computarizada - automatizada
que debería suplementar, añadir
funcionalidad y aún reemplazar a las
bibliotecas tradicionales”.

(Entonces el nombre es nuevo pero no el concepto)...

...¿Qué es una Biblioteca Digital?

Más recientemente otros términos como:

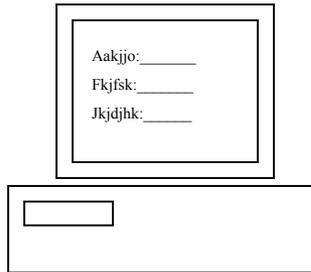
- Biblioteca Electrónica
- Biblioteca Virtual
- Biblioteca sin muros
- Librería Electrónica
- Bibliotecas sin papel

...han sido usados mas o menos con el mismo significado original y siguiendo las primeras ideas de algunos pensadores como Vanevar Bush, Ted Nelson, etc (años 40 al 70)...

Propiedades de una Biblioteca Digital

Visión Tradicional (basada en bibliotecas tradicionales)	Visión amplia (punto medio)	Visión más amplia (basada en lo que es Internet)
Objetos son recursos de información	Muchos de los objetos son recursos de información	Objetos pueden ser cualquier cosa
Objetos estan localizados en un lugar físico	Objetos estan ubicados en un lugar lógico	Objetos no estan ubicados ni en lugar físico ni lógico
Objetos estan organizados		No existe ninguna organización
Objetos sujetos al control de una autoridad	Se controla algunos aspectos	No existe ninguna autoridad de control
Los objetos son permanentes	La desaparición de objetos es controlada	Objetos pueden desaparecer en cualquier momento
Concepto de autor-autoría es muy importante	Concepto de autor es muy debil	No existe el concepto de autor
Se puede encontrar a especialistas llamados "bibliotecarios"		No existen los bibliotecarios
Existen grupos de usuarios bien definidos	Algunas clases de objetos tienen grupos de usuarios asociados	No hay grupos de usuarios definidos (o hay un numero infinito de ellos)

Catálogos Digitales y Contenidos Digitales



Catálogos Digitales

- Sist. de Clasificación
- Sist. de consulta
- SMBD
- Hardware, computadoras, sistemas de comunicación

Contenidos Digitales

- Sist. – mecanismos de digitalización
- Sist. de almacenamiento
- Hardware de almacenamiento
- Sist. de recuperación de información

Hasta ahora las bibliotecas se han concentrado mayormente en el componente “Catálogos Digitales”

Herramientas

- Herramientas de Catalogación y Recuperación de Información
- Herramientas de Digitalización
- Formatos de archivo
- Estándares de intercambio y representación de información, formato y lenguajes

Nos concentraremos aquí en:



Estándares y Lenguajes como Herramientas

Lenguajes de Marcas

Los “lenguajes de marcas” (*Markup Languages*) como HTML, SGML y XML, usan “etiquetas” u otros indicadores para decirle a la computadora como mostrar, reconocer, o manipular de cualquier manera, material “marcado”.

```
<?XML version="1.0" rmd="all" encoding="utf8" ?>
<!DOCTYPE tei.2 PUBLIC "-//TEI//DTD Historical
Editions//EN"> ...

<p>to recapitulate ... the greivances which America
<add>has long</add> labour'd under & which after
<del>long</del> <add>unexampled</add> patience &
forbearance on her part ...

.

<h1><a name="ID8">Why isn't SGML
universal?</a></h1><p>
SGML solves the problems of the Web, but ...
<UL>
<li>undeniable reputation for complexity</li>
<li>parsers tend to be large and hard to produce</li>
<li>hence they tend not to be incorporated into other
pieces of software like Web browsers</li>
</UL>
<p>So now what?</p>
```

HTML

- El *Hypertext Markup Language* (HTML) es el lenguaje de marcas más difundido.
- Casi todos hemos tenido algún contacto con él
 - *Recuperando información que usa este formato*
 - *Produciendo documentos HTML*
- Lenguaje con estructura muy simple (ventaja)
- Orientado a PRESENTACION
- Describe apariencia y composición del texto y otros elementos en la pantalla
- Funcionalidad muy limitada
- Diferentes versiones y soporte por parte de navegadores

...HTML

- La limitación real de HTML es su incapacidad para...

...**describir el contenido de un documento**

- HTML fue construido para “publicación” no para facilitar la recuperación o intercambio de información.
- Debido a su simplicidad, seguirá siendo el lenguaje escogido por muchos para codificar texto y gráficos así como para composición y formato básico de documentos.

XML

eXtensible Markup Language

- La funcionalidad del World Wide Web se ha ido restringiendo crecientemente por las limitaciones del HTML (y la complejidad del SGML).
- XML es un estándar emergente para Lenguajes de Marcas.
- Esta siendo desarrollado por un consorcio de casi 300 empresas bajo la guía del *World Wide Web Consortium* (W3C).
- 1^{ra} especificación: Febrero 1998.
- XML es visto por algunos como la herramienta que revolucionara al WWW (y de hecho lo esta haciendo).

...XML

- Es un metalenguaje: “un conjunto de reglas que gobiernan el desarrollo de etiquetas o marcas únicas para codificar documentos”
- Es “extensible” (se puede usar para muchos propósitos)
- Menos complejo que SGML
- XML no tiene la intención de reemplazar a HTML, los dos son complementarios.
- Los dos estándares hacen énfasis en diferentes elementos de un documento:
 - *HTML maneja “formato y presentación de datos” de texto e imágenes*
 - *XML resuelve el problema del “significado o semántica de los datos”*
- XML puede trabajar en términos de “fragmentos de contenido”
- XML introduce la posibilidad de manejo de información estructurada en el WWW
- **Suministra INFORMACIÓN, no solo páginas con texto e imágenes**

...XML. Ejemplo sencillo

<poema>

```
<titulo> Lento pero viene </titulo>
<autor>
  <nombre> Mario </nombre>
  <apellido> Benedetti </apellido>
</autor>

<fecha_creación> 1986 </fecha_creación>

<formato> text/txt </formato>
```

Metadatos

<contenido>

```
Lento pero viene
el futuro real
el mismo que inventamos
nosotros y el azar
```

```
    cada vez mas nosotros
    y menos el azar...
```

```
...lento pero viene
```

Datos

</contenido>

</poema>

XML y las Bibliotecas Digitales

- Muchas bibliotecas en Europa y Norteamérica están usando SGML
- Como XML ofrece más funcionalidad que HTML y menos complejidad que SGML, es natural que los bibliotecólogos consideren el uso de XML
- El primer campo donde las bibliotecas notaran y se beneficiaran de XML será en el manejo de recursos de información en el World Wide Web.
- Uso importante de XML dentro de las Bibliotecas:

Sistemas de Metadatos

...XML y las Bibliotecas Digitales

- Las colecciones electrónicas o digitales se volverán mas y mas importantes.
- A los bibliotecólogos se les presentara el reto de mantener actualizadas y organizadas estas crecientes colecciones de contenidos digitales.
- Allí entran en juego los estándares y lenguajes como XML

Compartimiento de Información

- Quedan aún muchos problemas por resolver en cuanto a estándares y lenguajes.
- Están apareciendo propuestas que organizarán el “desorden” actual

Por ejemplo:

Estándares para Metadatos y Contenidos Estructurados

- Dublin Core
- RDF

...que comienzan a usar y aprovechar las facilidades e infraestructura de lenguajes como XML.

...Compartimiento de Información

Para que? Si ya existen SGML, o estandares como MARC?

- Todo lleva a tratar que la información sea “entendible por computadoras”

(machine understandable)

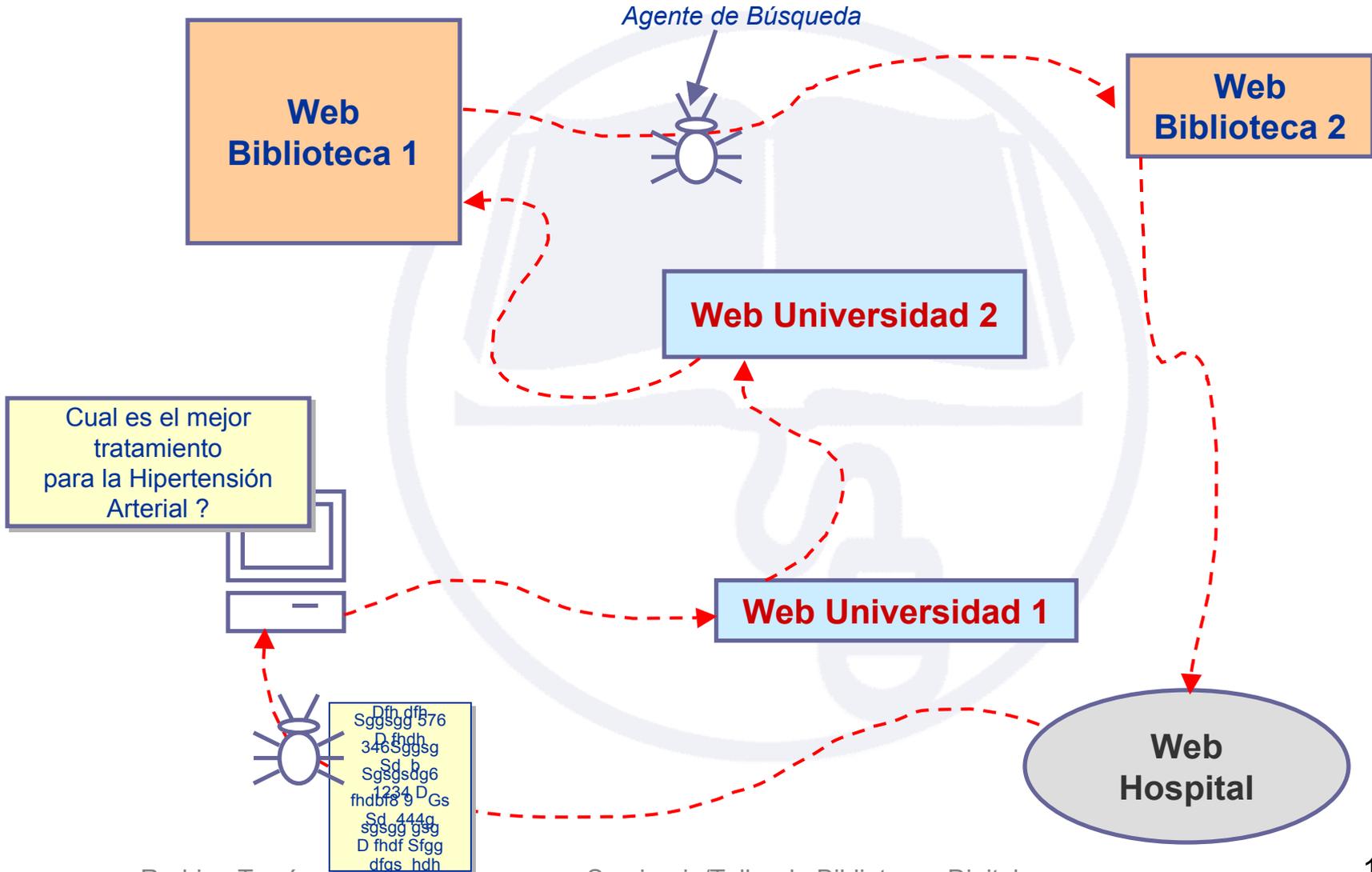
...lo que provocara nuevos niveles de procesamiento, intercambio y recuperación automatizada de recursos de información entre instituciones (como las que ahora llamamos “bibliotecas”)

Iremos acercandonos poco a poco al...

“Web Semántico”

(Suponiendo que continúe la tendencia actual a usar el Web como puente a todos los recursos de información disponibles en la Red)

El Web Semántico



Herramientas “Futuras”

(Que ayudaran a construir las Bibliotecas del Futuro)

- Nuevos estandares (o uso mas amplio de los propuestos actualmente)
 - Agentes de Software usaran principios de IA, redes neuronales, etc. para monitorear, administrar, gerenciar y asignar servicios y recursos de información.
 - Herramientas de comunicación permitirán el uso de Bibliotecas “ubicuas” (estaran presentes en todas partes y las usaremos sin que nos demos cuenta) ...
- ...(lo mismo que sucederá con el resto de Internet y los recursos que ofrece)

A manera de conclusión

Las verdaderas herramientas son los
estándares y protocolos
(maneras de comunicarse)

(Se oyen opiniones)...

Gracias por su atención !

Y por su asistencia a este evento y a nuestra ciudad.

