

Capítulo 2: Visualización de Información en Repositorios

Institucionales

Hoy en día, los usuarios consumen tiempo valioso buscando una información a través de un repositorio. De hecho, mucho del tiempo de cada usuario se desperdicia tratando de encontrar información para hacer su trabajo. Los métodos actuales para tener acceso a la información generan mucha pérdida de tiempo. Para solucionar este problema, crear un solo punto acceso para toda la información del repositorio es lo esencial.

En este capítulo analizaremos las herramientas que existen para la construcción de repositorios institucionales, analizaremos las características de cada uno y escogeremos el mejor. Por otro lado, analizamos las herramientas que existen para visualizar grandes cantidades de información y seleccionaremos la que mejor se adecue para ser implementada en un repositorio.

2.1 Herramientas para la construcción de repositorios institucionales

Para empezar, haremos un análisis de los siguientes repositorios institucionales: *Archimede*, *ARNO*, *CDSware*, *DSpace*, *Eprints*, *Fedora*, *i-Tor*, *MyCoRe*, y *OPUS*. Es importante mencionar que cada sistema tiene sus propias características, por lo que es necesario seleccionar el software que se adecue a los requerimientos particulares de la institución.

Las características de cada sistema, hacen énfasis a la filosofía de su diseño y se ofrece una descripción de los tipos de implementación en la que el software obtiene su mejor desempeño. La funcionalidad de cada sistema procura proporcionar un marco evaluativo

que compare las capacidades de estos sistemas. La característica particular del sistema, se debe considerar en el contexto del diseño y de los requisitos locales de la institución que lo solicita. Haciendo una comparación con las herramientas para la creación de repositorios institucionales [Crow 2004], observamos que DSpace obtiene la mayor ventaja como herramienta con las mejores características para ser utilizada para comprobar nuestra hipótesis. Entre las características sobresale que sus interfaces pueden ser modificadas, tiene la posibilidad de incluir módulos para realizar nuevas tareas, cuenta con distintas capacidades de búsqueda: por autor, por título, por día de publicación, por tema, por comunidades y colecciones.

Algo muy importante que hay que mencionar es que ninguna de las herramientas para crear repositorios institucionales cuentan con un mecanismo para poder visualizar todos sus documentos en una sola pantalla, de hecho todos los documentos son mostrados a través de hipervínculos y el usuario tiene que leer toda la información que se le presenta e ir accediendo diferentes páginas hasta encontrar el documento que desee.

2.2 Análisis de sistemas para visualizar grandes cantidades de información

Al hablar de visualización de información nos referimos a la representación gráfica de datos con la finalidad de facilitar al receptor su comprensión por medio de técnicas que los hagan visibles e inteligibles en el menor tiempo posible. Se busca presentar la información de grandes colecciones de datos de un repositorio permitiendo al usuario visualizar y tener acceso de los documentos del sistema en una sola pantalla en forma organizada, revisar

referencias o textos completos en muy poco tiempo a través de una interfaz con interacción dinámica.

La visualización ofrece grandes ventajas al evitar los tradicionales problemas que presenta el hipertexto: sobrecarga cognitiva, por la gran cantidad de información disponible y desorientación [Brunk 1999]. La solución reside en organizar esta información de manera que el usuario obtenga lo que está buscando en el menor tiempo posible. Para ello se necesitan incorporar al repositorio herramientas de organización y recuperación de la información.

Estas herramientas están pensadas para ofrecer una visión global al usuario y ayudar a encontrar información de interés. Estos programas crean una ayuda a la navegación en función de los datos que contienen las páginas y la presentan en modo texto o mediante un gráfico en una ventana aparte o integrada en la página. Lo más habitual es que la información que se presente sea interactiva, y precisamente este rasgo los hace ser más interesantes.

Las dimensiones de la pantalla en ocasiones no son suficientes para representar toda la información en un solo vistazo, por lo que se recurre a técnicas que presentan la información poco a poco, de lo general a lo específico, según requiera el usuario.

2.2.1 VizServer

VizServer, de Inxight Software, proporciona métodos intuitivos y eficaces para explorar grandes colecciones de información. Estos métodos son: *StarTree*, *TimeWall*, y *TableLens*, los cuales se describen en esta subsección. Los usuarios pueden localizar la información a través de patrones jerárquicos, con parentesco o basados en tiempo. VizServer permite que los usuarios encuentren la información específica más rápidamente que con métodos estándares de la navegación. Sin embargo, para usar este sistema, es necesario pagar por una licencia [Inxight Inc 2004].

VizServer es una aplicación que permite navegar, explorar y acceder a la información almacenada en una base de datos. VizServer es un servidor en J2EE que permite un despliegue rápido de visualización de la información.

2.2.1.1 *Star Tree*

Es un sistema de navegación interactivo basado en la idea del espacio hiperbólico. Usando la tecnología de StarTree, las jerarquías se presentan de una manera uniforme en un plano hiperbólico, sobre una región circular de forma interactiva con ayuda del ratón. Esto permite que el usuario tenga disponibles un gran número de opciones contextuales, en un espacio reducido. Inxight StarTree permite ligar archivos, documentos y las páginas Web. Proporciona un contexto visual para la información, en forma jerárquica. *StarTree* ofrece una gran variedad de APIs para programar en los ambientes en Java o en ActiveX. Estos

proporcionan una gama de acceso para navegar y visualizar grandes jerarquías de información [Inxight Inc 2004].

2.2.1.2 *TimeWall*

Este sistema simula a una pared tripartita que permite al usuario traer información al frente de la pantalla. Los datos se muestran en dos dimensiones y el usuario puede ordenar los documentos en el eje de abscisas (x) o por otro criterio distinto [Inxight Inc 2004].

2.2.1.3 *TableLens*

Table Lens es un software que permite visualizar tablas con gran cantidad de datos de una forma simultánea y global, al mismo tiempo, poder explorar sus datos concretos sin perder de vista el contexto en ningún momento [Rao y Stuard 1994]. *Table Lens* permite a programadores integrar una técnica visual patentada al análisis de datos en sus aplicaciones de software y sitios Web. Cuando se está desarrollando las aplicaciones para los datos o los sitios Web que exhiben grandes conjuntos de datos, se hace frente a un desafío significativo asegurándose de que los usuarios puedan entender fácilmente el contenido de dichos conjuntos. Si los usuarios no pueden entender los datos, entonces no les ayudará a hacer sus trabajos. La tecnología de Inxight es una manera para que los usuarios visualicen grandes cantidades de datos tabulares dibujando la información de los datos en tablas.

VizServer es una aplicación en Java J2EE basada en Web que publica los *applets* en el Web. Funciona en cualquier navegador. Se conecta con cualquier base de datos relacional JDBC

incluyendo Oracle, MySQL, el servidor SQL, y Sybase. Realiza actualizaciones en tiempo real de las visualizaciones de StarTree y de TableLens. Contiene mecanismos existentes de seguridad para el control de acceso a la información de la empresa. Sin embargo, este sistema no es de distribución gratuita. Hay que pagar una licencia por usa este producto.

2.2.2 Thinkmap

Thinkmap, de Plumb Design, es una plataforma para desarrollar interfaces de visualización de datos complejos. Se compone de un número de componentes que pueden ser reconfigurados para satisfacer diversas tareas de la visualización. Thinkmap viene con varias visualizaciones pre - configuradas: visualización araña, visualización jerárquica, visualización por ramificaciones, visualización por cronología. Estos componentes son ligeros, flexibles, y basados en estándares [Thinkmap Inc 2004].

Thinkmap trabaja flexible y dinámicamente con múltiples fuentes de datos: XML, Sql, y archivos de datos. Pueden presentar diferentes vistas de los mismos datos simultáneamente, así como incorporar múltiples fuentes de datos en una interfaz visual.

2.2.3 BrainEKP

BrainEKP, de la empresa de software TheBrain Technologies, es un sistema de organización de información que permite enlazar archivos, documentos y páginas Web entre aplicaciones y redes mostrando las relaciones entre ellos dentro de un entorno visual. Para ello se basa en las palabras clave que comparten los documentos.

BrainEKP utiliza una representación gráfica de información y de sus relaciones. Su interfaz de usuario permite que cualquier repositorio sea representado de una manera uniforme. Esta uniformidad permite acceder a cada artículo distinto y manipularlo de la misma manera, sin importar la fuente [Kirk 2004].

Los conectores de BrainEKP permiten la incorporación virtual de cualquier fuente de datos. Los conectores se ligan con la información de los repositorios y dicen al sistema BrainEKP qué relaciones existen en el repositorio. Al seleccionar un documento muestra todas las carpetas en las cuales aparece y en la que se relaciona el documento. Para la administración de base de datos y documentos, los conectores se crean en la base de datos Lotus Notes.

BrainEKP proporciona una interfaz gráfica que permite una navegación fácil y permite agrupar archivos en general, documentos y páginas Web mostrando las relaciones entre ellos dentro de un entorno visual. Para ello se basa en las palabras clave que comparten los documentos [Kirk 2004].

BrainEKP se representa por un término (llamado *thought*) en el centro y alrededor se muestran los documentos relacionados con él; si se pulsa sobre uno de los documentos éste pasa al centro y se reestructura la interfaz. Además de poder implantarlo en cualquier sistema. TheBrain ha creado su propio directorio, WebBrain, un buscador que presenta los resultados de la búsqueda en forma de un mapa. Sin embargo, para utilizar este sistema hay que pagar una licencia y únicamente recibe los datos de la base de datos Lotus Notes.

2.2.4 MAPA

La iniciativa de crear MAPA surgió en el año 1996 en IBM, empresa que en ese momento tenía un sitio Web de más de 75 mil páginas repartidas en una docena de URLs. Para manejar tal cantidad de información decidieron crear un sistema que pudiera hacer un mapa del sitio leyendo los enlaces, por lo que utilizaron el modelo desarrollado en la empresa de software dynamic diagrams a cargo de Kris Lenk, con él representaron las relaciones jerárquicas del sitio Web. Se hizo el software necesario para organizar la estructura de enlaces en una jerarquía, y se implementó un modelo visual en Java [Durand & Kahn, 1998].

El punto fuerte de MAPA es la transición animada de una vista a otra, porque permite mostrar una vista de la parte "recortada" del sitio completo y la parte del sitio a la que pertenece jerárquicamente la página; al moverse de una vista a otra es necesario que la transición sea animada para poder comprender la acción.

El sistema permite insertar etiquetas *meta* en los documentos para facilitar la organización en el mapa o para editar la estructura de enlaces en una base de datos.

Pasando el ratón sobre ellas se revela el título de cada página y pulsando el ratón en una página, se muestra el mapa jerárquico de esa página (si lo hay). Pulsando dos veces se traslada a la página en formato HTML.

Como ejemplos se muestran los mapas implementados en la enciclopedia Britannica Online, el sitio de IBM y al propio de Dynamic Diagrams.

2.2.5 WebTracer

WebTracer, es un Software desarrollado por la empresa NullPointer a cargo de Tom Betts para organizar documentos en HTML de un sitio en Web y permitir al usuario obtener información acerca de la estructura de un sitio, analizando las intenciones y principios utilizados en la construcción y diseño del mismo.

WebTracer es un proyecto basado en la visualización de la estructura de la red. Existen muchas aplicaciones que analizan la integridad estructural de los sitios, pero muy pocas revelan la estructura visual que el hipertexto crea en la Web. Webtracer representa esta estructura como un diagrama molecular en 3D, mostrando las páginas como nodos que se conectan entre si. Dentro del entorno de esta aplicación, su estructura puede ser examinada desde cualquier ángulo y distancia. Los nodos pueden ser seleccionados para acceder a más información acerca de su contenido, y puede ser ejecutado un navegador para mostrar el documento directamente. Este tipo de visualización refleja claramente la secuencia de vínculos entre las páginas y a la vez, la estructura interna del directorio del servidor a fin de obtener la organización más clara posible.

La licencia de este sistema es gratuita, sin embargo una de las desventajas de este sistema es que no cuenta con la posibilidad de crear una conexión para una base de datos que pueda ser utilizado para interactuar con el sistema.

2.2.6 EVA2D

Desarrollado en Java, EVA2D es un sistema de exploración visual basado en los conceptos de consultas dinámicas y visualización en campos de estrellas aplicados a una colección de libros [Sánchez et al. 2005].

Con EVA2D se obtiene información global sobre una colección así como información específica de cada libro; con la ventaja de que se pueden realizar incontables tipos de consulta en una sola pantalla y con retroalimentación constante sobre los resultados de tales consultas.

2.2.7 HyperTree Java Library

HypertTree es un software de código abierto en *applet* de java, muy similar a StarTree de Inxight. Permite ligar archivos, documentos y las páginas Web. HyperTree ofrece un API para que pueda ser ajustado al sistema en que se vaya ser implementado. Este sistema tiene la posibilidad de establecer una conexión con una base de datos, lo cual es muy importante debido a que puede ser una herramienta que califique para ser implementado en la herramienta para la creación de un repositorio institucional.

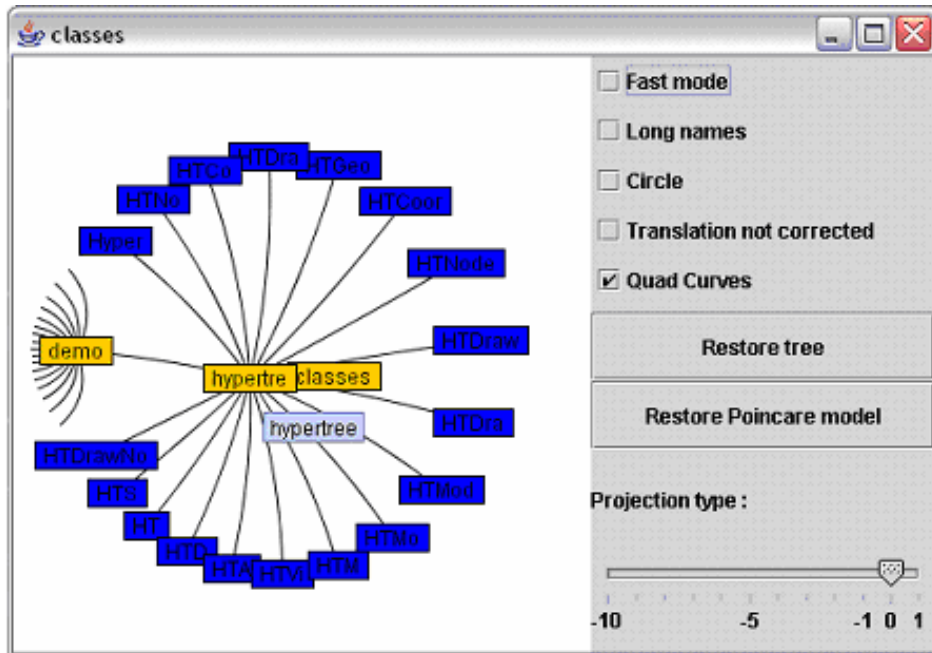


Figura 2.1 Sistema HyperTree Java Library generado en Appletviewer

Software	Creador	Se ejecuta en:			Distribución Libre	Lenguaje de Programación		¿Corre en plataforma Web?	Se puede modificar su código	Compatible con base de datos			Tipo de Visualización	
		Windows	Linux	Unix		Si	Java			OpenGL	Si	Si		Mysql
StarTree	Inxight	X	X	X		X		X					X	Árbol hiperbólico
TimeWall	Inxight	X	X	X		X		X					X	Pared tridimensional
TableLens	Inxight	X	X	X		X		X					X	Tabla tabular tridimensional
Thinkmap	Plumb Design	X	X	X		X		X		X				Árbol hiperbólico
BrainEKP	The Brain	X	X	X		X		X	X					Árbol hiperbólico
MAPA	Dynamic Diagrams	X	X	X		X		X	X				X	Diagrama tridimensional
WebTracer	Tom Betts	X			X		X				X			Árbol hiperbólico
EVA2D	Nabani Silva	X	X	X		X		X	X				X	Campo de estrellas
HyperTree	Christopher Bouthier	X	X	X	X	X		X	X				X	Árbol hiperbólico

Tabla 2.2 Comparación entre sistemas para visualizar grandes cantidades de información.

De acuerdo a las comparaciones realizadas en la Tabla 2.2 de las herramientas para visualizar grandes cantidades de información, observamos que HyperTree obtiene las mayores características para poder ser implementada en una herramienta para la construcción de repositorios institucionales. Sin embargo, no se descarta la posibilidad de que haya alguna herramienta de visualización que sea prominente como concepto o como técnica, pero debido a la limitante del tiempo, únicamente nos concentraremos en evaluar el sistema de visualización HyperTree.

2.3 Tendencias en representación de información

Podemos afirmar que las ventajas hacia las que deben dirigirse las nuevas presentaciones de información deberían ser las siguientes:

- Mayor habilidad para reunir una gran cantidad de información en un espacio reducido.
- Mayor capacidad de revelar las relaciones semánticas entre los documentos.
- Mayor facilidad para la búsqueda de documentos.

El resultado de las investigaciones en recuperación de información ha demostrado que la presentación global con interfaces que permiten la exploración de datos, facilita a los usuarios el reconocimiento de información relevante y permite interactuar con una gran cantidad de información en un espacio reducido de presentación.

Algunos programas de los anteriormente descritos ya optan por presentar el contenido y la estructura de los sitios Web con técnicas de visualización en lugar de ofrecer el tradicional índice jerárquico por hipervínculos.

La tendencia, por tanto, parece ir encaminada hacia una navegación que consiste en desarrollar una representación gráfica de la información almacenada, independientemente de su volumen o complejidad. La idea básica es encontrar la información a través de un mapa navegable donde prácticamente se pueda ver todo en una sola pantalla.