

VII.3 BIBLIOTECA AUTOMATIZADA

*Fernando Galindo Soria**

INTRODUCCIÓN

Cuando se observan los trabajos desarrollados en el área de Automatización de Bibliotecas, se detecta que la mayoría de estos trabajos versan realmente sobre la automatización de un catálogo de títulos, autores, etc., y sobre la explotación automática de estos catálogos y con esto el concepto de Biblioteca Automatizada se queda en el limbo, ya que lo que realmente se automatiza es la parte administrativa-operativa del manejo de información bibliográfica y la biblioteca en sí no es tocada.

En este trabajo se presenta una propuesta de automatización de bibliotecas, en la cual se plantea que la biblioteca debe estar dentro de la computadora y no en estantes o libreros. Esta idea no es nueva y realmente en la Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas (UPIICSA) del Instituto Politécnico Nacional (IPN) se han desarrollado algunos prototipos bajo este concepto, sin embargo, la mayoría de las personas involucradas en automatización de bibliotecas no la contemplan principalmente por dos motivos:

- 1) Rompe completamente con el esquema tradicional de una biblioteca, donde los libros se encuentran en libreros y se tiene un sistema de alta complejidad para localizarlos y mantenerlos en su lugar (es por eso, que la mayoría de aplicaciones de automatización de bibliotecas tienden a automatizar el proceso de catalogación).

* Fernando Galindo Soria IPN-UPIICSA

2) *No es fácil conceptualizar a una biblioteca dentro de la computadora* y aún más se marcan una serie de problemas para su logro. Un libro ocupa mucho espacio y no cabría en la computadora, esto ya no es un argumento si se considera que un libro típico tiene 150 páginas, 40 renglones por página, 50 caracteres/renglón = 300,000 caracteres lo cual cabe sin ningún problema por ejemplo en un disco flexible. Por lo que si tomamos en cuenta que una computadora típica actual puede tener 2 Mb en memoria principal y 1000 Mb en disco, la capacidad de almacenamiento es sin problema la de una biblioteca mediana (alrededor de 3000 títulos deferentes).

Otro problema que se marca mucho es el de la captación de la información, sin embargo, este problema es en sí una falacia, ya que la mayoría de las editoriales actuales capturan para su edición los libros en algún sistema de computo y posteriormente lo imprimen en papel (en la actualidad sale tan caro un ciento de hojas de papel como un disco flexible) por lo que en su momento se podría almacenar directamente el libro en la computadora, por otro lado, en el área de reconocimiento de imágenes se tienen desarrollos orientados a la “captura visual” de libros.

Poca gente tendría acceso a este tipo bibliotecas, esta es una verdad/mentira, ya que en la actualidad poca gente tiene un acceso real a cualquier tipo, de biblioteca y los libros son cada vez más caros por otra parte con introducción de las computadoras en las bibliotecas públicas (Proyecto Computación para Todos los Niños) y en las escuelas por un lado y por el otro con el acceso casero a la computadora, en algunos lugares, cada vez más se contará con herramientas para un acceso masivo.

Sin embargo, el argumento más fuerte en contra de este tipo de desarrollo se centra en la no disponibilidad de herramientas de programación orientadas a manejar una biblioteca automatizada. Es precisamente el objetivo de este trabajo, mostrar que una herramienta de este tipo es relativamente fácil de construir ya que se cuenta prácticamente con todos los ingredientes del menú y lo que se requiere es combinarlos en un buen platillo.

I ARQUITECTURA DEL SISTEMA

Al preguntarle a varias personas cómo se imaginan la interacción con una biblioteca automatizada, se ha llegado a la conclusión de que el diálogo con esta debe ser en Español y contener oraciones como:

- 1) *Dame los libros de Físicoquímica*
- 2) *Mira máquina quiero que me des la lista de todos los libros que escribió Cervantes*
- 3) *Imprime el contenido de "Suave Patria"*
- 4) *Dame tres temas relacionados con Histología*
- 5) *Dame los principales temas de Topología*
- 6) *Dame los fragmentos del libro de Biología Celular donde hablan de los átomos*
- 7) *Dame todos los fragmentos que tengas sobre Cuernavaca.*
- 8) *Dame diez fragmentos de temas relacionados con Pulsares*

Como se ve, este tipo de peticiones al sistema cubren una gama muy amplia y van desde un simple listado de catálogo hasta un acceso a bases de conocimiento con el fin de encontrar los temas relacionados con un tema dado y posteriormente un acceso a base de datos y la Biblioteca con el fin de acceder sólo los fragmentos solicitados.

En general la arquitectura de un Sistema de Biblioteca Automatizada capaz de soportar una interacción, en Lenguaje Natural y almacenar la biblioteca dentro del sistema es la que se muestra en la Figura 1.

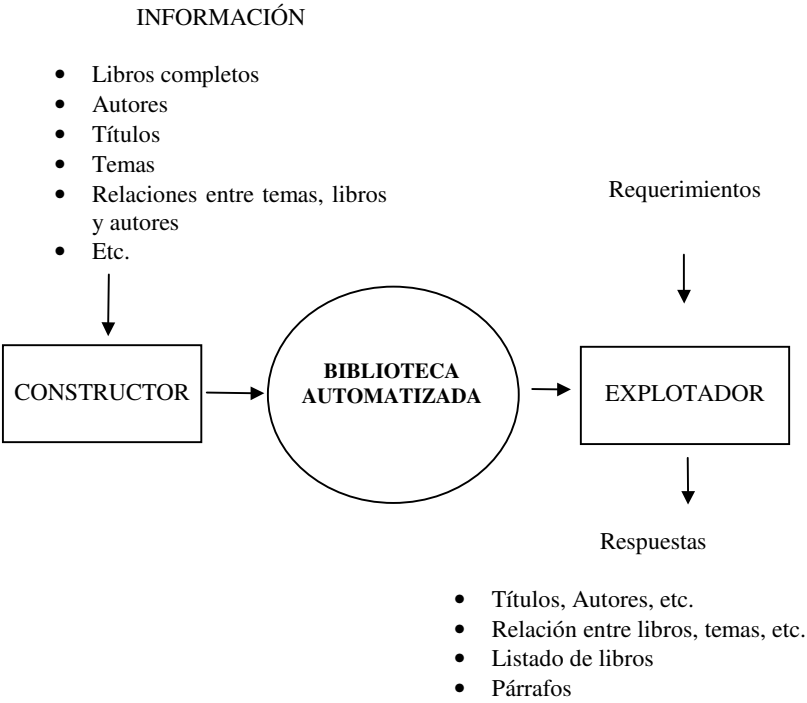


Figura 1. Componentes de un Sistema de Biblioteca Automatizada

El esquema de la Biblioteca Automatizada se muestra en la figura 2.

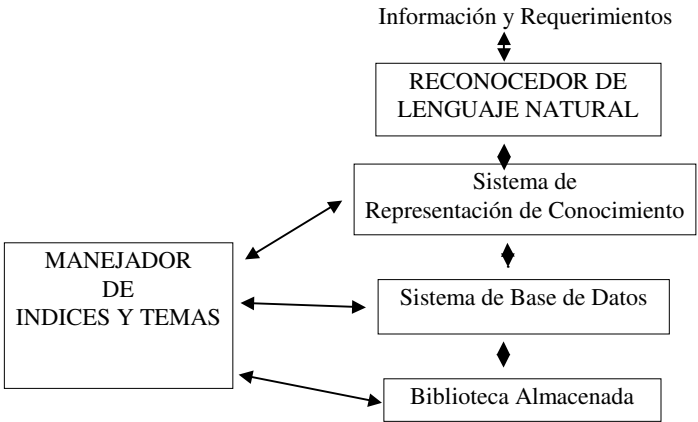


Figura 2. Esquema de la Biblioteca Automatizada

II RECONOCEDOR DE LENGUAJE

El reconocedor de Lenguaje Natural, es el responsable de detectar a partir de la información ó requerimientos del usuario, los entes y acciones involucradas.

Por ejemplo en la oración:

Dame todos los párrafos que tengas sobre temas relacionados con Antropología.

- 1) El sistema detecta una acción: **Dame**
- 2) Detecta un cuantificador: **todos**
- 3) Detecta lo que se quiere (párrafos, libro, área, etc.): **párrafos**
- 4) Detecta que no quiere el tema principal sino: **los temas relacionados**
- 5) Detecta el tema principal: **Antropología**

Este tipo de reconocedor es relativamente fácil de construir utilizando las técnicas de Programación Dirigida por Sintaxis.

III SISTEMA DE REPRESENTACIÓN DE CONOCIMIENTO

Ya que se detectaron los entes y acciones involucradas y para dar respuesta al requerimiento se interconectan tres niveles de almacenamiento de la información:

- 1) *El primero de ellos o Sistema de Representación de Conocimiento*, se tiene un mecanismo que nos permite almacenar conocimiento de la biblioteca a nivel de hechos y reglas de inferencia.

Para aclarar el punto, de las oraciones que siguen, se construye una red semántica que se muestra en la figura 3: La Citología está relacionada con la Histología. La Histología estudia los tejidos. La Citología estudia las células. Los tejidos están compuestos de células. La Histología es parte de la Biología. Biología Celular es un libro de Biología. C.U.M. Smith es el autor de Biología Celular.

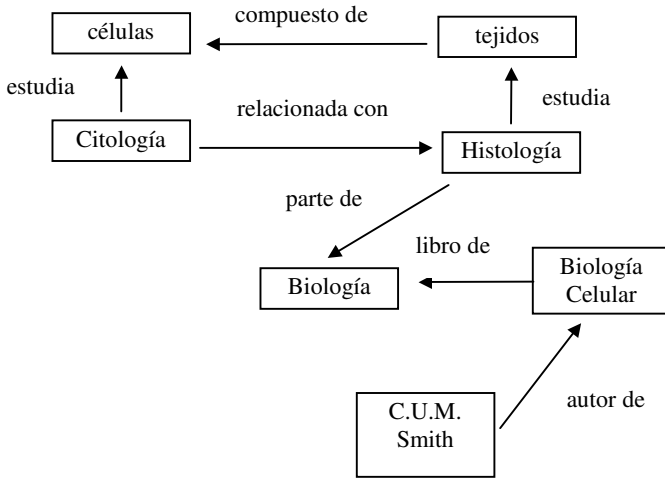


Figura 3. Una Red Semántica

Por lo que si alguien pregunta:

Dame un libro de Biología

el sistema puede responder: **Biología Celular**

Con la pregunta:

Dame un tema relacionado con Citología

el sistema puede responder: **Histología.**

Por otra parte si a lo anterior se le aumentan reglas de inferencia, se cuenta ya con una herramienta muy interesante.

Por ejemplo, si se agrega:

Si X es parte de Y, X estudia Z, entonces Y estudia Z

Si X está relacionado con Y entonces Y está relacionado con X

Si X estudia y entonces X está relacionado con Y

Si X está compuesto de Y entonces X esta relacionado con Y

Si X es parte de Y entonces X esta relacionado con Y

Por lo tanto, la pregunta:

Dime los temas relacionados con Biología
da como respuesta.

Histología **Tejidos**

IV BASE DE DATOS

Este primer nivel de almacenamiento es extremadamente poderoso y en su momento nos permite dar respuestas a múltiples interrogantes en las que participan distintos entes involucrados.

Sin embargo, y si recordamos el modelo de Chen de Base de Datos (de Entidad/Relación), la realidad se puede conceptualizar en términos de entidades relacionadas entre sí, donde cada entidad es a su vez descrita mediante un conjunto de atributos y hasta el momento, lo único que hemos manejado ha sido a las entidades y sus relaciones y no se ha involucrado a los atributos, por lo que oraciones como las siguientes:

- 1) *Pedro Pérez nació en la ciudad de Guadalajara*
- 2) *Pedro Pérez vive en cerritos #38, col. Chapulín*
- 3) *Pedro Pérez nació el 20 de enero de 1950*

no se han manejado.

Como un primer enfoque "se siente" que se podría aumentar la Base de Conocimientos, sin embargo, desde un punto de vista práctico, tal vez no sea la mejor solución, ya que los atributos de un ente pueden ser muchos y en su momento sólo se refieren a ese ente en particular por lo que no se requiere ni la capacidad de relacionar todo contra todo, ni los mecanismos de inferencia.

En particular, en este trabajo se plantea la introducción de un segundo mecanismo de manejo de información fuertemente relacionado con el primero.

Este mecanismo es un sistema manejador de Base de Datos de tipo Relacional.

Mediante este enfoque, se detectan los principales tipos de Entes involucrados en el sistema (en este caso pueden ser: autor, libro, editorial, etc.) y para cada caso, sus principales atributos, por ejemplo en el caso de autor pueden ser: nombre, fecha de nacimiento, lugar de nacimiento, estudios, etc.; y a partir de estos se puede diseñar la Base de Datos (figura 4), en donde a cada entidad le correspondería una parte del esquema.

AUTOR			LIBROS			
NOMBRE	FECHA-NAC	LUGAR-NAC	AUTOR	TITULO	EDICION	PAGINAS
Juan Pérez	20-ene-50	Guadalajara				

EDITORIAL			
NOMBRE	DIRECCION	TELEFONO	DIRECTOR

Figura 4. Ejemplo de la Base de Datos

mediante este mecanismo, los atributos de Juan Pérez quedan en la Base de Datos, por lo que si alguien pregunta:

¿Dónde nació Juan Pérez?

El sistema responde: **En Guadalajara**

Por otra parte, los dos mecanismos se encuentran fuertemente relacionados, ya que si por ejemplo se almacena:

Juan Pérez escribió “La Vida es Verde”

esta información se almacena en la Base de Conocimientos y si más adelante alguien pregunta:

1) *¿Quién escribió La Vida es Verde?*

el sistema contesta: **Juan Pérez** (tomándolo de la Base de Conocimientos).

2) ¿Dónde nació el autor de “La Vida es Verde”

En este caso el sistema primero tiene que ir a la Base de Conocimientos para enterarse que Juan Pérez escribió “La Vida es Verde” y con esta información accesa la Base de Datos para saber que Juan Pérez nació **en Guadalajara**, que es lo que responde.

V BIBLIOTECA ALMACENADA

Sin embargo, y a pesar de estos mecanismos, aún no se cuenta con una Biblioteca Automatizada, ya que los libros aún no están en el sistema pero esto se puede arreglar introduciendo el tercer mecanismo, con el cual se pretende almacenar precisamente los libros.

Este mecanismo es un almacén o Biblioteca Almacenada donde cada libro queda físicamente guardado.

Por lo que si alguna persona pide el contenido de un libro en particular, lo único que se requiere es desplegarlo.

Ahora bien, para lograr lo anterior, nuevamente los diferentes mecanismos deben estar fuertemente relacionados.

Por ejemplo, almacenado como atributos del libro dentro de la Base de Datos los siguientes:

- 1) *Nombre del Archivo Almacén*
- 2) *Posición Inicial del libro en el archivo*
- 3) *Posición Final de caracteres del libro*

con lo cual, si se pide un libro en particular se puede localizar.

Hasta este punto se puede decir que ya se cuenta con una Biblioteca Automatizada, sin embargo, es ilógico desaprovechar tan bonita oportunidad y quedarnos sólo en el umbral de lo que se puede obtener ya que, sí se involucra un nuevo mecanismo, el potencial del sistema aumenta considerablemente.

VI MANEJADOR DE ÍNDICES Y TEMAS

Este mecanismo es un manejador de índices y temas.

Mediante este mecanismo se puede dar respuesta a planteamientos como los siguientes:

- 1) *Dime en que partes del libro de Biología Molecular se estudian las células*
- 2) *Dame los párrafos del libro Biología Molecular que tratan de las células*
- 3) *Dame todos los párrafos que tengas sobre Vicente Guerrero*

En principio un mecanismo de este tipo está formado por un directorio o índice, donde para cada tema aparecen los libros donde se encuentra y se indica en que partes del libro se menciona el tema.

Al integrar este mecanismo con los anteriores, el potencial crece ya que nos permite un acceso prácticamente a la información y realmente **el concepto de libro desaparece** ya que lo que se tiene es un gran almacén de información con múltiples mecanismos de acceso a diferentes niveles de especificidad.

Uno de los problemas planteados a este último mecanismo es el de la dificultad para obtener y almacenar este tipo de información, sin embargo, ésta no es una limitante real ya que la mayoría de los libros trae un índice por capítulos y muchos cuentan con índices por temas.

Por otro lado, esta es otra de las ventajas de una Biblioteca Automatizada, ya que al tener los libros en la computadora se puede desarrollar una herramienta que permita listar todas las palabras diferentes con las se cuenta y su posición en los diferentes documentos (en su momento se le puede dar una lista de palabras relevantes para que no las almacene, y otra de palabras compuestas v.g.: Buenos Aires que es, importante manejar en forma compuesta).

CONCLUSIÓN: MEMORIA AUTOMATIZADA

En el punto anterior se menciona que el potencial de la herramienta trasciende al de una Biblioteca Automatizada y a lo que realmente se llega es a un Mecanismo para manejar información masiva a múltiples niveles de especificidad.

Este es un punto sustancial del trabajo, ya que lo que se plantea es un mecanismo que lo mismo permite manejar una Biblioteca general, que una especializada (Médica, Legal, Matemática, etc.) y con potencial de darnos acceso a la información específica, sin importar la fuente de acceso original, con lo que se cuenta con un mecanismo de representación de conocimiento capaz de entregarnos, por ejemplo, toda la información relacionada con cierto tipo de enfermedad, o también sobre un tema dado, con lo cual, el proceso tedioso de investigación documental se puede soslayar.

Aún requerimientos que se le hagan y no encuentre datos, los puede almacenar y solicitar a algún especialista la respuesta, que estaría disponible eventualmente.

En su extremo más interesante, un sistema de este tipo nos puede contestar todo lo que sabe sobre cualquier tema del que tenga información, independientemente de la fuente de consulta o de la fecha en que fue almacenada, ya que integra todo en una sola gran memoria (figura 5).

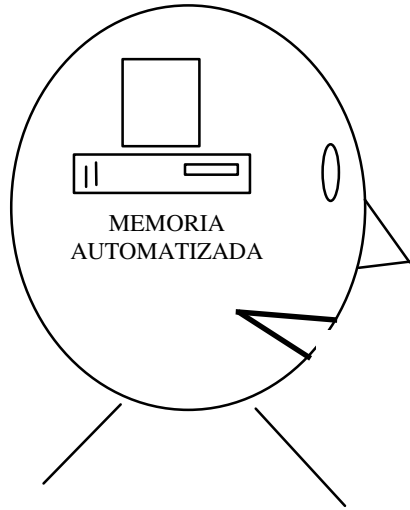


Figura 5. Conclusión: Memoria Automatizada