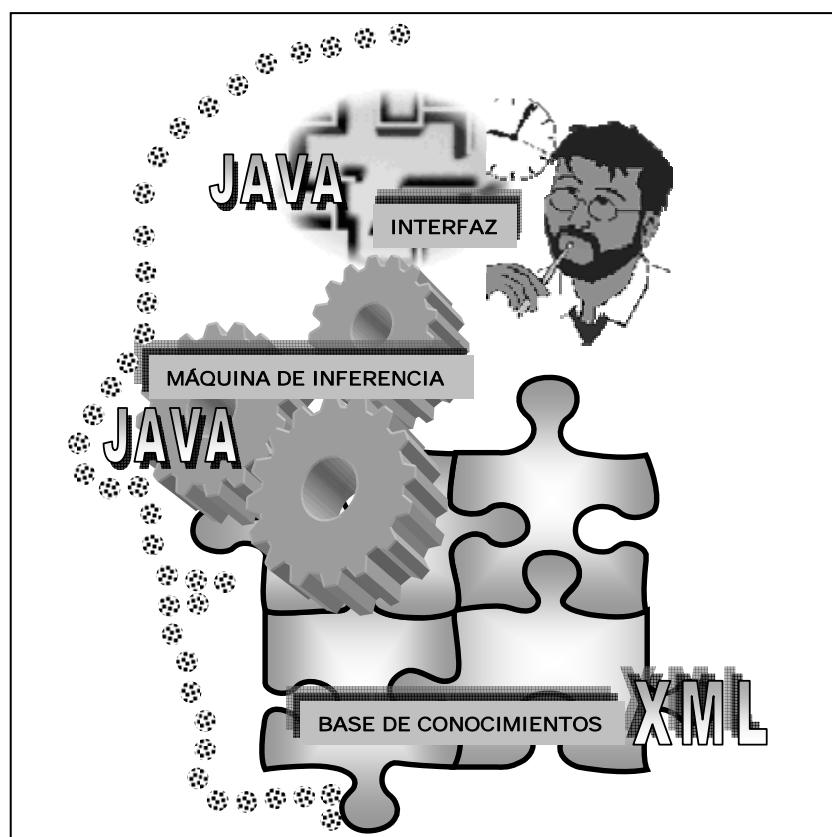


GESTIÓN DEL CONOCIMIENTO UTILIZACION DE XML PARA BASES DE CONOCIMIENTOS

XML y derivados: Estándares para un paradigma de Bases de Conocimientos



Omar Soriano Zárate
Licenciado en Informática
Instituto Tecnológico de Orizaba
Veracruz México

Ma. Antonieta Abud Figueroa
Maestra en Ciencias
Instituto Tecnológico de Orizaba
Coordinadora de proyectos de investigación de
sistemas y computación

RESUMEN

En la actualidad solamente un pequeño número de métodos de Inteligencia Artificial son útiles en los sistemas expertos, es decir, hay actualmente sólo un puñado de maneras para representar el conocimiento, o hacer deducciones, o generar explicaciones. El artículo describe una manera más de representar el conocimiento a través de un estándar abierto tipo XML.

XML es un nuevo estándar de almacenamiento semiestructurado en el que, a diferencia del HTML, se pueden definir etiquetas o marcas personales por medio de una DTD (las Definiciones de Tipo de Documento o Restricciones de Domino) o mediante el uso de Schema.

Nuestro objetivo fundamental es exponer la representación de una base de conocimientos mediante la utilización de XML y lenguajes de consulta del

mismo, como pueden ser: Xquery, Xpath, XSD y XSLT.

PALABRAS CLAVES

XML, Bases de Conocimiento, Sistemas Expertos

ABSTRACT

Actually only a small number of methods of artificial intelligence there're useful in the expert system, is telling, there're actuality only an handful of shapes for represent the knowlege, or to make deductions, or to generate explains. The article describe an shape more of represent the knowlege for an open standar kid XML.

XML is new standar of warehousement semistructure in the that diferent of HTML, it can define labels o marks persons for an DTD (the definitions of kid of documents o to keep out of dominion) or to used of Schema.

Our objective basic is showing the representation of knowlege, using of XML and languajes of consult selves, how can be. Xquery, Xpath, XSD y XSLT.

INTRODUCCIÓN

Los programas de Inteligencia Artificial que logran la capacidad experta de solucionar problemas aplicando las tareas específicas del conocimiento se llaman Sistemas Basado en Conocimiento o Sistemas Expertos. La construcción de un sistema experto se llama ingeniería del conocimiento y sus diseñadores son los ingenieros del conocimiento, el ingeniero del conocimiento debe cerciorarse de que el ordenador tenga todo el conocimiento necesario para solucionar un problema.

También debe elegir una o más formas en las cuales representar el conocimiento requerido en la memoria del ordenador, es decir, él debe elegir una representación del conocimiento, es aquí donde colocamos nuestro objetivo, el representar el conocimiento de un sistemas experto mediante XML y lenguajes de consulta del mismo, como pueden ser: Xquery, Xpath, XSD y XSLT y su manipulación mediante una maquina de inferencia realizada en JAVA.

1. SISTEMA EXPERTO

Los sistemas expertos forman parte de un firme y verdadero avance en inteligencia artificial, y la tarea principal de un sistema experto es mediante la realización de una serie de preguntas a una persona y la repuesta de ésta.

El sistema realiza la ardua labor de recorrer miles de reglas de conocimientos previamente

incorporados, hasta dar con la respuesta a fin del problema, o en su defecto, la más parecida a ésta.

De esta manera el sistema Experto trata de imitar el comportamiento de un experto humano en la solución de un problema. Cabe mencionar que en un sistema experto se pueden almacenar conocimientos de expertos para un campo determinado y solucionar un problema mediante deducción lógica de conclusiones.

La función de un Sistema Experto es la de aportar soluciones a problemas, como si de humanos se tratara, es decir capaz de mostrar soluciones inteligentes. Y esto es posible gracias a que el sistema lo crean con expertos (humanos), que intentan estructurar y formalizar conocimientos poniéndolos a disposición del sistema, para que este pueda resolver una función dentro del ámbito del problema, de igual forma que lo hubiera hecho un experto.

Acceder a los conocimientos adquiridos por experiencia es lo más difícil, ya que los expertos, al igual que otras personas, apenas los reconocen como tales. Son buscados con mucho esfuerzo y cuidado siendo descubiertos de uno en uno, poco a poco.

Una característica decisiva de los Sistemas Expertos es la separación entre conocimiento (reglas, hechos) por un lado y su procesamiento por el otro. A ello se añade una Interface de usuario y un componente explicativo fig1.

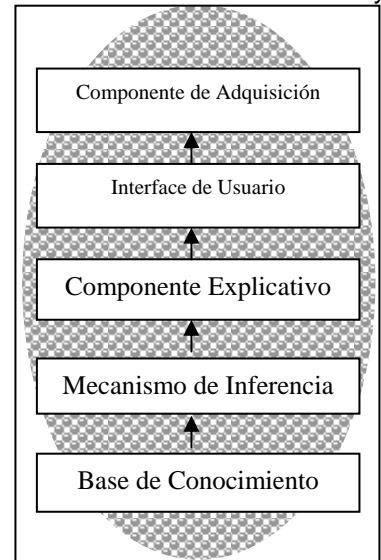


Fig. 1 componentes principales de un sistema experto

2. HERRAMIENTAS DE DESARROLLO

2.1. XML (eXtensible Markup Language)

[1] Hoy en día XML se ha convertido en uno de los estándares más conocidos, estamos ante un típico caso de una idea sencilla pero que abre perspectivas completamente nuevas.

XML no es un lenguaje más, es un metalenguaje, es decir, un lenguaje para poder definir lenguajes, gracias a esta característica, y a otras más que le acompañan, ha sido posible desarrollar una colección de lenguajes "XML-based" que se están convirtiendo en estándares internacionales para facilitar una nueva manera de desarrollar sistemas de información y de integrar los ya existentes, y si hablamos de sistemas de información por qué no aplicarlo a Sistemas Expertos.

Bien vale la pena mencionar los diferentes tipos de normas del estándar XML[2].

- Las extensiones del XML original (XML Schema, XPath, XLink, Xpointer, familia XSL, CSS2, Xquery, DOM, SAX, Namespaces, XInclude, XBase, ...) que le confieren la gran potencialidad que tiene actualmente.
- Los lenguajes horizontales que aportan cierta funcionalidad que es útil en general: seguridad (XML DSig, XML Enc, SAML, XACML, XKMS), servicios web (SOAP, WSDL, UDDI), metadatos y conocimiento (RDF, XMI), formularios web (Xforms), y otros (SVG, DC).
- Los lenguajes verticales dirigidos a un sector o utilidad particular.

En nuestro caso únicamente se van a utilizar las siguientes extensiones del XML original: XML, Schema, Xpath, XSL, DOM, SAX, Namespaces.

2.2. JAVA

[3] Java es un lenguaje de cuarta generación orientado a objetos que ofrece grandes ventajas, brinda acceso uniforme a bases de datos, Java no sólo provee la enorme ventaja de "escríbelo una vez, ejecútalo donde quieras", si no también permite integración con componentes heterogéneos, un modelo de seguridad que protege la información, permite crear aplicaciones escalables, confiables e integradas con los sistemas existentes en tiempos cada vez más cortos.

3. DESARROLLO

[4] En el sistema experto que se desarrollará se planteará la base de conocimientos en un conjunto de Reglas, Casos y procedimientos, a

continuación se mencionara en que consiste cada uno de estos:

Conocimiento en formas de reglas:

- Una regla tiene resultados o unos antecedentes y unos resultados o consecuencias
- Si los antecedentes se cumplen, se deduce el resultado

Conocimiento en formas de casos:

Se dice que un caso puede presentarse como una situación en especial, un problema, un conjunto de alternativas, una solución, unos resultados o en su defecto unas lecciones aprendidas.

[5] Identificamos los principales servicios de una base de conocimientos los cuales se mencionan a continuación:

- **Aprendizaje:** primero una base de conocimientos se debió de haber almacenado con el conocimiento de un experto.
- **Almacenaje:** el conocimiento se debe de almacenar bajo alguna estructura, para su posterior manipulación.
- **Análisis de similitud:** es donde se realizará el proceso para la maquina de inferencias.
- **Recuperación:** el conocimiento almacenado se empezará a utilizar y por lo tanto se estará recuperando continuamente.
- **Adaptación:** el conocimiento se adaptara al tipo de usuario con el que se esté tratando.

El comportamiento de lo antes mencionado se ilustra en la fig. 2.

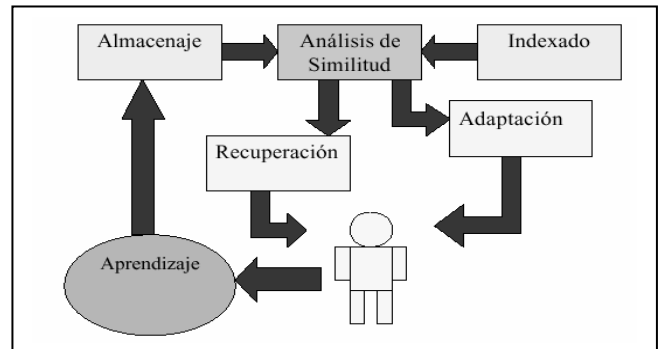


Fig. 2. Servicios de una base de conocimientos

Para la implementación de esta base de conocimientos, tomando en cuenta que el sistema experto a desarrollar debería ser multiplataforma, se decidió que se iba a utilizar tecnologías simples, que fuesen libres y fáciles de implementar.

[6] En primera instancia se planteó la base de datos bajo un modelo relacional, pero lo que se requería que no fuese dependiente de una tecnología en especial, por lo que se determinó que el estándar apropiado era XML (eXtensible Markup Language), por las características ya mencionadas, entonces se planteó una NXD (native XML DB).

Una NXD (native XML DB) es una alternativa limpia y sencilla de publicación de datos desde y hacia bases de datos relacionales. Ahora pensando en futuro que los conocimientos almacenados previamente pueden llegar a incrementarse, entonces NXD (native XML DB) se adecua a ser la herramienta para lograr este soporte.

Ahora bien las NXD no son bases de datos verdaderamente XML, en el sentido de que almacenan información en formato XML, son sólo un modelo (lógico) de un documento XML –en contraposición con los datos que un documento contiene- y almacena y recupera los documentos según este modelo. A lo mínimo, el modelo debe incluir elementos, atributos, PCDATA, y un orden para el documento, para ello es necesaria la utilización de el siguiente modelo Xpath, y los modelos implicados en el DOM y los eventos en SAX.

De esta manera el almacenar los hechos y las reglas en una NXD lleva consigo a la realización de un DOM, es decir nuestra base de conocimientos jugará el papel de un DOM, ya que se tiene un documento XML como unidad fundamental del almacenamiento (lógico), de igual forma que una base de datos relacional tienen la fila en una tabla como su unidad fundamental del almacenaje (lógico).

De esta manera con este modelo no se requiere tener ningún modelo físico de almacenamiento subyacente. Y se pudo haber construido sobre una base de datos relacional, jerárquica, u orientada a objetos, o utilizar un formato de almacenaje propietario tal como archivos índices o archivos comprimidos.

El modelo incluye niveles arbitrarios de anidamiento y de complejidad, así también como soporte de completo para datos de contenido mixto y datos semi-estructurados.

El modelo es automáticamente mapeado por la NXD dentro del mecanismo de almacenamiento (base de datos relacional, en ese caso.) El mapeo usado asegura que el modelo específico de datos XML se mantiene.

Ahora bien la NXD abstrae todos los detalles de cómo almacenan los hechos en XML, y lo único que necesitaremos para su manipulación de utilizar herramientas tales como SAX, Xpath y XSL-T.

[7] Xpath es el lenguaje de interrogación para NXDs. Para funcionar como un lenguaje de interrogación de NXD, Xpath se amplía ligeramente para permitir consultas a través de colecciones XML. Desafortunadamente Xpath realmente no fue diseñado como lenguaje de consultas de Bases de Datos y se encuentra muy limitado, debido a esto Xpath necesita ser ampliado como parte de un lenguaje más poderoso. Muchas de las carencias pueden ser resueltas utilizando XSLT para complementar los agujeros, pero un lenguaje más orientado a Bases de Datos está en desarrollo, en la forma de Xquery. Ya que actualmente varios vendedores han comenzado a lanzar implementaciones prototipo de Xquery para el uso de sus Bases de Datos.

Pero si se necesitara mejorar el funcionamiento de las consultas realizadas a la base de

conocimientos, NXD soporta también la creación de índices para los datos almacenados en colecciones. Con la creación de estos índices se pueden mejorar la velocidad de ejecución del acceso a las consultas a la base de conocimientos.

En la figura 3 se muestra un pequeño

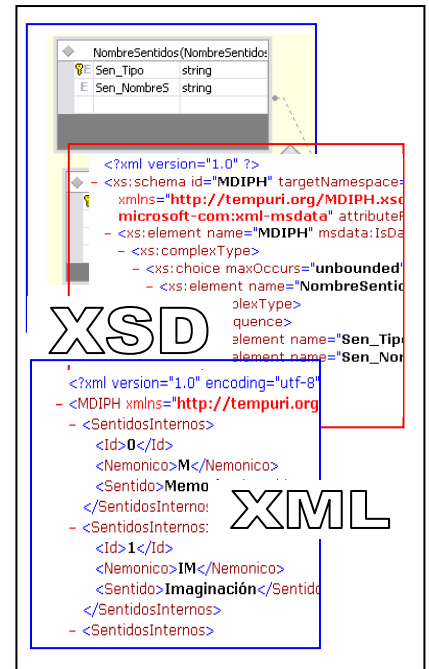


Fig. 3 Base de Conocimientos fragmento de la base de conocimientos.

Ahora bien hasta este momento hemos hablado sobre el punto específico que es la representación de los hechos en la base de conocimientos y los lenguajes para su manipulación, pero ahora vale la pena mencionar que sucede con JAVA.

[8] Como ya se mencionó anteriormente lo que se plantea es un sistema experto independiente de la plataforma, entonces el único lenguaje que me brinda esta ventaja actualmente es JAVA, que es en donde se realizará el motor de inferencia, aunque para el desarrollo efectivo de un sistema experto es posible hacer uso de un paquete adaptable, que es una aplicación preescrita en la cual se pueden realizar unos ajustes mínimos para adaptarla a unas necesidades concretas.

También es posible un desarrollo específico utilizando los lenguajes propios de la IA (Lisp y Prolog, generalmente) para construir un sistema con todos sus elementos y detalles (motor de inferencia, base de conocimientos y las interfaces), estos últimos no se adaptan a lo que se desea realizar, En este caso se trata de un sub-sistema que se elaborará un motor de inferencia, y se va a definir un único esquema de representación del conocimiento, también pre-fijado.

Así mismo en JAVA se realizarán las interfaces necesarias para obtener el conocimiento del experto humano e introducirlo en el la base de conocimientos de XML.

[9] En un inicio estamos planteando el motor de inferencia con un mecanismo de razonamiento de forward Chiang (encadenamiento hacia delante), por la misma naturaleza del modelo que se desea que valide el sistema experto.

Ahora describiremos en la figura 4 de manera global el modelo de sistema experto que estamos planteando.

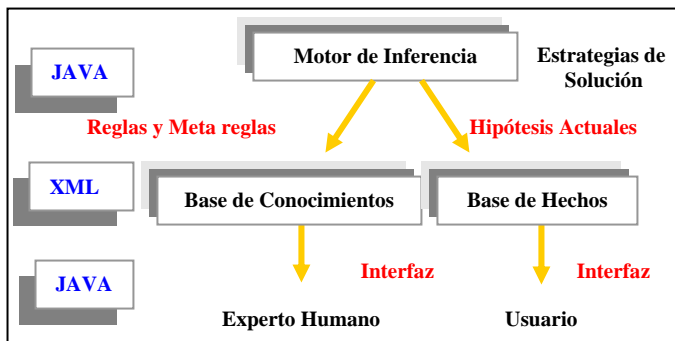


Fig. 4 Propuesta de Sistema Experto.

Los conocimientos se almacenan en la base de conocimientos y en la base de hechos que se encontrará en XML, y los procedimientos capaces de razonar se implementan en el motor de inferencia que estará en JAVA.

Es necesario la presencia de interfaces que permitan el acceso al sistema, ya sea por parte del experto humano que es el encargado de alimentar la base de conocimientos, como del usuario del sistema que proporciona los hechos que determinan una utilización concreta del sistema experto que al igual que el motor de inferencia, éstas se encontrarán en JAVA. El usuario obtiene las respuestas proporcionadas por el sistema que por las mismas restricciones de la interfaz del sistema experto que se requiere no es posible utilizar LISP o Prolog.

4. CONCLUSIONES

El por qué se optó por JAVA y XML para el desarrollo del sistema experto, la respuesta es sencilla, XML y Java son lenguajes que poco a poco tienen mayor aceptación y que se están convirtiendo en herramientas estándares en el mundo de desarrollo.

Así pues, el XML se está transformando en un medio vanguardista para la publicación electrónica en Internet al soportar y enriquecer al formato HTML, agregándole una mayor seguridad y versatilidad en el manejo de la información y, al mismo tiempo, Java ha resultado ser un gran apoyo para las mejoras de los sistemas aprovechando sus bondades, estándares y alto crecimiento.

La conjunción de XML, apoyándose en Java, será la punta de lanza para la publicación de contenido en multiplataformas a través de aplicaciones robustas en cuanto a sus capacidades, pero ligeras en cuanto al procesamiento. Ahora bien, tomando a Java como base principal, el uso de ambas será crucial para el desarrollo de aplicaciones diversas en medios de distribución, como son transacciones financieras, aplicaciones inalámbricas, aplicaciones cliente/servidor que permitan comunicar y compartir información, entre otras y ahora el otro paradigma que se está proponiendo que es enfocado a la inteligencia artificial, mas propiamente a Sistemas Expertos que es nuestro caso de estudio; las cuales apoyadas con XML permitirán visualizar en forma rápida cualquier cambio en el diseño o formato de la información, al utilizar DTDs (Document Type Definition) y esquemas, dando una mayor flexibilidad en el manejo de la estructura de datos, así como brindando ahorro de tiempo y dinero en el desarrollo y mantenimiento de los sistemas.

Ahora por lo que se refiere a la representación de la base de conocimientos en una NXD, cabe mencionar que las NXD son una alternativa limpia y sencilla de publicación de datos desde y hacia bases de datos relacionales.

En la actualidad esta tecnología cuenta con gran respaldo de los grandes en el negocio, que especialmente promete garantizar la compatibilidad entre diferentes formatos de almacenaje físico de datos, pero también por el nivel flexible de descripción de datos mediante el metalenguaje que es XML. Su facilidad de uso y fácil adaptación a las actuales condiciones en la WWW, la proyectan como la herramienta de elección popular.

Ahora si se llega a obtener los resultados esperados por que no decir que en un futuro podrían realizarse sistemas expertos que no únicamente se ejecuten en una PC de escritorio si no también en cualquier otro dispositivo móvil.

REFERENCIAS BIBLIOGRÁFICAS

- [1] http://www.baquia.com/index_com.html
[01 Febrero de 2004] XML, los Datos del futuro
- [2] <http://www.w3schools.com/default.asp>. La dirección de la escuela en línea de la W3C, la ultima palabra en todo lo relativo a normas y estándares para la Web.
- [3] JAVA
<http://www.javasoft.com/>
[Consulta: 18/03/04]
- [4] GARCÍA Martínez, R. 1994. Adquisición de Conocimiento. En Abecasis, S. y Heras, C. Metodología de la Investigación. 157 páginas. Editorial Nueva Librería. ISBN 950-9088-65-x.
- [5] DURKIN, John: Expert Systems. Design and Development 1994
Prentice Hall Inter-national.
- [6] <http://www.xmldb.org/XML:DB>
initiative for XML Databases
- [7] W3C
<http://www.ebxml.org/>
[Consulta: 23/03/04]
- [8] <http://java.sun.com/> Esta página Contiene todo lo que se desea saber acerca de java

[9] IBARREÑA Javier Mtz
<http://www.geocities.com/javierml.geo/doc/SistemasExpertos.html>
[Consulta: 20/03/04]



M.C. Ma. Antonieta Abud Figueroa

Maestra en Ciencias en Sistemas de Información por el ITESM Campus Morelos.

Actualmente es profesora del Instituto Tecnológico de Orizaba, y tiene a su cargo la coordinación de proyectos de investigación de sistemas y computación.

Áreas de interés: orientación a objetos, ingeniería de software y software educativo

Datos Generales:
Dirección: Sur 8 No. 605
Teléfono: 0127259316
Ciudad: Orizaba
Estado: Veracruz
País: México
Email: mabud@itorizaba.edu.mx



L.I. Omar Soriano Zárate

Licenciado en informática, egresado del Instituto Tecnológico de Orizaba

Actualmente se encuentra estudiando la maestría en ciencias en ciencias de Computación en el instituto tecnológico de Orizaba

Datos Generales:
Dirección Avenida Juárez No.243 Int. 5
Teléfono: 012721049074
Ciudad: Nogales
Estado: Veracruz
País: México
Email: omarsorianoz@hotmail.com