

BASE DE DATOS INTELIGENTE.

Ing. Juliet Díaz Lazo

Reserva Científica del Instituto Nacional de Ciencias Agrícolas. San José de las Lajas,
Mayabeque, Cuba.

juliet@inca.edu.cu

Resumen

Este artículo tiene como objetivo fundamental detallar el conjunto teórico mediante el cual pueda manipularse una base de datos relacional de manera transparente, extrayendo conocimientos a partir del análisis de estructuras simples, interrelacionando procesos de adquisición de datos con procesos de inferencia, en suma, consolidar un punto de partida en la representación y manipulación semántica de las aplicaciones avanzadas. En las bases de datos inteligentes se concentra la evolución e integración de las bases de datos relacionales - activas y las técnicas avanzadas para la búsqueda, recuperación y procesamiento de la información en el área de los sistemas expertos y la inteligencia artificial.

Palabras claves: base de datos inteligentes, sistema experto, sistema gestor de base de dato, representación del conocimiento, inteligencia artificial.

Introducción.

Las nuevas tecnologías de la información han permitido a los desarrolladores del software responsables de administrar base de datos estudiar e incorporar nuevas herramientas que le permitan acceder y manipular eficientemente los datos almacenados en las mismas multiplicado su nivel de exigencia, realizando un análisis mas profundo para responder eficientemente a consultas no planificadas realizando distintas operaciones entre los datos estructurados y los no estructurados. Dando lugar a las base de datos inteligentes en el que se relaciona la inteligencia artificial con los sistemas relacionales activos.

La Base de Datos Inteligente, precisa del desarrollo de un sistema que posea la capacidad de gestionar conocimiento complejo a través del procesamiento de datos simples, provenientes tanto de ficheros controlados de forma autónoma como de un SGDB (Sistema Gestor de Bases de Datos); incorporar la capacidad de utilizar la experiencia en un dominio particular de aplicación para resolver clases de problemas dentro del dominio.

En el desarrollo de nuestra investigación especificaremos sobre el conjunto de indicadores involucrados en la arquitectura de bases de datos, dando cuenta de la evolución de los SGDB, el impacto de las tecnologías aplicadas a la inteligencia artificial en la representación del conocimiento.

Sistema Gestor de Base de datos (SGBD)

Un sistema gestor de base de datos se define como el conjunto de programas que administran y gestionan la información contenida en una base de datos. Ayuda a realizar acciones como la definición, el mantenimiento de la integridad, el control de la seguridad y privacidad, y la manipulación de los datos dentro de la base de datos.[1]

Clasificación de los (SGBD)

SGBD Relacionales: El modelo de datos relacional fue presentado por Codd en 1970 y se basa en la representación del universo del discurso mediante el álgebra relacio-

nal. Codd, que era un experto matemático, utilizó una terminología perteneciente a las matemáticas, en concreto de la teoría de conjuntos y de la lógica de predicados [2]. Este modelo es fundamental porque dio origen a los primeros sistemas comerciales de SGBD-Relacional que son los que dominan el mercado de bases de datos.

Los SGBD construidos sobre el modelo relacional se caracterizan, por las estructuras de datos, los operadores asociados y los aspectos semánticos.

SGBD Semánticas: Cuando una base de datos es sometida a un proceso de reingeniería, la información en bruto debe considerarse consiste en los objetos existentes y los vínculos que pueda haber entre ellos. Por ejemplo, la relación *ALUMNO(Id, Nombre, IngresoFecha)* en una base de datos de una universidad consta en principio de 4 objetos (la relación y sus tres atributos) y los vínculos entre los objetos: por ejemplo, existe una vinculación "es atributo de" entre cada atributo y la relación. La especificación semántica se almacena en una *base de datos semántica (SDB)*, a los efectos de ser procesada por los algoritmos de conceptualización encargados de generar un modelo conceptual [3].

SGBD Activas: Un sistema de bases de datos activas es un sistema de gestión de bases de datos (SGBD) que contiene un subsistema que permite la definición y la gestión de reglas de producción, (reglas activas). Las reglas siguen el modelo evento-condición-acción (modelo ECA): cada regla reacciona ante un determinado evento, evalúa una condición y, si ésta es cierta, ejecuta una acción, es decir manejan la vigilancia de condiciones (con disparadores y alertas) cuando ocurren sucesos predefinidos. Desde el punto de vista funcional, vigila condiciones disparadas por sucesos que representan acciones de bases de datos. La ejecución de las reglas tiene lugar bajo el control de un subsistema autónomo, denominado motor de reglas, que se encarga de detectar los eventos que van sucediendo y de planificar las reglas para que se ejecuten [4]. Mediante los sistemas de bases de datos activas se hace posible la integración de distintos subsistemas (control de accesos, gestión de vistas, entre otros.) y se extiende el ámbito de aplicación de la tecnología de bases de datos a otro tipo de aplicaciones.

Uno de los problemas que ha limitado el uso extensivo de reglas activas, a pesar de su potencial para simplificar el desarrollo de bases de datos y de aplicaciones, es el hecho de que no hay técnicas fáciles de usar para diseñar, escribir y verificar reglas. Por ejemplo, es bastante difícil verificar que un conjunto de reglas es consistente, es decir, que no se contradice. También es difícil garantizar la terminación de un conjunto de reglas bajo cualquier circunstancia. Para que las reglas activas alcancen todo su potencial, es necesario desarrollar herramientas para diseñar, depurar y monitorizar reglas activas que puedan ayudar a los usuarios en el diseño y depuración de sus reglas.

SGBD Deductivas: Un Sistema de Bases de Datos que tenga la capacidad de definir reglas con las cuales deducir o inferir información adicional a partir de los hechos almacenados en las bases de datos se llama Sistema de Bases de Datos Deductivas. Una base de datos deductiva es, en esencia, un programa lógico; mapeo de relaciones base hacia hechos, y reglas que son usadas para definir nuevas relaciones en términos de las relaciones base y el procesamiento de consultas.



Una Base de Datos Deductiva (BDD) utiliza dos tipos de especificaciones:

- Los **hechos** se especifican de manera similar a como se especifican las relaciones, excepto que no es necesario incluir los nombres de los atributos. En una BDD, el significado del valor del atributo en una tupla queda determinado exclusivamente por su posición dentro de la misma.
- Las **reglas** se parecen un poco a las vistas relacionales. Especifican relaciones virtuales que no están almacenadas realmente, pero que se pueden formar a partir de los hechos aplicando mecanismos de inferencia basados en las especificaciones de las reglas. La principal diferencia entre las reglas y las vistas es que en las primeras puede haber recursividad y por tanto pueden producir vistas que no es posible definir en términos de las vistas relacionales estándar.

Las BDD buscan derivar nuevos conocimientos a partir de datos existentes proporcionando interrelaciones del mundo real en forma de reglas. Utilizan mecanismos internos para la evaluación y la optimización.[5].

SGBD Relacional Difusa: Una solución que aparece recurrentemente en los trabajos de investigación actuales en esta área es la fusión de los sistemas manejadores de bases de datos relacionales con la lógica difusa, lo que da lugar a los que se conoce como sistemas manejadores de bases de datos difusas o FRDBMS (por sus siglas en inglés, Fuzzy Relational Database Management System). Un sistema basado en reglas difusas es un sistema basado en reglas donde la lógica difusa es utilizada como una herramienta para representar diferentes formas de conocimiento acerca del problema a resolver, así como para modelar las interacciones y relaciones que existen entre sus variables. Debido a estas propiedades, los sistemas basados en reglas difusas han sido aplicados de forma exitosa en varios dominios en los que la información vaga o imprecisa emerge en diferentes formas [6].

Actualmente, el modelo relacional no permite el procesamiento de consultas del tipo *“Encontrar a todos los gerentes cuyo sueldo no sea muy alto”* dado que ni el cálculo ni el álgebra relacional, que establecen el resultado de cualquier consulta como una nueva relación, tienen la capacidad de permitir consultas de una manera difusa.

El problema de la implementación de los sistemas manejadores de bases de datos difusas ha sido tratado en dos vertientes principales:

1. Iniciar con un sistema manejador de bases de datos relacionales con información precisa y desarrollar una sintaxis que permita formular consultas imprecisas, lo cual da origen a extensiones SQL con capacidades de manejar la imprecisión.
2. Construir un FRDBMS prototipo que implemente un modelo concreto de base de datos relacional difusa en el que la información imprecisa pueda ser alma-

cenada. Dentro de esta vertiente existen dos grandes ramas: Los modelos a través de unificación por relaciones de semejanza y los modelos relacionales basados en distribuciones de probabilidades.

SGBD Orientado a Objetos: se han diseñado para que se puedan integrar directamente con aplicaciones desarrolladas con lenguajes orientados a objetos. También están diseñadas para simplificar la POO. Almacenan los objetos en la BD con las mismas estructuras y relaciones que los lenguajes de POO.

Una SGBDOO es el gestor de una colección persistente y compatible de objetos definida por un modelo de datos orientado a objetos. Debe satisfacer dos criterios: ser un sistema orientado a objetos, y ser un sistema de gestión de bases de datos

Surgen por la existencia de problemas para representar cierta información y modelar ciertos aspectos del mundo real, puesto que los modelos clásicos permiten representar gran cantidad de datos, pero las operaciones y representaciones que se pueden realizar sobre ellos son bastante simples [7].

Un objeto en BDOO como es una entidad identificable unívocamente que describe tanto el estado como el comportamiento de una entidad del 'mundo real'. El estado de un objeto es descrito mediante atributos mientras que su comportamiento es definido mediante métodos.

No existe una metodología estándar ampliamente aceptada que guíe el proceso de diseño ni un conjunto de reglas (como las reglas de normalización en el modelo relacional) para evaluar el diseño. Sin embargo, esta situación está mejorando. En 1989, un consorcio de fabricantes de ODBMS formó el Object Management Group (OMG) para garantizar la interoperabilidad entre todos los diferentes sistemas orientados a objetos[8].

SGBD Multidimensionales: En una base de datos multidimensional, la información se representa como matrices multidimensionales, cuadros de múltiples entradas o funciones de varias variables sobre conjuntos finitos. Cada una de estas matrices se denomina *Cubo*[9], es decir que en vez de un valor, encontramos varios dependiendo de los "ejes" definidos o una base de datos de estructura basada en dimensiones orientada a consultas complejas y alto rendimiento.

La estructura básica es un hipercubo compuesto por dos elementos: un conjunto de dimensiones y una función que mapea coordenadas formadas por valores de cada una de las dimensiones en tuplas o booleanos. Una dimensión es un nombre con un dominio asociado. Eso facilita el manejo de grandes cantidades de datos dentro de empresas, dándole a esto una amplia aplicación dentro de varias áreas y diferentes campos del conocimiento humano.

Representación del conocimiento Generalmente los sistemas basados en conocimiento se diseñan de forma diferente a los sistemas convencionales, debido a que, en general, los sistemas basados en conocimiento tratan de solucionar problemas que no tienen una solución algorítmica, entre ellos encontramos:

Redes semánticas: Es una representación gráfica de nodos y arcos interconectados que tiene como objetivo representar un determinado conocimiento» y distingue varias aproximaciones clasificadas principalmente en dos grupos: las de quienes se centran en las *redes de palabras* basadas en la presencia conjunta de los vocablos y la de quienes optan por *redes textuales*, que ponen su énfasis en el modo en el que se relacionan las palabras en los textos [10].

Reglas de Producción: Básicamente, una regla de producción consta de un par ordenado de símbolos (situación-acción), la sintaxis de las reglas es diversa y se estructura en función del proceso de selección y ejecución de éstas. Las Reglas son una forma natural de expresar el conocimiento. Tienen una estructura del tipo If <condi-

ción> then <acción>, donde el lado izquierdo de la regla se denomina antecedente o premisa, y el lado derecho se denomina consecuente o conclusión. El consecuente de una regla que se cumple puede generar un nuevo conjunto de hechos o añadir nuevos hechos a los que ya existían. La inferencia en estos sistemas se realiza encadenando reglas. Cuando el razonamiento se realiza desde hechos iniciales para llegar a unas conclusiones, se denomina encadenamiento hacia delante. Si, por el contrario, el razonamiento parte de las conclusiones para determinar los hechos que las producen, se denomina encadenamiento hacia atrás[11].

Objetos estructurados: Permite las capacidades de inferencia, lo que sirve de apoyo en la elaboración de la nueva información y mejora el recuerdo, provee lazos potenciales entre el conocimiento almacenado y la información nueva que llega, lo que facilita el proceso de solución de problemas [12].

El conocimiento del novato contiene reglas que por ellas mismas no son adecuadas a las soluciones de algunas tareas pero que pueden llegar a hacerla mediante la transformación dada por una correcta combinación.

El conocimiento del neófito posee reglas asociadas a situaciones específicas de solución de problemas

El conocimiento del experto: posee reglas de mayor eficiencia que operan sobre estructuras complejas.

Árboles de decisión: es un método no paramétrico ampliamente utilizado en el reconocimiento de patrones complejos, especialmente cuando en la tarea de clasificación se involucran varias clases de patrones y un gran número de atributos. En este método se emplean funciones de decisión, generalmente binarias, a fin de determinar la identidad de un patrón desconocido[13].

Sistemas de Expertos:

Los sistemas expertos, también llamados sistemas basados en conocimiento, pueden considerarse el primer producto verdaderamente operacional de la Inteligencia Artificial y son esquemas diseñados para emular a un especialista humano en un dominio particular o área de conocimiento[14]. Se dice que un SE, se basa en el conocimiento declarativo (hechos sobre objetos, situaciones) y el conocimiento de control (información sobre el seguimiento de una acción).

La ventaja de un sistema inteligente es el dominio completo (o casi completo) de la información, de manera que la decisión prácticamente se toma después de un análisis cualitativo de la misma, armando al decisor con varias opciones de buenas soluciones. Los sistemas con agentes inteligentes poseen diferentes características que son de gran utilidad alguna de las cuales son[15]

Estos sistemas permiten actuar y dialogar con los agentes, de manera que en el momento de realizarse acciones, estas pueden ser vistas de diferentes ángulos.

- Los usuarios pueden realizar preguntas en cualquier momento.
- El agente está en todo momento “observando” el actuar del usuario.
- El agente puede reconstruir y redefinir su actuar en cada momento a partir del actuar del usuario.
- El agente puede adaptar su actuar a situaciones inesperadas.
- El usuario puede tomar el control en cualquier momento.
- En caso de errores el agente ayuda a que el usuario aprenda de ellos.
- En cualquier momento, el usuario puede consultar en qué parte del proceso se encuentra y las tareas faltantes.

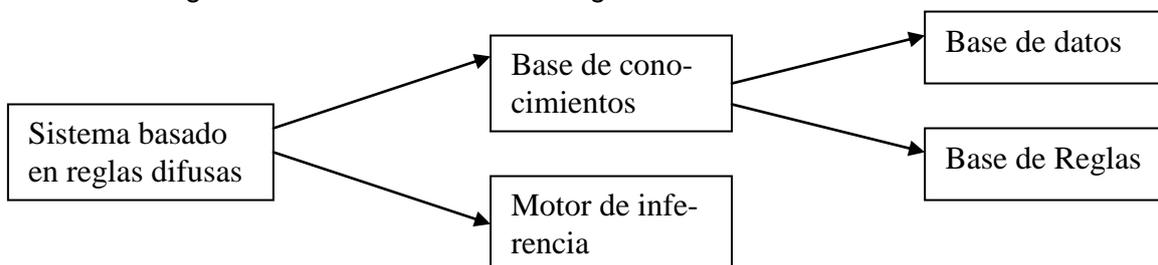
Existen otras variadas ventajas de estos agentes tales como el manejo de emociones, apoyo efectivo del trabajo colaborativo, interacciones pedagógicas adaptables, entre otros

Sistemas Expertos y Bases de Datos Inteligentes:

Las bases de datos inteligentes (BDI) recogen información con la que generan modelos de usuarios a los que recomendar recursos. En una BDI converge la arquitectura de un Sistema Experto (SE) y sistemas basados en conocimiento por lo que para en su desarrollo se incluye como aspecto fundamental en la programación del conocimiento la cual hace uso de la representación explícita del conocimiento a utilizar por el sistema y de su interpretación y manipulación lógica por medio de mecanismos de inferencia, que permitan deducir nuevo conocimiento a partir del que ya se conoce.

Con el transcurso de los años ha surgido un mayor número de investigaciones dentro de la inteligencia artificial enfocadas a la inducción de conocimiento en bases de datos. Consecuencia de esta creciente necesidad ha aparecido un nuevo campo de interés: la minería de datos (*data mining*), que incluye nuevos métodos matemáticos y técnicas software para el análisis inteligente de datos

La minería de datos y el descubrimiento de información dan lugar a los Sistemas basados en reglas difusas, que no es más que la extracción del conocimiento intrínseco contenido en grandes bases de datos con reglas de asociación difusa.



El descubrimiento del conocimiento es el que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados. El valor real de los datos reside en la información que se puede extraer de ellos, información que ayude a tomar decisiones o mejoren nuestra comprensión de los fenómenos que nos rodean. Empleando métodos analíticos avanzados para la explotación de datos se puede maximizar la eficiencia, reducir costos y mejorar las satisfacciones del cliente [16].

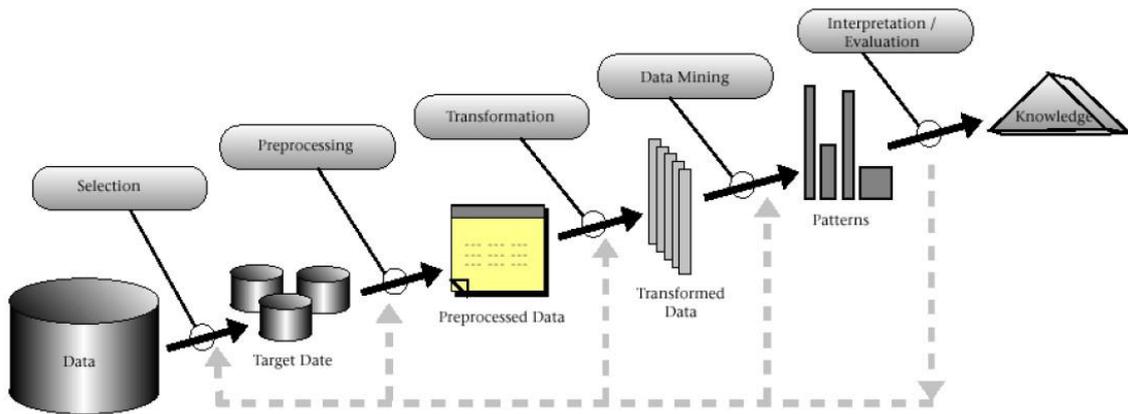
Las tareas comunes en el descubrimiento del conocimiento son la inducción de regla, los problemas de clasificación, el reconocimiento de patrones, la detección de dependencia entre otros.

Los datos recogen un conjunto de hechos (una base de datos) y los patrones son expresiones que describen un subconjunto de los datos.

Se denomina *descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Database)* al proceso global de búsqueda de nuevo conocimiento a partir de los datos de una base de datos. Este proceso incluye no sólo el análisis inteligente de los datos con técnicas de *minería de datos*, sino también diversos pasos previos de regular complejidad, como el filtrado y preprocesado de los datos, y los posteriores, como la interpretación y validación del conocimiento extraído en que se involucran técnicas de la ingeniería del conocimiento.[17]

El descubrimiento de conocimientos en base de datos entre otros aspectos incluye los siguientes:

- Cómo son almacenados y accedidos los datos,
- Cómo pueden escalarse los algoritmos para trabajar con cantidades de datos enormes y seguir siendo eficientes,
- Cómo pueden interpretarse y visualizarse los resultados,
- Cómo modelar y dar soporte a la interacción hombre-máquina durante todo el proceso.



Ciclo de descubrimiento del conocimiento en base de dato

Arquitectura de Base de Datos Inteligentes:

La arquitectura de las Base de Datos Inteligentes (BDI) consta de tres niveles: la herramienta de alto nivel, interfaces de usuarios de alto nivel y el motor de las BDI.

- Las herramientas de alto nivel proporcionan al usuario una serie de facilidades y son objetos orientados y deductivos
- Las Interfaces indican al usuario como manejar la información y la base de datos para obtener la respuesta a las preguntas planteadas. Estas se dividen en dos partes:
 - Un núcleo con información orientada a hipertexto
 - Un conjunto de herramientas que van a mejorar el funcionamiento del sistema para ciertas clases de usuarios
- El motor de la BDI permite que representaciones objetos orientados de información puedan ser expresados y tratados de distintas formas. Además incluye procedimientos de inferencia (hacia atrás y hacia adelante), manipuladores de versión, optimización de compiladores entre otros. Sus principales misiones son por un lado permitir la conexión a periféricos con mecanismos multimedia y por otro gestionar eficientemente esos mecanismos a través de diversas técnicas de software.

Un aspecto que no podemos pasar por alto es el tipo abstracto de datos a utilizar, lo más óptimo es tener en cuenta las características de la programación orientada a objetos, además de incorporar predicados y reglas declarativas que permiten la inferencia.[18] [18]

A continuación se listaran algunas de las herramientas de alto nivel que se pueden utilizar en uno de los niveles de las BDI:

- **Herramientas del descubrimiento del conocimiento:** Extrae el conocimiento acerca de los datos almacenados, para lo que tenemos que ser capaces de saber como recoger datos, preguntar por ellos e intentar comprender estos datos. Esta herramienta permite automatizar el proceso proporcionando el conocimiento así como los algoritmos utilizados para su localización.
- **Herramientas de control de integridad de datos y calidad:** Se utilizan para detectar errores indetectables que posteriormente pueden causar problemas. Permitiendo al usuario cambiar la integridad sin cambiar la aplicación del programa. Para ello el sistema busca automáticamente anomalías en valores basados en un conjunto de criterios proporcionados por el usuario.
- **Herramientas de presentación y visualización de datos:** Esta herramienta proporciona gráficos y distintas formas de presentación de datos
- **Herramientas de gestión de formato de datos:** Permite a los usuarios transformar los datos entre formatos y son indispensables para aplicaciones reales de análisis de datos

- **Herramientas de diseño de sistemas inteligentes:** Permite la integración fácilmente entre base de datos, sistemas de información y sistemas expertos y facilita el mantenimiento con mayor eficiencia en las BDI

Desafíos en la explotación inteligente de datos:

La tecnología de explotación de datos ha avanzado mucho en los últimos años, pero todavía se enfrenta a numerosos retos por resolver, que son objeto de discusión en numerosas conferencias y publicaciones especializadas, donde destacan los siguientes:

- Minería interactiva del conocimiento a vanos niveles de abstracción, para que el usuario pueda refinar de forma iterativa su petición.
- Minería de información de diversas fuentes.
- Protección de la seguridad y privacidad de los datos.
- Conjuntos de datos masivos y soporte de alta dimensionalidad así como de bases de datos de millones de registros.
- Inteligibilidad de los patrones, que desde un punto de vista matemático son correctos pero que a veces el usuario no llega a comprender.

Conclusiones

Este artículo muestra el avance que ha existido en los últimos años en las base de datos aplicando diversas técnicas de la inteligencia artificial como lo sistemas expertos para obtener nuevos conocimientos de los datos almacenados en las base de datos aplicando reglas difusas, la unión entre el campo de la Inteligencia Artificial y la base de datos conforman las base de datos inteligentes, proporcionándole al cliente facilidades a la hora de deducir e interpretar la información almacenada.

Se logro ampliar los conocimientos acerca de las bases de datos inteligente, la arquitectura y las herramientas de alto nivel que conforman uno de los niveles de las mismas.

Bibliografía.

1. Franco, A.C. *Sistemas de Gestión de Bases de Datos*. 2009-2010. España.
2. Elmasri, R. and S.B. Elmasri, *Sistemas de Bases de datos*". 1997. **2**.
3. Larriera, G. (1998) *Representación Semántica de un Esquema Relacional Obtenida Mediante Reingeniería de Bases de Datos*.
4. Marqués, M., *Diseño de Sistemas de Bases de Datos*. 2002.
5. Salgado, S.E.C. and R.M. Rodríguez., *VALORIZACIÓN DE LAS BASES DE DATOS DEDUCTIVAS Y DE LAS BASES DE DATOS ACTIVAS*. 2003.
6. Muela, T.M., J.M.M. Gómez, and J.L.H.T.d. Cepeda, *Modelos Avanzados de Bases de Datos*, in *BASES DE DATOS DEDUCTIVAS Y BASES DE DATOS DIFUSAS*. 2009.
7. Manzanque, A.A. and J.G. Díaz-Tendero. *Modelos Avanzados de Bases de Datos*. in *Bases de datos Orientadas a Objetos y Bases de Datos Objeto-Relacionales*. 2007.
8. Muñoz, V.A., et al., *Modelos de Bases de Datos Orientadas a Objetos y Bases de Datos Objeto-Relacionales*. Universidad de Castilla - la Mancha, 2008.
9. Carpani., F., *CMDM: Un Modelo Conceptual para la Especificación de Bases Multidimensionales.*, in *Facultad de Ingeniería*. 2000, Universidad de la República. Pedeciba Informática.
10. ESCOBAR, M., *Redes semánticas en textos periodísticos: propuestas técnicas para su representación*¹. 2009. **17**: p. 13-39.
11. Cuadrado, J.L.L., *DEFINICIÓN DE UN MODELO DE REPRESENTACIÓN DEL CONOCIMIENTO PARA PROCESOS DE ESTIMACIÓN DE PRESUPUESTOS*, in *DEPARTAMENTO DE INFORMÁTICA*. 2009: Leganés.

12. Yáñez, M.C. and C.E.A. Escobar, *Diseño instruccional: Método de presentación del conocimiento*. Universidad Nacional Autónoma de México, 1996.
13. A.E., G.J.C.C.J.M.M.F.M.M., *Redes Neuronales y Árboles de Decisión: Un Enfoque Híbrido*. 1995.
14. Ribas, M.M. and J.M.D. Alfonso, *Los agentes inteligentes o electrónicos: aceptar o no aceptar sus contratos*. 2002.
15. Farah, G.V., *Agentes Inteligentes en Educación*. Universidad de Santiago de Chile, 2003.
16. Asencios, V.V. (2004) *Data Mining y el descubrimiento del conocimiento*. Industrial Data Revista de Investigación p 83-86
17. Cepeda, S.A.B., *Bases de datos inteligentes.*, in *Facultad de Ingeniería Mecánica y Eléctrica*. 1999, Universidad de Colima: Coquimatlán, Colima. p. 145.
18. Acosta, M.P., J.F.G. Galvés, and E.B. Alonso, *Una aplicación educativa de las base de datos inteligentes*. Revista de la Asociación para el desarrollo de la informática educativa., 1996. **2**.

SOCIEDAD DE LA INFORMACION

www.sociedadelainformacion.com

Edita:



Director: José Ángel Ruiz Felipe

Jefe de publicaciones: Antero Soria Luján

D.L.: AB 293-2001

ISSN: 1578-326x