

PATROCINADORES



ciberSur

ENTIDADES ORGANIZADORAS

Comunidad de Docentes e Investigadores de las Escuelas y Facultades de Informática de distintas Universidades Españolas

ATC- ugr <http://atc.ugr.es>  
Departamento de Arquitectura y Tecnología de Computadores de la Universidad de Granada.

ADIE <http://chico.inf-cv.uclm.es:8080/adie/index.html>  
Asociación para el Desarrollo de la Informática Educativa.

SISTEDES  
Sociedad de Ingeniería del Software y Tecnologías de Desarrollo de Software.

AENUI <http://www.aenui.org/> "it" blank  
Asociación de Enseñantes Universitarios de la Informática.

AEPIA <http://www.aepla.org>  
Asociación Española para la Inteligencia Artificial.

AERFAI <http://www.aerfai.org/index2.shtml>  
Asociación Española de Reconocimiento de Formas y Análisis de Imágenes.

AIPD <http://grifo.udf.es:8080/alpo>  
Asociación Interacción Persona-Ordenador.

EUROGRAPHICS <http://www.eg.org>  
Capítulo Español de la European Association for Computer Graphics.

EUSFLAT <http://www.eusflat.org/index.htm>  
European Society for Fuzzy Logic and Technology.

SC of the IEEE CIS  
Capítulo Español de la IEEE Computational Intelligence Society.

RADISC  
Red Andaluza en Sistemas Complejos.

SEPLN <http://www.sepln.org>  
Sociedad Española para el Procesamiento del Lenguaje Natural.

TIN - MEC  
Programa Nacional de I+D en Tecnologías Informáticas, Ministerio de Educación y Ciencia.

CEA-IFAC <http://www.cea-ifac.es>  
Comité Español de Automática de la International Federation of Automatic Control.

ISTANET <http://www.istanet.net>  
Red Andaluza de Tecnología de Sistemas Inteligentes.

W3C <http://www.w3c.es>  
Consortio World Wide Web. Oficina Española.

RADI-AEB  
Red Andaluza de Algoritmos Evolutivos y Bioinspirados

THOMSON

# CEDI 2005

I CONGRESO ESPAÑOL DE INFORMÁTICA  
GRANADA DEL 13 AL 16 DE SEPTIEMBRE



## X Jornadas de Ingeniería del Software y Bases de Datos

[JISBD'2005]

EDITORES

Ambrósio Tova Álvarez - Juan Hernández Núñez

X Jornadas de Ingeniería del Software y Bases de Datos [JISBD'2005]

CEDI 2005

ISBN: 84-9732-434-X



Nuevos retos científicos y tecnológicos en Ingeniería Informática

<http://cedi2005.ugr.es>

THOMSON

LE INVITAMOS A COLABORAR CON

**THOMSON**



**¿Tiene algún proyecto...**

- ... editorial que se adapte a los planes actuales de estudio universitarios?
- ... editorial para desarrollar un libro de texto universitario enfocado a los nuevos planes de estudio?
- ... para desarrollar contenidos educativos de e-learning para la universidad?
- ... educativo dentro de su área de conocimiento?

**¿Quiere ser uno de nuestros colaboradores en la evaluación de libros en inglés, proyectos originales o contenidos electrónicos?**

Le invitamos a colaborar con el grupo editorial THOMSON para, entre todos, conseguir publicar los proyectos editoriales mejor adaptados a las necesidades educativas de profesores y estudiantes universitarios.

**¿Qué puede ofrecerle THOMSON?**

- Evaluar cualquier proyecto editorial en un plazo breve de tiempo.
- Colaborar con uno de los editoriales más importante del mundo a nivel universitario.
- Nuestra amplia experiencia editorial en la publicación de libros científicos y técnicos.
- Nuestros amplios equipos de promoción y marketing al servicio de los libros de Thomson.
- Una amplia distribución de los libros, tanto a nivel nacional, como en todas las países de habla hispana.
- Posibilidad de traducir sus libros a otros idiomas como el portugués.
- Pertenecer al club de autores y colaboradores de Thomson.

Si quiere conocernos con más detalle y proponernos algún tipo de colaboración, estaremos en el stand que el grupo Thomson tendrá instalado en JENUI 2005.

También puede contactar con nosotros en nuestras oficinas centrales de Madrid:

THOMSON PARANINFO

Magallanes, 25

28015 Madrid

Tel: 91-446-33-50

Fax: 91-445-62-18

andres.otero@paraninfo.es

carmen.roncero@paraninfo.es

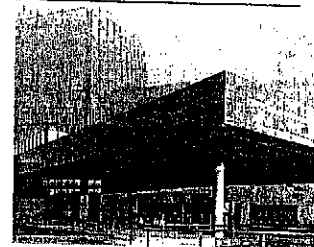
www.paraninfo.es

www.thomsonlearning.com

I CONGRESO ESPAÑOL  
DE INFORMÁTICA

**CEDI 2005**

Nuevos retos  
científicos y tecnológicos  
en Ingeniería Informática



ACTAS DE LAS  
**X Jornadas de Ingeniería  
del Software y Bases de Datos**  
[JISBD'2005]

EDITORES  
Ambrosio Toval Álvarez  
Juan Hernández Núñez

JORNADAS ORGANIZADAS POR  
Sociedad de Ingeniería del Software y Tecnologías de  
Desarrollo de Software

**THOMSON**

**THOMSON**

Actas de las X Jornadas de Ingeniería  
del Software y Bases de Datos [JISBD'2005]

© Los Autores



**Editores de la serie de Actas del CEDI**

Rafael Molina Soriano  
Antonio Díaz García  
Alberto Prieto Espinosa

**Editores de las Actas de las presentes Jornadas**

Ambrosio Toval Álvarez  
Juan Hernández Núñez

**Diseño de Cubiertas**



www.dixi-e.com

**Impresión**

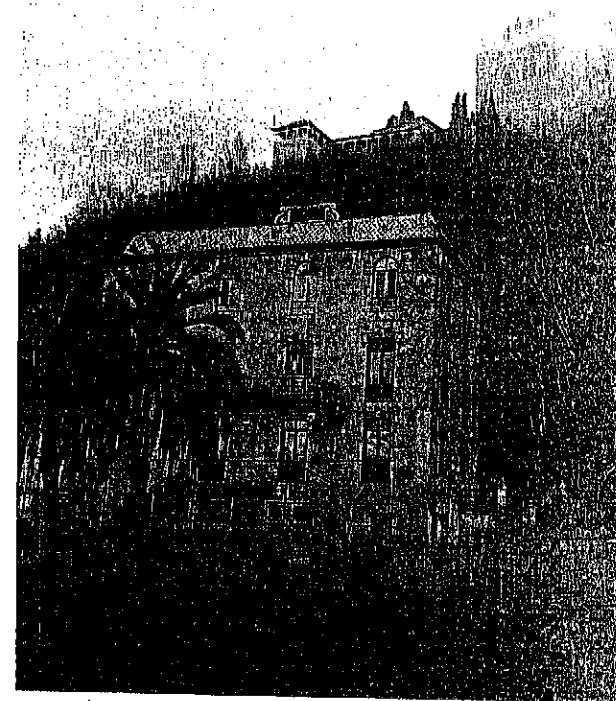
**THOMSON**

COPYRIGHT© 2005 International  
Thomson Editores Spain  
ParanInfo, S.A.  
Magallanes 25 · 28015 Madrid España  
Tel: 91 446 33 50 · Fax: 91 445 62 18  
clientes@paraninfo.es

**Impreso en España**  
Printed in Spain

ISBN: 84-9732-434-X  
Depósito legal: SE-4046-2005 European Union  
Printed by Publidisa

Reservados todos los derechos para todos los países de lengua española. De conformidad con lo dispuesto en el artículo 270 del código penal vigente, podrán ser castigados con penas de multa y privación de libertad quienes reprodujeran o plagiaran, en todo o en parte, una obra literaria, artística o científica fijada en cualquier tipo de soporte sin la preceptiva autorización.  
Ninguna parte de esta publicación, incluido el diseño de la cubierta, puede ser reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste electrónico, químico, electro-óptico, grabación, fotocopia o cualquier otro, sin la previa autorización escrita por parte de los autores.



**Antigo Hotel Reuma. Granada**  
Foto realizada para el CEDI2005 por DIXI

## CONTENIDOS

<b>Artículos.....</b>	<b>1</b>
Um Quadro de Referência para a Comparação de Metodologias Ágeis.....	3
Joao Carlos Ribeiro, Joao Araujo	
Búsqueda Tabú para la generación de casos de prueba de cobertura de bucles. I	14
María Eugenia Díaz Fernández, Raquel Blanco, Javier Tuya	
Providing platforms for developing pervasive systems with MDA. An OSCG metamodel.....	19
Javier Muñoz, Vicente Pelechano, Estefanía Serral	
PRISMANET middleware: Soporte a la Evolución Dinámica de Arquitecturas Software Orientadas a Aspectos.....	27
Cristóbal Costa Soria, Jennifer Pérez, Nour Ali, José Á. Carsí, Isidro Ramos	
Sistematizando la Especificación de Requisitos Safety: un Caso de Estudio sobre Aplicaciones Teleoperadas.....	35
Elena Navarro, Pedro Sánchez, Patricio Letelier, Juan A. Pastor, Isidro Ramos	
Una Arquitectura para la Integración de Portales Web basada en Servicios Web Semánticos.....	43
César J. Acuña, Juan M. Gómez, Esperanza Marcos, Christoph Bussler	
Producción Científica en Ingeniería de Requisitos en España: Un Análisis en el Contexto Europeo.....	51
Oscar Dieste, Natalia Juristo, Ana M. Moreno, Alan M. Davis, Ann Hickey	

Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones.....	59
Raúl Marticorena Sánchez, Yania Crespo González-Carvajal, Carlos López Nozal	
Supporting the Automatic Generation of Advanced Modelling Environments with Graph Transformation Techniques.....	67
Esther Guerra, Paloma Díaz, Juan de Lara	
Un servicio web de políticas de acceso basadas en roles para hipermedia.....	75
Daniel Sanz García, Ignacio Aedo, Paloma Díaz	
Síntesis de patrones de interacción a partir de diagramas de secuencia en UML.....	83
Miguel Ángel Pérez, Amparo Navasa Martínez, Juan Manuel Murillo, Carlos Canal Velasco	
Modelos estructurales de aspectos para arquitectura de software.....	91
Carlos E. Cuesta, M. Pilar Romay, Pablo de la Fuente, Manuel Barrio Solórzano	
Finding where to apply object-relational database schema refactorings: an ontology-guided approach.....	99
Coral Calero Muñoz, Aline Baroni, Fernando Brito e Abreu	
Do composite states improve the understanding of UML statechart diagrams?.....	107
José Antonio Cruz Lemus, Marcela Genero, Esperanza Manso, Mario Piattini	
Transformaciones MDA sobre especificaciones computacionales de UML 2.0 en Maude.....	115
José Raúl Romero Salguero, Nathalie Moreno, Antonio Vallecillo	

Improving automatic SQL translation for ROLAP tools.....	123
Oscar Romero Moral, Alberto Abelló Gamazo	
A Hybrid Method for Discovering Distance-Enhanced Inter-Transactional Rules.....	131
Pedro Gabriel Ferreira, Ronnie Alves, Paulo Azevedo, Orlando Belo	
The Effect of Coupling on Understanding and Modifying OCL Expressions: An Experimental Analysis.....	139
Luis Reynoso, Marcela Genero, Mario Piattini, Esperanza Manso	
Generación Automática de Aplicaciones Mixtas Sw/Hw mediante la Integración de Componentes COTS.....	147
Cristina Vicente Chicote, Ana Toledo Moreo, Carlos Fernández Andrés, Pedro Sánchez	
Método de unión de modelos independientes de plataforma en MDA.....	155
Álvato Prieto Ramos, Adolfo Lozano-Tello, Encarna Sosa Sánchez	
A product-line approach to database reporting.....	163
Felipe I. Anfurrutia, Oscar Diaz, Salvador Trujillo	
Un Enfoque Orientado a Procesos para la Especificación de Planes de Emergencia.....	171
Manuel Llavador, Patricio Letelier, Marcos R. S. Borges, José H. Canós, M <sup>a</sup> Carmen Penadés, Carlos Solís	
De la Arquitectura Software al Urbanismo Software: Hacia Nuevas Formas de concebir los Sistemas de Software Intensivo.....	179
Juan José Moreno-Navarro	

Adaptación de las normas ISO/IEC 12207:2002 e ISO/IEC 15504:2003 para la evaluación de la madurez de procesos software en países en desarrollo.....	187
Francisco J. Pino, Felix Garcia, Francisco Ruiz, Mario Piattini	
Un entorno integrado para la reingeniería.....	195
Ignacio García Rodríguez de Guzmán, Macario Polo Usaola, Mario Piattini	
PWSSEC: Proceso de Desarrollo para Seguridad de Servicios.....	203
Carlos Gutiérrez García, Eduardo Fernández-Medina, Mario Piattini	
Medidas de Usabilidad de Componentes Software.....	211
Manuel F. Bertoa, Antonio Vallecillo	
ORCDB: Arquitectura para la extensión de la semántica de SQL en bases de datos restrictivas orientadas a objetos con restricciones polinómicas de igualdad.....	221
M. Teresa Gómez-López, Rafael M. Gasca, Carmelo Del Valle, Víctor Cejudo	
Determinación de los requerimientos de calidad del producto software basados en normas internacionales.....	231
Abraham Eliseo Dávila Ramón, Karin Ana Melendez Llave, Luis Alberto Flores García	
<b>Artículos Cortos.....</b>	<b>239</b>
Una aproximación metodológica para soportar la evolución de requisitos a partir de un modelo arquitectónico OA.....	241
Amparo Navasa Martínez, Miguel Ángel Pérez, Juan Manuel Murillo	
Mejorando la accesibilidad de las aplicaciones GIS basadas en Web.....	247
Miguel R. Luaces, Nieves R. Brisaboa, Jose R. Parama, David Trillo, Jose R. R. Viqueira	

Del método formal a la aplicación industrial en Gestión de Modelos: Maude aplicado a Eclipse Modeling Framework.....	253
Artur Boronat, José Iborra, José Á. Carsí, Isidro Ramos, Abel Gómez	
Análisis de los Métodos de Selección de Componentes COTS desde una Perspectiva Ágil.....	259
Fredy Javier Navarrete Ramirez, Pere Botella, Xavier Franch	
Un Profile para el Modelado de Patrones de Software.....	265
José Luis Isla Montes, Francisco Luis Gutiérrez Vela, Patricia Paderewski Rodríguez	
Recuperación del conocimiento basada en contexto: Una aplicación en la Arqueología (ArqueOnto).....	271
Juan María Fernández González, Antonio Polo Márquez, Luis Jesús Arévalo Rosado, Enrique Cerrillo Cuenca	
Desarrollando aplicaciones hipermedia para la Web Semántica.....	277
Laura Montells Higuero, Susana Montero, Paloma Díaz, Ignacio Acdo	
Arquitectura para la Clasificación y Composición de Servicios Web.....	283
Ismael Navas Delgado, María del Mar Rojano-Muñoz, Jose F. Aldana-Montes	
Diagramas de casos de uso para el análisis de requisitos en almacenes de datos.....	289
Jose Norberto Mazón López, Juan Trujillo, Manuel Serrano, Mario Piattini	
Especificación de jerarquías de dimensión en un almacén de datos usando WordNet.....	295
Jose Norberto Mazón López, Juan Trujillo, Manuel Serrano, Mario Piattini	

MD a partir de estos diagramas de UC siguiendo nuestra aproximación para el diseño conceptual de AD con UML. Este esquema MD nos permitirá cumplir con los requisitos de usuario. En cada una de estas tres fases se describen, basándonos en nuestra experiencia en el diseño de AD reales, un conjunto de guías de diseño para crear de manera correcta los diagramas de UC y el esquema MD. La principal ventaja de nuestra propuesta es que forma parte de un aproximación completa en la cual se diseña cada parte de un AD (i.e. modelado conceptual MD, diseño de procesos ETL, etc.) con determinados diagramas UML.

Se pretende añadir métricas (empírica y formalmente validadas) a los diagramas de UC para obtener indicadores de calidad objetivos. Por otro lado, se diseñarán experimentos para analizar la comprensión del método [18]. Además, se añadirá la especificación de fuentes de datos operacionales, creando una metodología para el diseño de AD desde el análisis de requisitos.

#### Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto METASIGN (TIN2004-00779) del Ministerio de Educación y Ciencia de España y por el proyecto MESSENGER (PCC-03-003-1) de la Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla la Mancha.

#### Referencias

- [1] Abelló A., Samos J., Salto F. A framework for the classification and description of multidimensional data models. DEXA 2001.
- [2] Antón A.I., McCracken W.M., Potts C. Goal decomposition and scenario analysis in business process reengineering. CAISE 1994. LNCS. 811. pp. 94-104.
- [3] Böhmlein M., Ulbrich-vorn Ende, A. Business Process Oriented Development of Data Warehouse Structures. In: Proceedings of Data Warehousing 2000, Physica Verlag.
- [4] Cockburn A. Structuring use cases with goals. JOOP/ROAD; 10(5):56-62.
- [5] Freitas G.M., Laender A.H.F., Machado M.L. MD2: Getting Users Involved in the Development of Data Warehouse Applications. DMDW 2002: 3-12.

- [6] Inmon, W. Building the Data Warehouse (2nd Edition). New York, Wiley & Sons.
- [7] Jacobson I., Christenson M., Jonsson P., Overgaard G. Object-oriented Software Engineering : A use case driven approach. Addison Wesley 1992.
- [8] Jacobson I., Booch G, Rumbaugh J. The Unified Software Development Process. Addison Wesley Longman Inc. 1999.
- [9] Jarke, M., Lenzerini, M., Vassiliou, Y. and Vassiliadis, P. Fundamentals of Data Warehouses, Ed. Springer. (2000).
- [10] Kimball R., Ross M. The Data Warehouse Toolkit, 2<sup>nd</sup> ed., John Wiley & Sons. (2002).
- [11] Kruchten, P. Rational Unified Process: An Introduction. Addison-Wesley. 2003.
- [12] List B., Schiefer J., Tjoa AM. Process-Oriented Requirement Analysis Supporting the Data Warehouse Design Process: A Use Case Driven Approach. DEXA 2000. LNCS 1873 pp. 593-603. Springer-Verlag (2000).
- [13] Luján-Mora S., Trujillo J., Song I-Y. (2002). Multidimensional modeling with UML package diagrams. ER 2002. Vol. 2503 of LNCS. pp. 199-213. Springer, 2002.
- [14] Maciaszek L.A. Requirements analysis and system design. Addison Wesley. 2<sup>nd</sup> Ed. 2005.
- [15] Paim F.R.S., Castro J.B. DWARF: approach for requirements definition and management of data warehouse systems. RE 2003.
- [16] Prakash N., Singh Y., Gosain A.. Information scenarios for data warehouse requirements elicitation. ER 2004. LNCS 3288. pp. 205-216.
- [17] Rolland C., Souveyet C, Ben Achour, C. Guiding goal modeling using scenarios. IEEE Transactions on Software Engineering. vol. 24, no. 12, December 1998.
- [18] Siau K., Lee L. Are use case and class diagrams complementary in requirements analysis? An experimental study on use case and class diagrams in UML. Requirements Eng (2004) 9:229-237.
- [19] Trujillo J., Palomar M., Gómez J., Song I-Y. Designing data warehouses with OO conceptual models. IEEE Computer. special issue on Data Warehouses 34 (2001) 66-75.
- [20] Winter R., Strauch B. A method for demand-driven Information Requirements Analysis. In Data Warehousing Projects. 16th Hawaii International Conf. on System Sciences. 2003.

## Especificación de jerarquías de dimensión en un almacén de datos usando WordNet

Jose-Norberto Mazón, Juan Trujillo  
 Dept. de Lenguajes y Sistemas Informáticos  
 Universidad de Alicante  
 Apto. de correos 99 E-03080  
 {jnmazon, jtrujillo}@dlsi.ua.es

Manuel Serrano, Mario Piattini  
 Alarcos Research Group, Escuela Superior de Informática  
 Universidad de Castilla - La Mancha  
 Paseo de la Universidad, 4; 13071 Ciudad Real  
 {Manuel.Serrano, Mario.Piattini}@uclm.es

### Resumen

Las operaciones OLAP (*On-Line Analytical Processing*), tales como *roll-up* y *drill-down*, dependen de las jerarquías de dimensión del almacén de datos para poder agregar información a diferentes niveles y, de esta manera, apoyar el proceso de toma de decisiones requerido por los usuarios. Sin embargo, las fuentes de datos operacionales podrían no ser suficientes para construir jerarquías con los niveles de agregación requeridos. En este artículo, aplicamos el conocimiento lingüístico proporcionado por WordNet para completar las jerarquías de dimensión en un almacén de datos. De este modo, las herramientas OLAP serán capaces de mostrar aquella información necesaria para cumplir con los requisitos del usuario. Finalmente, mostramos los beneficios de nuestra propuesta con un caso de estudio, en el cual una jerarquía inicialmente pobre es enriquecida con nuevos niveles.

### 1. Introducción

Según la definición de Inmon [4], un almacén de datos (AD) es "una colección de datos orientados por tema, integrados, no volátil y variante en el tiempo, que sirve de apoyo a la toma de decisiones". Por lo tanto, el objetivo de un AD es ayudar en el proceso de toma de decisiones. Para lograr este objetivo, se utilizan herramientas OLAP (*On-Line Analytical Processing*). Estas herramientas permiten a los usuarios analizar la información contenida en el AD utilizando operaciones como *roll-up* o *drill-down* para agregar o desagregar datos según los niveles de agregación disponibles. Dichos niveles se definen organizando dimensiones en jerarquías [2,9,11].

Así pues, la especificación de jerarquías es una tarea esencial para mejorar el proceso de toma de decisiones: cuánto más completa sea una jerarquía, más significativas son las consultas que el usuario podrá realizar, permitiendo incrementar la calidad de sus decisiones [5]. Basándonos en nuestra experiencia en el diseño de AD, consideramos que la manera correcta de definir jerarquías es la siguiente: a partir de los requisitos del usuario, se construye un esquema conceptual, usando las fuentes de datos operacionales para completarlo. No obstante, es posible que las jerarquías no puedan completarse, ya que muchos términos y datos no se encuentran en las fuentes de datos. Por lo tanto, se necesita un método para especificar jerarquías de mejor calidad con el fin de cumplir con los requisitos de información que el usuario necesita para la toma de decisiones.

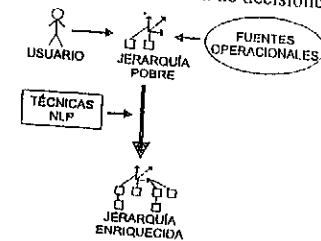


Figura 1. Visión general de nuestra propuesta.

En este artículo se presenta una propuesta (ver figura 1) para completar de manera automática las jerarquías usando técnicas de procesamiento de lenguaje natural (NLP). Se usa conocimiento lingüístico para proveer relaciones semánticas, obteniendo los sentidos de cada palabra y cómo se relacionan por medio de un concepto común. La razón radica en que las jerarquías de dimensión se derivan de abstracciones, representando relaciones



semánticas entre valores, como la generalización ("es-un-tipo-de") o la agregación ("es-una-parte-de") [14]. En nuestra propuesta, se usa WordNet, porque (i) proporciona conceptos procedentes de muchos dominios, (ii) puede ser entendido a varias lenguas por medio de EuroWordNet [19] y (iii) presenta jerarquías de relaciones semánticas entre conceptos que son fáciles de entender y usar.

El resto de este artículo se estructura como sigue. La sección 2 presenta algunos trabajos que aplican NLP a modelos multidimensionales. La sección 3 describe nuestra propuesta para el modelado de AD basada en UML (*Unified Modeling Language*). La sección 4 resume las relaciones de WordNet. La sección 5 define nuestra propuesta para enriquecer las jerarquías de dimensión. En la sección 6 se presenta un caso de estudio. La sección 7 muestra las conclusiones.

## 2. Trabajos relacionados

Las técnicas de NLP se han usado para mejorar la calidad de los sistemas de información, en concreto, el conocimiento lingüístico derivado de ontologías se ha usado para la integración de datos en entornos multidimensionales. Kedat y Métais [6] usan conocimiento lingüístico durante el proceso de limpieza de datos, solucionando problemas de terminología al integrar diferentes fuentes de datos. Toivonen y Niemi [17] presentan un método para encontrar datos a partir de diferentes fuentes e integrar semánticamente en un cubo OLAP aquellos que sean necesarios. Para la construcción del esquema lógico del cubo OLAP usan una ontología del área de aplicación requerida. Un resumen del uso de ontologías para la integración de fuentes heterogéneas de información puede encontrarse en [20].

En cuanto a las jerarquías de dimensión en el AD, varios trabajos [2,5,9,15] señalan su importancia en el modelado multidimensional. Sin embargo, no se ha realizado ningún trabajo para la mejora de su diseño utilizando técnicas de NLP.

## 3. Almacenes de datos, dimensiones y sus jerarquías

Los almacenes de datos, las bases de datos multidimensionales y las aplicaciones OLAP suministran a las empresas información histórica para ayudar en el proceso de toma de decisiones.

Estos sistemas están basados en el modelado multidimensional, estructurándose la información en hechos y dimensiones. Un hecho contiene medidas interesantes (atributos de hechos) de un proceso de negocio (ventas, entregas, etc.), mientras una dimensión representa el contexto de análisis de un hecho (producto, cliente, tiempo, etc.) por medio de atributos de dimensión organizados de manera jerárquica. En este artículo seguimos nuestro modelo conceptual para el diseño de AD [8,18]. Esta aproximación ha sido especificada por medio de un *profile* de UML que contiene aquellos estereotipos necesarios para llevar a cabo el modelado conceptual [8]. En nuestra propuesta, las propiedades estructurales del modelado multidimensional se definen mediante un diagrama de clases UML [13] organizando la información en hechos y dimensiones (clases hecho y clases dimensión).

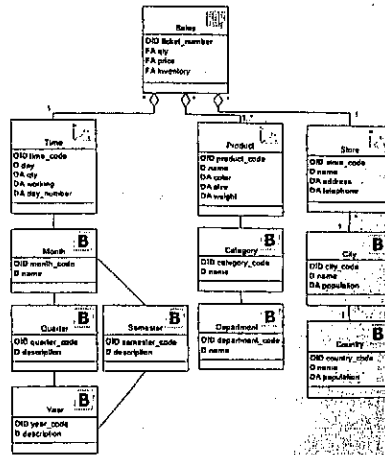


Figura 2. Modelo conceptual de AD con UML.

En cuanto a las dimensiones, existen dos tipos de jerarquías: de clasificación (representadas por relaciones de asociación) y de categorización (representadas por relaciones de generalización). Las jerarquías de clasificación, definidas sobre ciertos atributos de dimensión, son cruciales en el AD porque permiten el análisis de datos. Un atributo de dimensión puede también agregarse (relacionarse) con más de una jerarquía: jerarquías múltiple y de camino alternativo. Por esta razón la manera de representar y considerar dimensiones

con sus jerarquías de clasificación es usar grafos acíclicos dirigidos (DAG). No obstante, estas jerarquías de clasificación no son tan simples en la mayoría de los casos, y surgen los conceptos de jerarquía estricta y completa. Una jerarquía es estricta cuando un objeto de un nivel más bajo de jerarquía pertenece sólo a un nivel más alto, por ejemplo una ciudad está relacionada sólo con una provincia. Una jerarquía es completa cuando cada miembro pertenece a un objeto de clase más alta y dicho objeto está compuesto sólo de aquellos miembros (i.e. si la jerarquía entre provincia y ciudad es completa, una provincia está formada por todas las ciudades almacenadas y todas las ciudades que forman la provincia se almacenan).

En nuestro modelo conceptual de AD, cada nivel de una jerarquía de clasificación se especifica como una clase base. Una asociación de clases base relaciona dos niveles de una jerarquía de clasificación. El único prerrequisito es que estas clases deben definir un DAG con raíz en la clase dimensión. Cada clase base también debe contener un atributo identificador (OID) y un atributo descriptor (D). Debido a la flexibilidad de UML, podemos especificar jerarquías no estrictas y completas mediante la cardinalidad de los roles de las asociaciones. En la figura 2 se muestran ejemplos de jerarquías de clasificación.

Por último, la categorización de dimensiones es útil cuando se incrementa significativamente el número de dimensiones. Por lo que hay atributos que son válidos para todos los elementos dentro de una dimensión mientras que otros son válidos únicamente para un subconjunto de elementos.

Una vez que el modelo multidimensional ha sido definido, los usuarios finales pueden aplicar un conjunto de operaciones OLAP: *roll-up* (incrementar el nivel de agregación) y *drill-down* (decrementar el nivel de agregación) a través de una o más jerarquías de clasificación.

## 4. WordNet y sus relaciones

WordNet [10] es un recurso lingüístico que aporta información léxica sobre palabras y sus sentidos, además de proporcionar relaciones semánticas entre conceptos [11]. La categoría sintáctica de cada palabra determina sus posibles relaciones semánticas. En este artículo, nos centramos en las relaciones semánticas entre nombres (ya que los atributos de dimensión son usualmente nombres):

- *Simonimia*: indica que dos conceptos tienen un significado similar (i.e. *tubería* y *cañería*).
- *Antonimia*: indica que dos conceptos tienen un significado opuesto (i.e. *inferno* y *cielo*).
- *Hiponimia/Hiperonimia*: es una relación de subtipo/supertipo. Dados dos conceptos X e Y, se expresa como X "es-un-tipo-de" Y, donde X es un concepto más específico (hipónimo) e Y es un concepto más genérico (hiperónimo) (i.e. *queso* es-un-tipo-de *lácteo*).
- *Meronimia/Holonimia*: relaciones semánticas del tipo parte-todo. Dados dos conceptos X e Y, se expresa como X "es-una-parte-de" Y, donde X representa una parte (merónimo) de un concepto entero Y (holónimo). Por ejemplo: *rueda* es-una-parte-de *coche*.

Estas relaciones organizan conceptos en estructuras jerárquicas. Las relaciones usadas en nuestra propuesta son la hiperonimia y la meronimia, ya que son las relaciones más útiles para una jerarquía de dimensión [1,2,8,9,15].

## 5. Completando jerarquías de dimensión

Las jerarquías de dimensión en los AD muestran las relaciones entre los dominios de valores que toman los diferentes atributos de dimensión (reunidos en diferentes niveles de agregación) [14]. Como se describe anteriormente, WordNet también presenta relaciones de jerarquía entre conceptos, tales como hiperonimia/hiponimia y meronimia/holonimia. En nuestra propuesta nos ayudamos de esta organización jerárquica de WordNet para especificar automáticamente las jerarquías de dimensión en un AD.

En este artículo nos basamos en la definición de jerarquía de dimensión suministrada por [8], la cual usa UML para diseñar el AD, por lo que las jerarquías son modeladas usando relaciones UML (como se explica en la sección 3). Concretamente, para jerarquías de clasificación, se usan asociaciones entre niveles (incluyendo agregaciones), y para la categorización de dimensiones se usan generalizaciones. Es obvio que la generalización corresponde con la relación de hiperonimia/hiponimia de WordNet mientras que la relación de asociación de UML es más general, ya que sólo especifica la conexión entre dos elementos. Así pues, usamos las relaciones de hiperonimia/hiponimia o de meronimia/holonimia de WordNet dependiendo del dominio en el que



tomen valor los atributos de dimensión: si una asociación se considera como una agregación, se usa meronimia/holonimia, sino se usa hiperonimia/hiponimia. En el caso de la jerarquía ciudad-provincia-país, se usa la relación de meronimia/holonimia debido a que ciudad es parte de provincia y provincia es parte de país (i.e. Torrevieja es parte de Alicante y Alicante es parte de España). Sin embargo, si la jerarquía es producto-familia-clase, se usan relaciones de hiperonimia/hiponimia, porque cada producto es un tipo de familia y cada familia es un tipo de clase (i.e. queso es un tipo de lácteo y lácteo es un tipo de comida).

Con el fin de exponer de manera más comprensible nuestra propuesta, a partir de ahora, sólo se consideran jerarquías estrictas, sin tener en cuenta jerarquías de camino alternativo ni de camino múltiple. Se asume esto debido a las restricciones de WordNet en cuanto a relaciones, ya que existe normalmente un solo hiperónimo (o merónimo) por cada sentido de palabra [10].

Nuestro método consiste en agrupar sentidos de palabras cuyos hiperónimos (o merónimos) sean iguales en un conjunto superior de sentidos de palabras (correspondiente a un nivel de jerarquía). Cada conjunto de sentidos es descrito por su hiperónimo (o merónimo) común. Para crear otros niveles en la jerarquía, se realizan agrupaciones de conceptos de la misma manera hasta alcanzar el nivel de agregación requerido. Antes de comenzar con estas agrupaciones de conceptos, es necesario obtener el sentido correcto de cada concepto, realizando una desambiguación. Se ha utilizado un método de desambiguación del sentido de las palabras (WSD) basado en marcas de especificidad [12]. Esta técnica ofrece muy buenos resultados cuando cada palabra a desambiguar pertenece al mismo dominio. Se asume que cada valor posible de un cierto atributo de dimensión pertenece al mismo dominio. Por ejemplo, cada valor del atributo de dimensión ciudad, será un nombre de ciudad.

A continuación se explican en detalle todos los pasos de nuestro método (ver figura 3):  
**Prerrequisito 1.** Elegir un atributo de dimensión, a partir del cual se completa la jerarquía.  
**Prerrequisito 2.** Especificar un número de niveles de agregación a alcanzar (llamado  $a$ ).  
**Prerrequisito 3.** Crear la variable  $t$ , que indica cuántos niveles de agregación han sido creados. Esta variable se inicializa:  $t=0$ .

**Paso 1.** Extraer todos los valores (sin repetir ninguno) de los atributos de dimensión elegidos:

$W = \{w_1, w_2, \dots, w_n\}$ ; donde  $W$  denota el conjunto de atributos seleccionado.

**Paso 2.** Para cada palabra en el contexto ( $W$ ), se obtiene su sentido correcto usando el algoritmo de WSD de marcas de especificidad.

$S = \{s_1, s_2, \dots, s_n\}$ ;  $s_i$  es el sentido correcto de  $w_i$

**Paso 3.** Para cada sentido en  $S$ , se obtiene un hiperónimo/merónimo (sólo el de nivel más bajo):

$H_{s_i} = \{h_j \mid \forall s_j \in S, h_j \text{ es hiperónimo/merónimo para el sentido de } s_i\}$

También se deben obtener todos los hiperónimos/merónimos (sin repetir):  $H = \{h_1, h_2, \dots, h_n\}$

**Paso 4.** Se crea un nuevo nivel de jerarquía y se mapean los sentidos del hiperónimo/merónimo.

**Paso 5.** Tomar como nueva entrada los valores de los sentidos de los hiperónimos/merónimos:  $S=H$ .

**Terminación.**  $t=t+1$ . Si el nivel de agregación requerido es alcanzado ( $t=a$ ) o  $S$  tiene un único elemento, entonces se ha alcanzado el nivel máximo de agregación para los valores de entrada. En otro caso, se vuelve al paso 3.

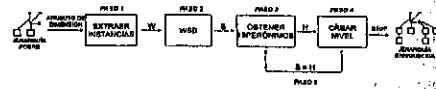


Figura 3. Fases de nuestra propuesta.

El método se ilustra en la figura 3: de una dimensión del modelo multidimensional que no cumple con los requisitos del usuario al no tener suficientes niveles de agregación, se elige un atributo de dimensión. Se obtienen los sentidos correctos para cada valor de los atributos de dimensión. Después, comienzan las iteraciones para obtener hiperónimos/merónimos de los valores y crear nuevos niveles en la jerarquía de dimensión hasta obtener el nivel de agregación requerido o llegar al nivel máximo de agregación.

6. Caso de estudio

En esta sección, mostramos los beneficios de nuestra propuesta mediante un pequeño caso de estudio. Con el fin de presentarlo en español, se utiliza EuroWordNet [19] en lugar de WordNet.

Este caso de estudio consiste en un negocio compuesto de varias tiendas situadas en varias localizaciones. En cada tienda se venden varios productos. Se pretende estudiar las ventas de

productos según la tienda y la fecha en la que se produce la venta, agregándolas por varios niveles (i.e. agregar las ventas por clases de productos). Sin embargo, sólo se dispone del nombre del producto en las fuentes operacionales, por lo que la jerarquía original (ver tabla 1), consiste en un único nivel: producto. Se construye un diagrama de clases multidimensional (ver sección 3) teniendo en cuenta las fuentes operacionales (ver figura 4). La dimensión producto no tiene niveles para cumplir los requisitos de usuario por lo que se aplica el método descrito, introduciendo nuevos niveles en la jerarquía de dimensión. La jerarquía original consiste en un único nivel de agregación, llamado producto (ver figura 4), pero el usuario necesita tres niveles de agregación más (figura 5): subtipo (nivel 1), tipo (nivel 2) y clase (nivel 3).

Producto	Nivel 1	Nivel 2	Nivel 3
Bourbon	Whisky	Licor	Alcohol
Chardonnay	Vino blanco	Vino	Alcohol
Cabernet	Vino tinto	Vino	Alcohol
Whisky escocés	Whisky	Licor	Alcohol

Tabla 1. Niveles de jerarquía creados y sus valores.

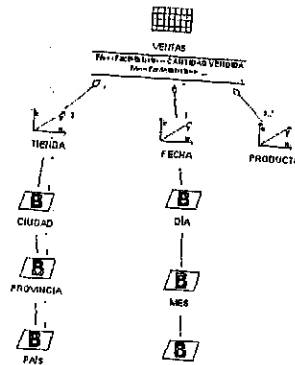


Figura 4. Diagrama de clases para el caso de estudio.

A continuación se describen cada uno de los pasos necesarios para la obtención de una jerarquía completa a partir de la jerarquía original:  
**Prerrequisito 1.** Se elige el atributo de dimensión producto (ver tabla 1).  
**Prerrequisito 2.** Se requieren tres nuevos niveles de agregación, por lo que  $a=3$ .  
**Prerrequisito 3.**  $t=0$ .

**Paso 1.** Los valores de entrada son:  $W = \{\text{bourbon}, \text{chardonnay}, \text{cabernet}, \text{whisky escocés}\}$ .

**Paso 2.** Para cada palabra de la entrada, se obtiene su sentido correcto:  $S = \{s_1, s_2, s_3, s_4\} = \{\text{bourbon}\#1, \text{chardonnay}\#2, \text{cabernet}\#1, \text{whisky escocés}\#1\}$ .

**Paso 3.** Obtener hiperónimos de  $S$  con WordNet.

$H_{\text{bourbon}\#1} = \{\text{whisky}\#1\}, H_{\text{chardonnay}\#2} = \{\text{vino blanco}\#1\}, H_{\text{cabernet}\#1} = \{\text{vino tinto}\#1\}, H_{\text{whisky escocés}\#1} = \{\text{whisky}\#1\}$

$H = \{\text{whisky}\#1, \text{vino tinto}\#1, \text{vino blanco}\#1\}$

**Paso 4.** Se añade el nivel 1 a la jerarquía (tabla 1).

**Paso 5.** Definición de nuevos valores para  $S$ :  $S=H = \{\text{whisky}\#1, \text{vino tinto}\#1, \text{vino blanco}\#1\}$

**Terminación.**  $t=1$ . No se ha llegado al nivel deseado ( $t < a$ ), entonces ir al paso 3.

**Paso 3.**  $H_{\text{whisky}\#1} = \{\text{licor}\#1\}, H_{\text{vino tinto}\#1} = \{\text{vino}\#1\}, H_{\text{vino blanco}\#1} = \{\text{vino}\#1\}; H = \{\text{licor}\#1, \text{vino}\#1\}$

**Paso 4.** Se añade el nivel 2 (ver tabla 1).

**Paso 5.**  $S=H = \{\text{licor}\#1, \text{vino}\#1\}$ .

**Terminación.**  $t=2, t < a$  entonces volver al paso 3.

**Paso 3.**  $H_{\text{licor}\#1} = \{\text{alcohol}\#2\}, H_{\text{vino}\#1} = \{\text{alcohol}\#2\}, H = \{\text{alcohol}\#2\}$

**Paso 4.** Añadir el nivel 3 (ver tabla 1).

**Paso 5.**  $S=H = \{\text{alcohol}\#2\}$ .

**Terminación.**  $t=3=a$  entonces stop.

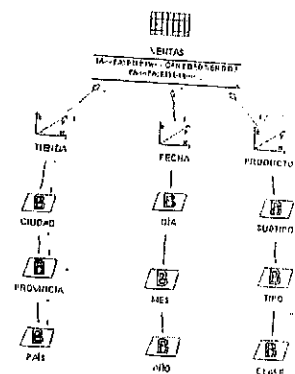


Figura 5. Diagrama que cumple los requisitos.

Después de aplicar nuestro método, se obtiene una jerarquía más completa, que cumple con los requisitos del usuario: analizar la cantidad de producto vendido agregando por varios niveles (subtipo, tipo, y clase de producto), a pesar de que sólo se disponía de un nivel de agregación (figura 4). En la figura 5 se muestra la jerarquía completa y las instancias en la tabla 1 y la figura 6.

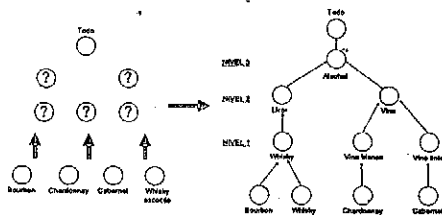


Figura 6. Instancias, antes y después del método.

## 7. Conclusiones y trabajos futuros

Usar jerarquías más completas cuando se consulta el AD con herramientas OLAP es crucial para mejorar el proceso de toma de decisiones. Nuestro método permite enriquecer automáticamente las jerarquías usadas en el modelo conceptual del AD desde el punto de vista de las relaciones semánticas de WordNet [3]. Además, se deben usar otro tipo de relaciones, tales como "es-un atributo-de" [16] para añadir nuevos atributos a cada nivel de jerarquía.

## Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto METASIGN (TIN2004-00779) del Ministerio de Educación y Ciencia de España y por el proyecto MESSENGER (PCC-03-003-1) de la Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla la Mancha.

## Referencias

[1] Abelló A., Samos J., Saltor F. Understanding Analysis Dimensions in a Multidimensional Object-Oriented Model. DMDW, 2001.  
 [2] Akoka J., Comyn-Wattiau I., Prat N. Dimension Hierarchies Design from UML Generalizations and Aggregations. ER2001. LNCS 2224, pp. 442-455, Springer, 2001.  
 [3] Gangemi A., Navigli R., Velardi P. The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. LNCS 2888, pp. 820-838, 2003.

[4] Inmon W., Building the Data warehouse, John-Wiley, 1996.  
 [5] Jagadish H. V., Lakshmanan L. V. S., Srivastava D. What can Hierarchies do for Data Warehouses?. 25<sup>th</sup> VLDB, 1999.  
 [6] Kedad Z., Métais E. Ontology-Based Data Cleaning. NLDB'02. LNCS 2553, p.137-149.  
 [7] Kimball R. Data Warehouse Toolkit: Practical Techniques For Building Dimensional Data Warehouse. John-Wiley, 1996.  
 [8] Luján-Mora, S., Trujillo, J., Il-Yeol Song. Extending UML for Multidimensional Modeling. UML'02, p. 290-304: LNCS 2460.  
 [9] Malinowski E., Zimányi E. OLAP Hierarchies: A Conceptual Perspective. CAISE 2004, LNCS 3084, pp. 477-491.  
 [10] Miller G.A., Beckwith R., Fellbaum C., Gross D., Miller K.J. WordNet: An on-line lexical database, International Journal of Lexicography, 3(4), 1990.  
 [11] Miller G.A., Fellbaum C. Semantic networks of english. Cognition, 41(1-3):197-229, 1991.  
 [12] Montoyo A., Palomar M. WSD Algorithm Applied to a NLP System. NLDB 2000. LNCS 1959, pp. 54-65, Springer, 2001.  
 [13] Object Management Group (OMG). Unified Modeling Language Specification 1.5. www.omg.org/cgi-bin/doc?formal/03-03-01.  
 [14] Pourabbas, E. Rafanelli M. Hierarchies and Relative Operators in the OLAP Environment. SIGMOD Record, Vol. 29, 1, March 2000.  
 [15] Schneider, M. Well-formed Data Warehouses Structures. DMDW, 2003.  
 [16] Storey V. Understanding Semantic Relationships. VLDB J. 2, 455-488, 1993.  
 [17] Toivonen S., Niemi T. Describing data sources semantically for facilitating efficient creation of OLAP cubes. ISWC, 2004.  
 [18] Trujillo J., Palomar M., Gómez J., Song I.Y. Designing Data Warehouses with OO Conceptual Models. IEEE Computer, 34(12):66-75, 2001.  
 [19] Vossen P. EuroWordNet: building a multilingual database with wordnets for European languages. The ELRA Newsletter February 1998, Vol. 3 n.1, pp. 7-10.  
 [20] Wache H., Vögele T., Visser U., Stuckenschmidt H., Schuster G., Neumann H., Hühner S. Ontology-based integration of information - A survey of existing approaches. IJCAI-01, 2001.

## ÍNDICE DE AUTORES

Abelló Gamazo, Alberto.....	123
Acuña, César J.....	43
Aedo, Ignacio.....	75,217
Aldana-Montes, Jose F.....	283
Ali, Nour.....	27
Alves, Ronnie.....	131
Araujo, Joao.....	3
Arévalo Rosado, Luis Jesús.....	271
Azevedo, Paulo.....	131
Baroni, Aline.....	99
Barrio Solórzano, Manuel.....	91
Belo, Orlando.....	131
Bertoa, Manuel F.....	211
Blanco, Raquel.....	11
Borges, Marcos R. S.....	171
Boronat, Artur.....	253
Botella, Pere.....	259
Brito e Abreu, Fernando.....	99
Bussler, Christoph.....	43
Calero Muñoz, Coral.....	99
Canal Velasco, Carlos.....	83
Canós, José H.....	171
Carsí, José Á.....	27,253
Cejudo, Victor.....	221
Cerrillo Cuenca, Enrique.....	271
Costa Soria, Cristóbal.....	27
Crespo González-Carvajal, Yania.....	59
Cruz Lemus, José Antonio.....	107
Cuesta, Carlos E.....	91
Dávila Ramón, Abraham Eliseo.....	231