



## **COMPATIBILIDAD ENTRE SISTEMAS DE CLASIFICACIÓN EN DIRECTORIOS COMERCIALES: ESTÁNDARES Y SOLUCIONES TÉCNICAS**

*Ricardo Eíto-Brun*

*Universidad Carlos III de Madrid, Biblioteconomía y Documentación,  
C/ Madrid, 126 - 28903 Getafe España, reito@bib.uc3m.es*

### **RESUMEN**

La generalización de los servicios de información accesibles a través de Internet ha supuesto un incremento en el número de servicios de directorios de empresas, en los que se puede acceder a información de contacto y a descripciones de los servicios que ofertan. En este contexto, los sistemas de clasificación de productos, servicios, y actividades industriales constituyen la principal alternativa para la recuperación en entornos globales caracterizados por el uso de múltiples idiomas.

El uso de distintos sistemas de clasificación acarrea las dificultades derivadas de la falta de compatibilidad entre los sistemas de clasificación y vocabularios controlados utilizados. En este contexto, es preciso complementar los servicios de información con servicios capaces de asegurar la compatibilidad entre distintos sistemas de consulta. Se presenta un prototipo basado en la tecnología de los servicios web XML y la especificación UDDI para facilitar la interrogación y el acceso a sistemas de información basados en sistemas de clasificación heterogéneos.

### **ABSTRACT**

Web-based information services to identify contact information about companies, and details of the services they offer have become widely available. In these systems, the use of classification systems and vocabularies to identify products, services and industrial activities are the main alternative to the weaknesses of the use of natural language and free text indexing, specially considering the international scope of these initiatives.

The adoption of different classification systems implies difficulties due to the lack of compatibility between their controlled vocabularies. It is necessary then to offer these information services with complementary services to ensure the compatibility between the different search facilities. The communication presents one prototype based on XML web services and in the UDDI standard to enable the access to information services based on heterogeneous product classification systems.



**PALABRAS CLAVES:**

UDDI, Directorios, Registros, Compatibilidad, Sistemas de clasificación de productos y servicios, Lenguajes controlados



## INTRODUCCIÓN

El uso de Internet como herramienta de promoción y como soporte de transacciones comerciales implica la necesidad de identificar información sobre posibles socios comerciales. Con este fin, se han diseñado distintos servicios de directorio, donde las organizaciones pueden dar a conocer su actividad y los productos y servicios que ofertan. A menudo, la descripción que ofrecen estos directorios consisten en breves descripciones textuales facilitadas por las propias organizaciones, situación que condiciona al uso de sistemas de recuperación básicos, donde el usuario debe prever de antemano los términos utilizados en la descripción de las organizaciones registradas en estos directorios. Tanto los servicios comerciales como los servicios públicos para la distribución y localización de información sobre empresas, precisan de sistemas de recuperación más sofisticados.

La aplicación de sistemas de clasificación basados en categorías preestablecidas y en el uso de términos controlados ofrece distintas ventajas, entre ellas garantizar la descripción homogénea de los servicios y facilitar al usuario la selección de categorías correctas a la hora de formular sus búsquedas. Por otra parte, los sistemas de clasificación incorporan notaciones para representar clases de forma abreviada e independiente de cualquier idioma en particular, lo que ayuda enormemente en la formulación de consultas en entornos globales donde los usuarios pueden proceder de distintas localizaciones geográficas.

## ESPECIFICACIONES TÉCNICAS Y SERVICIOS DE DIRECTORIO

Por otra parte, en los últimos años se han desarrollado tecnologías y estándares relevantes para la automatización de transacciones comerciales en la red – servicios web basados en XML y especificaciones para la consulta a directorios como UDDI (Universal Description, Discovery and Integration) o ebXML – diseñados en torno al concepto de directorio. Estas especificaciones ofrecen mecanismos para automatizar y establecer mecanismos de consulta entre distintos servicios de información de forma fácil y eficiente (BARON, 2000) (SCHMITZ, 2003).

Una de las características de los distintos protocolos brevemente descritos en los apartados anteriores es que normalizan aspectos técnicos y facetas relacionadas con la sintaxis de los mensajes de petición y respuesta. En este sentido, las especificaciones ofrecen pautas sobre cómo se deben formatear los mensajes XML que se intercambiarán entre el equipo que efectúa las consultas y el equipo que les dará respuesta, y la forma en la que se transmitirán dichos mensajes a través de la red usando protocolos como http y SOAP (Simple Object Access Protocol).

En el caso de UDDI, se establece un esquema XML que rige el formato de las entradas en el registro; también se ha añadido a la especificación UDDI un API (Application



Protocol Interface) en el que se definen las operaciones de lectura y escritura que deben soportar los registros. Este API establece una forma normalizada de acceder a un registro UDDI para identificar organizaciones que han publicado información sobre los servicios web que ofertan, así como la forma de añadir, modificar o borrar las entradas creadas en estos registros. UDDI incorpora un mecanismo para indicar que se está utilizando un sistema de clasificación o de identificadores particular en la descripción de una organización o servicio: los llamados tModel. Podemos afirmar que sin los tModel UDDI no sería más que una clasificación arbitraria de empresas que ofrecen servicios en Internet.

Un tModel se define como un conjunto de metadatos sobre cualquier concepto, que se define de forma única mediante una tModelKey. La unicidad del tModel no se basa en vincularlo a una URI, sino en un UUID (Universally Unique Identifier) que se asigna al tModel cuando éste se publica en el registro UDDI. El tModel también tendrá un nombre, una descripción, y puede apuntar a un documento externo disponible en una URL. Los tModel se usan en UDDI para hacer referencia sistemas de clasificación e identificación de productos y servicios, evitando cualquier posible ambigüedad en la clasificación de los mismos. En la descripción de una organización o servicio (concretamente, en el elemento categoryBag) se usará un atributo tModelKey para hacer referencia al sistema en cuestión que se esté utilizando. De la misma forma, el API para consulta ofrece la posibilidad de recuperar organizaciones o servicios utilizando la clase a la que han sido asignados como criterio de búsqueda. De esta forma la especificación incorpora un mecanismo para poder formular búsquedas usando lenguajes controlados sin ambigüedad.

## **EL PROBLEMA DE LA COMPATIBILIDAD**

Sin embargo, hay un aspecto clave en este tipo de soluciones no queda cubierto por las especificaciones, y se refiere a la compatibilidad de los sistemas de clasificación utilizados para la descripción de las entidades de las que se mantiene información en los registros.

En este sentido, las especificaciones permiten utilizar distintos sistemas para la clasificación de servicios y organizaciones, siempre que estos sistemas hayan sido previamente registrados y se les haya asignado un identificador (como un tModel en el caso de UDDI). Esto permite utilizar una amplia variedad de sistemas de clasificación. Los principales sistemas de clasificación de servicios y productos utilizados en entornos de comercio electrónico son la clasificación UNSPSC (United Nations Standard Products and Services Code System), NAICS (North America Industry Classification System) y ECLASS. Este último es una iniciativa europea, propuesta por el Cologne Institute for Business Research y varias empresas alemanas (Beneventano, 2004).

En el caso de UNSPSC, es el resultado de la fusión en 1000 del UNCCS (United Nation's Common Coding System) y SPSC (Dun & Bradstreet's Standard Product and



Service Codes). Ordenan las clases en cinco niveles jerárquicos (segment, family, class, commodity y Business function). Es un sistema que se puede utilizar libre y gratuitamente, y que disfruta de una gran aceptación en los Estados Unidos de América. NAICS se desarrolló en Estados Unidos para facilitar la captura de datos estadísticos de carácter económico por parte de las agencias federales. Respecto a ECLASS, el sistema de clasificación propuesto es más complejo: también se trata de un sistema jerárquico con cuatro niveles, y un total de doce mil términos (extraídos de los nombres de las clases). Junto a las clases, ECLASS incluye una serie de atributos que se pueden utilizar en la descripción de los productos y servicios, lo que le dota de una mayor flexibilidad y capacidad expresiva.

La posibilidad de utilizar distintos lenguajes de clasificación en los registros añade un nivel adicional de complejidad, y asegurar una conectividad real con múltiples registros dependerá de nuestra capacidad de lograr una compatibilidad a nivel conceptual entre los distintos lenguajes documentales en uso. La compatibilidad técnica y sintáctica que aseguran los estándares antes descritos, no es, por sí misma, suficiente. La solución pasa – forzosamente –, por el uso de sistemas de traducción que de una forma total o parcialmente automatizada permitan traducir una expresión de búsqueda formulada mediante un lenguaje controlado, en su expresión equivalente en un segundo vocabulario.

La compatibilidad y las equivalencias entre sistemas de clasificación de organizaciones y servicios ha sido tratada en la literatura, y encontramos interesantes trabajos como los de Bergamaschi (2001) y Beneventano (2003). Entre ellos, queremos destacar el completado por Beneventano (2003), en el que describió un sistema para la identificación semiautomática de equivalencias entre los sistemas de clasificación ECLASS y UNSPSC. Para establecer las equivalencias, utilizó el sistema WordNet con el fin de extraer el significado o la semántica asociada a las distintas clases. El objetivo de su sistema, desarrollado en el marco de un proyecto llamado MOMIS, era identificar equivalencias de forma automática y sugerir distintas posibilidades en aquellos casos en los que el sistema no sea capaz de resolver un caso de ambigüedad.

Estas iniciativas resultan especialmente relevantes, y no podemos dejar de señalar proyectos similares en el ámbito de los vocabularios controlados y lenguajes documentales orientados a la descripción de documentos. Ejemplos similares a los antes descritos en el ámbito de la documentación serían las tablas de equivalencia publicadas entre distintos lenguajes documentales. Otro hito relevante en la historia del estudio de la contabilidad entre los sistemas de clasificación fue el encuentro celebrado en septiembre de 1995 en Varsovia, en el que se sintetizó el estado del arte de la compatibilidad entre lenguajes documentales controlados.



## DESCRIPCIÓN DEL PROTOTIPO UDDINDEX

El objetivo del prototipo era elaborar un sistema capaz de interactuar con distintos servidores UDDI, actuar como traductor entre distintos sistemas de clasificación y redirigir peticiones de búsqueda a múltiples servidores UDDI remotos.

La principal característica o función del sistema es recibir peticiones UDDI, identificar el sistema de clasificación (tModel) utilizado para formularlas, y tras un proceso de traducción, redirigirlas a servidores UDDI tras reformular la consulta en el sistema de clasificación utilizado por el servidor al que va dirigida.

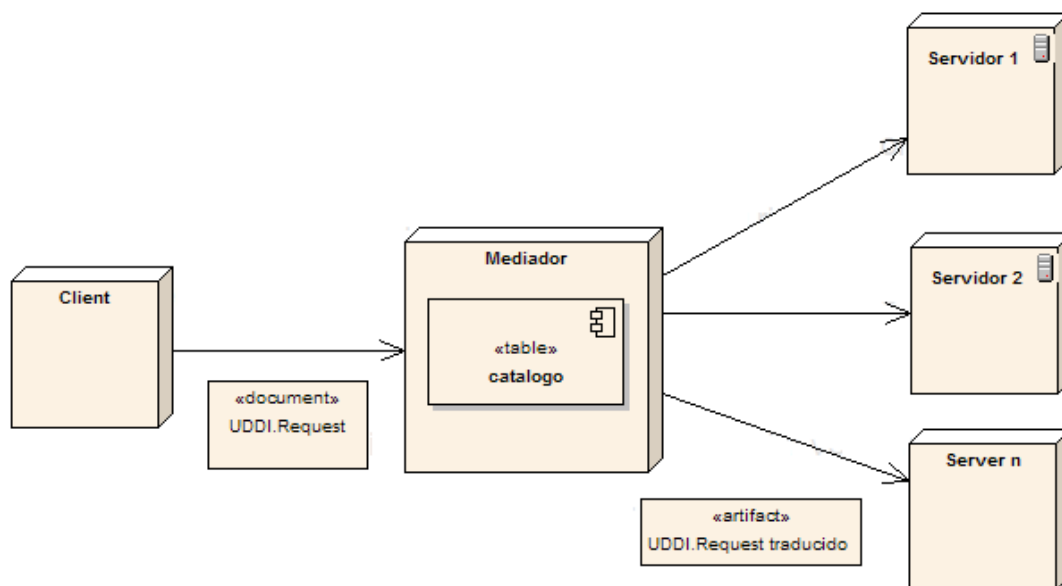


Figura 1.- UDDIndex como sistema mediador

Un componente clave en el modelo es la disponibilidad de un catálogo interno que mantiene información sobre distintos sistemas de clasificación, sus clases, notaciones y descripciones textuales, así como las equivalencias entre las clases de distintos sistemas. En el prototipo se ha trabajado con un subconjunto de 30 clases de UNSPSC y sus correspondencias en ECLASS y NAICS, previamente establecidas. Se añadieron 20 clases para las que no se estableció una equivalencia directa.

Inicialmente, el prototipo incorpora las siguientes capacidades:

1. Recepción de peticiones de búsqueda en formato UDDI (subconjunto del API Inquiry).
2. Extracción del lenguaje controlado utilizado para formularla.



3. Comprobar si el sistema mantiene información sobre dicho lenguaje.
4. En caso afirmativo, el sistema formula las peticiones de búsqueda en los vocabularios controlados de los otros dos sistemas, y construye los respectivos mensajes UDDI válidos que serán redirigidos a los servidores remotos. No se ha incluido en el alcance del prototipo el envío de las peticiones UDDI a los servidores destino y la posterior recepción e integración de resultados. Se ha tomado esta decisión porque el principal interés de este trabajo era analizar la compatibilidad a nivel conceptual entre directorios.

La formulación del prototipo es que las peticiones se tramitan a través de un formulario HTML, que recoge datos del usuario y los transforma a un mensaje UDDI válido que se envía al sistema mediador.

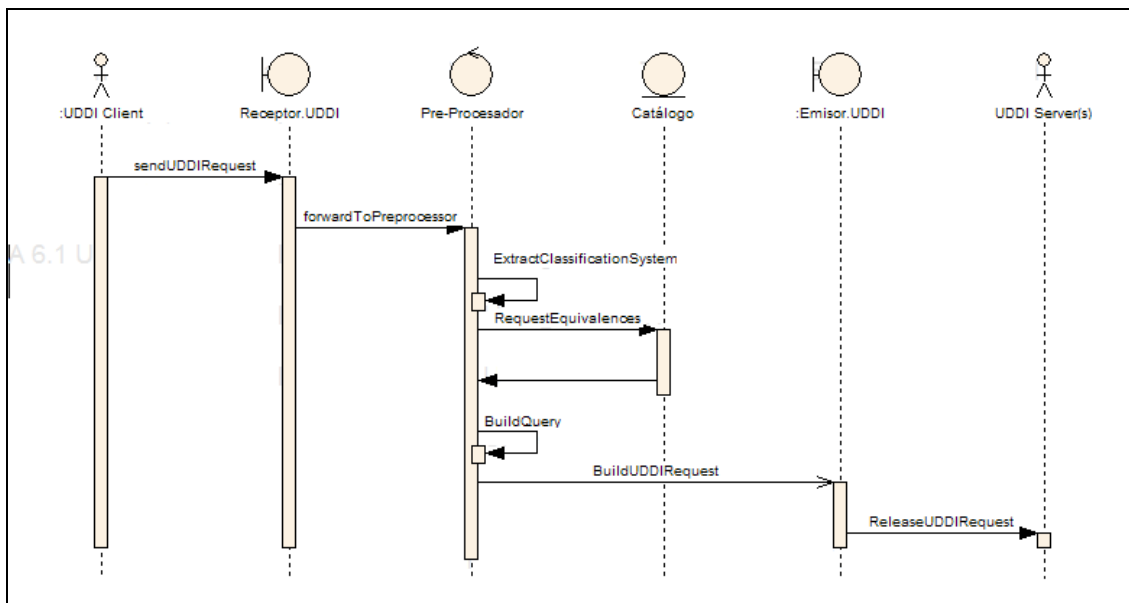
El principal problema encontrado durante el desarrollo del prototipo ha consistido en asegurar la compatibilidad entre clases para las que no existe una traducción directa. En líneas generales, las traducciones entre clases se pueden gestionar de la siguiente forma durante la construcción de una ecuación de búsqueda, dependiendo de si se cumplen una de estas condiciones:

- Existe una correspondencia directa entre las dos clases. En este caso el sistema únicamente debe recuperar la clase equivalente y reformular la ecuación de búsqueda con ella.
- Existe una correspondencia indirecta entre las dos clases. Por ejemplo, una clase corresponde a una combinación de dos o más clases. En este caso el sistema mediador debe construir la ecuación de búsqueda mediante la intersección o unión de dos o más clases, aplicando los mecanismos que ofrece para ello el API de UDDI.
- No existe una correspondencia para la clase seleccionada. En este caso, el sistema mediador busca en el catálogo una descripción de la clase que permita identificar un conjunto de términos sinónimos o equivalentes a la clase en cuestión. Se identifican las clases similares en el resto de sistemas de clasificación tomando como punto de partida los términos anteriores. En estos casos se ha identificado un problema cuando la descripción disponible sobre las distintas clases no es suficiente para identificar un conjunto de términos para los que existan equivalencias en los otros sistemas de clasificación. La solución implementada ha consistido en ampliar la búsqueda añadiendo términos procedentes de la superclase y de las clases situadas en el mismo nivel jerárquico de la clase tomada como punto de partida.

De la misma forma, el sistema permite explotar la organización jerárquica de clases en el sistema destino para identificar una superclase como equivalente, en aquellos casos en los que varias subclases de una misma clase presenten una similitud parcial con los términos usados como criterio de búsqueda.

Una posible mejora en el sistema consistiría en identificar – a partir de una colección de textos descriptivos de las entidades clasificadas en un grupo determinado -, un conjunto de términos representativo de cada clase (centroide) contra el que se podrían comparar los términos usados en la búsqueda. Esto ampliaría notablemente la capacidad de expandir las búsquedas y obtener resultados relevantes.

El siguiente diagrama muestra una vista resumen del funcionamiento del prototipo:



**Figura 1.-** Modelo de interacción básico





## CONCLUSIONES

La dificultad de establecer equivalencias semánticas entre los sistemas de clasificación utilizados por distintos directorios de organizaciones y servicios constituye uno de los problemas para la generalización de estas herramientas.

En los últimos años se han desarrollado modelos y especificaciones – en el ámbito técnico del comercio electrónico – para normalizar aspectos relacionados con la sintaxis de los mensajes y su transmisión física a través de redes. Estos sistemas permiten utilizar distintos sistemas de clasificación para describir organizaciones y servicios, lo que se traduce en problemas de compatibilidad, ya que un usuario (sea una persona o un agente software) debe utilizar distintos vocabularios y sistemas de clasificación para interrogar cada directorio.

El prototipo desarrollado analiza la viabilidad de un sistema capaz de actuar como mediador entre un usuario y varios directorios habilitados para enviar y recibir peticiones formuladas de acuerdo con la especificación UDDI. El prototipo establece equivalencias entre las clases utilizadas en la formulación de la búsqueda y las clases utilizadas en los directorios destino, aplicando el cálculo de similitud para obtener automáticamente la similitud entre clases.

## REFERENCIAS BIBLIOGRÁFICAS

BARON, J.P. [et al.] (2000) “Web-based e-catalog systems in B2B Procurement”. En: Communications of the ACM, vol. 43, no.5, pp. 93-100

BENEVENTANO, D. [et al.] (2003). “Synthesizing an integrated ontology”. Internet Computing, vol. 7, no. 5, p. 42-51

BERGAMASCHI, S. [et al.] (2001). “Semantic integration of heterogeneous information sources”, Special Issue on Intelligent Information Integration, Data and Knowledge Engineering, vol, 36, no 1, 215-224

Compatibility and Integration of Order Systems: Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September 1995 (1996). Warsaw: SBP

HUEMPEL, C.; SCHMITZ, V. (2000). “BMEcat: an XML standard for electronic product data interchange”. En: Proc. of the 1st German Conference XML. Heidelberg, p. 1-11



LI, H. (2000). "XML and industrial standards for Electronic Commerce". Knowledge and Information Systems, vol. 2, no. 4, pp. 487-497

SCHMITZ, Volker ; LEUKEL, Joerg ; DORLOFF, Frank-Dieter (2003). "XML Data Modelling Concepts in B2B Catalogue Standards". En: Proc. Of IADIS International Conference Society, June 3-6 2003, Lisbon, Portugal, pp. 227-234

UDDI Technical White Paper. Disponible en: <http://www.uddi.org>