

**Título:** CONTEXTUALIZING A DATA WAREHOUSE WITH DOCUMENTS

**Nombre:** Pérez Martínez, Juan Manuel

**Universidad:** Universidad Jaume I de Castellón

**Departamento:** METODOS INFORMATICOS AVANZADOS

**Fecha de lectura:** 26/02/2007

**Programa de doctorado:** DEPARTAMENTO DE LENGUAJES Y SISTEMAS INFORMATICOS

**Dirección:**

- > **Director:** Rafael Berlanga Llavorí
- > **Codirector:** MARÍA JOSÉ ARAMBURU CABO

**Tribunal:**

- > **presidente:** TORBEN BACH PEDERSEN
- > **secretario:** MICHAEL GOULD CARLSON
- > **vocal:** TRUJILLO MONDEJAR JUAN CARLOS
- > **vocal:** DAVID LOSADA
- > **vocal:** JOSÉ SAMOS JIMENEZ

**Descriptores:**

- > DISEÑO Y COMPONENTES DE SISTEMAS DE INFORMACION

**El fichero de tesis** ya ha sido incorporado al sistema

- > <http://hdl.handle.net/10803/10482>

**Localización:** BIBLIOTECA UNIVERSITAT JAUME I

**Resumen:** La tecnología actual de los almacenes de datos y las técnicas OLAP permite a las organizaciones analizar los datos estructurados que éstas recopilan en sus bases de datos. Las circunstancias que rodean a estos datos aparecen descritas en documentos, típicamente ricos en texto. Actualmente es posible encontrar muchos de estos documentos en formato XML. El objetivo de la tesis es proporcionar un marco formal para integrar un almacén de datos tradicional con un almacén de documentos XML y analizar los datos junto a su contexto. Llamamos al sistema resultante almacén contextualizado. En este nuevo tipo de almacén, el usuario especificará un contexto de análisis proporcionando una secuencia de palabras clave. El análisis se realizará en un nuevo tipo de cubo multidimensional, denominado R-cubo. Cada hecho de un R-cubo tendrá asociado el conjunto de documentos que describen su contexto. Asignaremos a cada hecho un valor numérico que representará su relevancia con respecto al contexto de análisis.

Esta tesis propone un nuevo modelo de recuperación para medir la relevancia de los hechos. Asimismo, se definen formalmente los R-cubos, proporcionando su modelo de datos y álgebra.

