

**Título:** AUTOMATIZACIÓN DE TESAUROS DOCUMENTALES PARA LA WEB SEMÁNTICA

**Nombre:** Perez Aguera, Jose Ramon

**Universidad:** Universidad Complutense de Madrid

**Departamento:** Biblioteconomía y documentación

**Fecha de lectura:** 29/10/2007

**Programa de doctorado:** DOCUMENTACIÓN: FUNDAMENTOS, TECNOLOGÍA Y APLICACIONES

**Dirección:**

> **Director:** ALFONSO LÓPEZ YEPES

**Tribunal:**

> **presidente:** FÉLIX DEL VALLE GASTAMINZA

> **secretario:** RODRIGO SÁNCHEZ JIMÉNEZ

> **vocal:** Eva María Méndez Rodríguez

> **vocal:** GARCÍA MARCO FRANCISCO JAVIER

> **vocal:** Antonio García Jiménez

**Descriptor:**

> BIBLIOMETRIA

> DOCUMENTACION

**El fichero de tesis** no ha sido incorporado al sistema.

**Resumen:** La presente tesis doctoral tienen como objetivo fundamental solucionar el problema de la automatización de tesauros documentales en entornos distribuidos de recuperación de información como la Web Semántica.

Este trabajo se divide en tres bloques interconectados pero claramente diferenciados. Cada uno de los cuales define un subproblema dentro de la automatización de tesauros y ofrece una solución al mismo que contribuya a la solución final del problema planteado en esta tesis doctoral.

En el primer bloque nos proponemos abordar el proceso de la generación automática de tesauros partiendo de lenguajes documentales ya generados manualmente para un dominio determinado, que pueden combinarse y enriquecerse con nuevas relaciones entre términos que no estaban relacionados o no aparecían en los tesauros de inicio, lo que constituye una contribución original. La selección de estos nuevos términos relacionados se realiza aplicando métodos estadísticos a la colección para la cual se genera el tesoro. Una vez que las nuevas relaciones han sido determinadas, es necesario establecer a qué tipo corresponden: equivalencia, jerarquía o asociatividad.

Una vez hemos generado el nuevo tesoro, se pasa al segundo bloque de esta tesis, donde afrontamos la

adaptación del tesoro generado a un entorno automatizado de recuperación de información, presentamos un lenguaje de marcado llamado SKOS Core adaptado a las necesidades de la Web semántica. El objetivo en este segundo bloque es el de ofrecer un lenguaje de marcado que garantice la interoperabilidad del tesoro en distintos entornos y su uso distribuido sin que ello suponga una pérdida de representatividad de la estructura del tesoro que hemos generado en el bloque anterior.

Finalmente en el tercer bloque definimos un marco de desarrollo basado en el paradigma software y proponemos una arquitectura que haga uso del tesoro y marcado en los bloques anteriores en un entorno distribuido de recuper