

**Título:** ENRIQUECIMIENTO DE FUENTES DE DATOS CON DESCRIPCIONES LINGÜÍSTICAS MEDIANTE BASES DE DATOS GEOGRÁFICAS EXTERNAS

**Nombre:** Sánchez Soriano, Javier

**Universidad:** Universidad Politécnica de Madrid

**Departamento:** Inteligencia artificial

**Fecha de lectura:** 20/01/2016

**Programa de doctorado:** Programa Oficial de Doctorado en Inteligencia Artificial

**Dirección:**

> **Director:** MARTÍN MOLINA GONZÁLEZ

> **Codirector:** ÓSCAR CORCHO GARCIA

**Tribunal:**

> **presidente:** LUIS GARROTE DE MARCOS

> **secretario:** DANIEL MANRIQUE GAMO

> **vocal:** DAVID ATAURI MEZQUIDA

> **vocal:** GEMMA BLASCO MORATILLA

> **vocal:** ENRIQUE PUERTAS SANZ

**Descriptores:**

> BASES DE DATOS

> SISTEMAS DE INFORMACION GEOGRAFICA

> INFORMATICA

> REDES DE COMUNICACIONES

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

> <http://oa.upm.es/39523/>

**Resumen:** El presente trabajo se ha centrado en la investigación de soluciones para automatizar la tarea del enriquecimiento de fuentes de datos sobre redes de sensores con descripciones lingüísticas, con el fin de facilitar la posterior generación de textos en lenguaje natural. El uso de descripciones en lenguaje natural facilita el acceso a los datos a una mayor diversidad de usuarios y, como consecuencia, permite aprovechar mejor las inversiones en redes de sensores.

En el trabajo se ha considerado el uso de bases de datos abiertas para abordar la necesidad de disponer de un gran volumen y diversidad de conocimiento geográfico. Se ha analizado también el enriquecimiento de datos dentro de enfoques metodológicos de curación de datos y métodos de generación de lenguaje natural. Como resultado del trabajo, se ha planteado un método general basado en una estrategia de generación y prueba que

incluye una forma de representación y uso del conocimiento heurístico con varias etapas de razonamiento para la construcción de descripciones lingüísticas de enriquecimiento de datos.

En la evaluación de la propuesta general se han manejado tres escenarios, dos de ellos para generación de referencias geográficas sobre redes de sensores complejas de dimensión real y otro para la generación de referencias temporales. Los resultados de la evaluación han mostrado la validez práctica de la propuesta general exhibiendo mejoras de rendimiento respecto a otros enfoques. Además, el análisis de los resultados ha permitido identificar y cuantificar el impacto previsible de diversas líneas de mejora en bases de datos abiertas.