
**GESTIÓN DE DATOS RDF USANDO UNA BASE DE DATOS ORIENTADA
A GRAFOS**

**SERGIO ARTURO SILVA RUBIO
INGENIERÍA CIVIL EN COMPUTACIÓN**

RESUMEN

La adopción de RDF como modelo de datos estándar para describir y publicar datos en la Web ha motivado el desarrollo de múltiples aplicaciones que emplean grandes cantidades de datos e información, para lo cual se han desarrollado sistemas de gestión de bases de datos comúnmente denominados RDF Stores. Los RDF Stores pueden clasificarse en nativos o adaptados. Un RDF Store nativo implementa estructuras de datos ad-hoc para gestionar datos RDF. Un RDF adaptado usa algún sistema de gestión de datos existente (por ejemplo, MySQL) como componente de bajo nivel encargado de la gestión de los datos. Las bases de datos orientadas a grafos son sistemas que permiten almacenar datos con estructura de grafos, y consultar estos datos usando operaciones típicas de grafos (por ejemplo, consultas de caminos). Actualmente existen múltiples aplicaciones, como por ejemplo las redes sociales o biológicas, las cuales han motivado el desarrollo de bases de datos para grafos. Teniendo en cuenta que RDF define un modelo de datos de grafo y que su lenguaje de consulta SPARQL implementa la noción de graph pattern matching, resulta intuitiva la idea de usar una base de datos de grafo para gestionar datos RDF. El objetivo principal de esta memoria fue el desarrollo de un RDF Store que adapte una base de datos de grafo para la gestión de datos RDF. Específicamente, primero se estudiaron diversos métodos para almacenar datos RDF en una base de datos de grafo. Luego de analizar las ventajas y desventajas de dos posibles métodos de almacenamiento, se definieron las reglas de transformación entre el modelo de datos RDF y el modelo de datos property graph. El método de almacenamiento seleccionado fue implementado sobre la base de datos Sparksee. En una segunda etapa, se estudió el problema de consultar los datos RDF almacenados en la base de datos de grafo usando el lenguaje de consulta SPARQL. Para esto se definió e

implementó un método que permite ejecutar una consulta SPARQL a través de los métodos disponibles en el API de Sparksee.

El RDF Store adaptado resultante fue evaluado con diversas pruebas. Adicionalmente, el sistema fue comparado con dos RDF Store nativos usando un benchmark para bases de datos de grafo denominado GDBench. La comparación mostro que la eficiencia (tiempo de respuesta) del sistema desarrollado se mantiene dentro del promedio en la mayoría de los casos, lo que es considerado bueno dada la escalabilidad que tiene el desarrollo de este sistema al ser la primera herramienta de este tipo.