

## Gestión semántica de recursos de información en sistemas de almacenamiento distribuido

**Palabras Clave:** Almacenamiento en la nube, datos abiertos enlazados, tecnologías semánticas.

### Responsable (s)

Nombre	Tiempo de dedicación
Dra. Reyna Carolina Medina Ramírez, Profesor de Tiempo Completo, RyT	10 hrs

### Participante (s)

Nombre	Tiempo de dedicación
Dr. Ricardo Marcelín Jiménez, Profesor de Tiempo Completo, RyT	10 hrs

### Área del responsable

**Área:** Redes y Telecomunicaciones

**Departamento:** Ingeniería Eléctrica

Proyecto de Investigación: Gestión semántica de recursos de información en sistemas de almacenamiento distribuido

## **Objetivo general**

Establecer un marco general para el diseño y construcción de sistemas de almacenamiento distribuido, orientados por el uso de sus contenidos documentales.

## **Objetivos particulares**

- Proponer modelos para caracterizar y gestionar la naturaleza de la información (documentos, personas) al interior de una organización.
- Desarrollar una arquitectura para la construcción de sistemas de almacenamiento masivo sensibles al contenido (SASC).
- Diseminar el conocimiento generado y la experiencia obtenida, participando en eventos especializados y publicando resultados en revistas de impacto académico.

## Descripción del proyecto

El almacenamiento distribuido (AD) es uno de los grandes temas emergentes de las tecnologías de la información. Este surge en la intersección entre las telecomunicaciones y la computación. Las posibilidades que abre este campo de las TI son enormes, como lo reconocen las empresas del ramo, los analistas económicos y los académicos. Sin embargo, los retos que plantea son también muy complejos y representan una oportunidad no sólo para desarrollar un proyecto de investigación básica, sino que posibilita la generación y transferencia de tecnología original que contribuya a la solución de problemas nacionales. Se sabe por ejemplo, que muchas instituciones públicas disponen de una infraestructura moderna de telecomunicaciones y cómputo que, sin embargo, está subutilizada. De poder coordinarse de forma cooperativa esta misma infraestructura les permitiría disponer de una capacidad de almacenamiento en el orden de los terabytes, sin incurrir en costos agregados. Se sabe también que cuando la información crece en volumen y en el número de usuarios que pueden beneficiarse de ella, no basta con la capacidad para almacenarla y recuperarla por un nombre. En muchos casos los usuarios buscan conceptos o ideas en documentos que ni siquiera saben si existen. Luego, se necesita clasificar y guardar los documentos a partir de criterios que reflejen un conjunto de nociones que articulan el conocimiento contenido en estos.

Hasta ahora, las diferentes soluciones de AD han abordado sólo la construcción de los niveles más bajos de lo que podría ser una incipiente arquitectura de almacenamiento. Para estos proyectos es transparente el tipo de datos con los que trabajan porque no existe ninguna función de interpretación de la información. Es el mismo caso de los switches y gateways de Internet, que sólo conmutan datos sin atender al contenido del segmento que reexpiden, que puede ser una secuencia de voz o un fragmento de un mail (cuando los nuevos enfoques sugieren tratamientos distintos en cada caso). Aquí existe un área de oportunidad: en la posibilidad de plantear una arquitectura de almacenamiento “inteligente” (flexible, adaptable, escalable) para acomodarse lo mismo sobre una pequeña red local, que sobre una red WAN (Wide Area Network) y que igualmente sirva para guardar recursos de información heterogénea tanto en contenido como en formato: multimedia, artículos científicos, bases de datos, por mencionar algunos. De lo anterior, proponemos en este proyecto de investigación establecer un marco general para el diseño y construcción de sistemas de almacenamiento distribuido, orientados por el uso de sus contenidos documentales que nos conduzca al diseño e implementación de una arquitectura de almacenamiento distribuido que reconozca todos los aspectos o dimensiones del problema y los jerarquice atendiendo a la complejidad de sus contenidos documentales y al nivel de agregación de la información que los soporta.

## Indicadores de desempeño y calendario

La siguiente tabla muestra los resultados esperados para el trienio 2014-17:

Componente	Productos de trabajo	Cantidad
------------	----------------------	----------

Proyecto de Investigación: Gestión semántica de recursos de información en sistemas de almacenamiento distribuido

Investigación	Artículos en revistas indizadas	2
	Artículos en memorias in extenso	3
	Presentaciones en congresos	3
	Artículos de divulgación	2
Docencia	Alumnos de licenciatura	3
	Alumnos de posgrado	3
	Desarrollos tecnológicos para apoyo a docencia	1