



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Integración Semántica de Recursos en una Memoria Corporativa

Idónea Comunicación de Resultados para obtener el grado de

MAESTRO EN CIENCIAS
(CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN)

por

Erik Alarcón Zamora

Asesores:

Dra. Reyna Carolina Medina Ramírez

Dr. Héctor Pérez Urbina

24 de septiembre de 2013

Resumen

El área de Redes y Telecomunicaciones (RyT) del departamento de Ingeniería Eléctrica (IE) de la Universidad Autónoma Metropolitana (UAM) tiene una amplia y rica variedad (heterogeneidad en formato, contenido y estructura) de recursos de información. Algunos ejemplos de estos recursos de información son: los profesores y alumnos del departamento IE, artículos científicos, notas de curso, bases de datos de los trabajadores del dpto. IE, libros, presentaciones, manuales, inventarios, especificaciones de circuitos eléctricos.

Cada recurso representa el conocimiento sobre investigaciones, colaboraciones, proyectos, cursos y temas de interés de los profesores y alumnos en el dominio RyT. Por ejemplo, los artículos científicos, presentaciones, notas de curso e inclusive el propio profesor autor de estos documentos y multimedia son fuentes de información. Todo el conocimiento de una organización representado a través de los recursos, se conoce como memoria corporativa [1].

Una adecuada gestión del conocimiento en una memoria corporativa (MC) se traduce en varias ventajas a nivel operacional, como: una organización bien informada y con mejores tomas de decisión, una herramienta de aprendizaje para las personas adscritas a la organización, una base de conocimiento persistente y accesible para estas personas, un instrumento para búsqueda, recuperación e intercambio de conocimiento entre personas, por mencionar algunas.

Para llevar a cabo esta gestión de los recursos en una MC, se necesitan dos operaciones: 1) la representación del conocimiento sobre los recursos y 2) la búsqueda sobre esta representación. En las tecnologías de la Información, hay varios enfoques tradicionales de representar/buscar el conocimiento de los recursos, como: motores de búsqueda sintácticos y bases de datos relacionales. Pero, el enfoque que nos llamó la atención, es el de las Tecnologías Semánticas.

Las Tecnologías Semánticas se basan en el uso de tecnologías, herramientas y estándares para: la representación de los recursos en un formato estándar, establecer un vocabulario conceptual, la explotación del conocimiento mediante reglas, la búsqueda y recuperación de la información a partir de la representación estándar, el uso de aplicaciones genéricas para la creación, manipulación y visualización de la información sobre los recursos, y para que los expertos en el dominio sean los encargados de suministrar y evaluar la información sobre los recursos.

En esta tesis de maestría, se propone una metodología para la representación, búsqueda, explotación e integración del conocimiento de los recursos de información en una memoria corporativa, mediante el uso de tecnologías semánticas. Esta metodología está guiada por

dos casos de uso base y la memoria corporativa es del área de RyT de la UAM.

- El primer caso de uso (Cartografía de competencias) consiste en la búsqueda de las personas (adscritas o relacionadas al depto. IE) a partir de sus características profesionales. En particular, se buscan a las personas por las competencias de profesionales, lingüísticas y sobre los temas que conocen de Redes y Telecomunicaciones. Por ejemplo, "todos los profesores de la UAM con conocimientos en radios cognitivos y que lean en inglés". Este primer caso también contempla la búsqueda de profesores que pueden impartir un curso, a partir de un conjunto de temas básicos que debe saber para dicho curso.
- El segundo caso de uso (Búsqueda de recursos digitales) consiste en la búsqueda de documentos y archivos multimedia, con base a uno o varios criterios de búsqueda (autor, título, año, temas de RyT, entre otros). Por ejemplo, "todos los artículos de Tim Berners Lee sobre Web Semántica y mayores al 2009".

La metodología para el desarrollo del modelo, la explotación y la integración del conocimiento sobre los recursos en una MC, se ha dividido en varias etapas que concuerdan con cada uno de los objetivos de la tesis. Los objetivos de la tesis son los siguientes:

- Un modelo (representación del conocimiento) de los recursos a partir de los dos casos de uso en un formato estándar.
- Un modelo coherente y del cual se explote el conocimiento sobre los recursos (ontología), a partir del uso de axiomas y un programa razonador.
- La búsqueda y recuperación (integración) de los recursos que satisfagan las necesidades informativas de los usuarios, a partir de un motor de consulta.
- Un prototipo (navegación y consultas específicas) para la interacción fácil y visual de los usuarios con el modelo .
- Evaluar los resultados devueltos y el tiempo de ejecución de las consultas a la ontología.

En las tecnologías de la web semántica, el marco de descripción de recursos (RDF) es la solución para la representación del conocimiento de manera formal sobre los recursos en la MC. La representación se basa en la descripción de las características significativas o relaciones semánticas de/entre los recursos. Por ejemplo, Jorge Aparicio Reyes tiene 29 años, vive en el Estado de México, lee en Inglés, conoce a Erik Alarcón, estudia en la UAM y tiene conocimientos en sistemas operativos, java y flash.

Si bien cada recurso de la MC tiene un nombre propio, en el marco RDF cada persona, documento, multimedia o concepto tiene un identificador único de recurso [2] (URI). Con la finalidad de no tener ambigüedades a la hora de referirse a un recurso. Por ejemplo, el URI de Jorge Aparicio es <http://www.mi-ejemplo.com/JorgeAparicio>. Para cada recurso

(identificado con URI) se describen las características/relaciones en forma de triples (sujeto-predicado-objeto) y cada elemento de un triple es un URI o en algunos casos el objeto es una Literal.

Esta representación de las características se encuentra en un formato estándar y para almacenar estos triples, se emplea un triplestore. En este trabajo de tesis se empleó el triplestore Apache Jena que proporciona almacenamiento, un motor de consulta y un razonador.

Las descripciones representan la información explícita de los recursos, pero, esta información explícita tiene conocimiento implícito. Por ejemplo, un alumno, niño, profesor, empleado, madre, hijo son personas, pero éstas como tal no tienen un triple que establezca que son personas. Entonces, para explotar este conocimiento implícito de los recursos, se proponen un conjunto de reglas o axiomas que permiten establecer estas relaciones. Aunque, para materializar estos triples a partir de los axiomas, es necesario un programa razonador que infiera estos triples. Este razonador también permite encontrar inconsistencias en el modelo. Algunos triplestores integran o permiten importar un razonador, en el caso de Jena permite las dos opciones.

El modelo que captura el conocimiento explícito (descripciones) de los recursos y los axiomas que completan el conocimiento sobre éstos, se denomina ontología. En esta tesis se hicieron dos ontologías; una para cada caso de uso, y también se modificó una ontología legada que tiene conceptos del área de RyT. Esta última ontología se emplea para vincular a personas, documentos y multimedia con los tópicos de RyT.

La consulta de los triples en el modelo, ya sea únicamente descripciones (triples explícitos) o una ontología con razonador (triples explícitos e inferidos), se hace con un motor de búsqueda (integrado en el triplestore) que compara los triples con un conjunto de patrones; aquellos triples que concuerden, se recuperará la información que se solicitó en la consulta.

Un motor de consulta y un razonador que materializa triples en una ontología, son una buena combinación, ya que permiten consultar el conocimiento inferido (triples inferidos) y reducir la complejidad de las consultas. Por ejemplo, se tienen seis individuos que afirman que son alumno, niño, profesor, empleado, madre, hijo respectivamente, también se tienen los axiomas que establecen que alumno, niño, profesor, empleado, madre, hijo son personas y se tiene la siguiente pregunta "¿Quiénes son personas?". Si se emplea solamente un motor de búsqueda, entonces no habrá ningún resultado, pero si se emplea la combinación motor y razonador, los seis individuos serán respuesta, porque estos seis individuos tienen el triple que afirma que son personas.

Los usuarios del área de RyT no están familiarizados con las tecnologías semánticas y en particular, al uso de la sintaxis de consulta. Entonces para facilitar a éstos la interacción y consulta del conocimiento de la ontología, se propone un prototipo que medie (interfaz) entre los usuarios y la ontología, específicamente este prototipo tiene los siguientes objetivos:

- Navegación a través de la información de los recursos; guiada por los casos de uso.
 - Estructurar la pregunta de un usuario.
 - Mapear las preguntas a consultas para el motor de consulta.
-

- Ejecutar la consulta con el motor de consulta, el razonador y la ontología.
- Publicar la información de los recursos respuesta en un formato visual agradable al usuario.

En esta tesis dos de los aspectos importantes a evaluar son: el desempeño de Apache Jena a la hora de consultar la ontología, así como el número y cuáles resultados responden estas consultas. Para llevar a cabo estas dos evaluaciones se obtuvieron un conjunto básico de preguntas para interrogar el modelo, para cada pregunta se sabe de ante manos el número y los recursos que la responden. En la primer evaluación, para cada consulta básica se calcula 20 veces el tiempo aproximado en milisegundos y se saca un tiempo promedio. Mientras, en la segunda evaluación, para cada consulta se compara el número/recursos que responde el motor con los recursos que previamente se sabe que la responden.

Las contribuciones de esta tesis son:

1. Una metodología para la Integración Semántica de Recursos en la MC de Redes y Telecomunicaciones.
 2. Identificación y descripción de los principales escenarios de búsqueda/recuperación de los recursos en la MC de RyT.
 3. Ontologías (Triples RDF + axiomas) que capturan el conocimiento de los recursos (apegados a los dos casos de uso) en la memoria corporativa RyT.
 4. Prototipo para la consulta interactiva de los usuarios con las ontologías de RyT.
 5. Evaluación del desempeño y calidad de resultados del triplestore Jena para la consulta de información.
-

Agradecimientos

Índice general

Índice de tablas	XI
Índice de figuras	XIII
Acrónimos	XV
1. Introducción	1
1.1. Problema a Tratar	1
1.2. Motivación Tecnologías Semánticas	3
1.3. Caso de Estudio (General)	4
1.4. Objetivos y Contribuciones	4
1.5. Estructura del Documento	5
2. Descripción del problema	7
2.1. Caso de Estudio	7
2.2. Fuentes de Conocimiento	8
2.3. Memoria Corporativa	8
2.4. Administración de una Memoria Corporativa	10
2.4.1. Administración por Fragmentos	11
2.4.2. Naturaleza de una Memoria Corporativa	12
2.4.3. Integración del Conocimiento	13
2.5. Casos de uso	14
3. Tecnologías Semánticas	15
3.1. Definiciones y descripciones	15
3.2. Marco de Descripción de Recursos (RDF)	15
3.3. Lenguaje de consulta sobre grafos RDF (SPARQL)	15
3.4. Reglas de inferencia (RDF(S)/OWL) y razonadores	15
3.5. Ventajas de las tecnologías Semánticas	15
4. Estado del arte	17
4.1. Integración semántica de recursos de información	17
4.2. Herramientas para la integración semántica de recursos	17

5. Integración semántica de recursos de información en una memoria corporativa	19
5.1. Representación del conocimiento de los recursos (modelo de datos)	19
5.2. Explotación del conocimiento en el modelo (axiomas)	19
5.3. Búsqueda y recuperación (consulta) del conocimiento de los recursos	20
6. Prototipo	21
7. Evaluación experimental	23
7.1. Escenarios de experimentación	23
7.2. Experimentación	23
7.3. Resultados	23
8. Conclusiones y Trabajo Futuro	25
Appendices	27
Apéndice A. Códigos interfaz de Usuario	29
Bibliografía	31

Índice de tablas

Índice de figuras

Acrónimos

Acrónimo	Descripción	Definición
RyT	Redes y Telecomunicaciones	7
IE	Ingeniería Eléctrica	7
UAMI	Universidad Autónoma Metropolitana Unidad Iztapalapa	7
MC	Memoria Corporativa	8
MO	Memoria Organizacional	8
TI	Tecnologías de la Información	11
TS	Tecnologías Semánticas	15
GBDR	Gestor de Bases de Datos Relacional	11
BD	Base de Datos	??

Capítulo 1

Introducción

Las personas todos los días están en contacto con diferentes organizaciones. Por ejemplo, el niño que asiste a la **escuela primaria**, el estudiante que asiste a la **universidad**, la ama de casa que compra productos en una **tienda departamental**, la persona que hace un depósito o cobrar en una **institución bancaria**, la personas que solicita un servicio en alguna **dependencia gubernamental**, el empleado trabaja en una **empresa**, inclusive una **familia** es una organización.

El concepto de organización tiene diferentes definiciones, nosotros elegimos la siguiente definición: “*una organización es una entidad a través de la cual las personas realizan actividades y de las cuales por lo menos algunas se dirigen a la consecución de fines comunes (metas) de las personas del grupo*” [3]. De esta definición, se tiene que una organización alcanza mayores logros, porque varias personas se coordinan y dirigen sus esfuerzos conjuntamente. Las organizaciones deben poner atención en las siguientes actividades para alcanzar sus metas y objetivos [4]:

1. Reunir recursos para alcanzar las metas y los resultados deseados.
2. Producir bienes y servicios de manera eficiente.
3. Buscar formas innovadoras de producir y distribuir con mayor eficiencia bienes y servicios.
4. Utilizar tecnologías de información y manufactura.
5. Adaptar, evolucionar e influir en un entorno que cambia con rapidez.
6. Crear valor para dueños, empleados y clientes.
7. Hacer frente y adaptarse a los cambios que plantea la diversidad del mundo laboral, problemas éticos, responsabilidad social y coordinación de los empleados.

1.1. Problema a Tratar

La **administración** es un concepto importante para una organización y éste se define como: “un conjunto de actividades dirigido a aprovechar los recursos de manera eficiente

y eficaz con el propósito de determinar y alcanzar los objetivos de la organización” [5]. A partir de esta definición, se tienen dos elementos importantes: actividades y recursos. Las **actividades** en una organización pueden ser *búsqueda de información, almacenamiento de los recursos, intercambio de información, control de bienes y materiales, control de inventario, colaboración con otras personas, solo por mencionar algunas*. Mientras, los **recursos** son “el medio que posee una organización para realizar las actividades que le permitan lograr los objetivos” [6]. Una organización puede tener los siguientes recursos: materiales o físicos, humanos (personas), financieros (dinero) e informáticos. La finalidad de la administración en una organización es que ésta sea estable, crezca y prospere.

La *administración en una organización* tiene diferentes enfoques que dependen de los principales elementos de la misma, por ejemplo: las metas, el proceso interno y los recursos. En particular, nuestro foco de atención son los recursos de información. Porque éstos son los instrumentos que representan y encapsulan el conocimiento de una organización. Algunos ejemplos de estos recursos son: una persona, una base de datos, un libro, un archivo multimedia, informes anuales, un equipo de cómputo, un servidor, por mencionar algunos. De esta manera, el enfoque para esta tesis es *con base en recursos* y éste se define como: “la capacidad de la organización para adquirir recursos valiosos, integrarlos y administrarlos exitosamente” [5].

La administración de los recursos puede realizarse con alguna herramienta de las Tecnologías de la Información. La finalidad de estas herramientas es facilitar, efficientar y agilizar las actividades relacionadas a la administración de los recursos. Los dos enfoques mediante el uso de estas tecnologías son: el manual y automático. Por un lado, el enfoque manual consiste en almacenar y organizar los recursos digitales (documentos, archivos de audio, presentaciones, documentos escaneados, etc) en carpetas que tienen cierta estructura. Por otro lado, el enfoque automático permite delegar ciertas tareas de gestión a programas computacionales; las dos herramientas comunes de este enfoque son: *sistemas gestores de bases de datos relacionales* o *motores de búsqueda basados en keywords*. Un *motor de búsqueda* es un sistema de recuperación de la información, que a partir de las palabras clave, realiza una búsqueda documental en la Web, y aquellos documentos que tengan las palabras clave, el motor los arrojará como respuestas al usuario. Por otro lado, un *gestor de bases de datos relacional* es un mecanismo para el almacenamiento y recuperación de la información sobre una Base de Datos, que se basa en la idea, que conjuntos de datos (bajo un mismo contexto limitado) son como tablas (relaciones). Un gestor para tareas de almacenamiento de información emplea un *esquema conceptual*. El esquema permite describir un conjunto de objetos, aspectos relevantes de los objetos, las interrelaciones entre estos y restricciones de integridad. Mientras, para fines de recuperación de la información, se emplean lenguajes de consulta sobre las bases de datos. En general, cualquiera de estas dos herramientas tiene un menor tiempo en el proceso de búsqueda, pero la calidad de los resultados va depender de los algoritmos y de la representación de los recursos.

Tanto el enfoque manual como las dos herramientas del enfoque automático no son del todo adecuadas para gestionar los recursos de una organización. En el caso de una solución manual, si hay un crecimiento explosivo de los archivos (recursos digitales), entonces

la búsqueda de recursos se vuelve un proceso tardado, pesado y cansado para las personas. Mientras, las dificultades de las dos herramientas son con respecto a la representación de los recursos, porque una inadecuada representación de éstos y del contexto, se traduce en menos resultados o resultados impertinentes.

1.2. Motivación Tecnologías Semánticas

Las tecnologías semánticas [7] son un conjunto de metodologías, lenguajes, aplicaciones, herramientas y estándares, para obtener y suministrar el significado de la información ¹. Estas tecnologías permiten la representación y la administración del conocimiento, por ello, son una solución interesante para la administración de los recursos en una organización. A continuación, se presentan los beneficios del uso de las mismas:

- **Formato estándar:** una persona, documento, objeto físico o digital, concepto, idea, en general, cualquier recurso posee información significativa y útil para las personas. Esta información puede estar embebida en el recurso o es referente a éste, por ejemplo, en un libro nos interesa saber sobre qué trata, el título, los autores, la fecha de edición, entre otros. Por otro lado, los datos de los recursos pueden ser de distintas formas: estructurados (bases de datos), semiestructurados (lenguajes de etiquetas, como XML y HTML) o sin estructura (orientados al texto). También, cada recurso puede estar almacenado en distintos tipos de archivo, por ejemplo, un documento digital puede ser un doc, pdf, odp, rtf, etc. Esta diversidad en los recursos hace difícil la administración de los mismos. Por ello, las tecnologías proponen representar los recursos a través de sus características significativas en un formato estándar, para que, los procesos automáticos puedan acceder, procesar, razonar, combinar, reutilizar y compartir esta información.
- **Explotar el conocimiento:** las características de los recursos es información explícita de los mismos. Si se introducen reglas para explotar el conocimiento implícito, entonces los procesos automáticos podrán aprovechar, completar y utilizar este conocimiento, para fines de búsqueda de la información. Por ejemplo, una persona, un perro y un gato pertenecen al campo semántico mamíferos, si se introduce la regla que establece que todo gato, perro o persona es un mamífero, entonces, un proceso automático podrá identificar quienes son mamíferos.
- **Flexibilidad e interoperabilidad:** las tecnologías semánticas pueden emplear a otros sistemas como fuentes de información. Por ejemplo, un gestor de base de datos es la principal herramienta que depende una organización, si ésta se reemplaza de golpe, entonces se puede desencadenar varios riesgos en la organización. A fin de que no suceda esto, se introduce una capa de interoperabilidad sobre estos datos, para que los datos sean transformado al formato estándar.

¹L. Feigenbaum, "Semantic Web vs. Semantic Technologies", Disponible:<http://www.cambridgesemantic.com/semantic-university/semantic-web-vs-semantic-technologies>

1.3. Caso de Estudio (General)

Existen distintos tipos de organizaciones que dependen del enfoque con el que se mira. Si es con respecto al alcance, se tienen corporaciones multinacionales, pequeños y medianos negocios, así como negocios familiares. Cuando el enfoque es el objeto final, se tienen organizaciones que fabrican productos o proveen servicios. Si es a partir de la naturaleza de la organización, se tienen instituciones económicas (empresas), fundaciones, organizaciones sin fines de lucro e instituciones públicas.

Esta tesis de maestría se enfoca en las organizaciones de investigación (institutos o universidades), porque tienen áreas o equipos de investigación. En concreto, la organización electa como caso de estudio es el grupo de investigación del **área de Redes y Telecomunicaciones** de la **Universidad Autónoma Metropolitana Unidad Iztapalapa**. Los recursos significativos en esta organización son: *personas (profesores, alumnos y colegas empleados), documentos (artículos científicos, libros, tesis), bases de datos, archivos multimedia (presentaciones, vídeos, imágenes), solo por mencionar algunos*. Porque representan el conocimiento de los profesores (miembros de esta organización) sobre sus investigaciones, colaboraciones, proyectos, actividades, cursos y temas de interés. Una adecuada administración de los recursos, se traduce en un grupo de investigación bien informado con mejores tomas de decisiones, así como una base de conocimiento persistente y accesible para los profesores y alumnos.

1.4. Objetivos y Contribuciones

El principal objetivo de esta tesis es *ver la viabilidad del uso de las tecnologías semánticas para la construcción de una base de conocimiento, con la finalidad de integrar la información sobre los recursos del área de RyT de la UAM, para responder las necesidades informativas de los usuarios de RyT*. Mientras, los objetivos particulares son:

- Desarrollar una metodología para la integración semántica de recursos en una memoria corporativa.
 - Determinar los casos de uso para la integración semántica de los recursos en una memoria corporativa.
 - Representar y explotar el conocimiento de los recursos de información (identificados en los casos de uso) del dominio de redes y telecomunicaciones de la UAM en una o varias ontologías.
 - Implementar un prototipo de interfaz de usuario que permita a éstos últimos una interacción amigable para la integración semántica de los recursos en una ontología.
 - Evaluar los resultados devueltos en la integración semántica para la memoria corporativa de redes y telecomunicaciones.
-

Mientras, las contribuciones de tesis son las siguientes:

1. Metodología para la integración semántica de recursos en una memoria corporativa.
2. Identificar los principales casos de uso para la integración semántica (cartografía de competencias y búsqueda de recursos digitales)
3. Estado del arte de la integración semántica de recursos y las herramientas para la generación de triples, editores de ontologías y triplestore.
4. Tres ontologías para modelar el conocimiento de los recursos en la memoria corporativa de redes y telecomunicaciones.
5. Prototipo de interfaz de usuario para la interacción amigable de los mismos en la integración semántica de los recursos de la memoria de redes y telecomunicaciones.
6. Evaluación del desempeño en el proceso de consulta y evaluación de la precisión de los resultados al triplestore Jena con nuestras ontologías.

1.5. Estructura del Documento

Al organizar esta tesis, hemos querido establecer un camino coherente para alcanzar cada uno de los objetivos planteados. Los capítulos se organizan de la siguiente manera:

El capítulo 2 describe la problemática principal de esta tesis, la cual es la integración semántica de recursos en una memoria corporativa, así como algunos conceptos básicos (memoria corporativa, integración, recurso). Los principales conceptos, definiciones, estándares de los elementos pertenecientes a las tecnologías semánticas, se presentan en el capítulo 3. En el capítulo 4 se presentan los dos estado del arte: el primero es sobre la integración semántica de los recursos en una memoria corporativa, mientras el segundo es sobre las herramientas para la generación de triples, editores de ontologías y triplestore. El capítulo 5 describe nuestra metodología para la integración semántica de recursos en una memoria corporativa. Primero, se presenta la arquitectura para la integración semántica de recursos. Segundo, se describe la representación (modelo) de los recursos en un formato estándar a partir del conocimiento explícito de éstos. Tercero, se describe los axiomas para completar el conocimiento y el uso de un razonador para explotar el conocimiento implícito de los recursos. Cuarto, se describe la manera de interrogar el conocimiento de los recursos en el modelo. El capítulo 6 describe los objetivos y características del prototipo para la integración semántica de recursos. Las pruebas y resultados (desempeño y calidad de las respuestas) hechos/obtenidos al triplestore Jena, así como al modelo para el área de redes y telecomunicaciones, se presentan en el capítulo 7. Finalmente, las conclusiones sobre la integración semántica de los recursos, el uso de las tecnologías semánticas y los resultados de nuestra experimentación, se presentan en el capítulo 8. En esta sección también se presentan algunos trabajos futuros que identificamos.

Descripción del problema

2.1. Caso de Estudio

El *área de Redes y Telecomunicaciones* (RyT) es una de las cinco áreas académicas en que se organiza el departamento de Ingeniería Eléctrica (IE) de la Universidad Autónoma Metropolitana Unidad Iztapalapa (UAM-I). En esta área se cultivan las siguientes líneas de investigación: *Redes y Servicios de Telecomunicaciones, Sistemas de Comunicación Digital, Sistemas Distribuidos y Web Semántica*.

El área de RyT es una organización que se constituye por un conjunto de personas. Ellas desempeñan las actividades de investigación, académicas, preservación y difusión de la cultura. Las personas de RyT pueden ser clasificadas en dos tipos: las que pertenecen al núcleo del área y las temporales. Las personas del núcleo del área son los **profesores-investigadores**. Ellos se encargan de realizar las siguientes actividades: *planear, definir, dirigir, coordinar y evaluar los cursos de las licenciaturas en Computación, Ingeniería Electrónica, Posgrado en Ciencias y Tecnologías de la Información, investigación, así como la investigación y desarrollo de proyectos asociados a sus líneas de investigación*. Ahora bien, las **personas temporales** trabajan con el personal del núcleo, ya sea en la investigación, colaboración, ayuda o servicios administrativos. Estas personas tienen un rol menos activo en el área. porque el tiempo en que trabajan es un periodo corto. Algunos ejemplos de este tipo de personas son: 1) *alumnos que realizan algún proyecto o servicios social y cuyo responsable de ellos es un profesor del núcleo*, 2) *profesores temporales que imparten cursos relacionados con los temas de Redes y Telecomunicaciones*, 3) *empleados de la universidad que proporcionan servicios administrativos a los profesores del núcleo* y 4) *empleados de otras organizaciones que colaboran con los profesores del núcleo*.

En cuanto a la cantidad de personas involucrada o que se han involucrado en el área RyT. Para el núcleo se tienen trece profesores-investigadores. Mientras el número de personas temporales, que han participado o participan con las personas del núcleo, no hay un número exacto de éstas. Porque cada profesor-investigador tiene su lista de personas conocidas (estudiantes y colegas) y cada trimestre estas listas se van incrementando.

2.2. Fuentes de Conocimiento

El conjunto de personas del área es el elemento más importante para ésta. Porque ellas realizan las actividades para lograr las metas y objetivos del área de RyT. Las personas al realizar sus actividades cotidianas y estructuradas, se convierten en las constructoras del conocimiento para la organización. Las etapas para la construcción del conocimiento son la *adquisición y representación*.

1. Las personas consiguen y hacen propio el conocimiento por distintas maneras, como: *la experiencia, al realizar sus actividades cotidianas; la observación, análisis, experimentación, evaluación y en general por distintas actividades de la investigación; la búsqueda, obtención, almacenamiento, recopilación, lectura, visualización y consulta de distintos soportes (documento, imagen, audio, vídeo); enseñanza y aprendizaje entre personas; por mencionar algunas*. Estas personas utilizan este conocimiento para ejecutar sus actividades y tareas en la organización.
2. Las personas realizan dos actividades con el conocimiento: *1) mantener el conocimiento en su mente o 2) hacer presente el conocimiento con palabras, imágenes, sonidos, símbolos en algún soporte como documento, imagen, audio, presentación, base datos, hoja de cálculo o vídeo*. En la primera actividad, la *representación del conocimiento es intangible*, como habilidades, destrezas profesionales, conocimiento privado o el conocimiento de la organización. La finalidad este conocimiento es que las personas sean instrumentos de conocimiento para realizar determinadas tareas o solucionar problemas específicos en la organización. Mientras, en la segunda actividad la *representación del conocimiento es tangible*. La finalidad de esta representación es que los objetos (recursos inanimados) conserven y transmitan la información a las personas de la organización.

Personas y recursos inanimados se agrupan bajo el concepto de **recurso de información o conocimiento**. En el área de RyT, los recursos de información son el conocimiento de *investigaciones, colaboraciones, proyectos, cursos, temas de interés, objetos, ideas o conceptos vinculados con los tópicos de Redes y Telecomunicaciones*. Esta área tiene las siguientes clases de recursos: *artículos científicos, presentaciones, libros, equipos de cómputo, bases de datos, tesis, reportes técnicos, audios, vídeo tutoriales, notas de curso, tareas, imágenes, páginas web, profesores, alumnos, empleados de otras organizaciones, servidores computacionales, programas y aplicaciones computacionales científicas-académicas*.

2.3. Memoria Corporativa

Los recursos de información expresan el conocimiento en la organización. A este conocimiento se denomina memoria corporativa (MC) o memoria organizacional (MO). Una definición formal de este concepto es la siguiente: “*una memoria corporativa es la representación explícita, tácita, consistente y persistente del conocimiento en una organización*” ???. Por explícita, se refiere a que el conocimiento se expresa de manera clara y formal. Representación

tácita significa que ciertas partes del conocimiento no se mencionan formalmente, sino que deben inferirse; por ejemplo, una mujer y un hombre son personas. Por consistente, se traduce en que el conocimiento es estable y no sufre grandes cambios. Persistente, es una cualidad temporal y se refiere a que el conocimiento debe durar por un tiempo prolongado.

Una memoria corporativa conserva y mantiene el conocimiento de una organización [8], con la finalidad de *facilitar el acceso, intercambio y difusión del mismo*. De esta manera, las personas adscritas o interesadas en la organización podrán *adquirir, reutilizar y razonar* con este conocimiento y realizar nuevas actividades o mejorarlas. Por ejemplo, aportar nuevas ideas, modificar ciertos aspectos en su trabajo, colaborar e intercambiar puntos de vista con sus colegas, abarcar otros mercados, generar mayor conocimiento, actualizar la información, por mencionar algunas.

En una organización existen distintas razones para tener una memoria corporativa. Rose Dieng et al. proponen una lista básica de razones [8]:

- Prevenir la pérdida del conocimiento de los expertos, cuando éstos salgan de la organización.
- Aprovechar las experiencias buenas y malas de trabajos pasados, con la finalidad de mejorar el trabajo y no caer en los mismos errores.
- Aprovechar el conocimiento global para mejores tomas de decisión en la organización.
- Mejorar las capacidades de la organización para reaccionar y adaptarse a los cambios.
- Mejorar la circulación de la información y la comunicación entre las personas de la organización.
- Mejorar el aprendizaje de las personas en la organización.
- Integrar el conocimiento fundamental de una organización, como flujos de trabajo, productos, técnicas, información secreta.

Una memoria corporativa representa todo el conocimiento de la organización: actividades de producción, datos técnicos, requisitos de productos, experiencia, habilidades, entre otros. Algunos autores hacen una clasificación de las memorias corporativas [8], con base a distintos puntos de vista de la administración:

- Memoria técnica: comprende el conocimiento, información y datos que obtienen los empleados en la organización.
 - Memoria gerencial: comprende el conocimiento de las estructuras organizativas de la organización.
 - Memoria de proyecto: comprende la definición del proyecto, actividades, historia y resultados
-

- Memoria de profesión: integra las referencias, documentos, herramientas y métodos para una determinada profesión.
- Memoria de empresa: comprende las actividades, productos, clientes, proveedores, contratistas, entre otras personas.
- Memoria individual: integra las competencias, estatus, conocimientos, actividades de una determinada persona en la organización.
- Memoria interna: comprende el conocimiento e información interna de la organización.
- Memoria externa: comprende el conocimiento e información útil para la organización que proviene del mundo externo.

El conocimiento del área de Redes y Telecomunicaciones no se clasifica en un determinado tipo de memoria corporativa. Porque este conocimiento representa más de un punto vista (multienfoque) de la clasificación de memorias (técnico, gerencial, profesional, individual, por mencionar). Este multienfoque del conocimiento de RyT se presenta en la sección ?? de este presente capítulo.

2.4. Administración de una Memoria Corporativa

Una memoria corporativa (MC) es uno de los principales elementos para una organización y las personas adscritas o interesadas en ésta, por esta razón, es importante la **administración de la memoria corporativa**. La *administración* es un concepto interesante para las organizaciones. Este concepto se define como: “*un conjunto de actividades dirigidas a aprovechar los recursos de manera eficiente y eficaz, con el propósito de determinar y alcanzar los objetivos en la organización*” [5].

La administración del conocimiento es un problema complejo que puede ser abordado de distintos enfoques: financieros y económicos, técnicos, metas, proceso interno, entre otros. En particular, el conocimiento prioritario para esta tesis son los **recursos de información**: 1) *elementos tangibles* como datos, procedimientos, planes, documentos, audios, vídeos, presentaciones, tesis, libros, por mencionar algunos y 2) *elementos intangibles* como habilidades, destrezas profesionales, conocimiento privado y el conocimiento del contexto en la organización.

Los **objetivos** en la administración de una memoria corporativa son: *integrar el conocimiento disperso en la organización, preservar y difundir el conocimiento, facilitar el acceso y visibilidad del conocimiento, tener con un instrumento para el aprendizaje, facilitar la búsqueda y recuperación del conocimiento, promover la comunicación y cooperación entre personas, emplear un lenguaje técnico entendido por todas las personas, promover el crecimiento e intercambio del conocimiento, facilitar la compartición de nuevas ideas, mejorar las tomas de decisión, por mencionar algunos*.

Un analogía de la administración de los *recursos de información* se presenta a continuación. Una biblioteca es una organización dedicada a la *adquisición, conservación, exposición*

y préstamo de libros. Para llevar a cabo estas tareas, la biblioteca realiza distintas actividades de administración con los libros. Las actividades básicas en la administración de los libros son: *caracterizar los libros, generar las fichas bibliográficas, clasificar las fichas de acuerdo a ciertos parámetros, asignar un identificador a cada libro, acomodar el libro de acuerdo a la clasificación y al identificador, generar un catálogo de todos los libros; consultar el catálogo, retirar el libro del estante, dar de baja un libro en el catálogo, indicar a quién se le presta el libro, indicar una fecha de devolución; dar de alta el libro en el catálogo y regresar el libro a su ubicación*. Este flujo de actividades las podemos agrupar en seis actividades generales: **representar, almacenar, clasificar, consultar, recuperar y actualizar**.

Una memoria corporativa debe administrar los recursos de información, de manera semejante a como, una biblioteca administra los libros. En la actualidad, la administración de los recursos se hace mediante el uso de las **Tecnologías de la Información** (TI). Esta tecnologías proporcionan un conjunto de herramientas, enfoques y aplicaciones para facilitar, agilizar y automatizar distintas actividades o procesos.

2.4.1. Administración por Fragmentos

En el área de Redes y Telecomunicaciones (RyT), la administración del conocimiento se hace de manera individual, es decir, cada profesor, alumno o empleado administra sus recursos de información. Porque cada persona tiene intereses particulares (líneas de investigación) y emplea la herramienta que más le conviene. Estas personas administran sus recursos mediante dos enfoques:

- El **enfoque manual** consiste en almacenar los recursos de información (recolectados o generados) en carpetas organizadas. Estas carpetas están estructuradas de forma jerárquica y cada recurso tiene un nombre significativo. Las personas para recuperar los recursos, tienen que buscar en las carpetas e identificar el recurso con base al nombre o al contenido de éste.
 - El **enfoque automático** consiste en emplear alguna aplicación para automatizar el almacenamiento, búsqueda y recuperación de los recursos. Los profesores emplean como aplicaciones a motores de búsqueda basados en keywords y gestores de bases de datos relacionales. Los *motores de búsqueda basados en keywords* hacen una búsqueda documental de acuerdo a las palabras (keywords) que un usuario escribe. Los resultados de esta búsqueda se presentan como un ranking de enlaces a los documentos fuente. Un motor de búsqueda no realiza actividades que se relacionan al almacenamiento de los documentos. Estos motores generan índices del contenido de los documentos, para facilitar el trabajo de búsquedas futuras. Mientras, un *gestor de bases de datos relacional* (GBDR) almacena, modifica y recupera la información en una base de datos (BD). La consulta de información se hace mediante un lenguaje de consulta estructurado. Los resultados asociados a las consultas, se presentan en forma de tabla. Un GBDR necesita de esquema relacional para almacenar y actualizar la información en la base de datos.
-

Estos dos enfoques en la administración de recursos de información se aplican a fragmentos de la memoria corporativa. Sin embargo, todos los recursos de la memoria corporativa no se administran bajo un mismo enfoque. Ahora bien, *cuál es el enfoque o herramienta para aprovechar los recursos de manera eficiente y eficaz*. Para tomar esta decisión, deben ser analizadas: 1) las características de una memoria corporativa y 2) los beneficios de los distintos enfoques de las Tecnologías de información.

2.4.2. Naturaleza de una Memoria Corporativa

En una memoria corporativa, los recursos de información tienen distintas cualidades que deben considerarse para administrar el conocimiento de éstos. Porque estas cualidades pueden causar dificultades en etapas tempranas del proceso de administración. Esta tesis presenta las principales características a considerar en la gestión de una memoria corporativa. En particular, las características de la memoria del área RyT.

Diversidad en formato

Esta característica tiene que ver con los recursos digitales. En el área de RyT, los recursos digitales se clasifican de acuerdo al soporte (documento, audio, vídeo, presentación, imagen, base de datos y código). Los recursos pertenecientes a un determinado soporte, no tienen el mismo formato que otros recursos pertenecientes a otros soportes. Inclusive, recursos pertenecientes al mismo soporte, no necesariamente todos tienen el mismo formato. Esto se debe a la gran ***diversidad de formatos*** que emplean las aplicaciones como: *procesadores de texto, hojas de cálculo, editores de código, bases de datos, por mencionar algunas*. Por ejemplo, los recursos de información que son documentos tienen los siguientes formatos: *pdf, doc, txt, docx, odp, tex y html*. Idealmente, se podría pensar que todos estos recursos sean guardados con el mismo formato. Sin embargo, esto no sucede porque las personas emplean distintas aplicaciones computacionales. En la gestión del conocimiento se debe considerar esta *diversidad en formato* que cambien se denomina ***heterogeneidad en formato***.

Diversidad en Contenido

El conocimiento del área de Redes y Telecomunicaciones se clasifica en las cuatro líneas de investigación: *Redes y Servicios de Telecomunicaciones, Sistemas de Comunicación Digital, Sistemas Distribuidos y Web Semántica*. Cada línea tiene un conjunto de temas que se relacionan a ésta. Por ejemplo, la línea de Sistemas Distribuidos tienen los siguientes temas: *p2p, middleware, estado global, sistema operativo, replicación, concurrencia, sincronización, por mencionar algunos*.

En una memoria corporativa, un recurso en su contenido representa el conocimiento de uno o más temas de una línea de investigación. Por ejemplo, un conjunto de documentos pueden tener el mismo formato, pueden pertenecer a la misma organización, pero éstos pueden representar distintos temas como: p2p, middleware o estado global. De esta manera, se puede

afirmar que una memoria corporativa tiene una *variedad en el contenido de los recursos*. Esta diversidad también se conoce como *heterogeneidad en contenido*.

Diversidad en la Estructura

Los datos en los recurso digitales aparecen en distintas formas. Éstos se pueden clasificar en tres formas: 1) **datos estructurados**: la información se apega a una estructura formal, como el modelo relacional en las bases de datos, 2) **datos semi-estructurados**: la información está contenida entre etiquetas para marcar el contenido de recurso, y 3) **datos sin estructura**: la información es orientada al texto. Ejemplos de estos tres tipos son los siguientes: una base de datos con los datos de los profesores del área de RyT es ejemplo de datos estructurados, páginas web son ejemplos de datos semi-estructurados, notas de un curso son ejemplos de datos sin estructura.

Significado de la Información

Los recursos de información contienen palabras (escritas o habladas), símbolos lingüísticos, expresiones o situaciones, en general, información. Esta información puede ser entendida e interpretada sin ningún problema. Sin embargo, la naturaleza de nuestro lenguaje (escrito y oral) puede llevar a confusiones y malas interpretaciones. En particular, se puede tener dificultades con las siguientes cualidades de las palabras: *homonimia y la sinonimia*. La **homonimia** es la relación entre palabras que se escriben o pronuncian igual y tienen distinto significado. La **sinonimia** es la relación entre palabras que se escriben o pronuncian diferente y tienen el mismo significado. Un ejemplo de homonimia es la palabra radio, ya que esta palabra tiene distintos significados que se asocian a la Química, Comunicación, Anatomía o Geometría. Mientras, un ejemplo de sinonimia son las palabras resumen, sumario, síntesis y recapitulación.

2.5. Integración del Conocimiento

La **administración en una memoria corporativa (MC)** contempla varias actividades (representar, almacenar, clasificar, consultar, recuperar, actualizar, entre otras) que puede prolongar el tiempo y la complejidad de ésta. Además, en esta administración se debe contemplar las características de una memoria corporativa. Por estas razones se debe limitar el conjunto de actividades a una menor cantidad, es decir, ajustar el alcance de esta administración.

En la administración en una memoria corporativa existen distintos objetivos que son los elementos prioritarios, para alcanzar la finalidad de ésta (promover el acceso, intercambio y difusión de conocimiento). En particular, los siguientes objetivos prioritarios tienen una relación cercana: integrar el conocimiento disperso en la organización, facilitar el acceso y visibilidad del conocimiento, tener con un instrumento para el aprendizaje y facilitar la búsqueda y recuperación del conocimiento.

El análisis de estos objetivos, nos lleva a un problema de integración de la información o del conocimiento. La ***integración del conocimiento*** es el proceso de utilizar el conocimiento en un dominio dado que previamente se ha representado y almacenado, con el fin de llevar a cabo actividades de búsqueda, recuperación y combinación de la información en los recursos. Esta integración debe proporcionar la información correcta a la consulta o pregunta del usuario.

2.5.1. Casos de uso

El objetivo de esta tesis es realizar la **integración de los recursos de información en una memoria corporativa**. En particular, se integra la memoria corporativa del área de Redes y Telecomunicaciones de la Universidad Autónoma Metropolitana.

En cualquier organización existe una gran cantidad de recursos de información. Esta cantidad de recursos hace difícil las actividades de integración del conocimiento. Para limitar la cantidad de los recursos a un conjunto manejable y representar formalmente el conocimiento del área de Redes y Telecomunicaciones. Nosotros proponemos descubrir y registrar los principales **casos de uso** para la integración de los recursos. La finalidad de éstos es identificar: los objetivos de los usuarios y los principales recursos de información.

En este trabajo, se identificaron dos casos de uso para la integración de los recursos. Estos casos se describen a continuación:

En el área de RyT los principales recursos de información son las personas y los recursos inanimados que están generan/recolectan. Por un lado, las personas del área se clasifican en profesores, investigadores, alumnos y empleados. Por otra lado, las personas del área emplean más a los recursos digitales. Ejemplos de estos recursos son: artículos científicos, libros, reportes técnicos, páginas web, tesis, otros documentos, audios, vídeos, presentaciones, imágenes y otros archivos multimedia.

Cartografía de Competencias

Las personas tienen actitudes, valores, conocimientos técnicos, habilidades individuales y colectivas. Estas características de una personas son importantes para una organización. Porque con base en esas características se identifican las personas que pueden realizar determinadas tareas, solucionar problemas específicos, hacer colaboraciones o tener un determinado cargo.

La **cartografía de competencias** consiste en la búsqueda y recuperación de las personas a partir de las características profesionales. Los principales parámetros en la búsqueda de personas son: las competencias profesionales (*trabajo en equipo, liderazgo, organizar, planificar*), conocimientos en temas de Redes y Telecomunicaciones (*sistemas operativos, capa enlace, filtros, ontologías, radios cognitivos*), capacidades lingüísticas (*lee en inglés, escribe en español, habla en francés*), relaciones profesionales (*colega, asesor o conocido*) y finalmente por la ocupación (*estudiante, empleado o profesor*).

Para cada resultado (persona) de una pregunta debe recuperarse información significativa de éste, con la finalidad de auxiliar al usuario, ya sea con información para localizar a la

persona, un medio de comunicación (correo o teléfono) y otras características útiles (sexo, edad, habilidades).

Búsqueda de Recursos Digitales

Los recursos digitales son la llave para solucionar las dificultades de una persona. Porque las personas emplean a estos recursos como objetos de aprendizaje. Los recursos digitales representan ideas, objetos, teorías, procesos, flujos de trabajo y el conocimiento en la organización. Por ello, es importante que una persona recupere los recursos digitales para responder sus necesidades de información.

La **búsqueda de recursos digitales** consiste en la búsqueda y recuperación de los documentos y archivos multimedia a partir del contenido de éstos. Los principales parámetros de búsqueda de los recursos digitales son: el autor, la extensión (*ppt, wav, mp3, mpg, jpg*), relaciones con los temas de Redes y Telecomunicaciones (*sistemas operativos, capa enlace, filtros, ontologías, radios cognitivos*), el idioma fuente (*inglés, español, francés, ruso, chino*), tipo de recurso digital (*artículos, reportes técnicos, páginas web, tesis, libros, audios, vídeos, imágenes y presentaciones*) y la organización a la que pertenece (*uam, unam, ipn, iee, acm, oracle*).

Para cada resultado (recurso digital) de una pregunta debe recuperarse información significativa de éste, con la finalidad de apoyar al usuario, ya sea con información para localizar el archivo (ruta), el autor, un resumen del recurso, el año y otras características útiles (número de páginas, extensión, lenguaje fuente).

Capítulo 3

Tecnologías Semánticas

(TS) La ontología es un nuevo objeto de IA, que recientemente llegó a la madurez y una potente herramienta de modelado conceptual del conocimiento. Proporciona una base coherente para construir, y una referencia común para alinearse con, en forma de un vocabulario conceptual consensuado, en el que se puede construir descripciones y comunicación del informe acts.This presenta el objeto que se llama ñna ontologíaz un estado de la técnica de técnicas de ingeniería para ontologías

3.1. Definiciones y descripciones

Algún texto...

3.2. Marco de Descripción de Recursos (RDF)

Más texto...

3.3. Lenguaje de consulta sobre grafos RDF (SPARQL)

Más texto...

3.4. Reglas de inferencia (RDF(S)/OWL) y razonadores

Más texto...

3.5. Ventajas de las tecnologías Semánticas

Más texto...

4.1. Integración semántica de recursos de información

Algún texto...

4.2. Herramientas para la integración semántica de recursos

Más texto...

Integración semántica de recursos de información en una memoria corporativa

Las reglas para la descripción multinivel son las siguientes: ? Descripción de lo general a lo particular ? Información pertinente para cada nivel de descripción ? Vinculación de las descripciones ? No repetición de la información

Como advierte en (Nonaka, 1991; Van Engers, Mathies, Leget y Dekker, 1995), la cadena del conocimiento consta de siete enlaces: lista de los conocimientos existentes, determinar los conocimientos, el desarrollo de nuevos conocimientos, la asignación de conocimiento nuevo y existente, la aplicación del conocimiento , el mantenimiento de los conocimientos, ya que dispone de los conocimientos.

En particular, permiten hacer el proceso de integración (búsqueda y recuperación) de información significativa de los recursos en una memoria corporativa. Para lograr esta integración, se deben efectuar las siguientes actividades: *1) modelar el conocimiento de los recursos en un formato estándar, 2) explotar el conocimiento implícito de los recursos y describir el vocabulario (conceptos y relaciones) de la memoria, y 3) buscar y recuperar la información sobre los recursos, para responder una pregunta dada.*

5.1. Representación del conocimiento de los recursos (modelo de datos)

Algún texto...

5.2. Explotación del conocimiento en el modelo (axiomas)

Más texto...

5.3. Búsqueda y recuperación (consulta) del conocimiento de los recursos

Más texto...

Capítulo 6

Prototipo

Evaluación experimental

7.1. Escenarios de experimentación

Algún texto...

7.2. Experimentación

Más texto...

7.3. Resultados

Más texto...

Conclusiones y Trabajo Futuro

Algún texto...

Appendices

Apéndice A

Códigos interfaz de Usuario

Bibliografía

- [1] L. Gandon, Fabien. Ontology Engineering: a Survey and a Return on Experience. Technical Report RR-4396, INRIA, March 2002.
- [2] T. Berners-Lee, R. Fielding, and L. Masinter. Uniform resource identifiers (uri): Generic syntax. 1998.
- [3] James G. March, Herbert A. Simon, and Harold S. Guetzkow. *Teoría de la Organización*. Ariel, 1987.
- [4] Richard L. Daft. *Teoría Y Diseño Organizacional*. Cengage Learning, 09 edition, 2007.
- [5] Reinaldo O. Silva. *Teorías de la administración*. Thomson, 01 edition, 2002.
- [6] Juan J. Gilli. *Diseño organizativo: estructura y procesos*. Granica, 2007.
- [7] S. Alfred, A. Arpah, L. H S Lim, and K. K S Sarinder. Semantic technology: An efficient approach to monogenean information retrieval. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, pages 591–594, 2010.
- [8] Rose Dieng, Olivier Corby, Alain Giboin, and Myriam Ribière. Methods and Tools for Corporate Knowledge Management. Technical Report RR-3485, INRIA, September 1998.