



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

16 de diciembre de 2013

Resumen

El área de Redes y Telecomunicaciones (RyT) pertenece al departamento de Ingeniería Eléctrica (IE) de la Universidad Autónoma Metropolitana (UAM). Esta área tiene una amplia y rica variedad de *recursos de información*. Estos *recurso de información* representan el conocimiento sobre investigaciones, colaboraciones, proyectos, cursos, también temas de interés de los profesores y alumnos en el dominio RyT. Ejemplos de *recursos de información* son: *profesores y alumnos del departamento IE, artículos científicos, notas de curso, bases de datos de los trabajadores del dpto. de IE, libros, presentaciones, manuales, inventarios, especificaciones de circuitos eléctricos*.

Este conocimiento sobre una organización, expresado a través de los recursos, se conoce como memoria corporativa [1]. Una gestión de una *memoria corporativa* se traduce en varias ventajas a nivel operacional, por ejemplo, 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.

La gestión de una memoria establece varias actividades para aprovechar los recursos de manera eficaz y eficiente. Ejemplos de estas actividades son: el almacenamiento, búsqueda, acceso, creación, mantenimiento, entre otras. En este trabajo, éstas son dos actividades críticas en la gestión de una *memoria corporativa*. Primero, la representación del conocimiento en los *recursos de información*. Segundo, la búsqueda y recuperación de información sobre esta representación.

La *representación y búsqueda* pueden llevarse a cabo con distintos enfoques tradicionales de las *tecnologías de la información*. Ejemplos de estos enfoques son: *motores de búsqueda sintácticos y bases de datos relacionales*. En particular, nosotros estamos interesados en el uso las *tecnologías semánticas*.

Las *tecnologías semánticas* son metodologías, herramientas, lenguajes y sintaxis estándares. Estas tecnologías permiten realizar estas actividades: 1) representar la información sobre los recursos en un formato estándar, 2) establecer un vocabulario conceptual, 3) enriquecer el conocimiento mediante reglas de inferencia, 4) buscar y recuperar la información a partir de la *representación estándar*, 5) usar aplicaciones genéricas para la creación, manipulación y visualización de la información sobre los recursos, así como 7) permitir a los expertos del dominio que suministren y evalúen la información en los recursos.

En este trabajo, se propone una metodología para la *integración semántica de recursos*

en una *memoria corporativa*. Esta metodología se implementa en el área de Redes y Telecomunicaciones. Pero, ésta puede ser implementada en otra memoria corporativa, por ejemplo, Biomédica, Química, Biología, entre otras. Porque éstas están compuestas por *recursos de información*.

Esta metodología está dividida en tres etapas. La primera etapa es la representación del conocimiento sobre los *recursos de información* en una *memoria corporativa*. La segunda etapa es la introducción de *reglas de inferencia*, para hacer explícito el conocimiento implícito de la memoria. La tercera etapa es la *búsqueda y recuperación* inteligente de información en la *representación del conocimiento*.

La metodología propuesta emplea dos *casos de uso* para guiar la realización del proceso de *integración semántica*. Pero, ésta puede extenderse a otros casos de uso. Éstos son los dos *casos de uso* que se emplean en este trabajo.

- 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".

En este trabajo, además de una metodología, se propone la obtención de distintos objetivos para la *integración semántica*, los cuales son:

- Una *representación del conocimiento* (modelo) acerca de los *recursos de información* en la memoria corporativa del área de Redes y Telecomunicaciones.
 - Un modelo construido mediante el uso de *tecnologías semánticas* y que tiene una semántica definida.
 - Un modelo guiado por dos casos de uso: *cartografía de competencias* y *búsqueda de recursos digitales*.
 - Un modelo que explote el conocimiento de los recursos, mediante la introducción de reglas de inferencia.
 - Búsqueda y recuperación de información en el modelo, para responder las preguntas de los usuarios.
-

- Un prototipo de *interfaz gráfica de usuario* que permita a cualquier usuario, realizar la navegación y consulta de información de manera fác.
- Una evaluación de: *la calidad de los resultados* y *el tiempo de consulta* sobre el modelo.

Agradecimientos

Contenido

Lista de Tablas	XI
Lista de Figuras	XIII
Acrónimos	XVIII
1. Introducción	1
2. Descripción del Problema	7
2.1. Memoria Corporativa	8
2.1.1. Administración de una Memoria Corporativa	9
2.1.2. Naturaleza de una Memoria Corporativa	11
2.1.3. Integración del Conocimiento	12
2.2. Casos de uso	13
2.2.1. Cartografía de Competencias	14
2.2.2. Búsqueda de Recursos Digitales	14
2.3. Hipótesis	15
3. Tecnologías Semánticas	17
3.1. Introducción y definiciones	17
3.2. Marco de Descripción de Recursos	17
3.3. Lenguaje de consulta sobre grafos RDF (SPARQL)	21
3.4. Reglas de inferencia (RDF(S)/OWL) y razonadores	21
3.5. Ventajas de las tecnologías Semánticas	25
4. Estado del arte	31
4.1. Integración semántica de recursos de información	31
4.2. Herramientas para la integración semántica de recursos	35
4.3. Resumen y Discusión	38
5. Integración semántica de recursos de información en una memoria corporativa	43
5.1. Representación del conocimiento en los recursos	46

5.2.	Enriquecimiento del conocimiento en el modelo	54
5.2.1.	Herencia de clases	55
5.2.2.	Herencia de propiedades	58
5.2.3.	Domingo y rango de propiedades	61
5.2.4.	Simetría en las propiedades	65
5.3.	Búsqueda y recuperación de información en el modelo	68
6.	Prototipo	81
7.	Desarrollo experimental	107
7.1.	Observaciones	107
7.2.	Hipótesis	107
7.3.	Experimentación	108
7.4.	Sujetos de experimentación	108
7.5.	Metodología	110
7.6.	Resultados	117
8.	Conclusiones y Trabajo Futuro	123
Appendices		127
Apéndice A.	Algoritmos para la generación de datos simulados	129
Apéndice B.	Transformación de preguntas a consultas SPARQL	133
Bibliografía		145

Lista de Tablas

3.1. Ejemplos de identificadores (URI) asignados a distintos recursos.	18
3.2. Ejemplos de identificadores asociados a distintas propiedades.	18
3.3. Ejemplos de tripletas que emplean la propiedad <i>rdf:type</i> para asignar un recurso a una determinada clase.	20
4.1. Características relevantes para la integración semántica de recursos en una memoria corporativa.	39
5.1. Ejemplos de identificadores URI asociados a los recursos persona para la cartografía de competencias.	50
5.2. Ejemplos de identificadores URI asociados a los documentos y archivos multimedia para la búsqueda de recursos digitales.	51
5.3. Ejemplos de identificadores URI de las propiedades pertenecientes a la cartografía de competencias.	51
5.4. Ejemplos de identificadores URI de las propiedades pertenecientes a la búsqueda de recursos digitales.	51
5.5. Dominio y Rango para las propiedades asociadas a la cartografía de competencias.	62
5.6. Dominio y Rango para las propiedades asociadas a la búsqueda de recursos digitales.	63
5.7. Identificadores de los recursos resultantes sin inferencia para la consulta de información.	78
5.8. Identificadores de los recursos resultantes a partir del uso de inferencia para la consulta de información.	80
7.1. Número de recursos persona por clase.	110
7.2. Número de recursos digitales por clase.	111
7.3. Preguntas y cantidad de personas que responden a éstas.	112
7.4. Preguntas y cantidad de recursos digitales que responden a éstas.	113
7.5. Recursos relevantes y no relevantes asociados a las preguntas de la cartografía de competencias.	117
7.6. Recursos relevantes y no relevantes asociados a las preguntas de la búsqueda de recursos digitales.	118

7.7. Exhaustividad y precisión para las consultas de la cartografía de competencias.	119
7.8. Exhaustividad y precisión para las consultas de la búsqueda de recursos digitales.	119
7.9. Tiempo promedio de procesamiento para las consultas de la cartografía de competencias.	120
7.10. Tiempo promedio de procesamiento para las consultas de la búsqueda de recursos digitales.	121

Lista de Figuras

2.1. Diagrama de casos de uso para la integración de los recursos de una memoria corporativa.	13
3.1. Ejemplos de tripletas asociadas a las declaraciones para los recursos Juan y libro de matemáticas discretas.	19
3.2. Ejemplo de un grafo RDF o grafo de conocimientos.	20
3.3. Estructura básica de una consulta SPARQL.	21
3.4. Regla para indicar que un Metal-Líquido pertenece a las clases Metal y Líquido. .	25
3.5. ABox y TBox para ejemplificar el beneficio de utilizar un razonador y un motor de búsqueda.	29
3.6. Consulta SPARQL para recuperar todos los individuos que son personas. . . .	29
3.7. Ontología con tripletas que han sido inferidas mediante un razonador.	29
5.1. Arquitectura general para la Integración Semántica de Recurso en una Memoria Corporativa.	45
5.2. Diagrama de Venn para visualizar las tres ontologías que conforman el modelo semántico	46
5.3. Recursos de información agrupados por casos de uso para nuestra memoria corporativa.	47
5.4. Diagrama de clases para la cartografía competencias.	48
5.5. Diagrama de clases para la búsqueda de recursos digitales.	49
5.6. Declaraciones del Dr. Ricardo Marcelin Jiménez en forma de tripletas RDF .	53
5.7. Declaraciones de vídeo “What is Linked Data?” en forma de tripletas RDF .	54
5.8. Jerarquía de clases para los recursos persona	56
5.9. Ejemplo de inferencia para los axiomas de jerarquía de clases y el uso del recurso Ricardo Marcelin.	56
5.10. Jerarquía de clases para los recursos digitales	57
5.11. Ejemplo de inferencia para los axiomas de jerarquía de clases y el uso del recurso What is Linked Data?	57
5.12. Jerarquía de propiedades para las habilidades lingüísticas.	58
5.13. Jerarquía de propiedades para el lugar de trabajo.	58
5.14. Jerarquía de propiedades para las relaciones profesionales entre personas. .	59
5.15. Jerarquía de propiedades para describir el contenido de un recurso digital. .	59

5.16. Jerarquía de propiedades para indicar el año de un recurso digital.	60
5.17. Jerarquía de propiedades para vincular a una organización con un recurso digital.	60
5.18. Axiomas de dominio y rango para modelar las relaciones profesionales entre personas del área de Redes y Telecomunicaciones.	61
5.19. Tripletas RDF que describen las relaciones profesionales entre personas del área de Redes y Telecomunicaciones.	64
5.20. Ejemplo de inferencia para los axiomas de Dominio y Rango que pertenecen a la cartografía de competencias.	64
5.21. Ejemplo del comportamiento unidireccional de una propiedad.	65
5.22. Ejemplo de simetría en una propiedad genérica.	65
5.23. Subgrafo RDF con tripletas que indican las relaciones profesionales entre profesores del área.	66
5.24. Ejemplo de inferencia a partir de la propiedad tiene-colega como propiedad simétrica.	66
5.25. Subgrafo RDF de las relaciones conoce-a entre personas del área de Redes y Telecomunicaciones.	67
5.26. Ejemplo de inferencia a partir de la propiedad conoce-a como propiedad simétrica.	67
5.27. Consulta SPARQL asociada a la pregunta Q1.1 de la cartografía de competencias.	71
5.28. Consulta SPARQL asociada a la pregunta Q1.18, en la cual no se da por hecho el uso del razonamiento.	72
5.29. Simplificación de la Consulta SPARQL asociada a la pregunta Q1.18 mediante la asunción de emplear razonamiento.	73
5.30. Consulta SPARQL asociada a la pregunta Q2.6, en la cual no se da por hecho el uso del razonamiento.	73
5.31. Simplificación de la Consulta SPARQL asociada a la pregunta Q2.6 mediante la asunción de emplear razonamiento.	74
5.32. Proceso básico de consulta de información para un motor de búsqueda SPARQL.	75
5.33. Ejemplo de modelo sin inferencia para el proceso de consulta de información.	77
5.34. Consulta de ejemplo para el proceso de consulta de información.	78
5.35. Ejemplo de modelo con inferencia para el proceso de consulta de información.	79
5.36. Consulta simplificada para el proceso de consulta de información.	80
6.1. Flujo básico de la interacción de un usuario con un servlet.	83
6.2. Validación del acceso de un usuario al prototipo de integración semántica de recursos.	87
6.3. Interfaces Web: navegación entre personas/documentos/multimedia.	89
6.4. Interfaces Web: filtrar los recursos por subclases de Persona, Documento y Multimedia.	91
6.5. Formulario Web: búsqueda avanzada de personas.	94
6.6. Página Dinámica: Resultados de búsqueda avanzada de personas.	95
6.7. Formulario Web: búsqueda avanzada de documentos.	98
6.8. Página Dinámica: Resultados de búsqueda avanzada de documentos.	99

6.9. Formulario Web: búsqueda avanzada de multimedia.	101
6.10. Página Dinámica: Resultados de búsqueda avanzada de multimedia.	102
6.11. Formulario Web para la búsqueda por temas en todos los recursos de información y la página Web mostrando los resultados de esta búsqueda.	103
6.12. Interfaces Web: mostrar detalles de un recurso de información.	105

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	10
GBDR	Gestor de Bases de Datos Relacional	10
BD	Base de Datos	10
IdI	Integración de la Información	13
TS	Tecnologías Semánticas	17
ABox	Componente Asertivo	23
TBox	Componente Terminológico	23
RDF	Resource Description Framework	17
URI	Identificador Único de Recursos	18
W3C	World Wide Web Consortium	21
RDF(S)	Schema RDF	23
OWL	Web Ontology Languages	23
FOAF	Friend Of A Friend	27
XML	Lenguaje de Marcado eXtensible	32
STAAR	Semantic Tourist informAtion Access and Recommending	33
GUI	Interfaz Gráfica de Usuario	35
IDE	Entorno de Desarrollo Integrado	37
API	Interfaz de Programación de Aplicaciones	37
ISR	Integración Semántica de los Recursos	43
CSV	Valores Separados por Coma	52
HTML	Lenguaje de Marcado de Hipertexto	82
URL	Localizador Único de Recursos	84
RAM	Random-Access Memory	84
RI	Recuperación de la Información	114

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**, las 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*” [2]. 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 [3]:

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.

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*” [4]. 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” [5]. 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” [4].

La administración de los recursos puede realizarse con alguna herramienta de las Tecnologías de la Información. La finalidad de estas herramientas es facilitar, eficientar 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* [6] es un sistema de recuperación de la información que a partir de las palabras clave, realiza una búsqueda documental. Este motor responde al usuario con aquellos documentos que tienen las palabras clave en su contenido. 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. Estos gestores se basan en esta idea: *la base de datos es percibida como un conjunto de tablas (relaciones) bajo un mismo contexto, donde, una tabla es una matriz que guarda datos* [7]. Un gestor emplea un *esquema conceptual* para las tareas de almacenamiento de información. El esquema permite describir un conjunto de objetos, aspectos relevantes y las interrelaciones de/entre estos, así como restricciones de integridad. Mientras, para fines de recuperación de la información, se emplean lenguajes de consulta sobre las bases de datos.

El enfoque manual y las dos herramientas del enfoque automático tienen algunos detalles que dificultan la gestión en 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: 1) *un motor de búsqueda en ocasiones recupera documentos innecesarios para los usuarios*, 2) *un motor proporciona resultados inadecuados por problemas de ambigüedad en las palabras*, 3) *una representación deficiente en una BD*

relacional, puede causar anomalías en los datos encontrados cuando el modelo crece, 4) un modelo relacional inadecuado propicia a tener datos inconsistentes, lo que provoca, problemas en la generación y validación de la información [7], y 5) perdida de información en el modelo cuando se representan las atributos de los recursos [7].

Las tecnologías semánticas [8] 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.
- **Enriquecer el conocimiento:** Las tecnologías semánticas permiten la introducción de reglas de inferencia para enriquecer el modelo de conocimiento implícito. La finalidad de estas reglas es que un programa especial realice inferencia sobre éstas para hacer explícito el conocimiento implícito. De esta manera, los procesos automáticos pueden aprovechar 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:** una característica importante en las tecnologías semánticas es la flexibilidad. Esta característica se refiere a la facilidad para representar y mantener el conocimiento de un dominio. Esta representación se basa en la descripción de los recursos a partir de sus características significativas y relaciones en un formato estándar. Otra característica relacionada a la flexibilidad, es la interoperabilidad. Este concepto se refiere a que gracias a los estándares pueden emplearse una variedad de herramientas y aplicaciones.

¹L. Feigenbaum, “Semantic Web vs. Semantic Technologies”, Disponible en: <http://www.cambridgesemantics.com/semantic-university/semantic-web-vs-semantic-technologies>

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, videos, 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.

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 de información pertenecientes a un dominio particular.
- Determinar los casos de uso para esta integración semántica de los recursos.
- Representar y enriquecer 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 representaciones de conocimiento.
- Implementar un prototipo de interfaz de usuario que permita a éstos últimos una interacción amigable para la integración semántica de los recursos.
- Evaluar los resultados devueltos en la integración semántica para el dominio de redes y telecomunicaciones.

Mientras, las contribuciones de tesis son las siguientes:

1. Metodología para la integración semántica de recursos en un dominio particular.
 2. Identificar los principales casos de uso para la integración semántica (cartografía de competencias y búsqueda de recursos digitales)
-

3. Estados del arte para la integración semántica de recursos y las herramientas semánticas.
4. Tres modelos de conocimiento que representan el conocimiento e información en los recursos del dominio de redes y telecomunicaciones.
5. Prototipo de interfaz de usuario para la interacción amigable de los usuarios con los modelos de conocimientos del dominio de redes y telecomunicaciones.
6. Evaluación del desempeño en el proceso de consulta y evaluación de la precisión de los resultados del gestor de modelos semánticos con nuestros modelos de conocimiento.

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 se 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 estados 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. 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 gestor del modelo semántico, 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.

Capítulo 2

Descripción del Problema

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) estudiantes 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 involucradas 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.

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 computo, bases de datos, tesis, reportes técnicos, audios, video tutoriales, notas de curso, tareas, imágenes, páginas web, profesores, estudiantes, empleados de otras organizaciones, servidores computacionales, programas y aplicaciones computacionales científicas-académicas*.

2.1. Memoria Corporativa

Los *recursos de información* expresan el conocimiento en una 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*” [1]. 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 [9], 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 [9]:

- 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.

2.1.1. 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*” [4].

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, videos, 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.

Unanalogía 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.

En el área de Redes y Telecomunicaciones (RyT), la administración del conocimiento se hace de manera individual, es decir, cada profesor, estudiante 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 sintácticos basados en keywords y gestores de bases de datos relacionales. Los *motores de búsqueda sintácticos basados en keywords* (MBSK) 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, *cúal 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.

2.1.2. 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 tambien 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 usualmente 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.

2.1.3. 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 de 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 (IdI) o del conocimiento. La *integración del conocimiento* es el proceso de representar y utilizar el conocimiento de un dominio dado (*Memoria Corporativa*), con el fin de llevar a cabo actividades de búsqueda, recuperación y combinación de la información de los recursos. Esta integración debe proporcionar información correcta a la consulta o pregunta del usuario.

2.2. Casos de uso

Esta tesis presenta la integración de la memoria corporativa del área de RyT. Los *principales usuarios* de la integración son: los profesores-investigadores del área RyT, estudiantes de Computación y Electrónica, así como personas interesadas en el área (colegas de los profesores).

La memoria corporativa de RyT tiene una gran cantidad de recursos de información. Esto hace difícil las actividades de integración del conocimiento. Por ello, se propone descubrir y registrar los principales *casos de uso*. La finalidad de éstos, es identificar las operaciones básicas o aspectos funcionales en la integración de los recursos, describir situaciones específicas, así como identificar los principales recursos de información y el contexto de éstos.

En este trabajo, los casos de uso se identificaron a través del análisis de los principales recursos de información. Los principales recursos del área son *las personas y los recursos digitales*. De esta manera, los casos de uso identificados son: *Cartografía de competencias* para personas y *Búsqueda de recursos digitales*. La Figura 2.1 presenta el *diagrama de casos de uso*, en la cual, se ve la interacción entre los usuarios y los dos casos de uso.

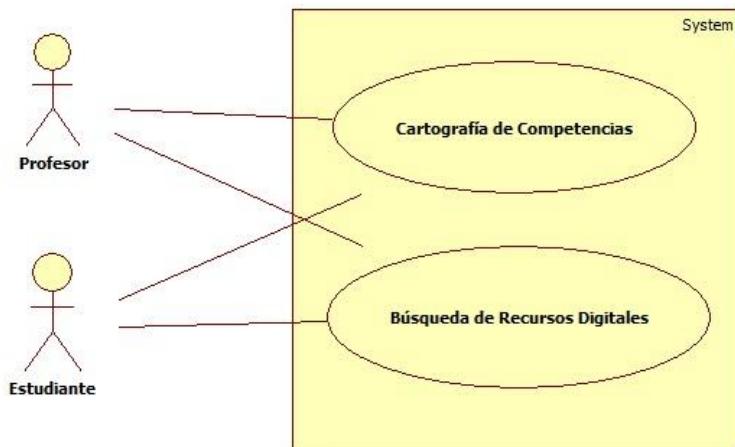


Figura 2.1: Diagrama de casos de uso para la integración de los recursos de una memoria corporativa.

2.2.1. Cartografía de Competencias

El elemento dinámico en el área de RyT es el conjunto de personas que se clasifican en: *profesores, investigadores, estudiantes y empleados*. Estas personas tienen actitudes, valores, conocimientos técnicos, habilidades individuales y colectivas. Las caracterizadoras profesionales son importantes para la organización. Porque con base en éstas se pueden identificar las personas para: *realizar determinadas tareas, solucionar problemas específicos, hacer colaboraciones o tener un determinado cargo*.

La **cartografía de competencias** es la búsqueda y recuperación de las personas a partir de las características profesionales. Los principales parámetros en la búsqueda de estas 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 *persona* identificada en la memoria corporativa, debe ser recuperada *información significativa* de ésta. La finalidad esto, es proporcionar al usuario mayor información, para que pueda localizar y contactar a la persona o filtrar los resultados de acuerdo a otros criterios (*sexo, edad, habilidades*).

2.2.2. Búsqueda de Recursos Digitales

En la memoria corporativa de RyT, los recursos digitales representan *ideas, objetos, teorías, procesos, flujos de trabajo y conocimiento estático de la organización* en un formato digital. Estos recursos se clasifican en: *artículos científicos, libros, reportes técnicos, páginas web, tesis, otros documentos, audios, videos, presentaciones, imágenes y otros archivos multimedia*. Las personas emplean a estos recursos como *objetos de aprendizaje*. Por esta razón, deben identificarse los recursos que solucionen las *necesidades informativas* de los usuarios.

La **búsqueda de recursos digitales** es 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, videos, imágenes y presentaciones*) y la organización a la que pertenece (*uam, unam, ipn, iee, acm, oracle*).

Para cada *recurso digital* identificado en la memoria corporativa, debe recuperarse información significativa de éste, con la finalidad de proporcionar al usuario mayor información. De esta manera, el usuario verifica la importancia del recursos filtrar los resultados de acuerdo a otros criterios (*número de páginas, extensión, lenguaje fuente*).

2.3. Hipótesis

En este capítulo, se ha descrito de manera explícita el alcance, los principales elementos y características para la integración de los recursos en una Memoria Corporativa. Esta integración se ha planteado de manera genérica con respecto al uso de una determinada tecnología, con la finalidad de poder desarrollar la integración con algún enfoque, herramienta, metodología o aplicación de las Tecnologías de la Información.

Nosotros no elegimos alguna de los dos herramientas que ocupan las personas del área (MBSK y GBDR). En cambio, seleccionamos a las Tecnologías Semánticas como enfoque para solucionar esta integración. De esta manera, *nuestra hipótesis de investigación* para esta tesis es: *¿Acaso es posible usar a las Tecnologías Semánticas para solucionar la integración de los recursos en una memoria corporativa?*

Capítulo 3

Tecnologías Semánticas

3.1. Introducción y definiciones

La *semántica* [10] es un subcampo de la lingüística que determina la relación entre palabras y el significado de éstas; así como el estudio de cómo las palabras, frases y otros símbolos lingüísticos, se relacionan entre sí para formar un significado estructurado.

Las *tecnologías semánticas* (TS) [8] son *un conjunto de metodologías, lenguajes, aplicaciones, herramientas y estándares para suministrar u obtener el significado de las palabras, información y las relaciones entre éstos*¹. En estas tecnologías existen varios enfoques para la aplicación del concepto. Estos enfoques se agrupan en dos categorías: 1) *mejorar las capacidades de los procesos automáticos para analizar y comprender el lenguaje* y 2) *técnicas para describir formalmente las palabras, información y el conocimiento para un dominio especializado*.

La categoría para la integración de la información (búsqueda) es la segunda (*técnicas para describir formalmente el conocimiento*), porque al describir formalmente la información y el conocimiento en los recursos, se crea una *capa de conocimiento* en los recursos. La finalidad de esta capa es que los procesos automáticos puedan *acceder, procesar, razonar, combinar, reutilizar y compartir la información y su significado* [11]. De esta manera, se podrá mejorar la búsqueda de información, ya que se evitan problemas de ambigüedad y las personas obtendrán resultados más significativos de acuerdo al contexto del dominio dado.

3.2. Marco de Descripción de Recursos

Las *tecnologías semánticas* proponen al *marco de descripción de recursos* (RDF²) como *marco de trabajo para representar el conocimiento e información acerca de los recursos en un formato estándar* [12]. La finalidad de expresar este conocimiento (*modelar*) es proveer a los recursos con un significado que sea comprensible por los procesos automáticos. Mientras,

¹L. Feigenbaum, “Semantic Web vs. Semantic Technologies,” Disponible en: <http://www.cambridge semantics.com/semantic-university/semantic-web-vs-semantic-technologies>

²W3C, “RDF 1.1 Concepts and Abstract Syntax,” Disponible en: <http://www.w3.org/TR/rdf11-concepts/>

la finalidad de un *formato estándar* es tener un formato compatible y universal para que los procesos automáticos interpreten, mezclen y compartan la información.

En el marco RDF se tienen tres conceptos claves [13]: **1) Recurso, 2) Propiedad y 3) Sentencia.**

El **recurso** es cualquier *persona, lugar, documento, página web, objeto abstracto o físico* que se quiera representar. Cualquier recurso en rdf debe tener un identificador único de recursos (URI), para distinguirlo de otros. Un URI es “*una cadena compacta de caracteres que proporciona un medio simple y extensible para la identificación de un recurso*” [14]. En la Tabla 3.1, se muestran algunos identificadores URI para cinco recursos.

Recurso	Identificador (URI)
Juan López	http://www.mi-ejemplo.com/Juan_Lopez
UAM	http://www.mi-ejemplo.com/UAM
kitty	http://www.mi-ejemplo.com/kitty
celda solar	http://www.mi-ejemplo.com/celda_sol
Mamífero	http://www.mi-ejemplo.com/mamifero

Tabla 3.1: Ejemplos de identificadores (URI) asignados a distintos recursos.

La **propiedad** es un *aspecto significativo, característica, metadato (datos de datos) o relación* que se describe del recurso. Por ejemplo, en una persona los metadatos y relaciones interesantes son: *nombre, edad, teléfono, correo electrónico, habilidad lingüística, nivel de estudios o relación amistad*; en un libro los metadatos interesantes son: *título, autor, isbn, resumen, edición, editorial, año de publicación, volumen o referencia*.

Estas propiedades indican acción entre dos recursos, por ello, es común que el *nombre de una propiedad* empiece por un verbo. Estas propiedades se identifican con URI y deben tener un significado bien definido, para expresar sin ambigüedad su funcionalidad. En la Tabla 3.2 se ejemplifican las propiedades asociadas a determinados metadatos.

Metadato/Relación	Propiedad (URI)
nombre	http://www.mi-ejemplo.com/tiene-nombre
conocido	http://www.mi-ejemplo.com/conoce-a
autor	http://www.mi-ejemplo.com/tiene-autor
referencia	http://www.mi-ejemplo.com/refiere-a

Tabla 3.2: Ejemplos de identificadores asociados a distintas propiedades.

Los identificadores URI de los recursos y propiedades son cadenas con una longitud larga. Para abbreviar estas cadenas se emplea un **prefijo**. Un prefijo sustituye la secuen-

cia de caracteres desde *http://* hasta el comienzo del *nombre del recurso o propiedad* por una *abreviación*. Por ejemplo, el prefijo “*exp*” es la abreviación de esta URI: *http://www.mi-ejemplo.com/*. De esta manera, los recursos y propiedades *Juan Lopez*, *Mamífero*, *nombre* y *conocido* de las Tablas 3.1 y 3.2 se escriben de la siguiente manera.

- exp:Juan_Lopez
- exp:Mamifero
- exp:tiene-nombre
- exp:conoce-a

La *declaración* [12] (sentencia o descripción) es una *afirmación de un hecho explícito* de un recurso, en términos de una *propiedad de objeto o dato* y el *valor* asignado a ella (otro recurso o literal). Estas declaraciones representan el conocimiento o información explícita de los recursos. La forma básica para escribir una declaración, es la *tripleta* [15]. La notación de una tripleta es: *sujeto-predicado-objeto*.

1. *Sujeto* es el recurso que se describe.
2. *Predicado* es la propiedad.
3. *Objeto* es otro recurso o una literal (cadena o entero) que describe el predicado.

En la Figura 3.1 se ejemplifican las tripletas que están asociadas a las siguientes declaraciones: 1) *Juan estudia en la UAM*, 2) *Juan tiene como mascota a kitty*, 3) *Juan es conocido de Jorge*, 4) *Jorge tiene 28 años* y 5) *El libro de matemáticas discretas fue escrito por Jorge*.

```

exp:Juan exp:estudia-en exp:UAM ..... (1)
exp:Juan exp:tiene-mascota exp:kitty ..... (2)
exp:Juan exp:conoce-a exp:Jorge ..... (3)
exp:Jorge exp:tiene-edad "28 años" ..... (4)
exp:mate_disc exp:tiene-autor exp:Jorge ..... (5)

```

Figura 3.1: Ejemplos de tripletas asociadas a las declaraciones para los recursos Juan y libro de matemáticas discretas.

El marco RDF proporciona la propiedad *tipo (type)* para indicar que *un recurso pertenece a una determinada clase*. Esta propiedad tiene el siguiente URI *http://www.w3.org/1999/02/22-rdf-syntax-ns#type* o en su forma compacta “*rdf:type*”. La propiedad *rdf:type* es una de las más importantes para describir y hacer declaraciones sobre los recursos, porque nos permite clasificar a los recursos. La tripleta asociada a esta descripción es: *prefijo:recurso*

Clase persona	Clase mascota
exp:Jorge rdf:type exp:Persona	exp:fido rdf:type exp:Mascota
exp:Juan rdf:type exp:Persona	exp:kitty rdf:type exp:Mascota
exp:Pablo rdf:type exp:Persona	exp:orion rdf:type exp:Mascota

Tabla 3.3: Ejemplos de tripletas que emplean la propiedad *rdf:type* para asignar un recurso a una determinada clase.

***rdf:type* prefijo:Clase.** Para exemplificar esto: los recursos Jorge, Juan, y Pablo son personas, mientras los recursos fido, kitty, orión son mascotas. Las respectivas tripletas de estos se muestran en la Tabla 3.3.

Un **grafo estructurado y dirigido** es la estructura para visualizar las tripletas. Este **grafo RDF** [13] está compuesto por nodos, aristas y etiquetas para representar las tripletas. El nodo origen es el sujeto, el nodo destino es objeto, mientras la etiqueta de la arista es la propiedad que vincula al nodo origen y al nodo destino.

La Figura 3.2 muestra un grafo RDF asociado a las tripletas de la Figura 3.1. En este grafo, los nodos circulares son recursos y los nodos rectangulares son literales, el nodo destino es aquel a quien apunta la *punta de flecha*, mientras el otro nodo es el origen.

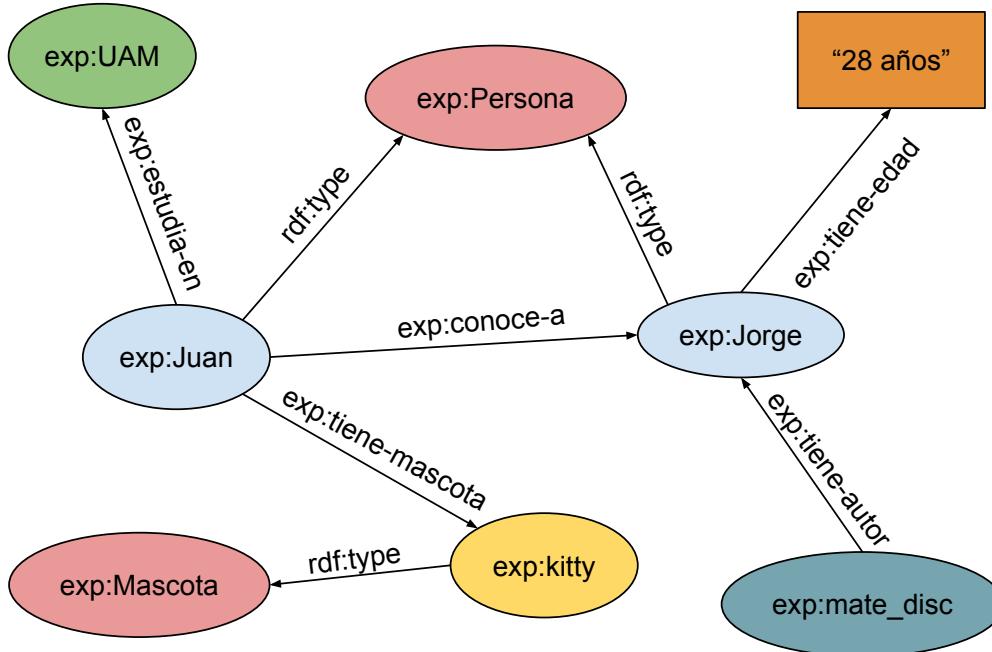


Figura 3.2: Ejemplo de un grafo RDF o grafo de conocimientos.

En el marco RDF, existen distintas sintaxis para escribir y almacenar las tripletas. Estas

sintaxis son: N3³, turtle⁴, RDF/XML⁵, N-triples⁶. El Consorcio de la Web (W3C) establece como sintaxis estándar al RDF/XML. Aunque, la sintaxis Turtle es equivalente a las tripletas de la Figura 3.1.

3.3. Lenguaje de consulta sobre grafos RDF (SPARQL)

Las tecnologías semánticas proponen al lenguaje **SPARQL** como *lenguaje de consulta y protocolo de acceso a RDF* [16], para la búsqueda y recuperación de la información en un grafo RDF.

La idea básica de una **consulta SPARQL** es encontrar conjuntos de tripletas en el grafo RDF que coincidan con un **patrón tripleta**. Un *patrón tripleta* es parecido a una *tripleta RDF*, excepto que el sujeto, predicado y objeto pueden ser una variable. La estructura genérica de una consulta SPARQL se presenta en la Figura 3.3.

```
###Lista de prefijos
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX exp: <http://www.mi-ejemplo.com/>

### Variables a recuperar
SELECT ?x
WHERE {
    ### Lista de patrones tripletas
    ?x exp:propiedad1 exp:objeto1.
    ?x exp:propiedad2 ?y.
}
```

Figura 3.3: Estructura básica de una consulta SPARQL.

Un motor de consulta SPARQL a partir de estas consultas básicas, realiza las siguientes operaciones: 1) interpretar una consulta SPARQL, 2) comparar los *patrones tripleta* con el *grafo RDF*, y 3) recuperar los valores asociados a las variables de la cláusula SELECT. Los resultados que proporciona este motor son *conjuntos de datos*.

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

En las tecnologías semánticas, el concepto clave es la **ontología** para representar (modelar) y gestionar el conocimiento de un dominio particular. Varios investigadores en las TI, como: Newell, Genesereth y Nilsson, Neches y Gruber, han definido este concepto. Nosotros

³W3C, “Notation3 (N3),” Disponible en: <http://www.w3.org/TeamSubmission/n3/>

⁴W3C, “Turtle,” Disponible en: <http://www.w3.org/TR/2013/CR-turtle-20130219/>

⁵W3C, “RDF/XML Syntax Specification,” Disponible en: <http://www.w3.org/TR/rdf-syntax-grammar/>

⁶W3C, “N-Triples,” Disponible en: <http://www.w3.org/2001/sw/RDFCore/ntriples/>

elegimos la siguiente definición: “*Una ontología es una especificación formal y explícita de una conceptualización compartida* [17]. En esta definición se tienen las siguiente características [11], [18].

- **Conceptualización** es una visión simplificada de algún fenómeno en el mundo que queremos representar a partir de los conceptos, funciones, relaciones, restricciones y otros objetos relevantes en dicho fenómeno.
- **Explícita** consiste en definir expresa y claramente los conceptos así como las restricciones sobre ellos.
- **Formal** significa que los elementos de una conceptualización deben ser representados en un lenguaje para que sea comprensible por los procesos automáticos.
- **Compartida** se refiere a que la conceptualización debe ser consensuada y aceptada por el grupo de personas.

La finalidad de una **ontología de un área investigación** es permitir encontrar información pertinente sobre temas especializados para los grupos de investigación. De esta manera, estas personas en vez de dedicar tiempo en la búsqueda, mejor pasen más tiempo en realizar sus actividades de investigación.

Los principales objetivos en el uso de una ontología son [17]: 1) *La construcción de un vocabulario conceptual formal y consensuado para un dominio dado.* 2) *Un conjunto de reglas para combinar los conceptos y relaciones, de esta manera, componer expresiones complejas en el vocabulario.* 3) *Un vocabulario para construir descripciones y comunicar hechos.* 4) *Personas y procesos automáticos interpreten sin ambigüedad el conocimiento y vocabulario de un dominio dado.* 5) *Personas y procesos intercambien y reutilicen el conocimiento para diferentes propósitos.* 6) *La inferencia de información a partir de un programa especializado (razonador) y los hechos en una ontología.* 7) *Personas y procesos consulten información mediante motores de búsqueda y razonadores*

Una ontología tiene tres elementos clave [19], [20]:

- Clase (Class) representa una colección de objetos que comparten características comunes.
- Individuo (Individual) es el nombre de un objeto específico que pertenece a alguna clase.
- Propiedad (Property) describe relaciones binarias entre los objetos.
 - Propiedad de Objeto (Object Property) son relaciones entre objetos.
 - Propiedad de Dato (Data Property) son relaciones entre un objeto y una literal (cadena, entero).

Una ontología tiene dos componentes [21]:

- Un componente assertivo (ABox) representa el conocimiento e información explícita en los recursos del dominio. Este componente está constituido por las declaraciones (descripciones o hechos verdaderos) de los recursos que afirman que los individuos son instancias de una clase o propiedad. Por ejemplo, puede afirmarse que: *el curso Temas Selectos de Bases de Datos pertenece al plan de estudios de la Licenciatura en Computación, el alumno Jorge Aparicio está cursando Temas Selectos de Bases de Datos o el Laboratorio de Análisis y Rendimiento de Teleservicios está en la Universidad Autónoma Metropolitana Unidad Iztapalapa.*
- Un componente terminológico (TBox) representa el conocimiento implícito en los recursos del dominio. Este componente describe las clases y propiedades relevantes, así como los axiomas que permiten aprovechar la manera en que las instancias se relacionan entre sí. Por ejemplo, se puede afirmar que: 1) *todo alumno está inscrito a un curso y pertenece a una universidad, 2) toda universidad es una institución educativa o 3) todo estudiante de universidad pertenece a la comunidad universitaria.*

Los axiomas [17] son expresiones para enriquecer el conocimiento explícito en el grafo RDF. Estos axiomas tienen diferentes propósitos [21], como son: describir relaciones entre clases, definir propiedades en términos de otras, definir relaciones entre conceptos, definir restricciones de cómo las propiedades se relacionan, por mencionar algunos.

Los axiomas deben serializarse en varias tripletas y los vocabularios para escribirlas, son el *esquema RDF* (RDF(S)⁷) y al *Lenguaje de Ontologías Web* (OWL⁸). Estos dos vocabularios son los estándares propuestos por las tecnologías semánticas. En una ontología el prefijo “owl” abrevia el siguiente URI “<http://www.w3.org/2002/07/owl#>”, mientras el prefijo “rdfs” abrevia al URI “<http://www.w3.org/2000/01/rdf-schema#>”.

Las funcionalidades de los axiomas para *relacionar clases en términos de otras*, se listan a continuación. Estos axiomas son los que generalmente se emplean en la construcción y mantenimiento de ontologías [22], [20].

- ***Subclase*** (*rdfs:subClassOf*) afirma que una *clase A* se subsume por una *clase B*, es decir, la clase A es un caso particular de la *clase B*. En este caso, las instancias de la clase A son instancias de la clase B. Este axioma permite especificar la jerarquía entre clases. Por ejemplo, *todo animal o planta es un ser vivo. esto significa que, las clases Animal y Planta son subclases de la clase Ser vivo.*
- ***Clases equivalente*** (*owl:equivalentClass*) afirma que la *clase A* y *clase B* representan al mismo conjunto de individuos y el significado de ambas clases es el mismo, es decir, son sinónimas. Por ejemplo, *todas las personas son humanos y todos los humanos son personas, esto significa que todas las instancias de la clase Persona deben ser instancias de la clase Humano y viceversa.*

⁷W3C, “RDF Vocabulary Description Language 1.0: RDF Schema,” Disponible en: <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>

⁸W3C, “OWL 2 Web Ontology Language Structural Specification and Functional Style Syntax,” Disponible en: <http://www.w3.org/TR/owl2-syntax/>

- **Clases disjuntas** (*owl:disjointWith*) afirma que la *clase A* y *clase B* no tienen instancias en común. Por ejemplo, *ninguna mujer es hombre*, esto significa que *ninguna instancia de la clase Mujer debe pertenecer a la clase Hombre y viceversa*.

Las funcionalidades de los axiomas para definir propiedades en términos de otras, se listan a continuación.

- **Subpropiedad** (*rdfs:subPropertyOf*) afirma que todos los recursos que se relacionan por la *propiedad X*, también se relacionan por la *propiedad Y*. Este axioma permite especificar la jerarquía entre propiedades. Por ejemplo, *la propiedad es padre es un caso particular de la propiedad es familiar, de esta manera, si Juan es padre de Pedro, entonces Juan es familiar de Pedro*.
- **Propiedad equivalente** (*owl:equivalentProperty*) afirma que la *propiedad X* y la *propiedad Y* relacionan a los mismos recursos y éstas tienen el mismo significado. Por ejemplo, *las propiedad tienen automóvil es sinónimo de la propiedad tienen carro, por ello, si Juan tiene un automóvil tipo sedan, entonces Juan tiene un carro tipo sedan y viceversa*.
- **Propiedad inversa** (*owl:inverseOf*) afirma que si la *propiedad X* relaciona al *individuo A* con el *individuo B*, entonces hay una *propiedad Y* que relaciona al *individuo B* con el *individuo A*. Por ejemplo, *la propiedad inversa de es abuelo, es la propiedad es nieto, por ello, si Juan es abuelo de Antonio, entonces Antonio es nieto de Juan*.
- **Propiedad transitiva** (*owl:TransitiveProperty*) afirma que si la *propiedad X* relaciona al *individuo A* con el *individuo B* y también ésta relaciona al *individuo B* con el *individuo C*, entonces debe relacionar a los *individuos A y C*. Por ejemplo, *si Juan tiene parentesco de consanguinidad con Pedro y Pedro tiene parentesco de consanguinidad con Arturo, entonces Juan tiene parentesco de consanguinidad con Arturo*.
- **Propiedad simétrica** (*owl:SymmetricProperty*) afirma que la *propiedad X* es su propia propiedad inversa, es decir, si la *propiedad X* relaciona al *individuo A* con el *individuo B*, entonces, esta propiedad debe relacionar al *individuo B* con el *individuo A*. Por ejemplo, *si Juan es familiar de Pedro, entonces Pedro es familiar de Juan*.
- **Propiedad reflexiva** (*owl:ReflexiveProperty*) afirma la *propiedad X* relaciona al *individuo A* consigo mismo. Por ejemplo, Juan se conoce a sí mismo.

Las funcionalidades de los axiomas para asociar restricciones a las propiedades, se listan a continuación.

- **Dominio** (*rdfs:domain*) especifica qué clase se aplica a una propiedad. Por ejemplo, *todo individuo que emplea la propiedad es madre, debe ser una Mujer, por ello, si Rocío es madre de Arturo, entonces Rocío es una instancia de la clase Mujer*.
-

- **Rango** (*rdfs:range*) especifica los valores (clase o tipo de literal) que puede asumir una propiedad. Por ejemplo, *toda persona que tiene abuelo debe vincularse con un individuo de la clase Hombre*, esto es, si María tiene por abuelo a Ramón, entonces Ramón es una instancia de la clase **Hombre**.

El vocabulario OWL ofrece otros axiomas que tienen otras funcionalidades⁹ (restricciones cardinalidad, valores de literales, existenciales o universales) para enriquecer el conocimiento en un dominio [23]. Los axiomas en los lenguajes OWL y RDF(S) pueden ser combinados para construir clases y propiedades complejas [21], [23]. Un ejemplo de esta combinación es el siguiente. *Todo metal líquido es aquel elemento que pertenece a la intersección de la clase Metal y la clase Líquido*. Este ejemplo, se representa en la Figura 3.4.

`exp:Metal-Liquido rdfs:subClassOf (exp:Metal and exp:Liquido)`

Figura 3.4: Regla para indicar que un Metal-Líquido pertenece a las clases Metal y Líquido.

En las tecnologías semánticas, un *razonador* [22], [24] es un programa que deduce declaraciones a partir de los axiomas y declaraciones explícitas en la ontología. Este programa también se denominan razonador semántico o motor de inferencias. Un razonador permite realizar dos actividades importantes con una ontología:

- Un razonador como *instrumento de validación de consistencia de una ontología*. La validación consiste en deducir información con los axiomas y encontrar si hay contradicciones en el modelo. Si no existen contradicciones en el modelo, entonces, éste es consistente. Por el contrario si hay una contradicción, entonces el modelo no es consistente. Por ejemplo, si en la ontología se establece que la clase Hombre y Mujer son disjuntas, y el recurso Antonio es instancia de estas dos clases, entonces el modelo tiene una contradicción, por tanto, el modelo no es consistente.
- Un razonador para *mejorar la búsqueda de la información en una ontología*. Las declaraciones explícitas y un motor SPARQL solamente permiten recuperar información explícita de los recursos. Un motor de búsqueda SPARQL junto con un razonador, permiten recuperar mejor información en el grafo RDF. Esto es, el razonador expande el grafo RDF con las declaraciones inferidas, donde esta expansión puede ser o no ser explícita. De esta manera, el motor consulta y recupera la información en el grafo.

3.5. Ventajas de las tecnologías Semánticas

Las tecnologías semánticas proporcionan varias características y funcionalidades que benefician la representación y gestión del conocimiento. Algunos de estos beneficios son *facilitar*

⁹W3C, “OWL 2 Web Ontology Language Primer,” Disponible en: <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>

la percepción y representación de dominios particulares, integrar el conocimiento de fuentes heterogéneas (formato, contenido, estructura) de información, compartir y reutilizar el conocimiento a partir modelos dados e inferir conocimiento a partir de los axiomas. A continuación, se describen estos y otros beneficios que ofrecen las tecnologías semánticas.

Las tecnologías semánticas proporcionan una **manera fácil y sencilla** de representar el conocimiento de un dominio particular en una ontología. Esta facilidad para modelar, se debe a los siguientes hechos: 1) todo recurso debe tener un URI, 2) las características y relaciones en los recursos se representan en forma de tripletas, 3) una tripleta se compone por un sujeto, verbo y un objeto, 4) las definiciones de clases, propiedades y axiomas se representan en forma de tripletas, 5) las tripletas del conocimiento explícito e implícito pueden ser visualizadas en un grafo (nodos, etiquetas y arcos) y 6) el grafo de conocimiento constituye la ontología de dominio. De esta manera, dominios particulares con una gran cantidad de objetos pueden representarse a partir de elementos básicos y sencillos en un formato estándar (tripletas). Esta **facilidad** en la construcción y mantenimiento de una ontología, se denomina **flexibilidad de las tecnologías semánticas** [1], [25], [26].

En una ontología el conocimiento no se limita a unas cuantas características sobre un recurso, sino que toda información significativa puede describirse para un determinado recurso. Por ejemplo, se desea modelar las personas dependientes de un empleado en una BD relacional de una organización. En esta BD pueden realizarse dos cosas: 1) para cada empleado se asigna un determinado número de dependientes en la tabla de información personal de los empleados, o bien, 2) se construye una nueva tabla con todos los dependientes de los empleados y se hace una relación entre esta tabla y la tabla de los empleados. Ahora bien, en una ontología para cada recurso empleado se representan los dependientes en forma de tripleta. Esta propiedad de desarrollar ontologías de forma creciente, se denomina **extensibilidad en una ontología** [1], [27].

El marco RDF es una herramienta para solucionar la heterogeneidad en formato, contenido y estructura en los recursos. Porque este marco permite representar cualquier recurso a partir de sus características significativas y relaciones con otros recursos. En concreto, el marco RDF permite realizar las siguientes actividades para cada tipo de heterogeneidad.

- **Formato** El marco RDF permite modelar cualquier recurso, sin importar si es un documento con extensión doc, pdf, odp, html, xml, o un archivo multimedia con extensión ppt, mp3, mpeg, jpg, o incluso si es una persona, organización o cualquier otro recurso físico. De hecho, una característica importante en un documento y archivo multimedia es la extensión del archivo. Otra característica importante es indicar a cuál clase pertenece un recurso, por ejemplo, Documento, Multimedia, Persona, Organización, por mencionar algunas.
 - **Contenido** Por definición el marco describe las características en torno o en los recursos. De esta manera, si los recursos hacen referencia a distintos temas, el marco RDF permite establecer las tripletas que vinculan a un recurso con uno o varios temas.
-

- **Estructura** La flexibilidad del marco, permite representar cualquier recurso, sin importar que este recurso sea estructurado, semi-estructurado o sin estructura. De hecho, el estándar R2RML¹⁰ es el lenguaje estándar para trasladar BD relacionales a modelos con tripletas RDF.

Las tecnologías semánticas solucionan problemas de *ambigüedad* en la representación de un dominio. Para empezar, una ontología soluciona el problema de homonimia. Porque todo recurso, clase y propiedad tiene un *identificador único*. Esto significa que si un recurso o propiedad tiene distintos significados, entonces para cada significado se le asigna un identificador único. Por ejemplo, el término **radio** tiene un significado distinto para cada uno de estos cuatro dominios: *Matemáticas, Anatomía, Geometría o Telecomunicaciones*. Este término por cada dominio se asigna un identificador único, es decir, *mat:radio, anat:radio, geo:radio y tel:radio*. De esta manera, si se emplea el recurso *mat:radio* en una tripleta, entonces, esta tripleta describe un objeto del dominio de Matemáticas.

En una ontología, puede definirse que un recurso, clase o propiedad es *sinónimo* de otro objeto del mismo tipo. Esta propiedad de sinonimia se hace con base en el uso de *axiomas*. Estos son los axiomas para definir objetos sinónimos: *clase equivalente (owl:equivalentClass)*, *propiedad equivalente (owl:equivalentProperty)* e *individuos idénticos (owl:sameAs)*. Esta propiedad de sinonimia es importante para fines de búsqueda. Porque, al hacer una consulta mediante un objeto que tiene un sinónimo, puede recuperarse mayor información del otro objeto, o bien, puede simplificarse la consulta. Por ejemplo, el recurso *computadora* tiene los siguientes sinónimos: *computador, ordenador, equipo de cómputo, por mencionar algunos*. Al hacer una consulta sobre alguna característica de una computadora, es importante, recuperar también la información de los recursos: computador, ordenador y equipo de cómputo. Si se emplea el axioma de equivalencia de clase entre estos recursos y un razonador, entonces, la información de estos recursos es recuperada por el motor de búsqueda.

Una utilidad importante en las tecnologías semánticas es la *interoperabilidad* [28], [26]. Este concepto se refiere a la *facilidad de reutilizar y compartir las ontologías entre personas o aplicaciones, gracias a que una ontología emplea y se elabora con varios estándares*. En concreto, el marco RDF propone la *estructura estándar* para describir el conocimiento, RDF(S) y OWL proponen los *lenguajes estándares* para escribir los axiomas, y SPARQL el *lenguaje de consulta estándar* sobre grafos de conocimiento.

Ejemplos de ontologías¹¹ (modelo de referencia) que proporcionan interoperabilidad son: 1) **Dublin Core** es un vocabulario genérico de metadatos que proporciona información descriptiva de cualquier documento en un sistema de información¹², y 2) **Friend Of A Friend** (FOAF) es un vocabulario para describir a las personas y las relaciones de éstas en la Web¹³.

La interoperabilidad promueve la realización de diversas actividades para mejorar de manera eficiente y eficaz la gestión del conocimiento. Estas actividades se listan a continuación:

- integrar el conocimiento desde distintas fuentes de información.

¹⁰W3C, “R2RML: RDB to RDF Mapping Language,” Disponible en: <http://www.w3.org/TR/r2rml/>

¹¹W3C, “Good Ontologies,” Disponible en: http://www.w3.org/wiki/Good_Ontologies

¹²Dublin Core Semantic Initiative, “Dublin Core,” Disponible en: <http://dublincore.org/>

¹³FOAF project, “The Friend of a Friend (FOAF) project,” Disponible en: <http://www.foaf-project.org/>

- independizar el uso de una única herramienta o sobre determinado sistema operativo.
- realizar tareas de inferencia a partir de los vocabularios OWL y RDF(S).
- recuperar información o construir subgrafos mediante consultas en una ontología.
- liberar a las organizaciones del uso de formatos propietarios que tienen un costo económico o de propiedad intelectual.
- producir ontologías genéricas para dominios particulares, como: Biomédica, Economía, Matemáticas, Ciencias de la Computación, Física, etc.
- construir rápidamente modelos de conocimiento a partir de ontologías básicas.
- mezclar ontologías y construir modelos de conocimiento complejos.

Una ventaja de tener un modelo flexible y estándar es desarrollar *aplicaciones genéricas* para aprovechar estos modelos. Los objetivos de estas herramientas son: *procesar datos, facilitar la visualización del grafo de conocimiento a los usuarios, incrementar el conocimiento en las ontologías mediante la introducción de descripciones o axiomas, facilitar el mantenimiento a una ontología, proporcionar tareas de inferencia en una ontología, mejorar la búsqueda de la información, integrar y mezclar ontologías, facilitar de uso y mejorar la integración de los usuarios*.

Estas herramientas genéricas posibilitan que personas expertas en el dominio sean las principales constructoras del grafo de conocimiento. De esta manera, la información en el grafo será confiable, ya que estas personas son las que tienen los conocimientos en el dominio. Mientras, los desarrolladores son los encargados de construir de estas aplicaciones genéricas.

En esta tesis, el beneficio más interesante es el *uso de un razonador y un motor de búsqueda* para *mejorar la búsqueda y recuperación de la información*. Un razonador a partir de los axiomas, expande el grafo RDF con las declaraciones inferidas, donde esta expansión puede ser o no ser explícita. A partir de este grafo de conocimiento, un motor de consulta puede compararlo para responder una consulta dada.

Para exemplificar el beneficio de combinar un motor de búsqueda y razonador. Primero, se parte de utilizar solamente un motor de búsqueda y la ontología. La Figura 3.5 muestra el TBox y ABox para la ontología de ejemplo.

Ahora bien, un usuario desea recuperar *todos los individuos que son personas*. La consulta en lenguaje SPARQL se presenta en la Figura 3.6.

Un motor de consulta SPARQL procesa esta consulta, seguido de esto, el motor no arroja ningún resultado. Porque en la ontología (Figura 3.5) no hay una tripleta explícita que establezca que un recurso pertenece a la clase Persona.

Esta ontología implícitamente tiene las tripletas que indican que un recurso pertenece a la clase Persona. Porque los axiomas establecen que los individuo de las clases Mujer, Hombre y Estudiante son instancias de la clase Persona. Ahora bien, un razonador infiere las tripletas a partir de los axiomas y descripciones en la ontología de la Figura 3.5. La Figura 3.7 muestra la ontología con las tripletas materializadas.

Ontología	
ABox = $\{$ <p style="margin-left: 20px;"><code>exp:Juan rdf:type exp:Hombre,</code></p> <p style="margin-left: 20px;"><code>exp:Laura rdf:type exp:Mujer,</code></p> <p style="margin-left: 20px;"><code>exp:Luis rdf:type exp:Estudiante.</code></p> $\}$	TBox = $\{$ <p style="margin-left: 20px;"><code>exp:Hombre rdfs:subClassOf exp:Persona,</code></p> <p style="margin-left: 20px;"><code>exp:Mujer rdfs:subClassOf exp:Persona,</code></p> <p style="margin-left: 20px;"><code>exp:Estudiante rdfs:subClassOf exp:Persona.</code></p> $\}$

Figura 3.5: ABox y TBox para ejemplificar el beneficio de utilizar un razonador y un motor de búsqueda.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX exp: <http://www.mi-ejemplo.com/>
SELECT ?x
WHERE
{
    ?x rdf:type exp:Persona.
}

```

Figura 3.6: Consulta SPARQL para recuperar todos los individuos que son personas.

Ontología+	
ABox+ = $\{$ <p style="margin-left: 20px;"><code>exp:Juan rdf:type exp:Hombre,</code></p> <p style="margin-left: 20px;"><code>exp:Laura rdf:type exp:Mujer,</code></p> <p style="margin-left: 20px;"><code>exp:Luis rdf:type exp:Estudiante,</code></p> <p style="margin-left: 20px;"><code>exp:Juan rdf:type exp:Persona,</code></p> <p style="margin-left: 20px;"><code>exp:Laura rdf:type exp:Persona,</code></p> <p style="margin-left: 20px;"><code>exp:Luis rdf:type exp:Persona.</code></p> $\}$	TBox = $\{$ <p style="margin-left: 20px;"><code>exp:Hombre rdfs:subClassOf exp:Persona,</code></p> <p style="margin-left: 20px;"><code>exp:Mujer rdfs:subClassOf exp:Persona,</code></p> <p style="margin-left: 20px;"><code>exp:Estudiante rdfs:subClassOf exp:Persona.</code></p> $\}$

Figura 3.7: Ontología con tripletes que han sido inferidas mediante un razonador.

Si de nuevo se hace la consulta de la Figura 3.6 con un motor de consulta SPARQL en la ontología de la Figura 3.7, se obtienen los siguientes resultados: `exp:Juan`, `exp:Laura` y `exp:Luis`. Por esta razón, la combinación de un razonador y motor de consulta es un mecanismo que permite recuperar más respuestas porque el conocimiento implícito se vuelve explícito.

Capítulo 4

Estado del arte

En este apartado se presenta la *revisión actual* de aquellos trabajos que están relacionados con la *integración semántica de recursos*. Esta revisión se basa en estos *ejes claves de investigación*:

1. Integración del conocimiento a partir del uso de tecnologías semánticas.
2. Búsqueda, recuperación y publicación de la información desde un modelo semántico.
3. Gestión de una memoria corporativa.

4.1. Integración semántica de recursos de información

La *integración semántica de recursos* tiene por objetivo la búsqueda y recuperación de información en los recursos, para responder las necesidades informativas de las personas. Esta integración emplea las tecnologías semánticas para representar, enriquecer y consultar la información, así como el conocimiento de/en los recursos de un dominio. El uso de una *memoria corporativa* se traduce en información y conocimiento de los recursos bajo un dominio particular.

Integración del conocimiento a partir del uso de tecnologías semánticas

El elemento clave en estos trabajos es la utilización de *tecnologías semánticas* para representar las características y relaciones de los recursos en un *modelo semántico* (formato estándar). Esta *representación semántica* permite la *integración de información* (IdI) desde los diversos *recursos de información*, con el fin de superar las dificultades en la heterogeneidad de formato, contenido y estructura en los recursos.

Moner et al. [29] proponen una *arquitectura dual* para la representación consistente y comprensible de la información clínica de cualquier persona. La finalidad es que profesionales de la salud puedan acceder al historial clínico de los pacientes.

La *arquitectura dual* se basa en un modelo que representa la información y el conocimiento. Por un lado, la información se describe en estructuras de datos comunes de los *historiales clínicos* que están distribuidos en varios sistemas independientes y heterogéneos. Por otro lado, en la representación del conocimiento se emplean arquetipos para describir el conocimiento formal de conceptos clínicos.

En este trabajo, la *información de los historiales* se representa en un modelo orientado a objetos. Mientras, el conocimiento se representa en arquetipos mediante el uso de un lenguaje de definición de arquetipos.

En esta *arquitectura dual*, los autores presentan a LinkEHR-Ed, una herramienta para desarrollar arquetipos de datos clínicos. La finalidad de LinkEHR-Ed es que los profesionales de la salud sean los constructores del conocimiento.

Jun Zhai et al. [30] proponen una arquitectura basada en ontologías y el **lenguaje de marcado extensible (XML)** para la *integración semántica* en **sistemas de información de energía eléctrica**. Estos *sistemas de información* son sistemas heterogéneos con funciones y organizaciones descentralizadas.

En esta arquitectura, el lenguaje XML se emplea como medio para el intercambio de datos entre los *sistemas de información de la electricidad*. Sin embargo, XML no expresa semántica, sino más bien la estructura. Por ello, esta arquitectura emplea una ontología para describir formalmente el conocimiento, a nivel conceptual, en el dominio de la electricidad. También, esta ontología permite compartir, reutilizar y usar el conocimiento.

La arquitectura de Jun Zhai et al está dividida en tres capas. La primera capa está constituida por varios *sistemas heterogéneos de información de electricidad*. La segunda genera ontologías locales y una ontología global a partir de la unión de las locales. La tercera capa es para el desarrollo de sistemas que consuman o utilicen los datos de la *integración semántica*.

K. Yang y R. Steele [31] proponen un *marco de integración semántica* como solución eficaz y flexible para la integración de la información en el dominio del alojamiento en-línea. El objetivo de este trabajo reunir y compartir la información en constante cambio que procede de páginas web de alojamiento en-línea.

Este marco propone estos elementos para representar e integrar el conocimiento procedente de distintos servidores. Primero, emplear un *proceso automático* para resolver la naturaleza dinámica de las fuentes de información. Segundo, emplear *ontologías* locales para facilitar el acceso a la *información integrada* y solucionar la heterogeneidad en la estructura de la información. Tercero, emplear un conjunto de *esquemas*, basados en la ontologías locales, para el intercambio de información. Finalmente, permitir a los *propietarios de la información* verificar la información que es integrada.

Búsqueda y recuperación de la información desde un modelo semántico

Un *modelo semántico* permite la realización de diversas actividades, para aprovechar el conocimiento que está residente en este modelo. Por ejemplo, visualización de la información, búsqueda y recuperación de la información, acceso a los recursos, tareas de inferencia, agregar

y extender el conocimiento, compartir la información, por mencionar algunas.

En particular, *la búsqueda y recuperación de información* son actividades de mucha importancia para las personas. Porque esas actividades permiten entregar los resultados que satisfagan las *necesidades informativas* de estas personas. La búsqueda y recuperación se hace a partir del *uso del lenguaje de consulta SPARQL* y un *motor de búsqueda para modelos semánticos*.

A continuación, se presentan los trabajos para la búsqueda y recuperación de información en un *modelo semántico*, con la finalidad de responder las *necesidades informativas* de las personas.

Tuan-Dung et al. [32] proponen desarrollar un sistema para la *recomendación y acceso a la información turística semántica* (STAAR). El propósito de este trabajo es aprovechar las ventajas de la tecnologías semánticas para auxiliar a los turistas en encontrar información turística mediante un smart-phone con Internet.

El sistema STAAR emplea una ontología para definir conceptos y propiedades asociadas a la información y conocimiento sobre el turismo. La ontología tiene un diseño abierto y para interoperar con repositorios que tienen grandes conjuntos de triplets.

Tuan-Dung et al. construyeron una aplicación móvil para la *búsqueda y recomendación de información en forma inteligente* a partir de aplicaciones que emparejan los *datos de viaje enriquecidos con semántica* y las *preferencias turísticas*.

Ha Inay et al. [33] proponen estas actividades. Primero, modelar una ontología, para *representar y recuperar* la información que los mecánicos necesitan sobre los contenidos de los *manuales de mantenimiento electrónicos*. Segundo, diseñar un *sistema de visualización* de la información en una ontología.

Ellos modelaron la ontología a partir del análisis dominio de *mantenimiento de aeronaves*, extrayendo los conceptos y definiendo las relaciones entre éstos. Posteriormente, para cada manual técnico se generaron instancias de estas clases. Todas las instancias son guardadas en un repositorio de triplets.

El objetivo del *sistema de visualización* es proporcionar a los mecánicos y trabajadores con una herramienta, para *comprender intuitivamente la información y obtener fácilmente información de tareas dadas*. De esta manera, estas personas reducen su tiempo para buscar *información relacionada*.

Suganyakala y Rajalaxmi [34] proponen un *marco de trabajo* para la recuperación de información basado en ontología. El objetivo de este marco es superar este problema: *la mayoría de los sistemas de búsqueda semánticos requieren que el usuario exprese su pregunta en forma de una consulta formal*.

Este *marco de trabajo* permite el diseño de una *interfaz de consulta* para que un usuario, sin conocimientos en un lenguaje de consulta formal, pueda escribir una consulta en lenguaje natural. Posteriormente, esta interfaz recupera las respuestas desde una ontología. Suganyakala y Rajalaxmi emplean la *ontología Película* para hacer pruebas de esta interfaz.

En el trabajo de **Salam** [35], él propone un método basado en ontologías para la *recuperación de información*. Este método emplea un *motor de búsqueda e indexación semántico* para recuperar *información contenida en un documento* a partir de los conceptos asociados.

dos que están en una ontología. De esta manera, se considera el contenido semántico de los documentos.

Este *motor de búsqueda e indexación* es una aplicación para 1) extraer *información textual, palabras y términos* que están relacionados, así como 2) almacenar esta información en un índice.

El enfoque de Salam se denomina MIRO, el cual está implementado en un *motor de indexación y búsqueda* sobre una ontología. MIRO permite una búsqueda semántica multilingüe de documentos, mediante una herramienta de búsqueda guiada. También, MIRO proporciona una herramienta para un enriquecimiento automático de la ontología.

El trabajo de Salam es una parte de un proyecto global sobre la investigación de información médica en la web para una clínica especialista en urología. Por ello, él y su grupo de investigación desarrollaron y construyeron una ontología para el dominio de Urología.

Gestión de una memoria corporativa

Una *memoria corporativa* es un elemento de vital importancia para las personas en una organización. Porque ésta *conserva y mantiene* el conocimiento de la organización, expresado mediante los *recursos de información*, para facilitar el acceso, intercambio y difusión del conocimiento por parte de las personas.

La *gestión de una memoria corporativa* es la realización de diversas actividades para aprovechar los *recursos de información* de manera eficiente y eficaz. El propósito de esta gestión es determinar y alcanzar los objetivos en la organización.

A continuación, se presenta la revisión de trabajos que utilizan una *memoria corporativa* para gestionar el conocimiento en una organización.

Xin y Guangleng [36] proponen el uso de una *memoria corporativa* para *capturar y proporcionar* información sobre la *justificación del diseño (design rationale)*. Donde, la *justificación del diseño* es un conocimiento que permite explicar: *qué y cómo se diseña un producto*. También, esta *justificación* apoya la reutilización, comunicación y verificación de diseños en empresas manufactureras. Este conocimiento se encuentra en los *recursos de información* heterogéneos.

En esta propuesta, Xin y Guangleng utilizan un enfoque basado en las **ontologías**, con el objetivo de describir los *objetos de justificación del diseño*. También, ellos emplean una ontología porque permiten acceder de manera uniforme e inteligente a los *recursos de información* heterogéneos.

Esta propuesta también proporciona un sistema que tiene estas tres funcionalidades: capturar, navegar y consultar en *justificación de diseño*. La *captura* es grabar e indexar la información de la *justificación del diseño*. La *navegación* es presentar aspectos del contenido, estructura y contexto sobre la *justificación del diseño*. Finalmente, la *consulta* es encontrar información de las instancias.

Chakhmoune et al. [37] proponen un entorno de colaboración para la gestión de memorias corporativas. La importancia de una *memoria corporativa* es para fines de representación, así como gestión del conocimiento y documentos en forma colaborativa.

En este trabajo se presenta al entorno colaborativo PCOGEME, cuya función es construir una *memoria corporativa* a partir de *lluvia de ideas* y un *mecanismo de toma de decisiones consensuadas*.

PCOGEME emplea una ontología como medio para crear y gestionar el conocimiento de una memoria corporativa. En particular, PCOGEME permite a los usuarios: 1) construir ontologías, 2) almacenar documentos indexados a partir de los conceptos de las ontologías, 3) solicitar preguntas y 4) navegar por las diferentes representaciones.

En este entorno, las ontologías juegan un papel importante. Éstas pueden resolver la falta de comunicación y comprensión en una organización. Porque, una ontología da detalles más precisos y proporciona menos ambigüedades de la terminología. De esta manera, se propicia la reutilización, división y interfuncionamiento del conocimiento en una memoria.

4.2. Herramientas para la integración semántica de recursos

Esta *revisión de la literatura* establece el estudio de aplicaciones y herramientas para desarrollar la *integración semántica de recursos en una memoria corporativa*. Esta *revisión de herramientas* se hace con base en estas funcionalidades:

1. Escribir y guardar automáticamente *tripletas RDF* en alguna sintaxis de serialización RDF.
2. Escribir, guardar y visualizar automáticamente los axiomas con los vocabularios OWL y RDF(S).
3. Gestionar las *tripletas RDF* con fines de almacenamiento, carga en memoria, búsqueda e inferencia a partir de éstas.

Descriptor semántico de recursos

Un *descriptor semántico de recursos* [38] es una herramienta que permite crear y almacenar tripletas RDF a partir de la *información explícita en los recursos*. El principal objetivo de un *descriptor* es construir instancias y relacionar éstas con determinados valores u otras instancias (*concepto de tripleta*).

Un descriptor proporciona una *interfaz gráfica de usuario* (GUI) para simplificar a los usuarios la creación y modificación de las declaraciones. En estos descriptores, las tripletas están escritas en una de las siguientes sintaxis: *RDF/XML*, *Turtle*, *N-triple* y *N3*.

Algunos descriptores requieren un TBox como *criterio de entrada*, con el fin de indicar a éstos *cuáles clases y propiedades pueden emplearse para la generación de tripletas*.

Pocos descriptores hacen sugerencias de la información para generar declaraciones. Ésto se hace mediante un proceso de aprendizaje en un corpus documental o de imágenes.

A continuación, se presentan los descriptores semánticos que nosotros estudiamos.

OntoMat Annotizer [39] es una herramienta para hacer anotaciones semánticas de páginas web, documentos basados en texto plano y lenguajes de marcado¹.

El objetivo de OntoMat es que el usuario cree de manera amigable instancias y declaraciones de éstas, mediante la funcionalidad de arrastrar y soltar (drag-and-drop).

MnM [38] es una herramienta que proporciona apoyo automatizado y semi-automatizado para describir páginas Web con contenido semántico². MnM tiene GUI que integra un editor de ontología, navegador Web, un editor de instancias y de propiedades.

El objetivo de esta herramienta es la descripción de documentos a partir de declaraciones derivadas de ontologías preexistentes.

GATE [40] es un entorno de desarrollo integrado (IDE) para el desarrollo de componentes en el procesamiento del lenguaje humano y el procesamiento de texto³. Las tareas en el procesamiento de texto son: *minería web, extracción de información y descripciones semánticas*.

Aktive Media [38] es una GUI para la descripción automática de una colección de imágenes o documentos (batch annotation) para un contexto específico.

El objetivo de Aktive es automatizar el proceso de descripción, mediante la sugerencia interactiva de la información al usuario, mientras éste está describiendo⁴. Estas sugerencias se hacen con base en axiomas y descripciones previas.

Editor de ontologías

Un **editor de ontología** [41] es una herramienta que proporciona una serie de interfaces amigables para la construcción y mantenimiento de ontologías.

Estos editores proporcionan las siguientes funcionalidades básicas a los usuarios: 1) *definir las clases, propiedades, instancias y axiomas*, 2) *cargar, almacenar, importar y exportar ontologías que son escritas con lenguajes estándar (RDF(S) y OWL)* y 3) *visualizar las clases, propiedades e individuos*.

A continuación se presentan las herramientas de nuestra revisión, para la construcción de ontologías:

Protégé [42] es una plataforma con herramientas para la creación, visualización y manipulación de ontologías en diversos formatos de representación⁵.

Esta plataforma proporciona al usuario una interfaz amigable para la definición de clase, propiedades y axiomas, así como la introducción de datos. La arquitectura de esta herramienta se puede extender a través de plug-ins y APIs. Esta herramienta tiene licencia open-source

¹M. Siroker, “OntoMat Annotizer,” Disponible en: <http://projects.semwebcentral.org/projects/ontomat/>

²The Open University, “MnM,” Disponible en: <http://projects.kmi.open.ac.uk/akt/MnM/>

³The University of Sheffield, “GATE,” Disponible en: <http://gate.ac.uk/>

⁴A. Chakravarthy, V. Lanfranchi , F. Ciravegna, “AKTive Media,” Disponible en: <http://www.aktors.org/technologies/aktivemedia/index.html>

⁵Stanford Center for Biomedical Informatics Research, “Protégé,” Disponible en: <http://protege.stanford.edu/>

Mozilla Public License⁶.

pOWL [43] es una herramienta para la visualización y edición de ontologías vía web⁷. Esta herramienta soporta la carga y edición de ontologías con vocabularios RDF(S) y OWL, generación de consultas y almacenamiento del modelo en una base de datos relacional.

TopBraid Composer [44] es un *entorno de desarrollo integrado*(IDE) para "desarrollar, gestionar y probar configuraciones de los modelos de conocimiento e instancias de las bases de conocimiento"⁸.

Esta herramienta proporciona un conjunto de editores para visualizar grafos RDF y diagramas de clase. Existen tres versiones de esta herramienta: maestro, estándar y gratuita. La versión gratuita permite crear y editar archivos OWL/XML, así como consultar con el lenguaje SPARQL.

SWOOP [45] es un editor para crear y editar ontologías, comprobar inconsistencias, navegar por las ontologías, compartir y reutilizar los datos existentes⁹.

Este editor ofrece un entorno con aspecto de navegador web para facilitar la navegación y edición de ontologías OWL. Este editor provee una interfaz amigable y eficaz para los usuarios web promedios.

Gestión de tripletas RDF

Un **triplestore** [46] es un programa para *el almacenamiento e indexación de tripletas RDF*, con el fin de permitir la consulta eficiente de información sobre estas tripletas. Estos triplestores emplean el estándar SPARQL como lenguaje de consulta para consultar el grafo RDF. Algunos triplestores soportan la capacidad de inferir en el grafo RDF a partir de axiomas, mediante la incorporación o importación de un razonador para ello. Los triplestores se idealizan como *sistema gestor de bases de datos para modelos basados en tripletas RDF*.

En el siguiente listado, se describen cuatro triplestores que estudiamos de muchos que hay disponibles¹⁰.

Apache Jena [47] es un *marco de trabajo* que proporciona un conjunto de interfaces de programación de aplicaciones (API) para Java. Estas APIs ofrecen las siguientes funcionalidades: *lectura, procesamiento y escritura de triples RDF, así como axiomas RDF(S) y OWL, un motor de inferencia y un motor de consulta SPARQL*. La finalidad de Jena es desarrollar aplicaciones que usan las tecnologías semánticas para la representación del conocimiento¹¹.

Las APIs de Jena tienen las siguientes funcionalidades: 1) *lectura, procesamiento y escritura de triples RDF en alguna sintaxis estándar (RDF/XML, N-triples y Turtle)*. 2) soporte de axiomas en los lenguajes OWL y RDF(S), 3) *un motor de inferencia que soporta axiomas*

⁶ Mozilla, “Mozilla Public License,” Disponible en: <http://www.mozilla.org/MPL/>

⁷ Sören Auer, “pOWL,” Disponible en: <http://aksw.org/Projects/Powl.html>

⁸ TopQuadrant, Inc., “TopBraid Composer,” Disponible en: <http://www.topquadrant.com/products/TB-Composer.html>

⁹ University of Maryland, “SWOOP,” Disponible en: <https://code.google.com/p/swoop/downloads/list>

¹⁰ W3C, “Triple Store,” Disponible en: http://www.w3.org/2001/sw/wiki/Category:Triple_Store

¹¹ The Apache Software Foundation, “Apache Jena,” Disponible en: <http://jena.apache.org/>

en *OWL* y *RDF(S)*, 4) almacenamiento eficiente de los triples en el disco duro y 5) un motor de consultas con soporte para el lenguaje *SPARQL*.

Stardog [48] es una base de datos para modelos semánticos. El propósito de esta herramienta es la ejecución de consultas sobre los datos RDF que están bajo su gestión directa¹². Esta herramienta emplea los protocolos *HTTP* y *SNARL* para acceder y controlar de manera remota el modelo de datos *RDF*, inferencia a partir de axiomas en lenguaje *OWL* y consultas *SPARQL*.

Stardog soporta el lenguaje estándar *SPARQL* y emplea los protocolos *HTTP* y *SNARL* para: acceder y controlar de manera remota el modelo de datos *RDF*, la inferencia y el análisis de datos en el lenguaje *OWL*. *Stardog* apoyado en el motor de búsqueda de texto *Lucene*, proporciona la capacidad de búsquedas semánticas que consiste en la indexación de literales *RDF*.

4store [49] es un sistema para el almacenamiento *RDF* que incorpora un motor de consultas *SPARQL*¹³. Las principales fortalezas de esta herramienta son el rendimiento, seguridad, escalabilidad y estabilidad.

Sesame [50] es un marco de trabajo estándar de facto para el análisis, almacenamiento, inferencia y consulta de datos *RDF*¹⁴. Este marco proporciona una API que puede emplearse sobre los distintos sistemas de almacenamiento *RDF* para consultar y acceder a esta información de manera remota.

4.3. Resumen y Discusión

Integración Semántica Recursos

Nuestro trabajo establece seis características básicas para realizar la *integración semántica de recursos* de la memoria corporativa de *Redes y Telecomunicaciones*. Primero, representar el conocimiento del dominio de las *Redes y Telecomunicaciones*. Segundo, emplear una ontología como modelo de información y conocimiento. Tercero, utilizar a las *tecnologías semánticas* para la *representación del conocimiento* o la *búsqueda y recuperación de información*. Cuarto, debe *integrar la información* de los distintos *recursos de información*, los cuales son heterogéneos en formato, contenido y estructura. Quinto, permitir la búsqueda y recuperación de información en una ontología, para responder las preguntas de un usuario. Sexto, emplear un motor de búsqueda en combinación de un motor de inferencia para la búsqueda de información.

En la revisión de la literatura, varios trabajos asociados a la *integración semántica de recursos*, cumplen con alguna de nuestras seis características básicas. Nosotros sintetizamos estos trabajos, mostrando los valores de esas seis características. Esta información se presenta en la Tabla 4.1.

¹²Clark & Parsia, LLC, “*Stardog*,” Disponible en: <http://stardog.com/>

¹³Garlik, “*4store*,” Disponible en: <http://4store.org/>

¹⁴Aduna, “*Sesame*,” Disponible en: <http://www.openrdf.org/index.jsp>

Autor	Objetivo	Dominio	Modelo	TS ^a	IdP ^b	ByRI ^c	MBel ^d
Moner et al.	Arquitectura para representar la información clínica de las personas	Salud	Orientado a objetos y Arquetipos	No	Sí	No	No
K. Yang y R. Steele	Marco de integración semántica para la información del alojamiento en-línea.	Alojamiento en-línea	Ontología	Sí	Sí	No	No
Jun Zhai et al.	Arquitectura basada en ontologías para la IdI en SIdEE ^e	Electricidad	Ontología	Sí	Sí	No	No
Tuan-Dung et al.	Sistema de recomendación y acceso a la información turística	Turismo	Ontología	Sí	No	Sí	No
Ha Inay et al.	Herramienta para visualizar la información en una ontología	Mantenimiento de aeronaves	Ontología	Sí	No	Sí	No
Suganyakala y Rajalaxmi	Marco de trabajo para la recuperación de información basado en ontologías	Películas	Ontología	Sí	No	Sí	No
Salam	Método basado en ontologías para la recuperación de información	Urología	Ontología	Sí	No	Sí	No
Xin y Guangleng	Uso de una memoria corporativa para capturar y proporcionar información del diseño	Justificación	Ontología	Sí	Sí	Sí	No
Chakhmoune et al.	Entorno de colaboración para la gestión de memorias corporativas	Memoria Documental	Ontología	Sí	Sí	No	No

^a Tecnologías Semánticas.^b Integración de la Información.^c Búsqueda y Recuperación de la Información.^d Motor de búsqueda e Inferencia.^e Sistemas de Información de Energía Eléctrica.

Tabla 4.1: Características relevantes para la integración semántica de recursos en una memoria corporativa.

La estructura de la Tabla 4.1 es: la primera columna tiene los autores de la *revisión de literatura*, la segunda tiene los objetivos finales de los trabajos, la tercera indica los dominios estudiados para cada trabajo, la cuarta tiene los modelos empleados para representar el conocimiento e información de su dominio, la quinta indica el uso o no de las *tecnologías semánticas*, la sexta indica la realización de la *integración de la información*, la séptima indica la realización de la *búsqueda y recuperación de información* y la octava indica el uso de un motor de búsqueda con razonamiento.

En esta Tabla 4.1, el valor ‘Sí’ indica que el *trabajo en cuestión* cumple con dicha característica. Mientras, el valor ‘No’ indica que no emplea esa característica.

Con base en nuestra revisión y la Tabla 4.1, encontramos que ningún trabajo cumple con todas las características que proponemos para la *integración semántica de recursos*. Por esta razón, nuestro trabajo propone utilizar las *tecnologías semánticas* para construir una ontología que represente el conocimiento explícito e implícito de una *memoria corporativa*. El objetivo de usar ontologías es que permiten superar las dificultades en la heterogeneidad en formato, contenido y estructura que tienen los recursos de una *memoria corporativa*.

En nuestra propuesta, la ontología representa el conocimiento del área de Redes y Telecomunicaciones, así como el conocimiento de los *recursos de información* personas y recursos digitales. Estos recursos se obtienen mediante el análisis de los *casos de uso*: *cartografía de competencias y búsqueda de recursos digitales*.

Nuestro trabajo permite la *búsqueda y recuperación de información* en una ontología, con la finalidad de responder las preguntas de los usuarios del área de RyT. Esta búsqueda semántica se hace mediante el uso de normas de las *tecnologías semánticas*, como SPARQL. Por ello, nosotros empleamos un motor de búsqueda basado en triplets, así como un motor de inferencia o razonador para mejorar la calidad de los resultados.

El proceso de *búsqueda y recuperación* está guiado por los *casos de uso*, con el objetivo de estar seguros que nuestra propuesta responde un conjunto de preguntas básica. Estas preguntas base se obtienen a partir del análisis de los *casos de uso*.

Herramientas para la integración semántica de recursos

Nosotros obtenemos varias conclusiones sobre la revisión de aplicaciones para el desarrollo de la *integración semántica de recursos en una memoria corporativa*. Estas conclusiones tienen que ver con la elección de las aplicaciones para *descripciones semánticas, edición de ontologías y gestión de triplets RDF*.

La finalidad de un *descriptor semántico de recursos* es construir triplets RDF a partir de la información que capture un usuario. En nuestra revisión encontramos y probamos cuatro descriptores: *MnM, OntoMat Annotizer, GATE y Aktive Media*.

Estas cuatro herramientas proporcionan entornos orientados a los usuarios. El objetivo de los entornos es facilitar la tarea de descripción semántica de los recursos. Ya que estos entornos proporcionan una interfaz de visualización de documentos y una interfaz para navegar entre una ontología. Las funciones de la *interfaz de ontología* son: 1) visualizar la jerarquía de clases y 2) crear instancias de las clases. Mientras, la función de la *interfaz de visualización* es que un usuario vea la información de un documento de texto plano, de esta manera, él selecciona

o marca la información para una instancia. Para guardar las tripletas RDF, el usuario guarda las descripciones sobre los documentos.

Si bien, estos entornos proporcionan una manera simple e intuitiva de generar el ABox de una ontología. Éstos tienen distintas características negativas que no nos permitieron elegir un descriptor semántico.

Primero, estos entornos permiten describir sólo documentos en *texto plano* e imágenes. De esta manera, varios recursos de información se descartan para describirlos semánticamente. Segundo, estos entornos no interpretan ontologías escritas en lenguajes OWL y RDF(S). Tercero, si nuestras tripletas tienen un prefijo establecido, algunas herramientas cambian este prefijo por uno propio. Cuarto, Gate es un entorno difícil de emplear para alguien sin conocimientos en tecnologías semánticas. Quinto, la mayoría de los entornos están descontinuados y las versiones de éstos son viejas e inclusive sin un repositorio o web oficial. Finalmente, al ser entornos obsoletos, no se tiene suficiente documentación para: 1) *solucionar problemas* o 2) *efectuar tareas más complejas* (descripción automatizada de los recursos o descripción por lotes).

Nosotros empleamos un par de scripts y formularios para generar *tripletas RDF* de los recursos de información en nuestra memoria corporativa. Con el propósito de facilitar y agilizar el proceso de generación de tripletas en alguna sintaxis estándar.

En nuestra solución, los formularios son empleados para que las personas capturen la información sobre los recursos. Mientras, los scripts transforman esta información en tripletas RDF. Las Secciones 5.1 y 7.4 presentan el funcionamiento general de los formularios y scripts. Mientras, el funcionamiento detallado de los scripts se presenta en el Anexo A.

El objetivo de editor de ontología es que un usuario pueda escribir, guardar y visualizar los axiomas en lenguajes OWL y RDF(S). En nuestra revisión examinamos estos editores: Protégé, pOWL, TopBraid Composer y SWOOP. Nosotros elegimos a Protégé para representar los axiomas en nuestra ontología. Éstas son las razones de elegir este entorno:

- Una interfaz amigable e intuitiva para el usuario.
- Amplia documentación y tutoriales, así como una comunidad de desarrolladores.
- Facilidad de extender la funcionalidad de esta herramienta, gracias a su arquitectura de plug-ins.
- Variedad de sintaxis para almacenar las ontologías, como: Turtle, Manchester, OWL/XML o XML/RDF.
- Visualización del grafo (axiomas, clases, propiedades e instancias)
- El entorno permite incorporar distintos razonadores, como: Pellet¹⁵, Fact++¹⁶ y Hermit¹⁷.

¹⁵Clark & Parsia, LLC, “Pellet,” Disponible en: <http://clarkparsia.com/pellet/>

¹⁶Clark & Parsia, LLC, “Pellet,” Disponible en: <http://clarkparsia.com/pellet/>

¹⁷Oxford University, “Hermit,” Disponible en: <http://hermit-reasoner.com/>

Un triplestore tiene distintas funcionalidades, las cuales son: almacenar conjuntos grandes de tripletas RDF, indexación de tripletas RDF, cargar un grafo RDF en memoria, buscar y recuperar información en los modelos, realizar tareas de inferencia, entre otras.

En nuestra revisión, se analizaron estos triplestores: Jena, Stardog, 4store y Sesame. De estos cuatro triplestores, nosotros elegimos a Jena para efectuar las tareas de almacenamiento, carga, inferencia, búsqueda y recuperación de información en grafos RDF.

Nosotros elegimos a Jena por estas razones:

- Existencia una amplia documentación y tutoriales para el desarrollo de modelos semánticos.
- Permite la integración de Jena en IDEs para el lenguaje Java, como Eclipse¹⁸.
- Es un proyecto open-source bajo la licencia Apache versión 2¹⁹.
- Tiene un conjunto de librerías para crear, cargar, almacenar y consultar declaraciones, así como axiomas en OWL y RDF(S).
- Incorpora un *motor de inferencia*²⁰ que soporta los axiomas OWL y RDF(S) de nuestra ontología (Ver Sección 5.2).
- Tiene una amplia comunidad de desarrolladores

¹⁸The Eclipse Foundation, “Eclipse IDE,” Disponible en: <http://www.eclipse.org/>

¹⁹The Eclipse Foundation, “Licencia Apache v. 2.0 ,” Disponible en: <http://www.apache.org/licenses/LICENSE-2.0.html>

²⁰The Apache Software Foundation, “OWL coverage,” Disponible en: <http://jena.apache.org/documentation/inference/index.html#OWLcoverage>

Capítulo 5

Integración semántica de recursos de información en una memoria corporativa

La *integración de los recursos* es el proceso de búsqueda y recuperación significativa de información existente en los recursos, para responder una consulta dada por un usuario. Si esta integración se hace mediante el uso de herramientas, estándares, metodologías y aplicaciones pertenecientes a las *tecnologías semánticas*, entonces, se dice que ésta es una *integración semántica de los recursos (ISR)*.

La *integración semántica de recursos* puede implementarse en una *memoria corporativa (MC)*. Porque una memoria tiene un conjunto diverso de recursos de información, los cuales representan el conocimiento en una organización (dominio particular). Estas son las principales razones de esta *integración en una memoria corporativa*: 1) *solucionar la heterogeneidad de los recursos y la ambigüedad de la información en una memoria corporativa*, 2) *adaptar el conocimiento cambiante o explosivo en los recursos*, 3) *extender y mantener un modelo (representación) del conocimiento*, 4) *permitir consultas específicas a partir de las características y relaciones de los recursos*, 5) *recuperar información significativa de los recursos para que respondan las preguntas de las personas adscritas en la organización* y 6) *emplear herramientas, aplicaciones, vocabularios y formatos estándar*.

El desarrollo de la *integración semántica de recurso* se hace con base en una *secuencia ordenada de métodos (metodología)*. Esta tesis describe una *propuesta de metodología* para la *integración semántica de recursos en una memoria corporativa*, la cual está guiada por dos *casos de uso*.

La *finalidad* de esta propuesta es *facilitar y guiar a los desarrolladores en estas dos tareas: 1) construir un modelo semántico (ontologías) y 2) consultar información en este modelo*. Mientras, los principales objetivos de esta integración semántica son:

- Realizar la ISR en cualquier memoria corporativa, por ejemplo *Biomédica, Química, Biología, Computación, Economía, Zoología, por mencionar algunas*
- Emplear distintos *casos de uso* para la ISR y no limitar el número de éstos.
- Representar una MC en un formato estándar con un vocabulario consensuado y asociado al contexto de la MC.

- Utilizar vocabularios estándares para los axiomas, así como el uso del lenguaje SPARQL para las consultas.

Esta metodología está organizada en tres etapas principales:

1. **Representación del conocimiento en los recursos** consiste en identificar los recursos de la memoria corporativa y representar los metadatos (conocimiento explícito) de estos recursos mediante el marco RDF.
2. **Enriquecimiento del conocimiento en el modelo** consiste en introducir axiomas en OWL y RDF(S) para extender, completar y adaptar el conocimiento explícito de los recursos.
3. **Búsqueda y recuperación de la información en el modelo** consisten en identificar las principales consultas de los usuarios en el dominio y ejecutar éstas mediante el uso de un *motor de búsqueda SPARQL* junto con un *razonador*, para recuperar información de los recursos.

La primera y segunda etapa de esta metodología es para construir el *modelo semántico* a partir del conocimiento explícito e implícito en los *recursos de información*. La tercera etapa consiste en identificar y efectuar las principales consultas sobre el modelo semántico.

La Figura 5.1 muestra nuestra arquitectura para la *integración semántica de recursos*. Esta arquitectura tiene un diseño de tres capas (*usuario, negocio y datos*). El objetivo de ésta es poder desarrollarla e implementarla en cualquier memoria corporativa. Los componentes de esta arquitectura se construyen, utilizan e implementan a lo largo de nuestra de metodología.

- En la capa de usuario: se tiene un conjunto de páginas Web dinámicas y estáticas que proporcionan la interfaz visual. Esta interfaz proporciona una manera fácil y sencilla de estructurar las preguntas de los usuarios, así como la visualización de los resultados vinculados a estas preguntas. Las páginas estáticas proporcionan los formularios para que los usuarios *estructuren las preguntas y capturen la información* a buscar en la MC. Mientras, las páginas dinámicas proporcionan la información que responde las preguntas en un formato visual agradable al usuario. La Sección 6 muestra los detalles de esta interfaz gráfica de usuario para nuestros modelos semánticos.
- En la capa de negocios: una aplicación transforma la información recopilada de las páginas estáticas en patrones triplets y construir una consulta SPARQL. Posteriormente, esta aplicación invoca al triplestore para efectuar estas actividades: 1) solicitar y cargar la ontología, 2) hacer inferencia en una ontología mediante el uso de un razonador y 3) buscar y recuperar la información en el modelo inferido mediante el uso de motor de búsqueda SPARQL y la consulta SPARQL.
- En la capa de datos (conocimiento): la ontología modela el conocimiento de los recursos de una memoria corporativa en un formato estándar y con un vocabulario consensual. El componente assertivo contiene las descripciones de las características y relaciones

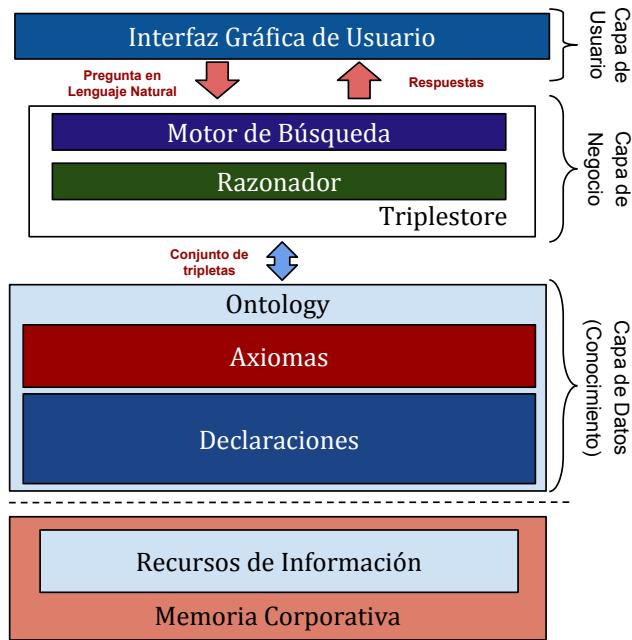


Figura 5.1: Arquitectura general para la Integración Semántica de Recurso en una Memoria Corporativa.

explícitas de los recursos. Mientras, el componente terminológico contiene los axiomas que definen y restringen la manera en que se relacionan los recursos.

Esta propuesta de metodología se pone en práctica para la *memoria corporativa* del grupo de investigación perteneciente al área de Redes y Telecomunicaciones del departamento de Ingeniería Eléctrica de la Universidad Autónoma Metropolitana Unidad Iztapalapa. Los principales usuarios en la integración son los *profesores-investigadores* del núcleo del área de Redes y Telecomunicaciones, así como los *estudiantes* que realizan algún proyecto o servicios social y están a cargo de profesor del núcleo.

En esta metodología, el *elemento clave* es el *caso de uso*. Porque un caso de uso encamina en el desarrollo de la integración semántica. En concreto, los *casos de uso* permiten encontrar: 1) *qué características y relaciones son significativas*, 2) *qué reglas de inferencia son necesarias* y 3) *cuáles consultas son importantes*. Los *casos de uso* básicos en esta metodología son *la cartografía de competencias y la búsqueda de recursos digitales*.

1. La cartografía de competencias es la búsqueda y recuperación de las personas a partir de las características personales y profesionales (competencias, capacidades, conocimientos en los temas del dominio).
2. La búsqueda de recursos digitales es la búsqueda y recuperación de los documentos y archivos multimedia a partir del contenido de éstos (temas del dominio, autor, año).

Estos dos *casos de uso* son independientes entre ellos, por tal razón, cada uno tiene una respectiva ontología. La ontología de la cartografía de competencias modela el conocimiento explícito e implícito de los recursos persona, con base en las características personales y profesionales de éstos. Mientras, la ontología de los recursos digitales modela el conocimiento explícito e implícito del contenido y acerca de éstos.

En ambos casos de uso, un aspecto importante es que tanto personas como recursos digitales se vinculan con los temas del área de Redes y Telecomunicaciones(RyT). Específicamente, los conocimientos de las personas son relaciones entre personas y temas de RyT. Mientras, los tópicos pertenecientes en los recursos digitales son las relaciones entre recursos digitales y temas de RyT. Por tal razón, se construye una tercera ontología para modelar el vocabulario consensual del área de Redes y Telecomunicaciones.

La Figura 5.2 muestra el modelo semántico en forma de un diagrama de venn. En este diagrama, las circunferencias representan las tres ontologías: cartografía de competencias, recursos digitales y vocabulario de RyT. Estas circunferencias no tienen intersección, porque cada ontología representa un determinado recurso de información (personas, documentos y multimedia, así como conceptos de RyT). En esta figura, las flechas representan el vínculo entre: *personas personas - conceptos de RyT* y *recursos digitales - conceptos de RyT*.

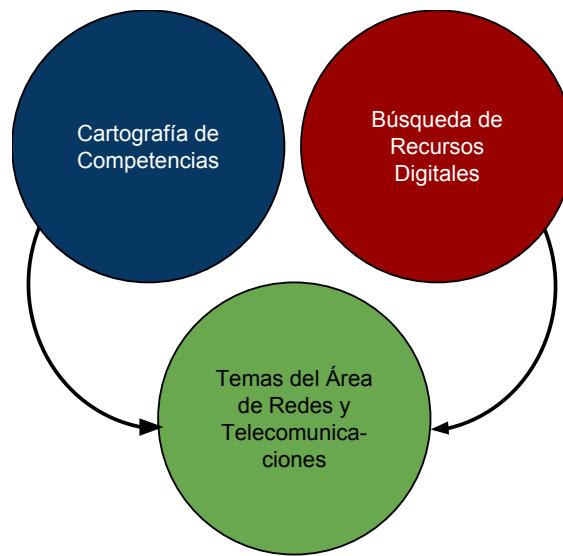


Figura 5.2: Diagrama de Venn para visualizar las tres ontologías que conforman el modelo semántico

5.1. Representación del conocimiento en los recursos

La primera actividad es *identificar los principales recursos de información* para construir la memoria corporativa. Esta identificación se hace a partir del análisis de los *casos de uso*. Los recursos asociados al primer caso de uso son: profesores adscritos al área de RyT,

estudiantes asociados a uno de éstos profesores, empleados de otras organizaciones que colaboran con los profesores. Mientras, los recursos digitales son: artículos científicos relacionados a los temas de investigación, libros y páginas Web de referencia, tesis de maestría y doctorado de los alumnos, reportes técnicos de los profesores y sus estudiantes, presentaciones de cursos o congresos, audios de reuniones o clases, vídeo tutoriales e imágenes de referencia.

La Figura 5.3 muestra el esquema de la memoria corporativa de área de Redes y Telecomunicaciones, donde los *recursos de información* están clasificados por caso de uso.

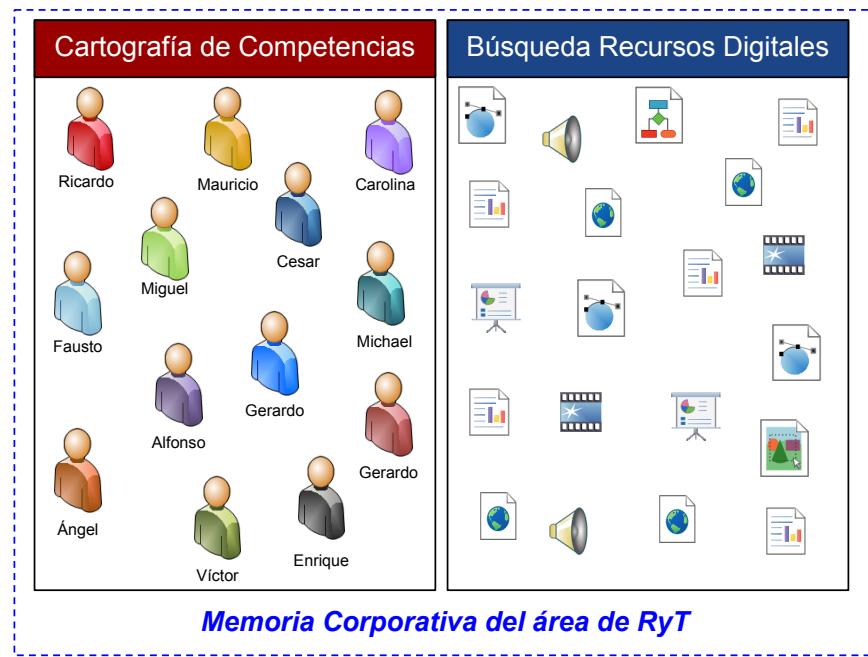


Figura 5.3: Recursos de información agrupados por casos de uso para nuestra memoria corporativa.

La siguiente actividad es *adquirir el conocimiento o información* en los recursos de información, mediante la utilización de los dos *casos de uso*. En esta adquisición debe considerarse un hecho importante del *marco de trabajo RDF*, el cual es “*cualquier persona, lugar, documento, objeto abstracto o físico se representa a partir de una serie de características y relaciones significativas de éste*”. Por esta razón, la *adquisición del conocimiento* se hace con base en las características y relaciones de los *recursos de información*.

Un *diagrama de clases* es una manera visual para mostrar un conjunto de clases, colaboraciones y sus relaciones [51]. Los diagramas de clases se utilizan para especificar y formalizar (modelar) las abstracciones y sus relaciones en un momento dado.

En la Figura 5.4 se presentan las clases, atributos(características) y relaciones entre éstas para visualizar y especificar el comportamiento de los *recursos de información* que pertenecen a la cartografía de competencias. Mientras, la Figura 5.5 muestra las clases y sus relaciones para los documentos y archivos multimedia pertenecientes a la búsqueda de recursos digitales.

La siguiente actividad es la *representación del conocimiento* e información mediante

5.1. Representación del conocimiento en los recursos

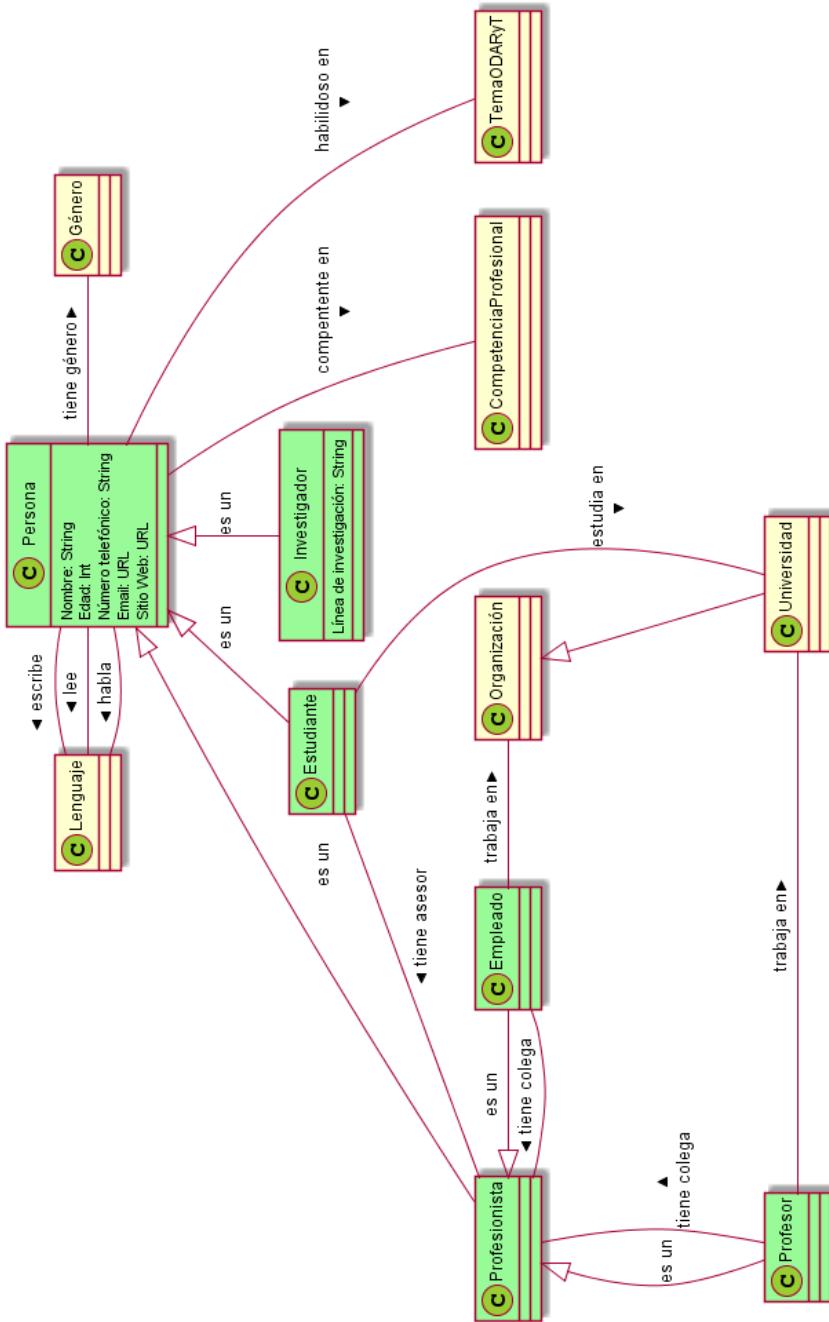


Figura 5.4: Diagrama de clases para la cartografía competencias.

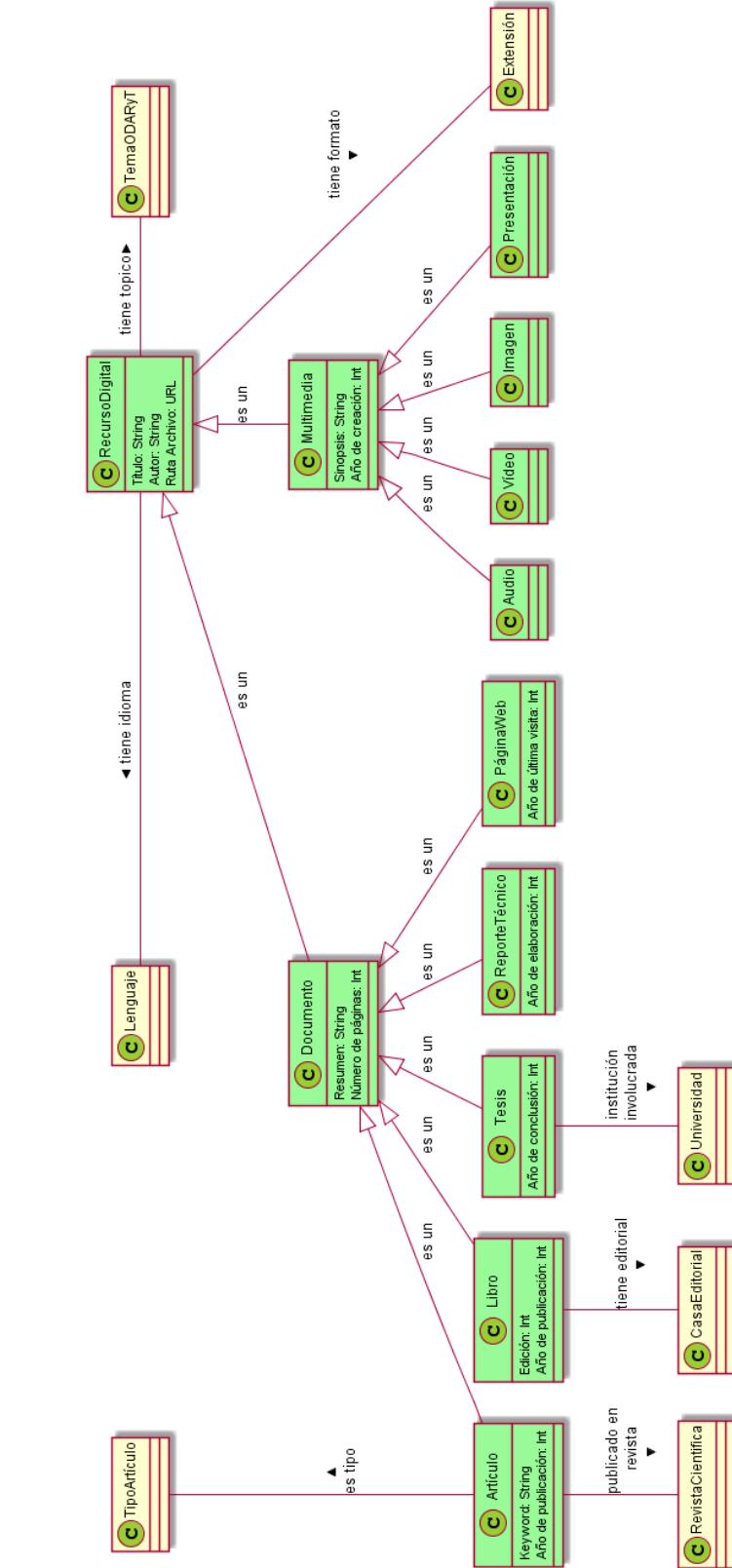


Figura 5.5: Diagrama de clases para la búsqueda de recursos digitales.

el **marco de trabajo RDF**. Esta representación tiene cuatro pasos: 1) *asignar identificadores a los recursos*, 2) *asignar identificadores a las propiedades*, 3) *reconocer si los valores de las propiedades son otros recursos o literales* y 4) *construir triplets*.

El primer paso, es asignar un URI para cada recurso de información. Esta asignación de identificadores URI se hace con base en los casos de uso:

- Los identificadores de los recursos en la cartografía de competencia emplean el prefijo “<http://arte.itzt.uam.mx/ontologies/personRyT.owl#>”, el cual se abrevia “*sirp*”.
- Los identificadores de los recursos pertenecientes a la búsqueda de recursos digitales, utilizan el prefijo “<http://arte.itzt.uam.mx/ontologies/digiResourceRyT.owl#>”, el cual se abrevia “*sird*”.

La Tabla 5.1 enuncia los identificadores URI de algunos profesores del área de RyT. En esta Tabla, la primera columna tiene los nombres de los profesores y la segunda columna enuncia los identificadores URI de éstos.

Nombre	Identificador
Alfonso Prieto	<i>sirp:AlfonsoPrieto</i>
Michael Pascoe	<i>sirp:MichaelPascoe</i>
Reyna Carolina Medina	<i>sirp:CarolinaMedinaRamirez</i>
Ricardo Marcelin	<i>sirp:RicardoMarcelinJimenez</i>
Miguel Lopez	<i>sirp:MiguelLopez</i>
Victor Manuel Ramos	<i>sirp:VictorRamosVictorRamos</i>
Fausto Marcos Casco	<i>sirp:FaustoCasco</i>
Cesar Jalpa	<i>sirp:CesarJalpa</i>
Enrique Rodriguez	<i>sirp:EnriqueRodriguez</i>

Tabla 5.1: Ejemplos de identificadores URI asociados a los recursos persona para la cartografía de competencias.

La Tabla 5.2 enuncia los identificadores URI de algunos recursos digitales pertenecientes a los profesores de RyT. En la primer columna de esta tabla se enuncia el nombre completo del recursos digital y la segunda columna presenta los identificadores (URI) de estos recursos.

El siguiente paso es asignar un identificador URI para cada propiedad. Estos identificadores so construidos con base en las características y relaciones en los diagramas de clases (Figuras 5.4 y 5.5). Los identificadores URI de estas propiedades dependen del caso de uso. Por un lado, si éstas pertenecen a la cartografía de competencias, entonces emplean el prefijo “*sirp*”. Por otro lado, si pertenecen a la búsqueda de recursos digitales, entonces emplean el prefijo “*sird*”.

La Tabla 5.3 presenta algunos identificadores de las propiedades que pertenecen a la cartografía de competencias. La primera columna enuncia las características o relaciones, mientras la segunda columna presenta los identificadores URI de estas característica.

Nombre	Identificador
Ontology engineering	<i>sird:RR-4396-2002-pdf</i>
A Description Logic Primer	<i>sird:A-DescriptionLogicPrimer-2012-pdf</i>
Introduction to Ontologies and OWL	<i>sird:Introduction2Ontol-2005-pdf</i>
The Semantic Web - An Overview	<i>sird:TheSemanticWeb-AnOverview-2011-flv</i>
What is Linked Data?	<i>sird:What-isLinkedData-2012-flv</i>

Tabla 5.2: Ejemplos de identificadores URI asociados a los documentos y archivos multimedia para la búsqueda de recursos digitales.

Característica o relación	Identificador
Nombre	<i>sirp:has-name</i>
Sitio Web	<i>sirp:has-webSite</i>
Lugar de trabajo	<i>sirp:worksIn</i>
Línea de investigación	<i>sirp:researchesOn</i>
Colega	<i>sirp:has-colleague</i>
Competencias	<i>sirp:competentIn</i>
Habilidades en de Redes y Telecom.	<i>sirp:expertiseIn</i>

Tabla 5.3: Ejemplos de identificadores URI de las propiedades pertenecientes a la cartografía de competencias.

La Tabla 5.4 presenta algunos identificadores de las propiedades de la búsqueda de recursos digitales. La primera columna enumera las características o relaciones, mientras la segunda columna presenta los identificadores URI de estas característica.

Característica o relación	Identificador
Título	<i>sird:has-title</i>
Autor	<i>sird:has-author</i>
Ruta Archivo	<i>sird:has-filePath</i>
Año de creación	<i>sird:has-yearOfCreation</i>
Lenguaje Fuente	<i>sird:has-languageSource</i>
Temas de Redes y Telecom.	<i>sird:has-topic</i>

Tabla 5.4: Ejemplos de identificadores URI de las propiedades pertenecientes a la búsqueda de recursos digitales.

El siguiente paso es *identificar el tipo de valor* de las características en el *diagrama de clases*. Por un lado, si el *objeto* es una cadena o un número, entonces es una *literal*. Por otro lado, si el *objeto* es otro recurso, entonces es un *identificador URI*. En el caso de las *relaciones*, los valores son otros recursos, por ello, éstos deben ser *identificadores URI*.

En la representación del conocimiento explícito, el último paso es la *generación* de *tripletas RDF* asociadas a las declaraciones de los recursos.

En esta tesis, la generación de tripletas se hace mediante la combinación de formularios y scripts. Por un lado, los formularios son las herramientas para la recuperación y almacenamiento de la información acerca de los recursos en *hojas de cálculo*. Por el otro lado, los scripts mapean la información de las *hojas de cálculo* en forma de tripletas RDF para que posteriormente éstas sean almacenadas en archivos “*.rdf*”.

Este es el procedimiento de generación de tripletas RDF mediante el uso de formularios y scripts:

1. Identificar la información que debe ser adquirida en los recursos con base en los diagramas de clases.
2. Construir los formularios para los recursos de información (persona, documento y multimedia) mediante Google Form¹, con el propósito de agilizar el proceso de recopilación de la información en los recursos.
3. Enviar los formularios vía email a los profesores o alumnos, para que ellos escriban la información sobre las características y relaciones de los recursos.
4. Recuperar y almacenar las respuestas de cada formulario en una de tres *hojas de cálculo* (persona, documento y multimedia).
5. Descargar la información de cada *hoja de cálculo* en un archivo CSV (persona, documento y multimedia).
6. Transformar cada fila de un archivo csv a un conjunto de tripletas RDF, mediante los scripts que están escritos en Java y con la librería Jena.
7. Almacenar las tripletas RDF asociadas a una fila (descripción semántica de un recurso) en un archivo “*rdf*” con la sintaxis de serialización Turtle.

En la Figura 5.6 se muestra la representación del Dr. Ricardo Marcellin Jiménez en forma de tripletas RDF. En la parte izquierda de esta figura, se presenta una foto representativa del Dr. Ricardo. Mientras, la parte derecha se dan a conocer las tripletas RDF en formato de serialización Turtle que describen las características y relaciones básicas de esta persona.

El identificador asociado al recurso *Ricardo Marcellin Jiménez* es “*sirp:RicardoMarcellinJimenez*”. En la Figura 5.6, el conjunto de tripletas RDF indican las siguientes declaraciones:

- En el primer recuadro, la tripla establece que este recurso pertenece a la clase Profesor.
- En el segundo recuadro, las tripletas establecen los valores para los atributos *nombre*, *email* y *sitio Web*.

¹Google, “Formularios,” Disponible en: https://support.google.com/drive/topic/1360904?hl=es&ref_topic=2811744

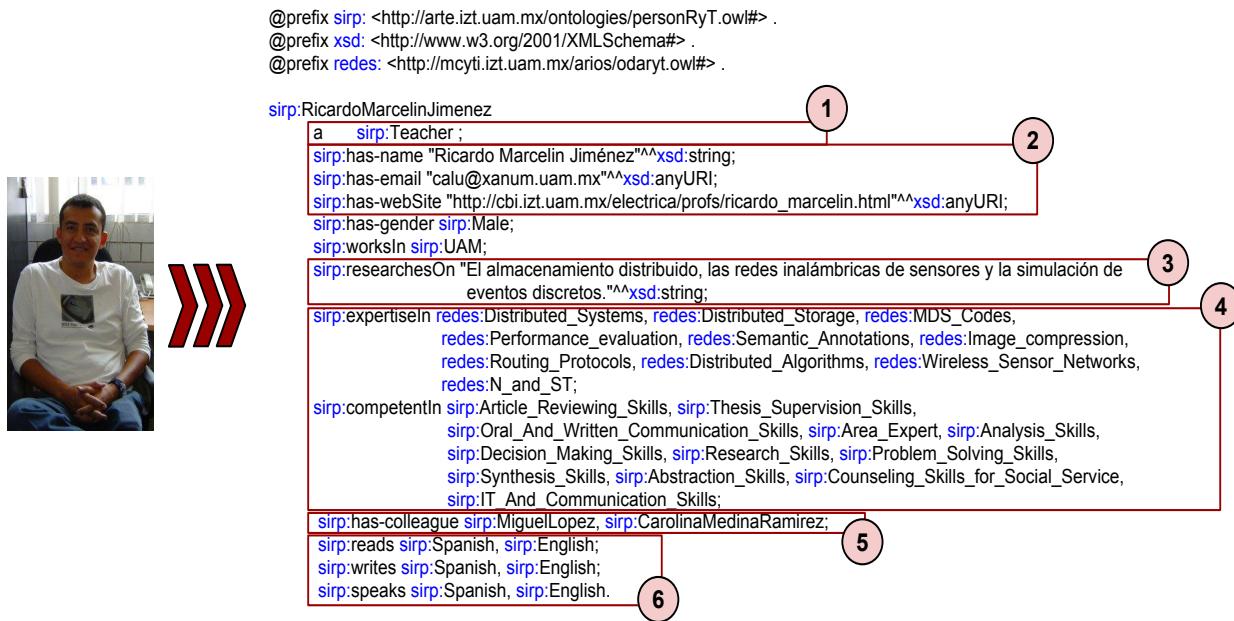


Figura 5.6: Declaraciones del Dr. Ricardo Marcelin Jiménez en forma de tripletas RDF

- En el tercer recuadro, la tripla RDF indica el tema o línea de investigación del Dr. Ricardo.
- En el cuarto recuadro, se establecen las habilidades en los temas de Redes y Telecomunicaciones, así como las competencias profesionales que se obtienen a partir de las competencias propuestas en el proyecto Tuning en latinoamérica.²
- El quinto recuadro indica las tripletas RDF que vinculan al Dr. Ricardo con sus colegas: *Miguel López* y *Carolina Medina*.
- El sexto recuadro establece las habilidades lingüísticas (*lee, habla y escribe*) para distintos idiomas.

La Figura 5.7 presenta la representación del vídeo “*What is Linked Data?*” en forma de tripletas RDF. La parte izquierda de esta figura, muestra una captura de pantalla de éste vídeo, mientras la parte derecha muestra las declaraciones de este vídeo en forma de tripletas RDF.

Las tripletas RDF de la Figura 5.7 establecen las siguientes características y relaciones para el vídeo *What is Linked Data?*.

- En el primer recuadro, la tripla RDF indica que el recurso *What is Linked Data?* pertenece a la clase Video.

²L. Victorino y G. Medina , “Educación basada en competencias y el proyecto tuning en Europa y Latinoamericana,” Disponible en: <http://www.observatorio.org/colaboraciones/2007/TuningEuropayAL-LiberoVictorionoRamirez%202011oct07.pdf>

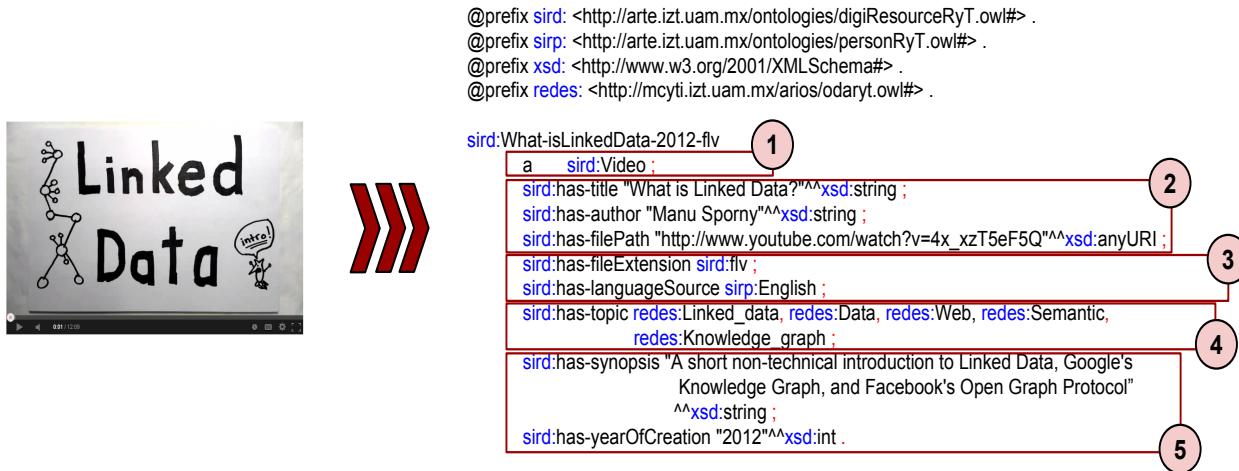


Figura 5.7: Declaraciones de vídeo “What is Linked Data?” en forma de triplets RDF

- En el segundo recuadro se indican los valores asociados a las características *título*, *autor* y *ruta del archivo*.
- En el tercer recuadro, la tripleta RDF establece la extensión o formato del recurso digital, así como el idioma fuente en que está escrito o que se habla en este recurso.
- En el cuarto recuadro, se establecen los temas de Redes y Telecomunicaciones que están en el contenido de este recurso digital.
- En el quinto recuadro, se establecen las tripletas para indicar la sinopsis y el año de creación de este vídeo.

Todas las *triplets RDF* que están asociadas a las declaraciones de los recursos persona, conforman el *componente asertivo* de la ontología *cartografía de competencias*. De igual manera, todas las *triplets RDF* asociadas a las declaraciones de los documentos y archivos multimedia, constituyen el *componente asertivo* de la ontología *búsqueda de recursos digitales*.

5.2. Enriquecimiento del conocimiento en el modelo

La etapa de representación del conocimiento, nos permite describir el conocimiento explícito en los *recursos de información*. Ahora bien, este conocimiento puede ser enriquecido mediante la introducción de axiomas o reglas de inferencia. Los axiomas permiten representar el conocimiento implícito sobre: los recursos y las relaciones de éstos. Por ejemplo, los profesores, empleados y estudiantes son personas, por ello, deben tenerse tres axiomas que establezcan que un profesor es una persona, un empleado es una persona y un estudiante es una persona.

Para cada *caso de uso* debe encontrarse el respectivo conjunto de axiomas (TBox). Este proceso de búsqueda de axiomas debe guiarse a partir de los siguientes elementos: 1) *diagramas de clase*, 2) *cualidades en las relaciones* y 3) *operaciones de la teoría de grupos*. A continuación, se describen y argumentan todos los *axiomas* que se identificaron para las ontologías de *cartografía de competencias y recursos digitales*.

Protégé es el editor que se emplea, para construir, visualizar y manipular los axiomas (TBox) en nuestra ontologías. Esta herramienta permite a los usuarios manipular las clases, propiedades y axiomas desde una interfaz gráfica de usuario. También, esta herramienta permite escribir a los axiomas en distintas sintaxis de serialización (XML/RDF, Sintaxis Manchester, OWL/XML, Sintaxis Funcional o Turtle³).

5.2.1. Herencia de clases

El primer *conocimiento implícito* a representar, es la jerarquía de clases para cada *caso de uso*. El objetivo de ésto es construir modelos de organización jerárquicos del conocimiento para la cartografía de competencias y los recursos digitales. La búsqueda y construcción de las jerarquías se hace mediante el análisis de los *diagramas de clases* de las Figuras 5.4 y 5.3.

A continuación, se describen las dos *jerarquías de clases* para la cartografía de competencias y recursos digitales.

- En el área RyT, las personas pueden agruparse en cuatro clases básicas: *Estudiante (Student)*, *Profesor (Teacher)*, *Investigador (Researcher)* y *Empleado (Employee)*. Las personas pueden pertenecer a más de una clase, por ejemplo, un profesor puede ser un investigador o un empleado, así como un estudiante puede ser un profesor o un empleado. Los profesores y empleados son profesionistas, por ello, la clase *Profesionista (Professional)* es super-clase de *Profesor* y *Empleado*. Finalmente, cualquier individuo perteneciente a una de estas cinco clases, pertenece es una persona. De esta manera, las cinco clases tienen como super-clase la clase *Persona (Person)*. La Figura 5.8 muestra la jerarquía de clases para los recursos persona.

En la Figura 5.9 se presenta un ejemplo de inferencia a partir del uso de axiomas de *jerarquía de clases*. En la parte izquierda de la Figura 5.9, se muestra una rama de la ontología *cartografía de competencias*. Esta rama tiene dos axiomas que establecen que un *Profesor es un Profesionista* y un *Profesionista es una Persona*. Esta rama también indica que el recurso Ricardo Marcellin Jiménez pertenece a la clase Profesor. Mientras, en la parte de la derecha de la Figura 5.9, se muestra la misma rama con inferencia, la cual indica que el recurso Ricardo Marcellin Jiménez pertenece las clases: Profesor, Profesionista y Persona.

- En el área RyT, los principales recursos digitales se pueden agrupar en ocho clases básicas: *Artículo (Paper)*, *Libro (Book)*, *Tesis (Thesis)*, *Página Web (Webpage)*,

³W3C, “OWL 2 Web Ontology Language Document Overview (Second Edition),” Disponible en: <http://www.w3.org/TR/owl2-overview/>

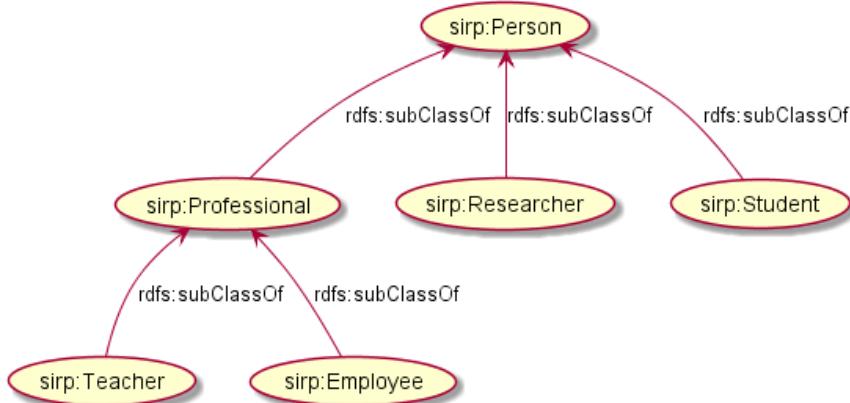


Figura 5.8: Jerarquía de clases para los recursos persona

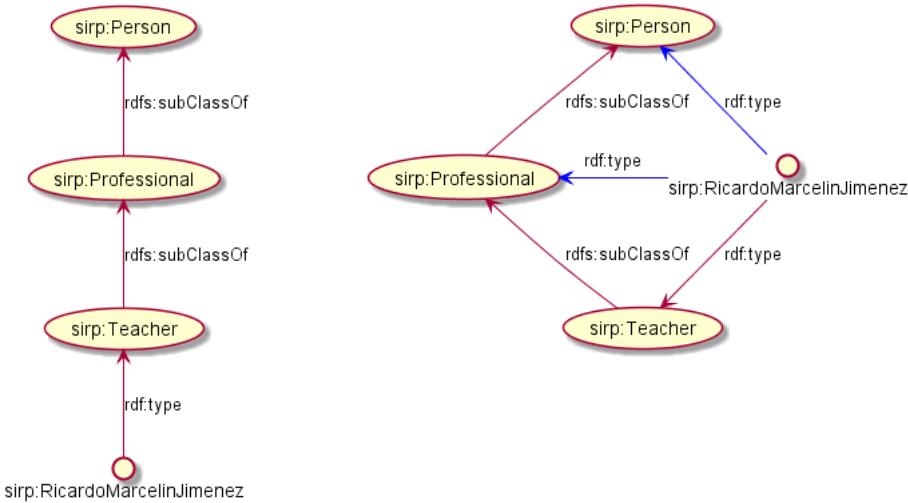


Figura 5.9: Ejemplo de inferencia para los axiomas de jerarquía de clases y el uso del recurso Ricardo Marcelin.

Reporte Técnico (TechnicalReport), Audio (Audio), Vídeo (Video), Presentación (Presentation) e Imagen (Image). Estas nueve clases son disjuntas, por ello, no tienen individuos en común. Los recursos digitales se agrupan en dos clases generales: **Documento (Document)** y **Mutimedia (Multimedia)**. Las primeras cinco clases Artículo, Libro, Tesis, Página Web, Reporte Técnico tienen como super-clase a la clase **Documento**. Mientras las otras cuatro clases básicas tienen como super-clase a la clase **Mutimedia**. Finalmente, cualquier individuo de estas once clases es un recurso digital, por ello, las once clases son subclases de la clase **Recurso Digital (DigitalResource)**. La Figura 5.10 muestra la jerarquía de clases para los recursos digitales.

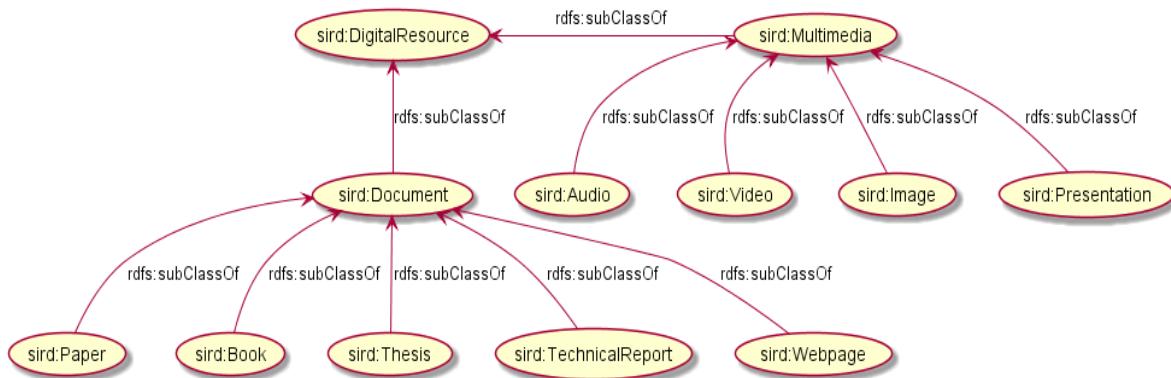


Figura 5.10: Jerarquía de clases para los recursos digitales

En la Figura 5.11 se presenta un ejemplo para la inferencia a partir de axiomas de *jerarquía de clases*. En esta figura, los axiomas son tomados de la ontología *búsqueda de recursos digitales*. La parte izquierda de esta Figura 5.11 ilustra una rama de la *ontología recursos digitales*. En esta rama, se exalta la asignación del recurso *What is Linked Data?* a la clase `Video`. Mientras, la parte derecha de la Figura 5.11, se presenta esta rama después de realizar inferencia. En esta rama, el recurso *What is Linked Data?* está asignado a estas clases: `Video`, `Multimedia` y `Recurso Digital`.

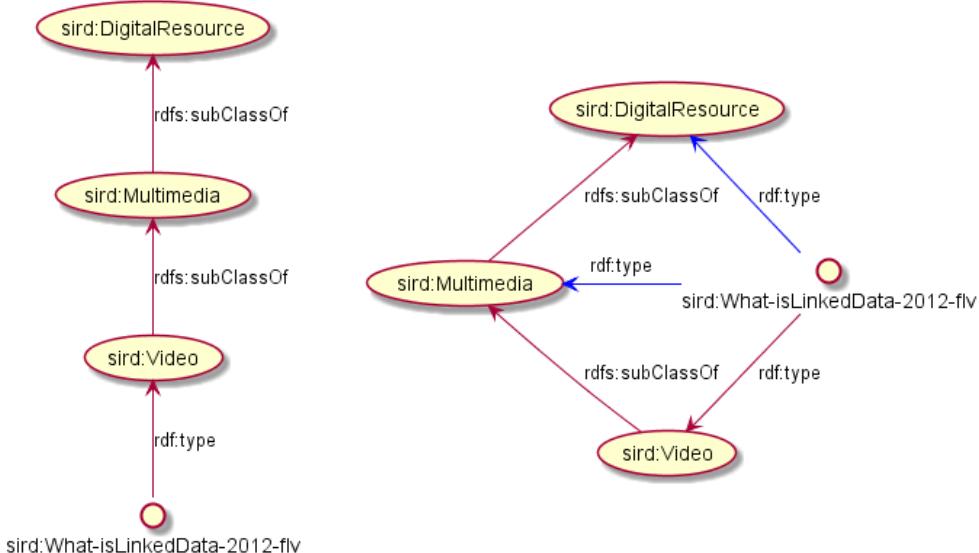


Figura 5.11: Ejemplo de inferencia para los axiomas de jerarquía de clases y el uso del recurso *What is Linked Data?*

5.2.2. Herencia de propiedades

Las propiedades al igual que las clases, pueden generalizarse mediante propiedades comunes. Esta jerarquización se hace mediante el análisis de los atributos en los caso de uso; averiguando que propiedades tienen un significado común.

La siguiente lista describe las *jerarquías de propiedades* que se identificaron para la *cartografía de competencias*. Para cada ítem de esta lista, se presenta un diagrama que muestra esta jerarquía en forma de un grafo.

- **Lee, habla y escribe** son propiedades para indicar las habilidades lingüísticas de una persona. Estas propiedades pueden generalizarse a partir de la propiedad **tiene lenguaje**, con el fin de indicar que una persona tiene algún conocimiento lingüístico en un idioma. Por ello, las propiedades *lee, habla y escribe* son subpropiedades de la propiedad *tiene lenguaje*. En la Figura 5.12 se muestra la jerarquía de propiedades para describir las habilidades lingüísticas de las personas.

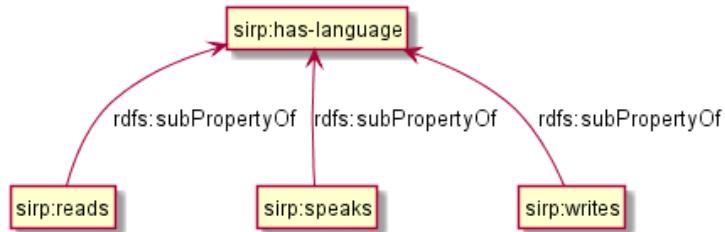


Figura 5.12: Jerarquía de propiedades para las habilidades lingüísticas.

- Las propiedades **trabaja en** y **estudia en** son relaciones que permiten vincular a una persona con el lugar donde labora (trabajo o estudio), por ello, estas dos propiedades son generalizadas a partir de la propiedad *tiene lugar de trabajo*. En la Figura 5.13 se muestra la jerarquía de propiedades para el lugar de trabajo.

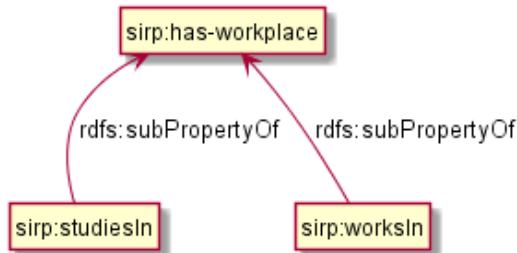


Figura 5.13: Jerarquía de propiedades para el lugar de trabajo.

- Las propiedades *tiene asesor* y *tiene colega* son relaciones que vinculan a dos personas. La propiedad *tiene asesor* vincula a un estudiante con un profesor. Mientras, la propiedad *tiene colega* vincula a dos profesionistas. Estas dos propiedades pueden generalizarse mediante la propiedad *conoce a*, es decir, *tiene colega* y *tiene asesor* son subpropiedades de la propiedad *conoce a*. En la Figura 5.14 se muestra la jerarquía de propiedades que representan las relaciones profesionales entre las personas del área RyT.

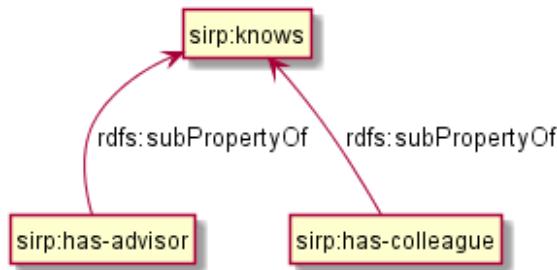


Figura 5.14: Jerarquía de propiedades para las relaciones profesionales entre personas.

De la misma manera que en la cartografía de competencias, el siguiente listado presenta las tres *jerarquías de propiedades* que se identificaron para el TBox de la *búsqueda de recursos digitales*. Al final de cada ítem, se presenta una imagen que muestra la respectiva jerarquía en forma de un grafo.

- La propiedad *tiene resumen* se emplea para describir el contenido básico de un documento. Mientras, la propiedad *tiene sinopsis* describe el contenido básico de un recurso multimedia. Estas dos propiedades pueden generalizarse mediante la propiedad *tiene compendio*, la cual vincula a cualquier *recurso digital* con la descripción básica del contenido de éste. Por esta razón, *tiene resumen* y *tiene sinopsis* son subpropiedades de *tiene compendio*. Esta jerarquía de propiedades se presenta en la Figura 5.15.

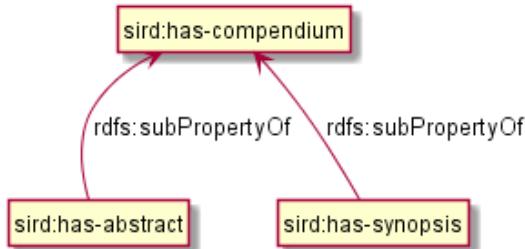


Figura 5.15: Jerarquía de propiedades para describir el contenido de un recurso digital.

- La propiedad **tiene año de conclusión** indica el año de conclusión de una tesis. La propiedad **tiene año de última visita** indica el año de la última visita para una página Web. La propiedad **tiene año de publicación** indica el año en que fue publicado un artículo científico o un libro. Mientras, las propiedades **tiene año de creación** y **tiene año de elaboración** se emplean para indicar el año de creación o elaboración de un recurso multimedia. Estas cinco propiedades pueden ser generalizadas a partir de la propiedad **tiene año**. De esta manera, la propiedad **tiene año** es la superpropiedad de las otras cinco propiedades. La Figura 5.16 muestra gráficamente esta jerarquía de propiedades para el año de construcción de un recurso.

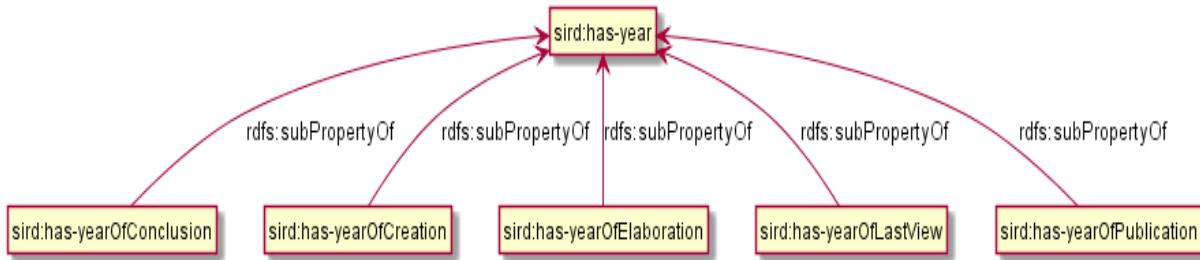


Figura 5.16: Jerarquía de propiedades para indicar el año de un recurso digital.

- La propiedad **publicado en revista** indica que un artículo se publica en una revista científica. La propiedad **tiene editorial** establece que un libro se publica por una determinada casa editorial. La propiedad **tiene institución involucrada** indica que una tesis pertenece a una determinada universidad o institución educativa. Estas tres propiedades se generalizan a partir de la propiedad **publicado en**, por tal razón **publicado en revista**, **tiene editorial** y **tiene institución involucrada** son subpropiedades de la propiedad **publicado en**. Esta jerarquía de propiedades gráficamente se muestra en la Figura 5.17.

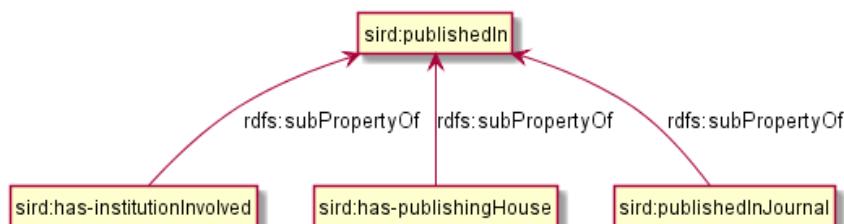


Figura 5.17: Jerarquía de propiedades para vincular a una organización con un recurso digital.

5.2.3. Domingo y rango de propiedades

Los axiomas de **Dominio** y **Rango** permiten definir qué recursos o recurso-literal pueden relacionarse en una propiedad. Esta idea es parecida a una función matemática, donde el dominio es el conjunto formado por *los valores que puede tomar una función*. Mientras, el rango es el conjunto de *los valores que resultan de evaluar la función*.

La Tabla 5.5 presenta el **dominio** y **rango** de las propiedades que pertenecen a la cartografía de competencias. En esta Tabla, la primer columna enumera las propiedades de la cartografía, la segunda columna muestra las clases para el dominio y la tercera columna enumera las clases o tipo de literales del rango. De la misma manera, la Tabla 5.6 presenta el **dominio** y **rango** para las propiedades en el TBox de la *búsqueda de recursos digitales*. La primera columna enumera las propiedades, la segunda las clases del dominio y la tercera las literales o clases del rango.

La Figura 5.18, 5.19 y 5.20 ejemplifican el proceso de inferencia a partir del uso de axiomas de Dominio y Rango. Estos axiomas pertenecen a la cartografía de competencias y son utilizados para restringir las relaciones profesionales entre personas.

En la Figura 5.18 se presentan los axiomas de dominio y rango para las propiedades “tiene asesor” y “tiene colega”. Por un lado, la clase *Profesionista* es tanto el **dominio** como el **rango** de la propiedad *tiene colega*. Por otro lado, las clases *Estudiante* y *Profesionista* son respectivamente el **dominio** y **rango** de la propiedad *tiene asesor*.

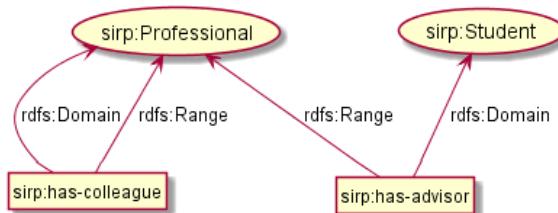


Figura 5.18: Axiomas de dominio y rango para modelar las relaciones profesionales entre personas del área de Redes y Telecomunicaciones.

En la Figura 5.19 se muestra un grafo RDF, en el cual las triplets indican lo siguiente: *la Dra. Carolina Medina y el Dr. Héctor Pérez son asesores de Erik Alarcón, Laura Mendez tiene por asesores a la Dra. Carolina Medina y al Dr. Enrique Rodriguez, el Dr. Ricardo Marcellin y el Dr. Enrique Rodriguez son colegas de la Dra. Carolina Medina y la Dra. Carolina Medina es colega del Dr. Héctor Pérez*.

En la Figura 5.20, se presentan las tripletas de vinculación profesional entre individuos, así como las tripletas de asignación a una de estas dos clases: *Estudiante* (*sirp:Student*) y *Profesionista* (*sirp:Professional*). Las tripletas de asignación son obtenidas después de realizar el proceso de inferencia. Los individuos *Ricardo Marcellin, Enrique Rodriguez, Carolina Medina y Héctor Pérez son Profesionistas*, porque el **dominio** y **rango** de la función *tiene colega* (*sirp:has-colleague*) es la clase *Profesionista*, también porque el **rango** de la función *tiene asesor* (*sirp:has-advisor*) es la clase *Profesionista*. Mientras, los individuos *Erik Alarcón*

Propiedad	Dominio	Rango
<i>sirp:has-name</i>	<i>sirp:Person</i>	<i>xsd:string</i>
<i>sirp:has-age</i>	<i>sirp:Person</i>	<i>xsd:integer</i>
<i>sirp:has-email</i>	<i>sirp:Person</i>	<i>xsd:anyURI</i>
<i>sirp:has-webSite</i>	<i>sirp:Person</i>	<i>xsd:anyURI</i>
<i>sirp:competentIn</i>	<i>sirp:Person</i>	<i>xsd:Competence</i>
<i>sirp:expertiseIn</i>	<i>sirp:Person</i>	<i>redes:TopicRyT</i>
<i>sirp:has-gender</i>	<i>sirp:Person</i>	<i>sirp:Gender</i>
<i>sirp:reads</i>	<i>sirp:Person</i>	<i>sirp:Language</i>
<i>sirp:speaks</i>	<i>sirp:Person</i>	<i>sirp:Language</i>
<i>sirp:writes</i>	<i>sirp:Person</i>	<i>sirp:Language</i>
<i>sirp:researchesOn</i>	<i>sirp:Researcher</i>	<i>xsd:string</i>
<i>sirp:has-advisor</i>	<i>sirp:Student</i>	<i>sirp:Professional</i>
<i>sirp:has-colleague</i>	<i>sirp:Professional</i>	<i>sirp:Professional</i>
<i>sirp:studiesIn</i>	<i>sirp:Student</i>	<i>sirp:University</i>
<i>sirp:worksIn</i>	<i>sirp:Professional</i>	<i>sirp:Organization</i>

Tabla 5.5: Dominio y Rango para las propiedades asociadas a la cartografía de competencias.

Propiedad	Dominio	Rango
<i>sird:has-title</i>	<i>sird:DigitalResource</i>	<i>xsd:string</i>
<i>sird:has-author</i>	<i>sird:DigitalResource</i>	<i>xsd:string</i>
<i>sird:has-filePath</i>	<i>sird:DigitalResource</i>	<i>xsd:anyURI</i>
<i>sird:has-fileExtension</i>	<i>sird:DigitalResource</i>	<i>sird:Extension</i>
<i>sird:has-languageSource</i>	<i>sird:DigitalResource</i>	<i>sird:Language</i>
<i>sird:has-topic</i>	<i>sird:DigitalResource</i>	<i>redes:TopicRyT</i>
<i>sird:has-abstract</i>	<i>sird:Document</i>	<i>xsd:string</i>
<i>sird:has-numberOfPages</i>	<i>sird:Document</i>	<i>xsd:integer</i>
<i>sird:has-description</i>	<i>sird:Multimedia</i>	<i>xsd:string</i>
<i>sird:has-edition</i>	<i>sird:Book</i>	<i>xsd:integer</i>
<i>sird:has-publishingHouse</i>	<i>sird:Book</i>	<i>sirp:PublishingHouse</i>
<i>sird:publishedInJournal</i>	<i>sird:Paper</i>	<i>sirp:Journal</i>
<i>sird:has-institutionInvolved</i>	<i>sird:Thesis</i>	<i>sirp:University</i>
<i>sird:isTypeOfArticle</i>	<i>sird:Paper</i>	<i>sirp>TypeArticle</i>
<i>sird:has-yearOfConclusion</i>	<i>sird:Thesis</i>	<i>xsd:integer</i>
<i>sird:has-yearOfElaboration</i>		<i>xsd:integer</i>
<i>sird:has-yearOfLastView</i>	<i>sird:Webpage</i>	<i>xsd:integer</i>
<i>sird:has-yearOfPublication</i>	<i>sird:Book</i>	<i>xsd:integer</i>
<i>sird:has-yearOfPublication</i>	<i>sird:Paper</i>	<i>xsd:integer</i>

Tabla 5.6: Dominio y Rango para las propiedades asociadas a la búsqueda de recursos digitales.

5.2. Enriquecimiento del conocimiento en el modelo

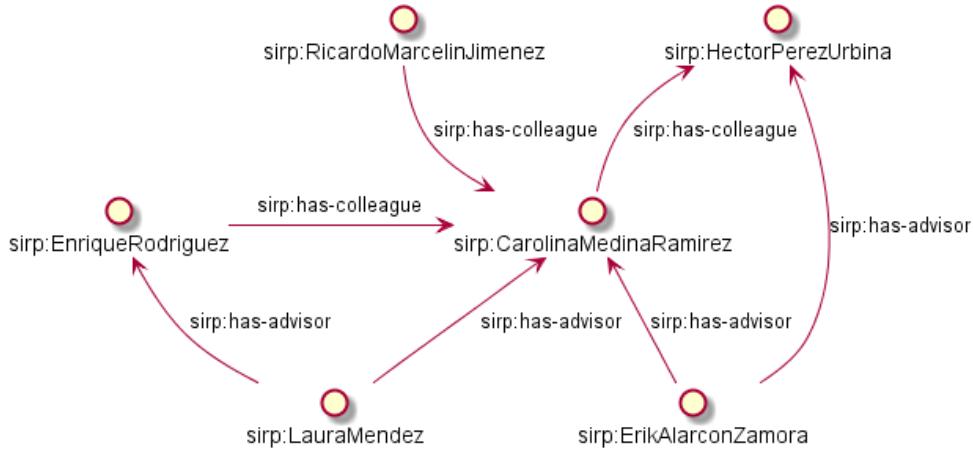


Figura 5.19: Tripletas RDF que describen las relaciones profesionales entre personas del área de Redes y Telecomunicaciones.

y *Laura Mendez* son *Estudiantes*, porque el *dominio* de la propiedad *tiene asesor* es la clase *Estudiante*.

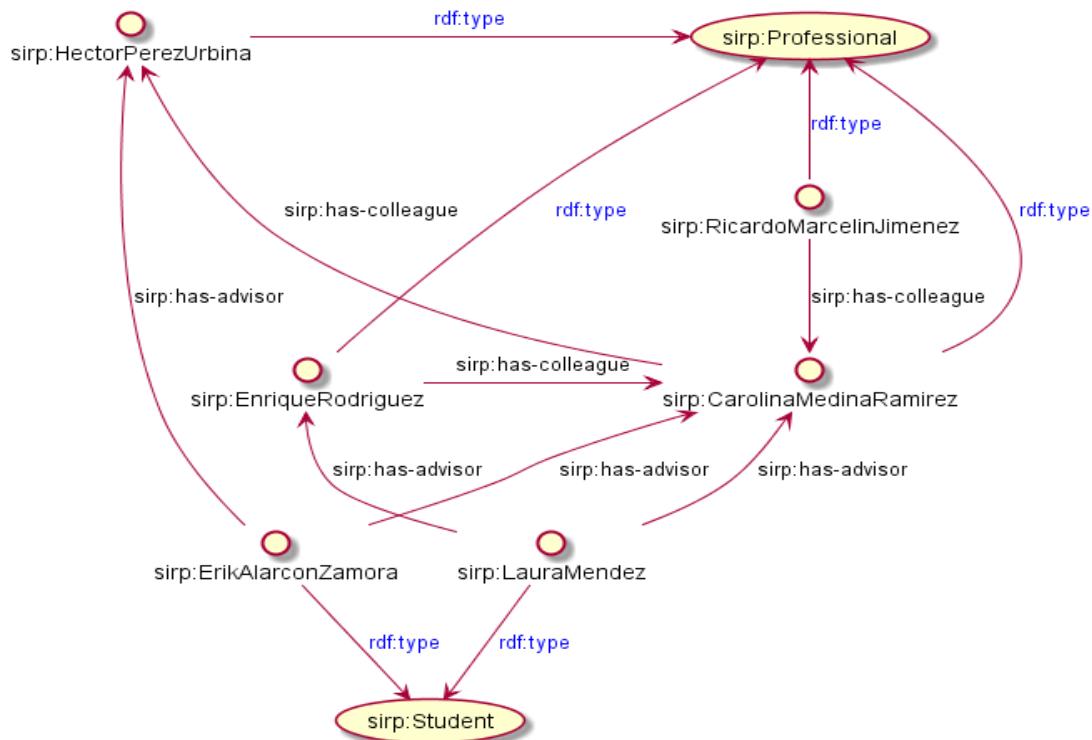


Figura 5.20: Ejemplo de inferencia para los axiomas de Dominio y Rango que pertenecen a la cartografía de competencias.

5.2.4. Simetría en las propiedades

Una tripleta RDF es una relación unidireccional, es decir, parte de un *individuo p* y termina en *individuo q*. En la Figura 5.21 se ejemplifica este comportamiento unidireccional, donde la relación *p tiene género q*, indica que todos los *individuos p* se relacionan con un *individuo q*.

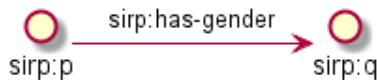


Figura 5.21: Ejemplo del comportamiento unidireccional de una propiedad.

Sin embargo, existen algunas propiedades que requieren tener un comportamiento bidireccional, con la finalidad de tener una mejor representación del dominio. En la Figura 5.22 se presenta un ejemplo genérico del uso de relaciones simétricas (“relación bidireccional”). En este ejemplo, los individuos *p* y *q* son el nodo origen y el nodo destino, es decir, en una tripleta RDF pueden ser el sujeto o el objeto.

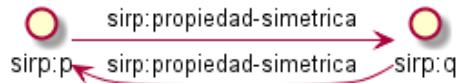


Figura 5.22: Ejemplo de simetría en una propiedad genérica.

A continuación, se describen las propiedades simétricas que se identificaron para la *cartografía de competencias*. Con respecto al caso de la búsqueda de recursos digitales, no se identificó alguna propiedad que tenga la característica simétrica.

- La propiedad *tiene colega (sir:has-colleague)* es una relación del tipo laboral que vincula a dos Profesionistas (Profesor o Empleado). La idea básica de esta relación es indicar quiénes son los colegas de trabajo o con quiénes ha hecho colaboración un *Profesionista*.

La Figura 5.23 muestra un subgrafo RDF con relaciones *tiene colega* entre cuatro profesores del área de Redes y Telecomunicaciones (RyT). En este subgrafo las tripletas indican las siguientes declaraciones: *el Dr. Ricardo Marcelin tiene por colegas a la Dra. Carolina Medina, Dr. Enrique Rodríguez y Dr. Miguel López*, *el Dr. Enrique Rodríguez tiene por colegas a la Dra. Carolina Medina y al Dr. Miguel López* y *la Dra. Carolina Medina tiene por colega al Dr. Miguel López*.

En la Figura 5.23, el recurso “*sirp:MiguelLopez*” es el **objeto** en todas las tripletas del subgrafo RDF, es decir, el Dr. Miguel López no tiene alguna declaración propia que indique quienes son sus colegas. El comportamiento de la propiedad *tiene colega*

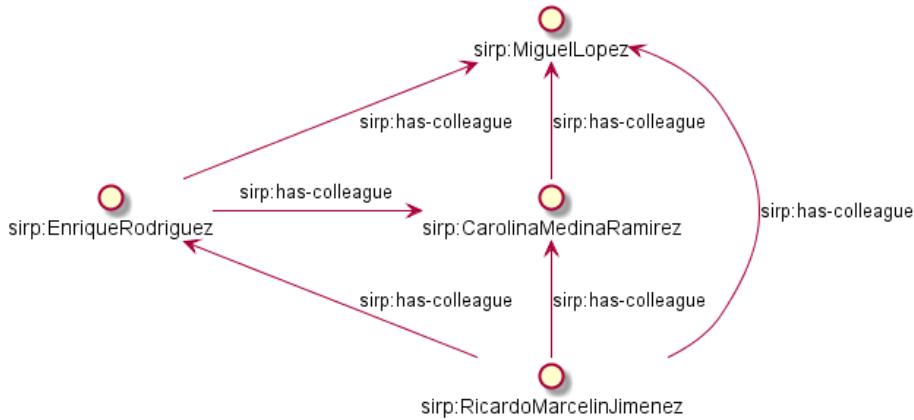


Figura 5.23: Subgrafo RDF con tripletas que indican las relaciones profesionales entre profesores del área.

indica una relación unidireccional. Sin embargo, esta propiedad debe tener un comportamiento bidireccional. Por ello, la propiedad *tiene colega* (*sirp:has-colleague*) tiene la *característica simétrica*.

La Figura 5.24 muestra el subgrafo RDF inferido. Este subgrafo se obtiene a partir de las tripletas en la Figura 5.23 y del axioma que establece que la propiedad *tiene colega* es simétrica. En esta Figura, el recurso “*sirp:MiguelLopez*” tiene el rol de *objeto*, así como el rol de *sujeto* para otras tres tripletas RDF. Ésto significa que el *Dr. Miguel López tiene por colegas a la Dra. Carolina, Dr. Enrique y al Dr. Ricardo*. De la misma manera, los otros recursos (Doctores) carentes de su propia afirmación, ahora tienen esta relación.

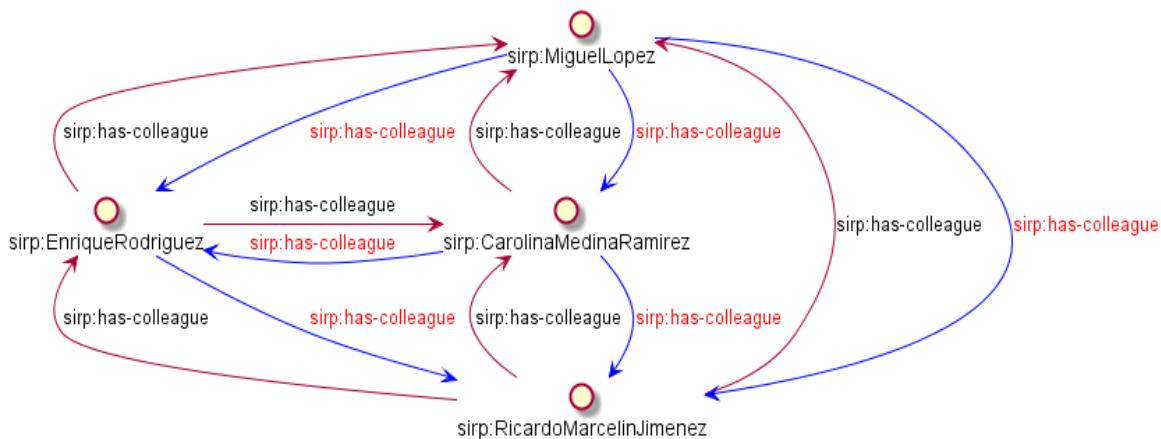


Figura 5.24: Ejemplo de inferencia a partir de la propiedad tiene-colega como propiedad simétrica.

- La propiedad **conoce a** (*sirp:knows*) es una relación profesional para la interacción entre las personas (profesor, empleado o estudiante) del área de Redes y Telecomunicaciones. Esta propiedad, al igual que la propiedad *tiene colega*, tiene un comportamiento bidireccional. Por ello, la propiedad *conoce a* es una propiedad simétrica.

La Figura 5.25 ejemplifica un subgrafo RDF, en el cual las relaciones *conoce a* tienen solo una dirección. En esta Figura, las tripletas indican lo siguiente: *Erik Alarcón conoce a la Dra. Carolina Medina, Dr. Héctor Pérez y Laura Méndez, Laura Méndez conoce a la Dra. Carolina Medina y Dr. Enrique Rodríguez, la Dra. Carolina conoce al Dr. Héctor Pérez y el Dr. Enrique Rodríguez conoce a la Dra. Carolina Medina.*

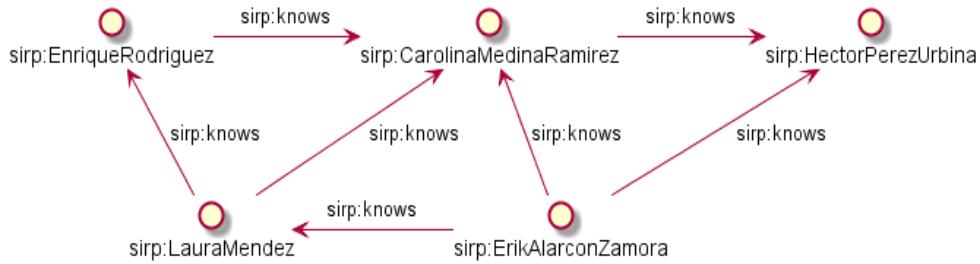


Figura 5.25: Subgrafo RDF de las relaciones *conoce-a* entre personas del área de Redes y Telecomunicaciones.

La Figura 5.26 presenta el subgrafo de la Figura 5.25 después de realizar el proceso de inferencia, donde la propiedad *conoce a* tiene la característica de ser simétrica. En esta Figura 5.26, todos los recursos (personas) son sujeto y objeto en las tripletas RDF. Por ejemplo, el Dr. Héctor Pérez en la Figura 5.25 no tiene sus propias declaraciones que indican a quienes conoce. Mientras, el Dr. Héctor Pérez en la Figura 5.26 sí tiene sus propias declaraciones que indican que conoce a la Dra. Carolina Medina y Erik Alarcón.

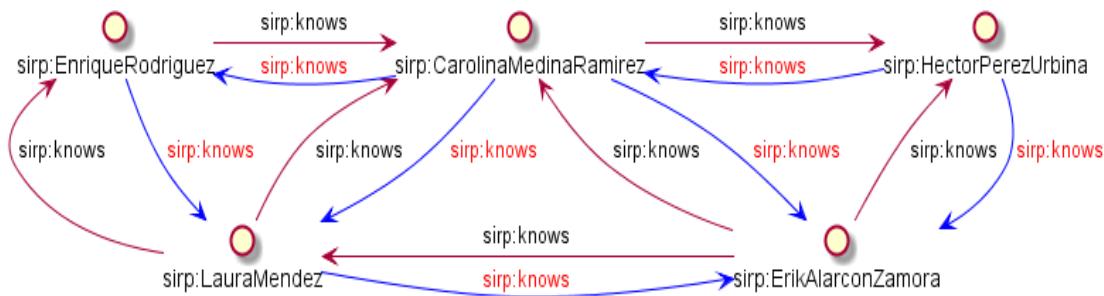


Figura 5.26: Ejemplo de inferencia a partir de la propiedad *conoce-a* como propiedad simétrica.

5.3. Búsqueda y recuperación de información en el modelo

Las etapas *representación y enriquecimiento del conocimiento* permiten construir nuestras bases de conocimiento, es decir, las ontologías para la cartografía de competencias, búsqueda de recursos digitales, así como el vocabulario de Redes y Telecomunicaciones. Estas ontologías son contenedores de conocimiento con un significado bien definido.

En esta tesis, las ontologías tienen como finalidad la búsqueda y recuperación de la información para responder las preguntas o necesidades informativas de los usuarios del área de Redes y Telecomunicaciones (RyT). Esta búsqueda y recuperación se hace a partir de tres puntos clave: 1) *identificar las preguntas en lenguaje natural*, 2) *transformar las preguntas a una consultas SPARQL* y 3) *ejecutar las consultas mediante un motor de búsqueda SPARQL*.

La *identificación de las preguntas en lenguaje natural* es el proceso de análisis y hallazgo de las principales preguntas a partir de los casos de uso.

Nuestros *casos de uso* son: 1) la *cartografía de competencias* que es la búsqueda de personas a partir de sus habilidades profesionales y conocimientos en RyT, así como 2) *la búsqueda de recursos digitales* que es el hallazgo de documentos y archivos multimedia a partir de los metadatos de éstos y los temas que están vinculados con el área de RyT.

Los parámetros de búsqueda en la cartografía competencias son:

- Habilidades profesionales como trabajo en equipo, creatividad, liderazgo, administración de proyectos, auto-aprendizaje, toma de decisiones, por mencionar algunas.
- Habilidades lingüísticas como hablar en español, inglés y francés; escribir en español, inglés y japones; leer en español y francés, entre otros idiomas.
- Conocimientos de Redes y Telecomunicaciones como sabiduría en sistemas operativos, capa enlace, filtros, ontologías, hilos de procesamiento, radios cognitivos, datos ligados, por mencionar algunos.
- Relaciones profesionales como conoce a la Dra. Carolina, Dr. Héctor, Dr. Ricardo, Dr. Miguel, Erik, Laura, entre otros.
- Clase de persona en RyT como profesionista, profesor, estudiante o empleado.
- Intervalo de edad como como 18 a 40, <40, >18, 18 a 18, por mencionar.
- Lugar de trabajo como la Universidad Autónoma Metropolitana, Clark & Parsia LLC, Infotec, Canonical Ltd, entre otras.
- Género, masculino, femenino, hombre y mujer.

A partir de estos parámetros de búsqueda, se hallaron un conjunto de preguntas básicas. El siguiente listado presenta las diecinueve preguntas básicas para la *cartografía de competencias*, donde cada pregunta tiene un respectivo *identificador de consulta*.

- *Q1.1.- ¿Cuáles son los nombres, correos, sitios web, géneros y edades de las personas del área de RyT?*
- *Q1.2.- ¿Cuáles son los nombres, sitios web y los lugares donde laboran las personas del RyT?*
- *Q1.3.- ¿Quiénes son mayores de 20 años y menores de 45 años?*
- *Q1.4.- ¿Cuáles son los nombres y sitios web de los profesionistas del área de RyT?*
- *Q1.5.- ¿Quiénes trabajan en la Clark & Parsia y son del sexo Masculino?*
- *Q1.6.- ¿Quiénes son estudiantes y leen en inglés?*
- *Q1.7.- ¿Quienes hablan, leen y escriben en inglés?*
- *Q1.8.- ¿Qué estudiantes saben algo de inglés?*
- *Q1.9.- ¿Qué profesores tienen la capacidad de síntesis?*
- *Q1.10.- ¿Qué profesionistas tienen conocimiento en los temas de Web Semántica?*
- *Q1.11.- ¿Qué profesores tienen conocimientos en Sistemas Distribuidos?*
- *Q1.12.- ¿Quiénes tienen conocimiento en Java, OWL, RDF, Threads, C, OpenMP?*
- *Q1.13.- ¿Qué estudiantes tienen algún conocimiento en los subtemas de Sistemas Operativos?*
- *Q1.14.- ¿Quiénes trabajan en una Universidad?*
- *Q1.15.- ¿Quienes laboran en la UAM y tienen algún conocimiento en Web Semántica?*
- *Q1.16.- ¿Qué personas tienen como asesor a Carolina Medina?*
- *Q1.17.- ¿Quiénes son los colegas de Ricardo Marcellin?*
- *Q1.18.- ¿Cuáles son los nombres y correos de las personas que conocen a Carolina Medina Ramírez?*
- *Q1.19.- ¿Qué personas son profesores-investigadores?*

De la misma manera que la cartografía, los principales criterios de búsqueda para los recursos digitales son:

- Nombre del autor como Carolina Medina, Héctor Pérez, Ricardo Baeza, Tim Berners Lee, por mencionar algunos.
- Lenguaje como inglés, español, francés, chino, ruso, entre otros.

- Extensión del archivo como pdf, doc, odp, txt, html, xml, ppt, flv, mpg, mp3, wav, jpg, entre otras.
- Institución involucrada como editorial (Mc Graw Hill, Grijalbo, Pearson), universidad (UNAM, UAM, IPN), revista científica (IEEE, ACM, Springer).
- Año de cota inferior, por ejemplo >1980, >2000, >2012.
- Clase de recurso digital en RyT como libro, artículo, reporte técnico, tesis, página Web, audio, vídeo, imagen, presentación, documento y multimedia.
- Temas vinculados al área de RyT como sistemas operativos, capa enlace, filtros, ontologías, hilos de procesamiento, radios cognitivos, datos ligados, por mencionar algunos.

Estos parámetros de búsqueda dieron pauta a la búsqueda de las principales preguntas para este *caso de uso*. A continuación, se listan las diez preguntas principales que se hallaron para la búsqueda de recursos digitales.

- ***Q2.1.*** - ¿Cuáles son los títulos, rutas, extensión, idioma de todos los recursos digitales de RyT?
- ***Q2.2.*** - ¿Cuáles libros tratan sobre algunos temas de Sistemas Distribuidos?
- ***Q2.3.*** - ¿Qué recursos fueron publicados por la UAM?
- ***Q2.4.*** - ¿Qué documentos sirven para dar un curso de Sistemas P2P?
- ***Q2.5.*** - ¿Qué recursos multimedia son mayores al año 2009?
- ***Q2.6.*** - ¿Cuáles documentos tratan sobre Ontologías?
- ***Q2.7.*** - ¿Qué recursos fueron publicados en una Revista científica?
- ***Q2.8.*** - ¿Qué recursos tienen en su descripción las palabras "linked data"?
- ***Q2.9.*** - ¿Cuáles documentos en Inglés y mayores al año 2000 son de autoría de Erik Alarcon Zamora?
- ***Q2.10.*** - ¿Cuál son las tesis de Samuel Hernandez Maza?

Las preguntas para los dos casos de uso fueron electas, porque algunas utilizan el conocimiento explícito y otras aprovechan el conocimiento implícito.

Con base en las preguntas de ambos listados, el siguiente paso es *transformar estas preguntas a consultas SPARQL*. Esta transformación para cada *pregunta-consulta* se hace de esta manera.

1. Analizar e identificar estos componentes: a) los valores que se desean recuperar, a los cuales se les asocia una *variable resultado*, b) las *propiedades* y c) los criterios de búsqueda que son *variables auxiliares* o *valores específicos* (recursos o literales).

2. Escribir las *variables resultado* en la cláusula SELECT.
3. Construir y escribir los *patrones tripletas* a partir de las variable resultado, propiedades, variables auxiliares y valores específicos.
4. Introducir los *patrones tripletas* en la cláusula WHERE.

A continuación, se presentan y describen algunos ejemplos para transformar las preguntas en *lenguaje natural* a sus respectivas consultas SPARQL. Estas preguntas son electas del listado de preguntas para la cartografía de competencias y la búsqueda de recursos digitales.

La Figura 5.27 presenta la consulta SPARQL que está asociada a la pregunta (Q1.1) *¿Cuáles son los nombres, correos, sitios web, géneros y edades de las personas del área de RyT?*. En esta pregunta, los valores de recuperación son: *nombre*, *correo*, *sitio Web*, *género* y *edad* para cualquier recurso persona. Cada valor tiene asociado una respectiva variable: *?name*, *?mail*, *?ws*, *?gender* y *?age*. Las propiedades asociadas a estos valores son: *tiene nombre* (*sirp:has-name*), *tiene email* (*sirp:has-email*), *tiene sitio Web* (*sirp:has-webSite*) y *tiene género* (*sirp:has-gender*). Finalmente, la variable auxiliar “*?x*” establece que sin importar el identificador URI del recurso, deben recuperarse los valores de las variables respuesta.

¿Cuáles son los **nombres, correos, sitios web, géneros y edades** de las personas del área de RyT?



```
PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>

SELECT ?name ?mail ?ws ?gender ?age
WHERE
{
    ?x sirp:has-name ?name;
        sirp:has-email ?mail;
        sirp:has-webSite ?ws;
        sirp:has-gender ?gender;
        sirp:has-age ?age.
}
```

Figura 5.27: Consulta SPARQL asociada a la pregunta Q1.1 de la cartografía de competencias.

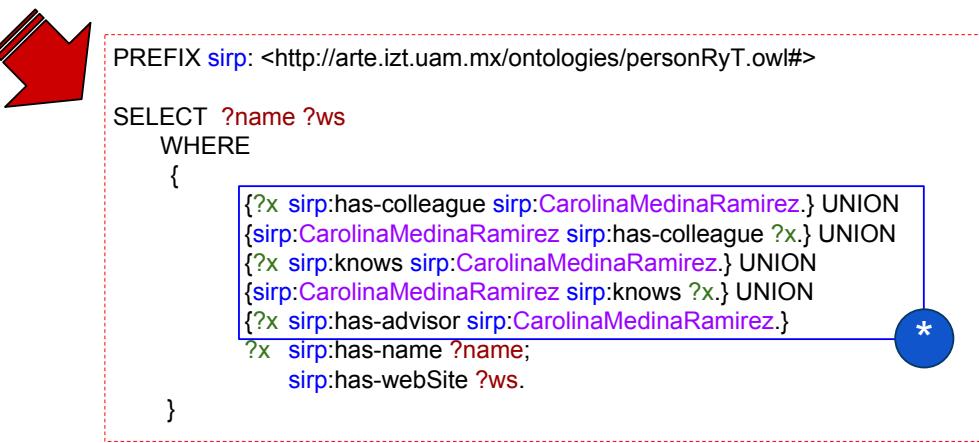
La consulta asociada a la pregunta Q1.1, es un ejemplo sencillo de construcción de consultas SPARQL. Ahora bien, existen preguntas que al transformarse a consultas SPARQL, emplean varios *patrones tripleta*. Pero, estas consultas pueden reducirse a una menor cantidad de *patrones tripleta*, si se da por hecho que hay razonamiento en la ontología. Un ejemplo de este tipo de consultas se da para la pregunta (Q1.18) *¿Cuáles son los nombres y sitios Web de las personas que conocen a Carolina Medina Ramírez?*. La Figura 5.28 muestra la consulta asociada a la pregunta Q1.18, en la cual no se da por hecho el uso de razonamiento. Mientras, la Figura 5.29 presenta la consulta para la misma pregunta, donde se da por hecho el uso del razonamiento.

A continuación, se listan los elementos de la consulta que están asociados a la pregunta Q1.18 de la Figura 5.28.

- Variable resultado: `?name` - nombre y `?ws` - sitio Web.
- Propiedades: *tiene nombre* (`sirp:has-name`), *tiene colega* (`sirp:has-colleague`) *tiene asesor* (`sirp:has-advisor`), *conoce a* (`sirp:knows`) y *tiene sitio Web* (`sirp:has-webSite`).
- Variables auxiliares: `?x` - identificador del recurso.
- Valores específicos: `sirp:CarolinaMedinaRamirez` - identificador de la Dra. Carolina Medina Ramírez.

En esta consulta SPARQL, la parte importante es el conjunto de *patrones triplete* que indican las relaciones profesionales de cualquier individuo con la Dra. Carolina. Porque, para indicar *a quiénes conoce* o *quiénes conocen a* la Dra. Carolina se deben emplear cinco *patrones triplete*, los cuales aparecen en el recuadro con el símbolo del asterisco.

¿Cuáles son los **nombres** y **sitios Web** de las personas que conocen a **Carolina Medina Ramírez?**



```

PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>

SELECT ?name ?ws
WHERE
{
  {?x sirp:has-colleague sirp:CarolinaMedinaRamirez} UNION
  {sirp:CarolinaMedinaRamirez sirp:has-colleague ?x} UNION
  {?x sirp:knows sirp:CarolinaMedinaRamirez} UNION
  {sirp:CarolinaMedinaRamirez sirp:knows ?x} UNION
  {?x sirp:has-advisor sirp:CarolinaMedinaRamirez}
  ?x sirp:has-name ?name;
  sirp:has-webSite ?ws.
}

```

Figura 5.28: Consulta SPARQL asociada a la pregunta Q1.18, en la cual no se da por hecho el uso del razonamiento.

La consulta de la pregunta Q1.18, puede simplificarse mediante la asunción de emplear razonamiento en la ontología. La Figura 5.29 presenta esta consulta SPARQL simplificada. En la cual, el número de patrones tripletes del recuadro con asterisco de la Figura 5.28 se reduce a un solo patrón marcado con el símbolo más. Porque, se da por hecho que las propiedades *tiene colega* (`has-colleague`) y *tiene asesor* son subpropiedades de la propiedad *conoce a* (`knows`), además esta superpropiedad tiene la característica simétrica.

En el caso de uso *búsqueda de recursos digitales*, existen consultas que mediante la inferencia en una ontología, pueden reducirse el número de patrones en sus respectivas cláusulas WHERE. A continuación, se presenta un ejemplo de este tipo de consultas. Este ejemplo esta

¿Cuáles son los **nombres** y **sitios Web** de las personas que conocen a **Carolina Medina Ramírez?**



PREFIX sirp: <<http://arte.itz.uam.mx/ontologies/personRyT.owl#>>

```
SELECT ?name ?ws
WHERE
{
    ?x sirp:knows sirp:CarolinaMedinaRamirez;
        sirp:has-name ?name;
        sirp:has-webSite ?ws.
}
```



Figura 5.29: Simplificación de la Consulta SPARQL asociada a la pregunta Q1.18 mediante la asunción de emplear razonamiento.

dividido en dos partes: 1) *la consulta se construye sin dar por hecho el uso de un razonador* y 2) la consulta se construye a partir de la aseveración de emplear razonamiento en una ontología.

La Figura 5.30 presenta la consulta asociada a la pregunta (Q2.6) *¿Cuáles documentos tratan sobre Ontologías?*. Esta consulta se construye a partir de las triplets explícitas en la ontología de *recursos digitales*.

¿Cuáles documentos tratan sobre **Ontologías**?



PREFIX sird: <<http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>>
PREFIX redes: <<http://mcycl.itz.uam.mx/arios/odaryt.owl#>>
PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

```
SELECT ?x
WHERE
{
    {?x rdf:type sird:Paper.} UNION
    {?x rdf:type sird:Book.} UNION
    {?x rdf:type sird:TechnicalReport.} UNION
    {?x rdf:type sird:Thesis.} UNION
    {?x rdf:type sird:Webpage.} UNION
    {?x rdf:type sird:Document.}
    ?x sird:has-topic redes:Ontology.
}
```



Figura 5.30: Consulta SPARQL asociada a la pregunta Q2.6, en la cual no se da por hecho el uso del razonamiento.

Los elementos importantes en la consulta (Q2.6) de la Figura 5.30, se listan a continuación:

- Variable resultado: $?x$ - identificador del recurso.

- Propiedades: *tipo* (*rdf:type*) y *tiene tópico* (*sird:has-topic*).
- Variables auxiliares: *?x* - identificador del recurso.
- Valores específicos: *sird:Book* - identificador de la clase Libro, *sird:Paper* - identificador de la clase Artículo, *sird:Thesis* - identificador de la clase Tesis, *Webpage* - identificador de la clase Página Web, *sird: TechnicalReport* - identificador de la clase Reporte Técnico, *sird: Document* - identificador de la clase Documento y *redes: Ontology* identificador del tema Ontología.

En esta consulta SPARQL, se resaltan los patrones del recuadro marcado con dos asteriscos. Estos seis patrones indican los recursos que son documentos. Porque, si solamente se emplea un patrón (*?x rdf:type sird: Document*), entonces se omiten artículos, libros, reportes técnicos, tesis y páginas Web que no tienen la asignación a la clase *Documento* (*sird: Document*).

La consulta SPARQL de la Figura 5.30 puede simplificarse mediante la asunción de emplear el proceso de inferencia en la ontología de recursos digitales. La Figura 5.31 presenta la consulta simplificada para la pregunta Q2.6. En esta consulta los seis patrones son reducidos a un patrón que está marcado con el doble signo de más. Porque, se da por hecho que los *individuos* de las clases *Artículo* (*sird: Paper*), *Libro* (*sird: Book*), *Reporte Técnico* (*sird: TechnicalReport*), *Tesis* (*sird: Thesis*) y *Página Web* (*sird: Webpage*) son individuos de la clase *Documento* (*sird: Document*).

¿Cuáles documentos tratan sobre Ontologías?



```

PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?x
WHERE
{
    ?x rdf:type sird:Document;
        sird:has-topic redes:Ontology.
}
  
```

Figura 5.31: Simplificación de la Consulta SPARQL asociada a la pregunta Q2.6 mediante la asunción de emplear razonamiento.

Estos son cinco ejemplos de transformación de preguntas a consultas SPARQL. Algunas consultas se construyen mediante triplets del ABox, porque la información a recuperar pertenece al conocimiento explícito; por ejemplo, la consulta de la Figura 5.27. Otras consultas se construyen mediante la agrupación de varios patrones, porque éstos permiten recuperar mejor información en el grafo RDF; por ejemplo las consultas de las Figuras 5.28 y 5.30.

Finalmente, existen consultas que aprovechan el uso de inferencia, para reducir el número de patrones y obtener mejores resultados; por ejemplo, las consultas SPARQL de las Figuras 5.29 y 5.31.

En la recuperación de la información, el *tercer punto clave* es la *ejecución de las consultas*. Esta ejecución se basa en el uso de un *motor de búsqueda* SPARQL, para buscar y recuperar la información en los grafos RDF de nuestras ontologías. El funcionamiento general de un motor SPARQL es el siguiente. Un motor de consulta SPARQL a partir de una consulta SPARQL y un grafo RDF, interpreta esta consulta SPARQL. Posteriormente, este motor compara los patrones de la cláusula WHERE con todos los triples en un *grafo RDF* (modelo). Aquellas tripletas que concuerden con los patrones, el motor recupera la información de las *variables resultado*. Finalmente, el motor regresa la información de estas variables al usuario. Esta información comúnmente se presenta forma de tabla. En la Figura 5.32 se presenta este funcionamiento de un motor SPARQL.

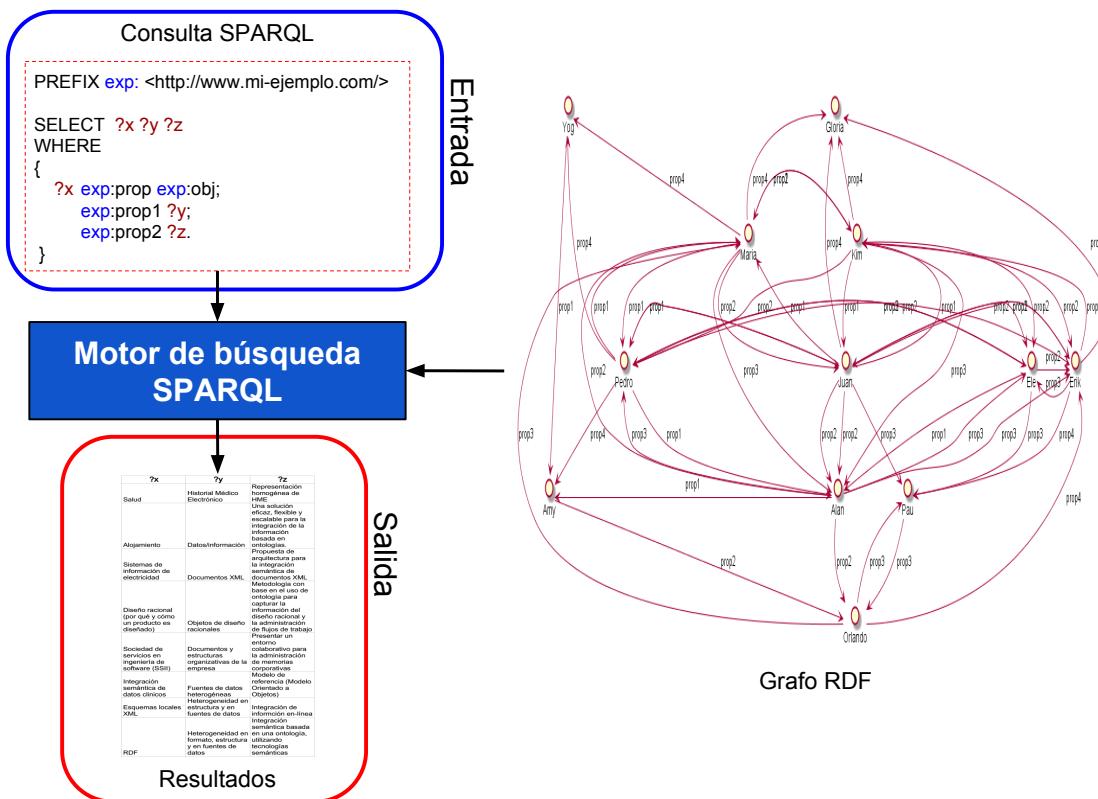


Figura 5.32: Proceso básico de consulta de información para un motor de búsqueda SPARQL.

Un motor SPARQL puede encontrarse en un triplestore. El triplestore Jena⁴ posee el motor de consulta (ARQ), el cual soporta el lenguaje SPARQL. Este motor ARQ permite recuperar los resultados de una consulta para mostrarlos en pantalla en forma de una tabla u obtenerlos de forma iterativa mediante JAVA.

⁴The Apache Software Foundation, “Apache Jena,” Disponible en: <http://jena.apache.org/>

Los resultados proporcionados por un motor SPARQL, dependen del uso o no del proceso de inferencia en una ontología. Si el motor emplea una ontología sin inferencia, entonces los resultados pueden ser menores a los esperados. Para ejemplificar ésto, supongamos que el grafo en la Figura 5.33 es el modelo de entrada para el motor SPARQL. Este modelo es un subgrafo que se tomo de la ontología *búsqueda de recursos digitales*.

A continuación, se listan las declaraciones que están asociadas a las tripletas de la Figura 5.33.

- El recurso *tesis 01* (*sird:tesis01-pdf*) tiene dos declaraciones; la primera establece que este recurso pertenece a la clase *tesis* (*sird:Thesis*) y la segunda indica que la *UAM* (*sirp:UAM*) es la institución involucrada en este recurso.
- Los recursos *tesis 02* (*sird:tesis02-doc*), *tesis 03* (*sird:tesis03-odp*) y *tesis 05* (*sird:tesis05-pdf*) tienen estas declaraciones. La *UAM* (*sirp:UAM*) es la institución involucrada en la *tesis 02* y *tesis 05*. Mientras, el *IPN* (*sird:IPN*) es la institución en la *tesis 03*.
- El recurso *tesis 04* (*sird:tesis04-pdf*) tiene estas declaraciones. Este recurso pertenece a la clase *Tesis* (*sird:Thesis*) y el *IPN* (*sirp:IPN*) es la institución involucrada en esta *tesis 04*.
- Los recursos *libro 01* (*sird:book01-pdf*), *libro 03* (*sird:book03-odp*) y *libro 05* (*sird:book05-pdf*) tienen las siguientes declaraciones. *Mc Graw Hill* (*sirp:McgrawHill*) es la editorial del recurso *libro 01*. Mientras, los recursos *libro 03* y *libro 05* tienen a la editorial *Oceano* (*sirp:Oceano*).
- El recurso *libro 02* (*sird:book02-doc*) es un recurso que pertenece a la clase *libro* (*sird:Book*) y la editorial de éste es *Mc Graw Hill* (*sirp:McgrawHill*).
- El recurso *libro 04* (*sird:book04-docx*) tiene estas declaraciones. Este recurso pertenece a la clase *libro* (*sird:Book*) y *Oceano* (*sirp:Oceano*) es la editorial de éste.
- Los recursos *artículo 01* (*sird:paper01-pdf*) y *artículo 02* (*sird:paper02-odp*) son publicados en la *IEEE* (*sirp:IEEE*) y la *ACM* (*sirp:ACM*) respectivamente.
- El recurso *artículo 03* (*sird:paper03-doc*) tiene dos declaraciones. La primera establece que este recurso se publicó en la *IEEE* (*sirp:IEEE*). Mientras, la segunda indica que éste pertenece a la clase *Artículo* (*sird:Paper*).
- El recurso *artículo 04* (*sird:paper04-docx*) pertenece a la clase *artículo* (*sird:Paper*) y se publicó en la *ACM* (*sirp:ACM*).
- La clase *Documento* (*sird:Document*) es superclase de las clases *artículo* (*sird:Paper*), *libro* (*sird:Book*) y *Tesis* (*sird:Thesis*).

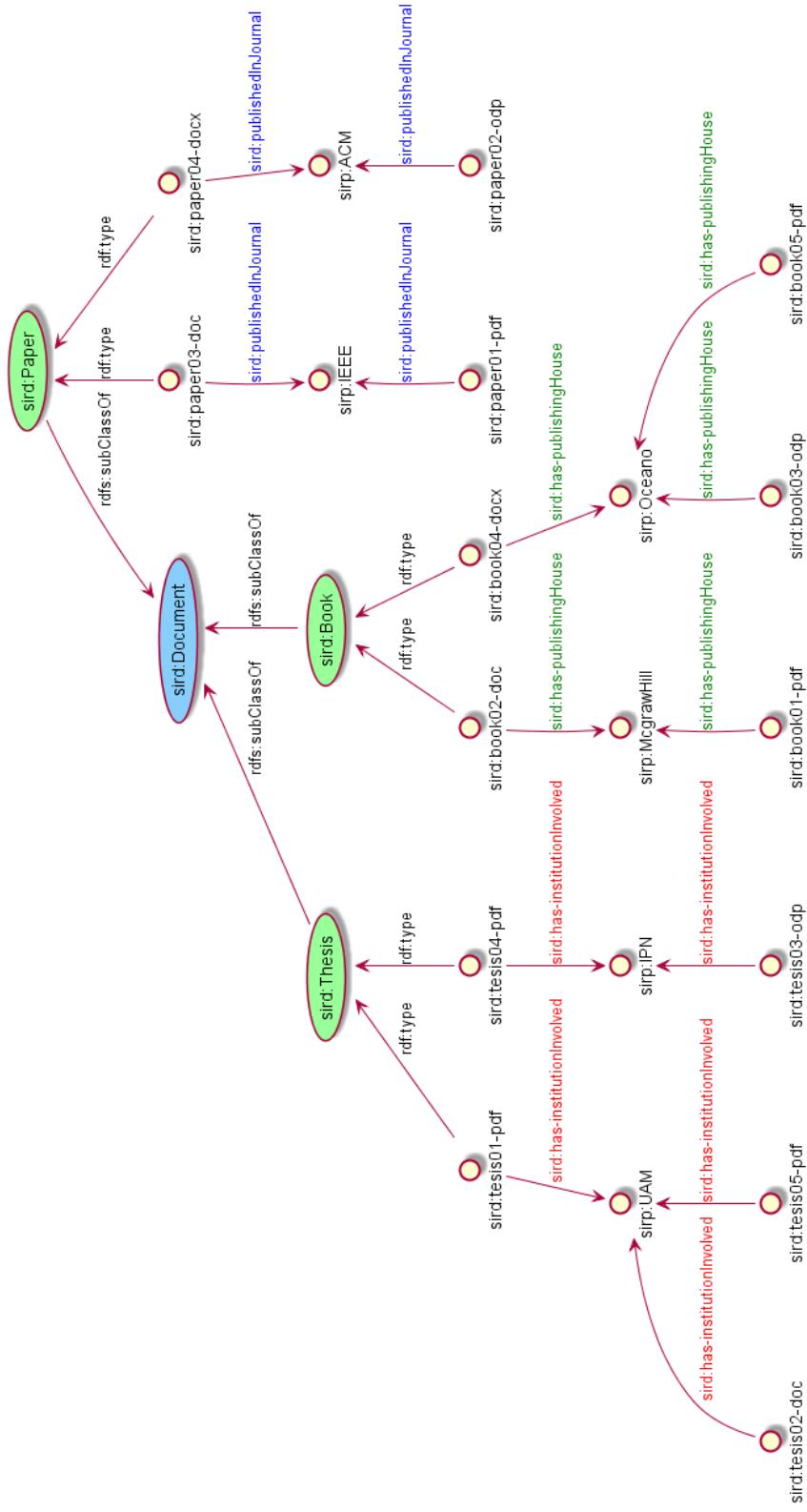


Figura 5.33: Ejemplo de modelo sin inferencia para el proceso de consulta de información.

5.3. Búsqueda y recuperación de información en el modelo

```

PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?x
WHERE
{
    {?x rdf:type sird:Paper.} UNION
    {?x rdf:type sird:Book.} UNION
    {?x rdf:type sird:TechnicalReport.} UNION
    {?x rdf:type sird:Thesis.} UNION
    {?x rdf:type sird:Webpage.} UNION
    {?x rdf:type sird:Document.}
}

```

Figura 5.34: Consulta de ejemplo para el proceso de consulta de información.

La consulta a interpretar por el motor SPARQL, es la asociada a la pregunta *¿Cuáles son los recursos que son documentos?*. Esta consulta se presenta en la Figura 5.34 y emplea seis patrones para establecer cuales recursos son documentos (artículos, libros, tesis, reportes técnicos, páginas web).

Un motor SPARQL interpreta esta consulta, compara estos patrones con el grafo de la Figura 5.33 y arroja una serie de resultados. La Tabla 5.7 muestra los identificadores de los recursos que son documentos.

?x
<i>sird:tesis01-pdf</i>
<i>sird:tesis04-pdf</i>
<i>sird:book02-doc</i>
<i>sird:book04-docx</i>
<i>sird:paper01-pdf</i>
<i>sird:paper02-odp</i>

Tabla 5.7: Identificadores de los recursos resultantes sin inferencia para la consulta de información.

Esta ejecución de la consulta por un motor SPARQL arrojo seis resultados. Sin embargo, existen otros recursos que son documentos y no tienen la declaración para indicar que pertenece a la clase *Documento* o a las subclases de ésta. De tal manera, es necesario el uso de un modelo con inferencia. La Figura 5.35 presenta el modelo que se obtiene por inferencia. En este modelo, todos los recursos están asignado a la clase *Documento* (*sird:Document*), por estas razones:

- Todos los recursos que pertenecen a las clases *artículo* (*sird:Paper*), *libro* (*sird:Book*) y *tesis* (*Thesis*), también pertenecen a la clase *documento* (*Documento*) , por los axiomas de jerarquía (*rdfs:subClassOf*).

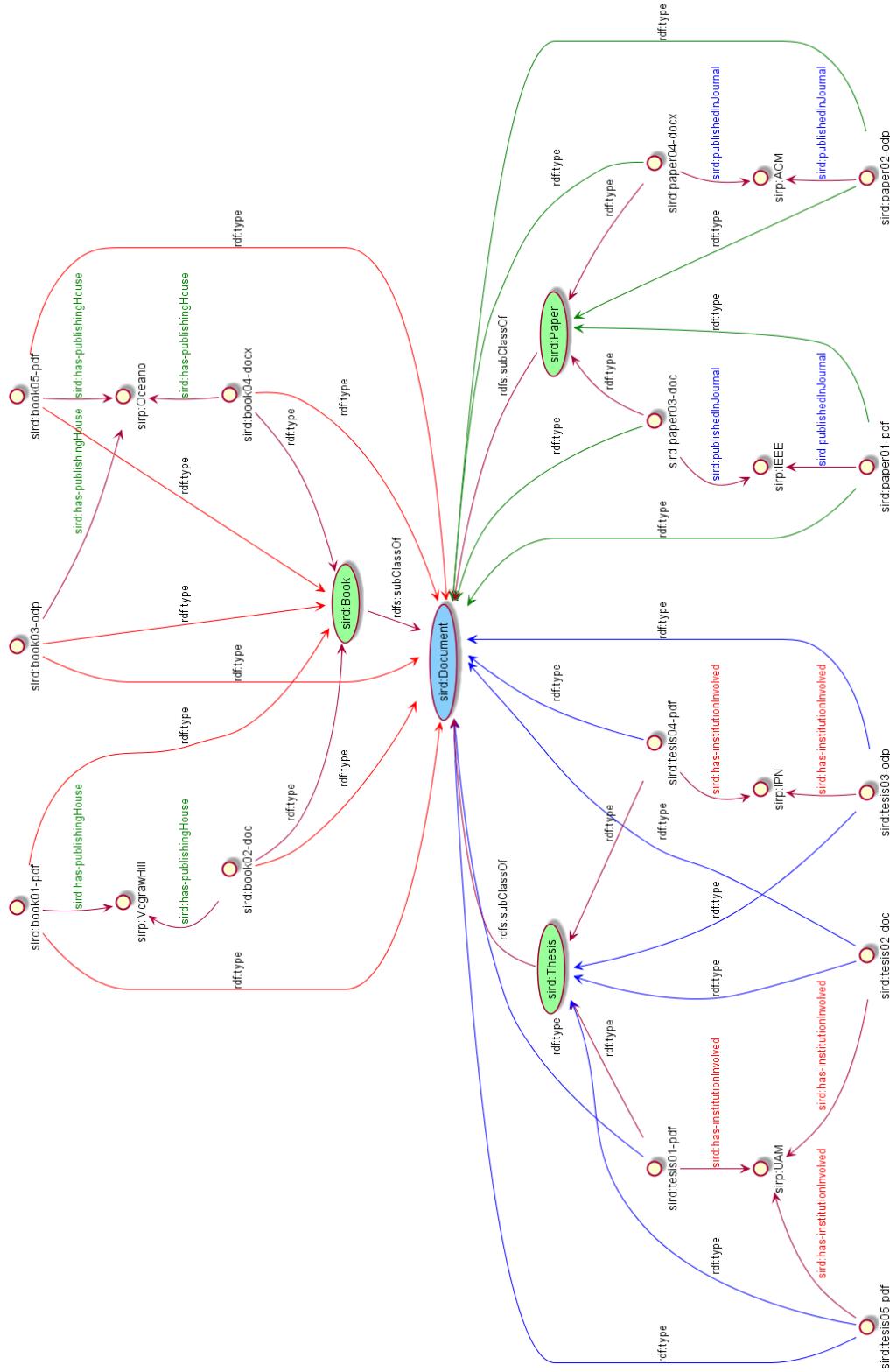


Figura 5.35: Ejemplo de modelo con inferencia para el proceso de consulta de información.

- La propiedad *publicado en la revista* (*publishedInJournal*) tiene por *dominio* la clase *artículo* (*sird:Paper*). Por ello, los recursos que utilizan esta propiedad, pertenecen a la clase Artículo, además la clase *artículo* es subclase de *documento*.
- De igual manera, las propiedades *tiene institución involucrada* (*sird:has-institutionInvolved*) y *tiene editorial* (*sird:has-publishingHouse*) tienen por dominio las clases *Tesis* (*sird:Thesis*) y *libro* (*sird:Book*) respectivamente. Estas dos clases son subclases de *documento*.

Al darse por hecho el uso de inferencia, la consulta 5.34 puede simplificarse a un solo patrón *tripleta*. Este patrón indica que se busca cualquier recurso que sea del tipo *documento* (*sird:Document*). En la Figura 5.36 se presenta la consulta simplificada para la pregunta *¿Cuáles son los recursos que son documentos?*

```
PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?x
WHERE
{ ?x rdf:type sird:Document. }
```

Figura 5.36: Consulta simplificada para el proceso de consulta de información.

Finalmente, la Tabla 5.8 presenta los resultados arrojados por un motor SPARQL para la consulta de la Figura 5.36 y empleando el grafo de la Figura 5.35. En esta Tabla, todos los identificadores de los recursos son los recursos que se esperan para responder esta pregunta.

?x
<i>sird:tesis01-pdf</i>
<i>sird:tesis02-doc</i>
<i>sird:tesis03-odp</i>
<i>sird:tesis04-pdf</i>
<i>sird:tesis05-pdf</i>
<i>sird:book01-pdf</i>
<i>sird:book02-doc</i>
<i>sird:book03-odp</i>
<i>sird:book04-docx</i>
<i>sird:book05-pdf</i>
<i>sird:paper01-pdf</i>
<i>sird:paper02-odp</i>
<i>sird:paper03-doc</i>
<i>sird:paper04-docx</i>

Tabla 5.8: Identificadores de los recursos resultantes a partir del uso de inferencia para la consulta de información.

Capítulo 6

Prototipo

La finalidad de la *integración semántica de recursos* es la *búsqueda y recuperación de información* en un modelo semántico, para responder las consultas de los usuarios del área de redes y telecomunicaciones. Aunque, un aspecto importante en esta *búsqueda y recuperación* es éste: *un usuario puede consultar información en un modelo, si éste tiene conocimientos en las tecnologías semánticas y en los vocabularios de las ontologías*. En concreto, él debe tener estas nociones:

- **Tecnologías semánticas:** tripletas rdf, axiomas, consultas SPARQL, patrones, variables resultado o auxiliares, inferencia.
- **Herramientas y operaciones de gestión de tripletas:** triplestore, motor de inferencia, motor de consulta, carga de modelo, forma de recuperación de la información, inferencia en un modelo.
- **Representación del dominio:** vocabularios (URIs), propiedades(URIs), tripletas de los axiomas, prefijos.

Sin embargo, no todos los *profesores-investigadores y estudiantes* tienen estos conocimientos. Ésto significa que no todos los usuarios del área de redes y telecomunicaciones utilizan la *integración semántica de recursos*. Por esta razón, se propone construir un prototipo de *interfaz gráfica de usuario* (GUI).

La finalidad de nuestro prototipo es proporcionar una manera fácil, sencilla y amigable de interacción entre un usuario y la *integración semántica de los recursos*. De esta forma, un usuario puede consultar y visualizar la información que responde a su consulta, sin que éste tenga conocimientos en las tecnologías semántica o los modelos.

Nuestro prototipo es una aplicación Web para consultar y visualizar información de los recursos de información. Estos recursos pertenecen a nuestros dos *casos de uso*: cartografía de competencias y búsqueda de recursos digitales.

El diseño del prototipo es una interfaz web, para facilitar el acceso de cualquier usuario a la información que proviene de los modelos semánticos. De esta manera, un usuario mediante un navegador puede acceder al prototipo y hacer consultas. Además, esta interfaz no requiere que el usuario instale alguna aplicación, porque todo el proceso de integración semántica se

hace del lado de un servidor. Con el objetivo de utilizar este prototipo en cualquier plataforma (Windows, Linux, Mac).

Los objetivos funcionales de nuestra *interfaz de usuario* son:

- Visualizar *información descriptiva* (Ver Figuras 5.4 y 5.5) de los *recursos de información*.
- Permitir búsquedas complejas con base en los parámetros de búsqueda en nuestros casos de uso (Ver Sección 5.3).
- Recuperar *información descriptiva* de los *recursos de información* en los modelos semánticos.
- Mostrar la información en un entorno amigable al usuario.
- Utilizar modelos semánticos con inferencia.
- Proporcionar un servicio transparente con respecto al uso de tecnologías semánticas.

Nosotros para alcanzar estos objetivos, construimos este prototipo como un aplicación Web dinámica. Porque, para cada pregunta de los usuarios se tienen distintos listados de resultados. Además, el contenido de una ontología puede cambiar (aumentar, disminuir o modificar) con el tiempo, de esta manera, los resultados se mantendrán actualizados.

Este prototipo está constituido por un conjunto de servlets y páginas dinámicas. Las páginas estáticas son formularios que permiten estructurar las consultas de los usuarios. Estos formularios reciben la información de los usuarios y la proporcionan a los servlets. Mientras, los servlets son los encargados de: recopilar la información para la consulta, así como procesar y ejecutar la consulta en el modelo semántico, recuperar la información de ésta, publicar la información en el navegador del usuario.

El motivo de usar formularios, es proporcionar al usuario un medio amigable para escribir los valores de su consulta. Mientras, los servlets proporcionan la generación dinámica de contenidos Web. Los servlets se programan en Java y el triplestore Jena es una librería para Java. Por esta razón, es fácil la interacción de estos dos elementos. En concreto, los servlets permiten utilizar el triplestore Jena para: cargar un modelo semántico, inferir en un modelo, aplicar una consulta SPARQL, consultar un modelo semántico, así como publicar ésta en lenguaje HTML.

La interfaz Web al implementarse a partir del uso de servlets y páginas estáticas, requiere un servidor apropiado que soporte estos dos elementos. Apache Tomcat¹ es una servidor que soporta el uso de servlets Java y páginas basadas en Java. Por ello, se utiliza Tomcat como servidor para nuestras páginas estáticas y dinámicas. Otra razón del uso de Tomcat, éste es un software open source e implementado bajo la licencia “Apache License version 2”.

La Figura 6.1 muestra de manera general la interacción de un usuario con una interfaz web del prototipo (formularios y servlets). En esta Figura, también se presentan los procesos internos del servlet para: cargar, inferir, consultar en un modelo semántico, así como la visualización de los resultados en una página dinámica.

¹The Apache Software Foundation, “Apache Tomcat,” Available: <http://tomcat.apache.org/>

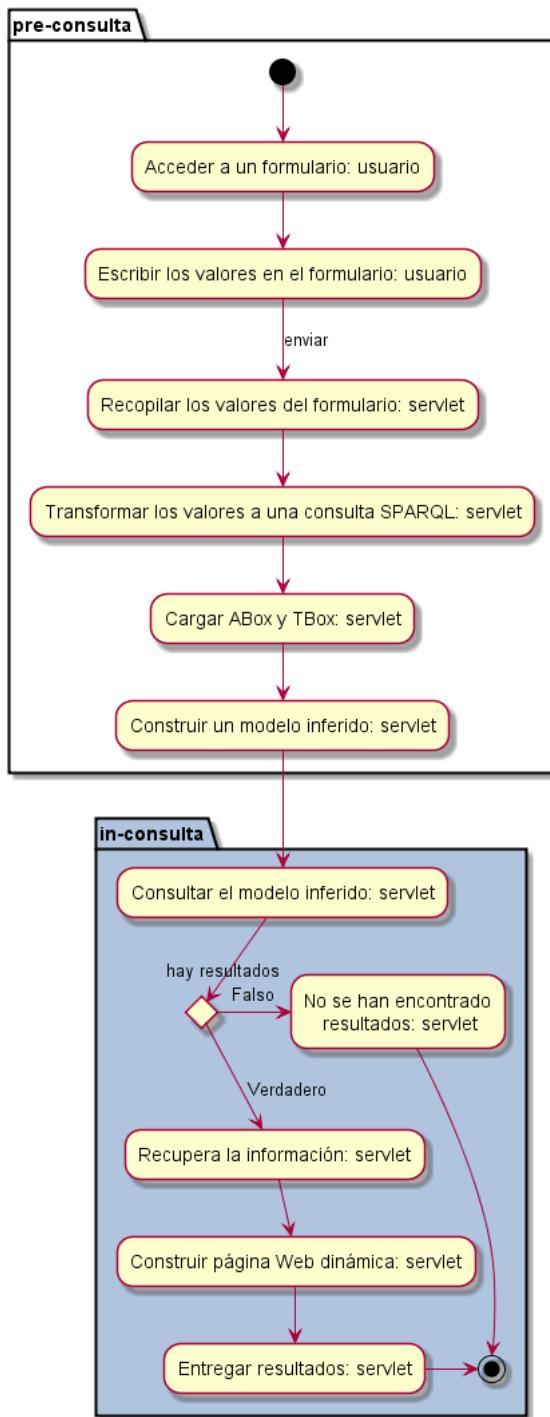


Figura 6.1: Flujo básico de la interacción de un usuario con un servlet.

Las actividades de la Figura 6.1 se describen en dos etapas: *pre-consulta* e *in-consulta*. Las actividades de la *pre-consulta* son las siguientes. El usuario escribe la dirección URL del formulario en su navegador. El sistema proporciona el formulario al navegador del usuario. El usuario escribe los valores de su consulta en el formulario y da clic en enviar. El sistema invoca al servlet y el sistema le proporciona los valores del formulario. El servlet transforma los valores del formulario a una consulta SPARQL. El servlet carga el ABox del modelo (*Recursos Persona* o *Recursos Digitales*) y los TBoxes (*Recursos Persona-Vocabulario RyT* o *Recursos Digitales-Vocabulario RyT*) en memoria RAM. El servlet invoca al *motor de inferencia* y genera un modelo inferido a partir del ABox y los TBoxes.

Mientras las etapas de la *in-consulta* son éstas. El servlet ejecuta la consulta SPARQL a partir del modelo inferido. Si esta consulta tiene respuestas, entonces el servlet recupera de manera iterativa los valores resultantes de la consulta y los incrusta entre etiquetas HTML. El servlet agrega las etiquetas necesarias (cabeceras, cuerpo, etc.) para que la página dinámica esté bien formada. El servlet libera memoria RAM del ABox, los TBoxes y el modelo inferido. El servlet envía la página dinámica (HTML) al navegador del usuario. Ahora bien, si la consulta no tiene respuestas, el servlet manda una página dinámica con el mensaje: “No se han encontrado resultados”.

Este prototipo se centra en tres operaciones, destinadas para los usuarios. En este listado, se describen estas operaciones:

- Navegación entre *información descriptiva* de personas, documentos y recursos multimedia. El objetivo es proporcionar a los *usuarios sin conocimientos en los recursos*, con un conocimiento general de estos *recursos*. De esta manera, los usuarios en el futuro puedan hacer búsquedas precisas de recursos.
- Búsqueda avanzada de los recursos persona, documentos y multimedia. La finalidad de ésto es encontrar aquellos recursos que respondan la consulta de un usuario. En esta búsqueda, los usuarios tienen establecidos uno o más *criterios de búsqueda* para su consulta.
- Opción para mostrar todos los detalles de un *recurso de información*. El objetivo de esta opción es mostrar a los usuarios con toda la información descriptiva que posee un recurso. La analogía de esta opción es la *ficha descriptiva* de un libro en una biblioteca.

Nuestro prototipo proporciona estas interfaces Web para los usuarios:

1. Validación de Usuario.
 2. Navegación entre Personas.
 3. Navegación entre Documentos.
 4. Navegación entre Multimedia.
 5. Filtrar por subclases de Persona.
-

6. Filtrar por subclases de Documento.
7. Filtrar por subclases de Multimedia.
8. Búsqueda Avanzada de Personas.
9. Búsqueda Avanzada de Documentos.
10. Búsqueda Avanzada de Multimedia.
11. Búsqueda por Tópicos de RyT.
12. Más detalles del Recurso Persona.
13. Más detalles del Recurso Documento.
14. Más detalles del Recurso Multimedia.
15. Más detalles del Recurso de Información.

Las especificaciones de estas interfaces son descritas en las siguientes Secciones de este Capítulo 6. Las interfaces asociadas a las *navegaciones* (ítem 2, 3 y 4 del listado de interfaces), se unieron en una sola especificación, porque el funcionamiento es similar entre estas tres navegaciones. Las diferencias son: la elección de los *recursos de información* (persona, documento, multimedia) y la *recuperación de información* para estos recursos.

De la misma manera, las especificaciones de las interfaces *Filtrar por subclases de Persona* (ítem 5), *Documento* (ítem 6 y *Multimedia* (ítem 7 se unieron en una especificación. Esta misma situación sucede para las interfaces de *Más detalles del Recurso de Información* (ítem 12), *Persona*, *Documento* (ítem 13) y *Multimedia* (ítem 14).

Especificación para validar a un usuario

Objetivo: Validar a un usuario para el uso de la interfaz Web de la integración semántica de recursos.

Actor: Usuario.

Pre-condición: El usuario debe estar registrado en el prototipo.

Post-condición: El usuario es valido y la interfaz Web presenta la página Web principal.

Flujo principal:

1. El usuario introduce el URL del Login en su navegador, por ejemplo: <http://localhost:8080/appsir/login.html>.
2. El usuario escribe su *nombre de usuario* y *contraseña*, después da clic en Enviar (Send).
3. La interfaz Web cambia la página Web Login a la página principal que se denomina Home.

Flujos secundarios:

- 3'. El usuario escribe erróneamente su *nombre de usuario* o *contraseña*, después da clic en Enviar (Send).
- 4'. La interfaz Web recarga la página Web *Login* y borra los datos de los campos *nombre de usuario* y *contraseña*.

La Figura 6.2 muestra dos interfaces del prototipo de integración semántica. La interfaz 6.2(a) presenta el *Login* para que un usuario escriba su nombre y contraseña. Mientras, la interfaz 6.2(b) muestra la página Home, como interfaz de presentación para el usuario valido.

Especificación para la navegación entre personas, documentos o multimedia

Objetivo: Desplazarse y visualizar las *características significativas* de las *personas, documentos o recursos multimedia* del área de Redes y Telecomunicaciones.

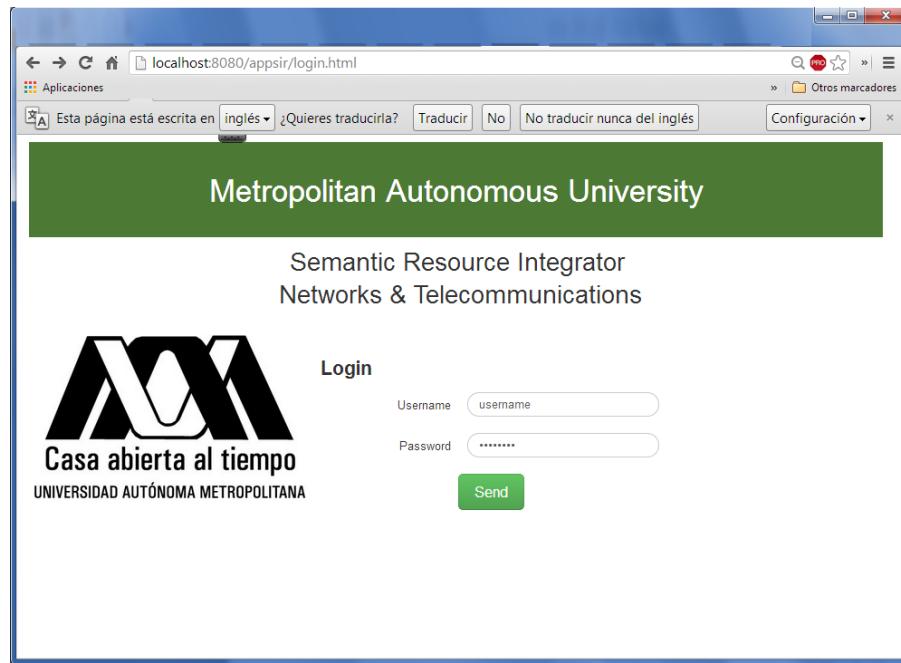
Actor: Usuario.

Pre-condición: El usuario está en la página *Home*.

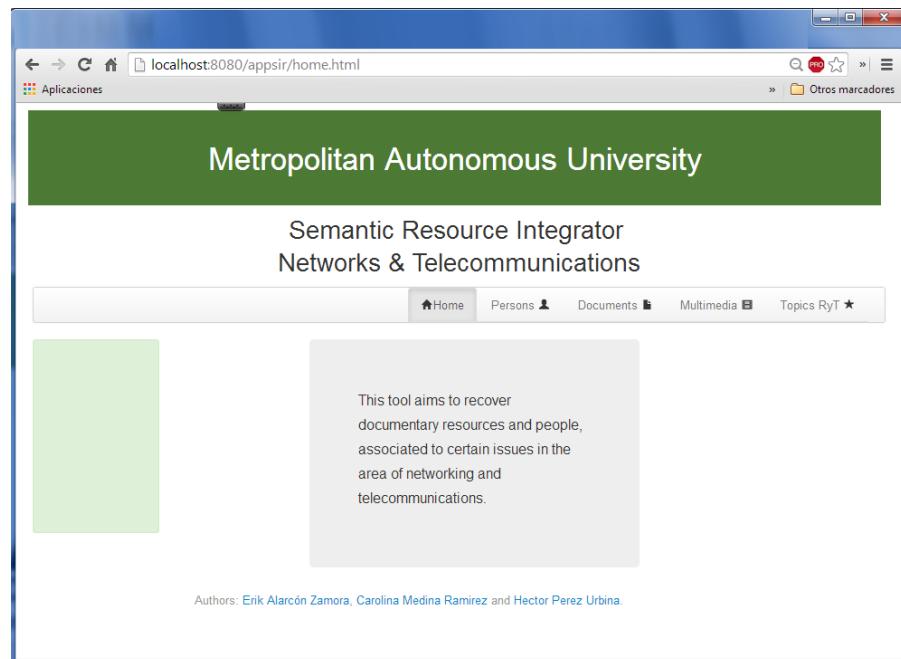
Post-condición: La interfaz Web muestra una sección para filtrar los recursos Persona, Documento o Multimedia a partir de las subclases de éstos. También, dependiendo la clase Persona, Documento o Multimedia, la *interfaz Web* muestra un botón de *búsqueda avanzada por clase* (*Advanced Search Person, Document, Multimedia*), y para cada *recurso de información* se muestra un botón de *más detalles* (*More Details*).

Flujo principal:

1. El usuario selecciona una de estas tres opciones: *Personas (Persons)*, *Documentos (Documents)*, *Multimedia (Multimedia)*.
2. La interfaz Web muestra algunas características significativas que dependen del tipo de *recursos de información*.
 - 2.1. Si la opción es **Persona**, entonces son mostrados el *nombre (Name)*, *edad (Age)*, *ocupación (Occupation)*, *género (Gender)*, *email (Email)* y *página personal (Web-Site)* para todas las *personas* en la *memoria corporativa* del área RyT.
 - 2.2. Si la opción es **Documento**, entonces son mostrados el *título (Title)*, *idioma (Language)*, *año (Year)* y *ruta (Path)* para todos los *documentos* en la *memoria corporativa* del área RyT.
 - 2.3. Si la opción es **Multimedia**, entonces son mostrados el *título (Title)*, *autores (Authors)*, *Año (Year)*, *tipo recurso (Resource Type)*, *idioma (Language)* y *ruta (Path)* para todos los *archivos multimedia* en la *memoria corporativa* del área RyT.



(a) Formulario Web: Validar al usuario



(b) Interfaz Web: Home

Figura 6.2: Validación del acceso de un usuario al prototipo de integración semántica de recursos.

La Figura 6.3 muestra las tres interfaces para la navegación entre la *información* de los *recursos de información*. La interfaz 6.3(a) muestra pequeñas *fichas descriptivas* de todos los recursos persona en la ontología de *cartografía de competencias*. De la misma manera, las interfaces 6.3(b) y 6.3(c) muestran respectivamente las *fichas descriptivas* de los documentos y archivos multimedia que pertenecen a la ontología de *recursos digitales*.

Especificación para filtrar los recursos por subclases de Persona, Documento o Multimedia

Objetivo: Buscar los *recursos de información* a partir de las subclases de Persona, Documento o Multimedia y mostrar un conjunto de *características significativas* de éstos.

Actor: Usuario.

Pre-condición: El usuario está en alguna de estas páginas: *Navegación entre personas*, *Navegación entre documentos* o *Navegación entre multimedia*.

Post-condiciones:

Si los recursos son persona, entonces la interfaz Web muestra la sección para filtrar los recursos persona, también la interfaz muestra un botón de *búsqueda avanzada de personas* (*Advanced Search Person*).

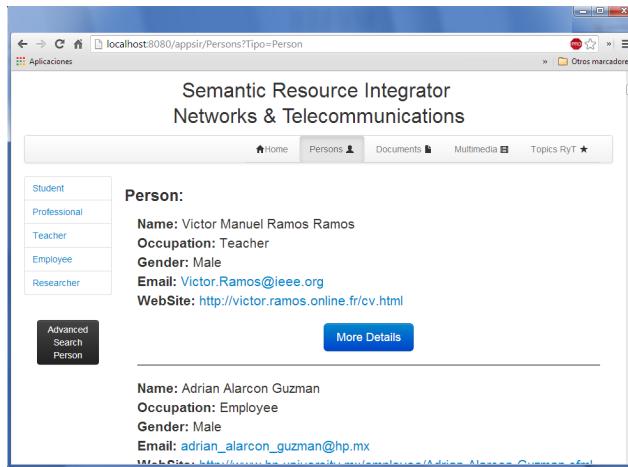
Si los recursos son documentos, entonces la interfaz Web muestra la sección para filtrar los recursos documento, también la interfaz muestra un botón de *búsqueda avanzada de documentos* (*Advanced Search Document*).

Si los recursos son documentos, entonces la interfaz Web muestra la sección para filtrar los recursos multimedia, también la interfaz muestra un botón de *búsqueda avanzada de multimedia* (*Advanced Search Multimedia*)

Además, la interfaz muestra para cada *recurso respuesta* un botón de *más detalles* (*More Details*).

Flujo principal:

1. El usuario selecciona la opción para filtrar los recursos a partir de las subclases de las clases Persona, Documento o Multimedia.
 - 1.1. Si la navegación es entre personas, entonces el usuario puede elegir las opciones: *estudiante* (*Student*), *profesionista* (*Professional*), *profesor* (*Teacher*), *empleado* (*Employee*) e *investigador* (*Researcher*).
 - 1.2. Si la navegación es entre documentos, entonces el usuario puede elegir las opciones: *libro* (*Book*), *artículo* (*Paper*), *tesis* (*Thesis*), *reporte técnico* (*TechnicalReport*) e *página web* (*Webpage*).
 - 1.3. Si la navegación es entre multimedia, entonces el usuario puede elegir las opciones: *audio* (*Audio*), *imagen* (*Image*), *presentación* (*Presentation*) y *vídeo* (*Video*).



(a) Interfaz Web: navegación entre personas del área RyT



(b) Interfaz Web: navegación entre documentos de la memoria corporativa del área RyT



(c) Interfaz Web: navegación entre recursos multimedia de la memoria corporativa del área RyT

Figura 6.3: Interfaces Web: navegación entre personas/documentos/multimedia.

2. La interfaz Web muestra algunas *características significativas* de los *recursos resultantes*.
 - 2.1. Si son personas, entonces son mostrados el *nombre* (*Name*), *edad* (*Age*), *ocupación* (*Occupation*), *género* (*Gender*), *email* (*Email*) y *página personal* (*WebSite*) para los *recursos persona* que pertenecen a la subclase electa.
 - 2.2. Si son documentos, entonces son mostrados el *título* (*Title*), *idioma* (*Language*), *año* (*Year*) y *ruta* (*Path*) para los *recursos documento* que pertenecen a la subclase electa.
 - 2.3. Si son multimedia, entonces son mostrados el *título* (*Title*), *autores* (*Authors*), *Año* (*Year*), *tipo recurso* (*Resource Type*), *idioma* (*Language*) y *ruta* (*Path*) para los *recursos multimedia* que pertenecen a la subclase electa.

En la Figura 6.4 se presentan tres interfaces para filtrar los *recursos de información* a partir de las subclases de Persona, Documento y Multimedia. La interfaz 6.4(a) presenta las *fichas descriptivas* de los *recursos persona* que pertenecen a la clase *Estudiante* (*Student*). La interfaz 6.4(a) muestra las fichas de los *recursos documentos* que pertenecen a la clase *Artículo* (*Paper*). De la misma manera, la interfaz 6.4(a) muestra las fichas de los *recursos multimedia* que son Vídeos.

Especificación para la búsqueda avanzada de personas

Objetivos: Estructurar la consulta de un usuario, así como buscar y recuperar *información significativa* de las personas del área de RyT para responder esta consulta.

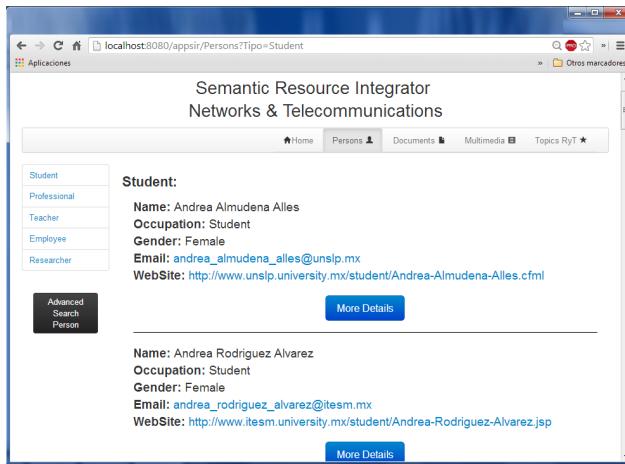
Actor: Usuario.

Pre-condición: El usuario está en la página de *navegación entre personas*.

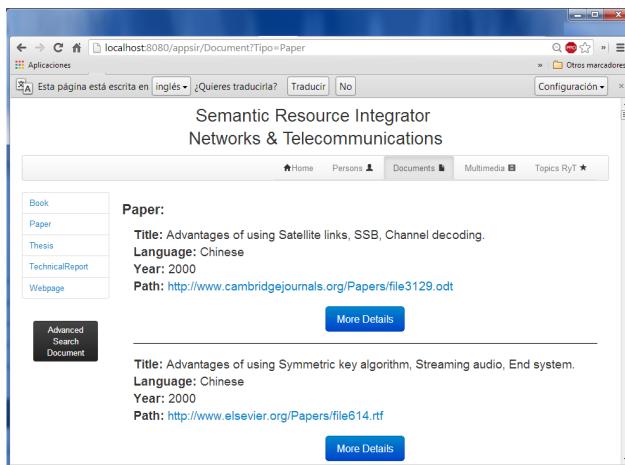
Post-condiciones: La interfaz Web muestra un conjunto de *fichas descriptivas de personas* que responden la consulta de un usuario. También, la interfaz Web muestra para cada *ficha descriptiva* el botón de *más detalles* (*More Details*).

Flujo principal:

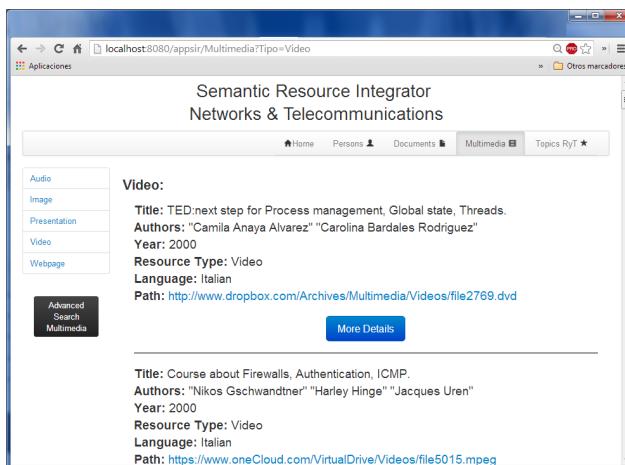
1. El usuario da clic en el botón *búsqueda avanzada de personas* (*Advanced Search Person*).
2. La interfaz Web muestra un formulario para estructurar la pregunta asociada a los *recursos persona*.
3. El usuario establece como criterio de búsqueda: *los conocimientos en los tópicos de Redes y Telecomunicaciones*.
 - 3.1. El usuario selecciona el *check box Tópicos* (*Topics*).
 - 3.2. El usuario escribe los tópicos de RyT, separando cada uno de éstos con una coma.



(a) Interfaz Web: filtrar personas por estudiantes



(b) Interfaz Web: filtrar documentos por artículos científicos



(c) Interfaz Web: filtrar recursos multimedia por videos

Figura 6.4: Interfaces Web: filtrar los recursos por subclases de Persona, Documento y Multimedia.

- 3.3. El usuario selecciona el operador *y* (*and*) u *o* (*or*) para comparar los tópicos de RyT.
 - 3.4. El usuario selecciona una opción de dos; *exactamente estos tópicos escritos o tópicos asociados con los tópicos escritos*.
 4. El usuario selecciona el tipo de *recurso persona*: *Persona (Person)*, *Estudiante (Student)*, *Profesionista (Professional)*, *Profesor (Teacher)*, *Empleado (Employee)* e *Investigador (Researcher)*.
 5. El usuario establece como criterio de búsqueda: ***las habilidades lingüísticas de una persona***.
 - 5.1. El usuario selecciona un idioma del *combo box*.
 - 5.2. Si el criterio es la habilidad de leer en el idioma electo, entonces el usuario selecciona el *check box lee (Reads)*.
 - 5.3. Si el criterio es la habilidad de escribir en el idioma electo, entonces el usuario selecciona el *check box escribe (Writes)*.
 - 5.4. Si el criterio es la habilidad de hablar en el idioma electo, entonces el usuario selecciona el *check box habla (Speaks)*.
 6. El usuario establece como criterio de búsqueda: ***el lugar de trabajo***.
 - 6.1. El usuario selecciona el *check box Lugar de trabajo (Workplace)*.
 - 6.2. Si el lugar es una universidad, entonces el usuario selecciona el *radio button universidad (University)* y selecciona la universidad del *combo box*.
 - 6.3. Si el lugar es una empresa, entonces el usuario selecciona el *radio button empresa (Company)* y selecciona la institución del *combo box*.
 7. El usuario establece como criterio de búsqueda: ***las relaciones profesionales o el conocimiento personal entre personas***.
 - 7.1. El usuario selecciona el *check box conoce (Knows)*.
 - 7.2. El usuario escribe el *nombre de una persona* que puede conocer a las personas que se buscan.
 8. El usuario establece como criterio de búsqueda: ***el género de las personas***.
 - 8.1. El usuario selecciona el *check box género (Gender)*.
 - 8.2. El usuario selecciona el sexo o género del *combo box*.
 9. El usuario establece como criterio de búsqueda: ***un intervalo de edad en las personas***.
 - 9.1. El usuario selecciona el *check box edad (Age)*.
-

- 9.2. El usuario escribe la edad mínima (número).
- 9.3. El usuario escribe la edad máxima (número).
10. El usuario establece como criterio de búsqueda: *las competencias profesionales*.
 - 10.1. El usuario selecciona el *check box competencias (Competences)*.
 - 10.2. Si se quiere una habilidad profesional, entonces el usuario selecciona la habilidad del primer *combo box*.
 - 10.3. Si se quieren dos habilidades profesional, entonces el usuario selecciona dos habilidades del primer y segundo *combo box*.
 - 10.4. Si se quieren tres habilidades profesional, entonces el usuario selecciona tres habilidades del primer, segundo y tercer *combo box*.
11. El usuario elige una opción (en el *combo box*) para ordenar los resultados de la búsqueda.
12. El usuario establece la cantidad de resultados a mostrar
 - 12.1. El usuario selecciona el *check box número de resultados (Number of results)*.
 - 12.2. El usuario escribe el número de recursos a recuperar.
13. El usuario da clic da en Enviar (Send).
14. La interfaz Web muestra el *nombre (Name)*, *edad (Age)*, *ocupación (Occupation)*, *género (Gender)*, *email (Email)* y *página personal (WebSite)*, para cada *recurso persona* resultante de esta búsqueda.

La Figura 6.5 muestra los distintos parámetros de búsqueda que un usuario puede elegir y escribir, para estructurar su consulta de *recursos persona*. Mientras, la Figura 6.6 muestra la *fichas descriptivas* de los *recursos persona* que responden esta búsqueda, los cuales se obtienen de la ontología *cartografía de competencias*.

Especificación para la búsqueda avanzada de documentos

Objetivos: Estructurar la consulta de un usuario, así como buscar y recuperar *información significativa* de los documentos del área de RyT para responder esta consulta.

Actor: Usuario.

Pre-condición: El usuario está en la página de *navegación entre documentos*.

Post-condiciones: La interfaz Web muestra un conjunto de *fichas descriptivas de documentos* que responden la consulta de un usuario. También, la interfaz Web muestra para cada *ficha descriptiva* el botón de *más detalles (More Details)*.

Flujo principal:

The screenshot shows a web application window titled "Metropolitan Autonomous University" with a green header bar. Below the header, the text "Semantic Resource Integrator" and "Networks & Telecommunications" is displayed. The main content area is titled "Advanced Search Person".

Topics: Topics Separate topics with commas And Exactly these issues Associated with these issues

Resource Type:

- Person Student Professional
- Teacher Employee Researcher

Language: English Reads Writes Speaks

Workplace: University UAM Company Clark & Parsia

Knows: Knows Name of person

Gender: Gender Male

Age: Age Min 0 Max 100

Competences: Competences Team Player

Order results by: Name Number of results Enter a positive number

Send

Figura 6.5: Formulario Web: búsqueda avanzada de personas.

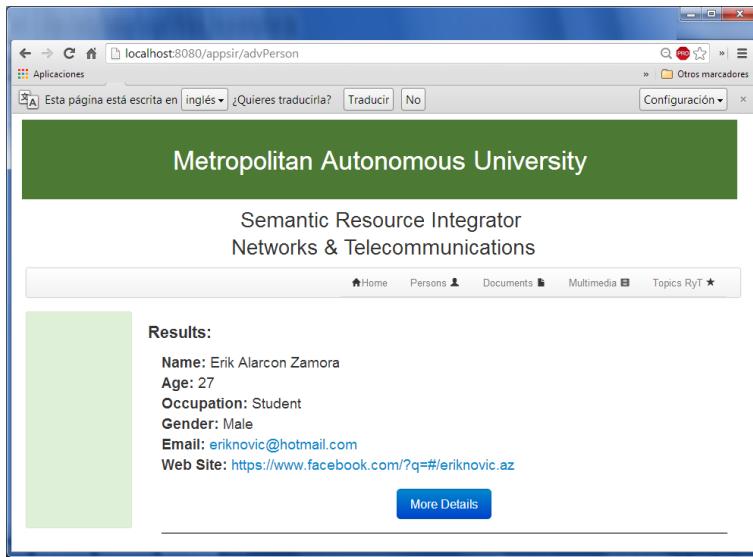


Figura 6.6: Página Dinámica: Resultados de búsqueda avanzada de personas.

1. El usuario da clic en el botón *búsqueda avanzada de documentos* (*Advanced Search Document*).
2. La interfaz Web muestra un formulario para estructurar la pregunta asociada a los *recursos documento*.
3. El usuario establece como criterio de búsqueda: ***los tópicos de Redes y Telecomunicaciones***.
 - 3.1. El usuario selecciona el *check box Tópicos* (*Topics*).
 - 3.2. El usuario escribe los tópicos de RyT, separando cada uno de éstos con una coma.
 - 3.3. El usuario selecciona el operador *y* (*and*) u *o* (*or*) para comparar los tópicos de RyT.
 - 3.4. El usuario selecciona una opción de dos; *exactamente estos tópicos escritos o tópicos asociados con los tópicos escritos*.
4. El usuario selecciona el tipo de *recurso documento*: *Documento* (*Document*), *Libro* (*Book*), *Artículo* (*Paper*), *Página Web* (*Webpage*), *Tesis* (*Thesis*) y *Reporte Técnico* (*Technical-Report*).
5. El usuario establece como criterio de búsqueda: ***el idioma o lenguaje del documento***.
 - 5.1. El usuario selecciona el *check box lenguaje* (*Language*).
 - 5.2. El usuario selecciona un lenguaje del *combo box*.

6. El usuario establece como criterio de búsqueda: *el nombre del autor*.
 - 6.1. El usuario selecciona el *check box autor (Author)*.
 - 6.2. El usuario escribe el *nombre o nombres y apellidos* del autor.
 7. El usuario establece como criterio de búsqueda: *el tipo de extensión de los documentos*.
 - 7.1. El usuario selecciona el *check box extensión del archivo (File Extensions)*.
 - 7.2. El usuario escribe la *extensión* o conjunto de extensiones separando éstas con una coma (pdf, doc, html).
 8. El usuario establece como criterio de búsqueda: *el año del documento* que puede ser año de publicación, año de creación o de ultima visita.
 - 8.1. El usuario selecciona el *check box año (Year)*.
 - 8.2. El usuario escribe el *año*.
 9. El usuario establece como criterio de búsqueda: *la organización donde se publicó o de procedencia del documento*.
 - 9.1. El usuario selecciona el *check box publicado (Published)*.
 - 9.2. Si la organización es una universidad, entonces el usuario selecciona el *radio button universidad (University)* y selecciona la universidad del *combo box*.
 - 9.3. Si la organización es una editorial, entonces el usuario selecciona el *radio button editorial (Publishing house)* y selecciona la editorial del *combo box*.
 - 9.4. Si el lugar es una revista científica, entonces el usuario selecciona el *radio button revista científica (Journal)* y selecciona la revista del *combo box*.
 10. Si el documento es un artículo científico, entonces el usuario puede establecer como criterio de búsqueda: *el tipo de artículo*.
 - 10.1. El usuario selecciona el *check box tipo de artículo (Type of paper)*.
 - 10.2. El usuario selecciona el *tipo de artículo* del *combo box*.
 11. El usuario elige una opción (en el *combo box*) para ordenar los resultados de la búsqueda.
 12. El usuario establece la cantidad de resultados a mostrar
 - 12.1. El usuario selecciona el *check box número de resultados (Number of results)*.
 - 12.2. El usuario escribe el número de recursos a recuperar.
 13. El usuario da clic da en Enviar (Send).
-

14. La interfaz Web muestra el *título (Title)*, *idioma (Language)*, *año (Year)* y *ruta (Path)*, para cada *recurso documento* resultante de esta búsqueda.

La Figura 6.7 muestra los distintos parámetros de búsqueda que un usuario puede elegir y escribir, para estructurar su consulta de *recursos documento*. Mientras, la Figura 6.8 muestra la *fichas descriptivas* de los *documentos* que responden esta búsqueda, los cuales se obtienen de la ontología *búsqueda de recursos digitales*.

Especificación para la búsqueda avanzada de recursos multimedia

Objetivos: Estructurar la consulta de un usuario, así como buscar y recuperar *información significativa* de los *recursos multimedia* del área de RyT para responder esta consulta.

Actor: Usuario.

Pre-condición: El usuario está en la página de *navegación entre multimedia*.

Post-condiciones: La interfaz Web muestra un conjunto de *fichas descriptivas de recursos multimedia* que responden la consulta de un usuario. También, la interfaz Web muestra para cada *ficha descriptiva* el botón de *más detalles (More Details)*.

Flujo principal:

1. El usuario da clic en el botón *búsqueda avanzada de multimedia (Advanced Search Multimedia)*.
 2. La interfaz Web muestra un formulario para estructurar la pregunta asociada a los *recursos multimedia*.
 3. El usuario establece como criterio de búsqueda: *los tópicos de Redes y Telecomunicaciones*.
 - 3.1. El usuario selecciona el *check box Tópicos (Topics)*.
 - 3.2. El usuario escribe los tópicos de RyT, separando cada uno de éstos con una coma.
 - 3.3. El usuario selecciona el operador *y (and)* u *o (or)* para comparar los tópicos de RyT.
 - 3.4. El usuario selecciona una opción de dos; *exactamente estos tópicos escritos o tópicos asociados con los tópicos escritos*.
 4. El usuario selecciona el tipo de *recurso multimedia: multimedia (Multimedia), Audio (Audio), Imagen (Image), Presentación (presentation) y Vídeo (Video)*.
 5. El usuario establece como criterio de búsqueda: *el idioma o lenguaje del recurso multimedia*.
-

The screenshot shows a web browser window with the URL `localhost:8080/appcir/AdvanceDocument.html`. The page title is "Metropolitan Autonomous University" and the subtitle is "Semantic Resource Integrator Networks & Telecommunications". The main content is a "Advanced Search Document" form.

Topics: Topics Exactly these issues Associated with these issues

Resource Type: Document Publication Book Paper
 Webpage Thesis Technical-Report

Language: Language English

Author: Author Peiyun Zhang

File: File PDF
Extensions:

Year: Year year

Published: University UAM Publishing house McGraw Hill Journal IEEE

Type of paper: Conference Publications

Order results by: Year Number of results 10

Send

Figura 6.7: Formulario Web: búsqueda avanzada de documentos.

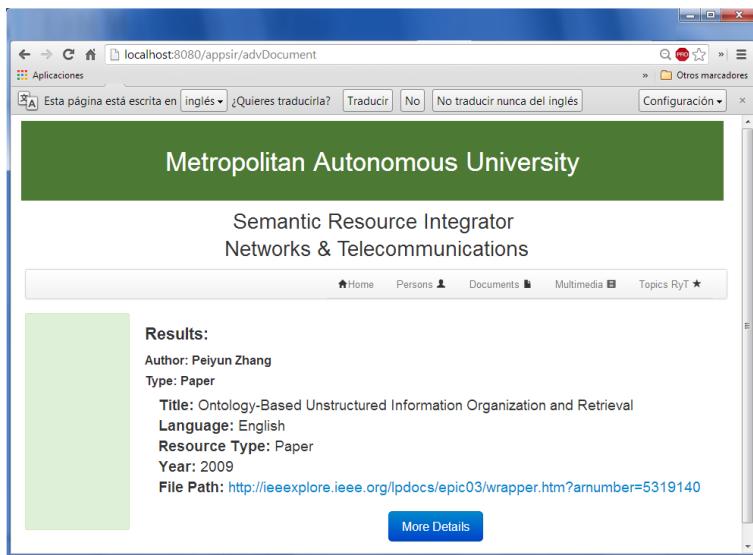


Figura 6.8: Página Dinámica: Resultados de búsqueda avanzada de documentos.

- 5.1. El usuario selecciona el *check box lenguaje (Language)*.
- 5.2. El usuario selecciona un lenguaje del *combo box*.
6. El usuario establece como criterio de búsqueda: *el nombre del autor*.
 - 6.1. El usuario selecciona el *check box autor (Author)*.
 - 6.2. El usuario escribe el *nombre o nombres y apellidos* del autor.
7. El usuario establece como criterio de búsqueda: *el tipo de extensión de los archivos multimedia*.
 - 7.1. El usuario selecciona el *check box extensión del archivo (File Extensions)*.
 - 7.2. El usuario escribe la *extensión* o conjunto de extensiones separando éstas con una coma (swf, wav, mpg).
8. El usuario establece como criterio de búsqueda: *el año del recurso multimedia* que puede ser año de creación o de exposición.
 - 8.1. El usuario selecciona el *check box año (Year)*.
 - 8.2. El usuario escribe el *año*.
9. El usuario elige una opción (en el *combo box*) para ordenar los resultados de la búsqueda.
10. El usuario establece la cantidad de resultados a mostrar
 - 10.1. El usuario selecciona el *check box número de resultados (Number of results)*.

- 10.2. El usuario escribe el número de recursos a recuperar.
11. El usuario da clic da en Enviar (Send).
12. La interfaz Web muestra el *título* (*Title*), *autores* (*Authors*), *Año* (*Year*), *tipo recurso* (*Resource Type*), *idioma* (*Language*) y *ruta* (*Path*), para cada *recurso multimedia* resultante de esta búsqueda.

La Figura 6.9 muestra los distintos parámetros de búsqueda que un usuario puede elegir y escribir, para estructurar su consulta de *recursos multimedia*. Mientras, la Figura 6.10 muestra la *fichas descriptivas* de los *recursos multimedia* que responden esta búsqueda, los cuales se obtienen de la ontología *búsqueda de recursos digitales*.

Especificación para la búsqueda por tópico(s) de redes y telecomunicaciones

Objetivo: Buscar *información significativa* en los *recursos de información* a partir de los tópicos de redes y telecomunicaciones.

Actor: Usuario.

Pre-condición: El usuario está en la página *Home*.

Post-condición: La interfaz Web muestra la *fichas descriptivas* de los recursos de información que responden la consulta de un usuario. También, la interfaz Web muestra para cada recurso el botón de *más detalles* (*More Details*).

Flujo principal:

1. El usuario selecciona la opción *Tópicos de redes* (*Topics RyT*).
2. La interfaz Web muestra un formulario para estructurar la pregunta asociada a cualquier *recurso de información* (Persona, Documento, Multimedia).
3. El usuario escribe los tópicos de RyT, separando cada uno de éstos con una coma.
4. El usuario selecciona el operador *y* (*and*) u *o* (*or*) para comparar los tópicos de RyT.
5. El usuario selecciona una opción de dos; *exactamente estos tópicos escritos o tópicos asociados con los tópicos escritos*.
6. El usuario selecciona el tipo de *recurso de información*: *Todos*, *Persona*, *Documento*, *Multimedia*, *Artículo*, *Libro* y *Presentación*.
7. El usuario establece como criterio de búsqueda: el idioma o lenguaje.
 - 7.1. El usuario selecciona el *check box lenguaje* (*Language*).
 - 7.2. El usuario selecciona un lenguaje del *combo box*.

The screenshot shows a web browser window with the URL `localhost:8080/appssir/AdvanceMultimedia.html`. The page title is "Metropolitan Autonomous University" and the subtitle is "Semantic Resource Integrator Networks & Telecommunications". The main content is an "Advanced Search Multimedia" form.

Topics: Semantic Web And
 Exactly these issues Associated with these issues

Resource Type: Multimedia Audio Image Presentation Video

Language: English

Author: Name of person

File Extensions: FLV

Year: 2007

Order results by: Year Number of results Enter a positive number

Send

Figura 6.9: Formulario Web: búsqueda avanzada de multimedia.

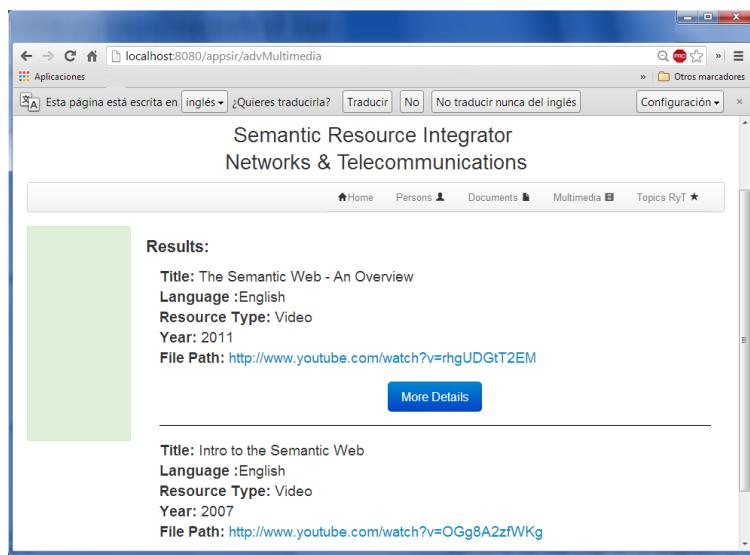


Figura 6.10: Página Dinámica: Resultados de búsqueda avanzada de multimedia.

8. El usuario elige una opción (en el *combo box*) para ordenar los resultados de la búsqueda.
9. El usuario establece la cantidad de resultados a mostrar.
 - 9.1. El usuario selecciona el *check box número de resultados (Number of results)*.
 - 9.2. El usuario escribe el número de recursos a recuperar.
10. El usuario da clic da en Enviar (Send).
11. La interfaz Web muestra el *nombre (Name)*, *tipo de recurso (Resource Type)*, *ruta o página web (URL)* y *tópicos de redes (Topics RyT)*, para cada recurso resultante.

La Figura 6.11 muestra dos interfaces. La interfaz 6.11(a) muestra el formulario para la búsqueda en todos los *recursos digitales*. Los campos de este formulario son genéricos para realizar la búsqueda en las ontologías de *cartografía de competencias* y *recursos digitales*. Mientras, la interfaz 6.11(b) muestra las fichas descriptivas de los *recursos de información* que repoden la consulta del usuario.

Especificación para mostrar los detalles de los recursos de información

Objetivo: Mostrar toda la *información descriptiva* de un determinado recurso de información: *persona*, *documento* o *multimedia*.

Actor: Usuario.

Pre-condiciones: Si se buscan detalles de un *recurso persona*, entonces, el usuario está en una de estas páginas: 1) Navegación entre personas, 2) Resultados de Filtrar los recursos

The screenshot shows a web browser window with the URL `localhost:8080/appssir/formAllResources.html`. The page title is "Search". It contains several input fields and radio buttons:

- Topics:** A text input field with placeholder text "Separate topics with commas" and a dropdown menu with options "And" and "Or". Below it are two radio button groups:
 - "Exactly these issues" (selected)
 - "Associated with these issues"
- Resource Type:** A radio button group with "Any" selected, and other options like "Person", "Document", "Multimedia", "Paper", "Book", and "Presentation".
- Language:** A dropdown menu set to "Any".
- Order results by:** A dropdown menu set to "Name" and a text input field for "Number of results" with placeholder text "Enter a positive number".
- A green "Send" button at the bottom.

(a) Formulario Web: búsqueda por tópicos

The screenshot shows a web browser window with the URL `localhost:8080/appssir/AllResources`. The page title is "Semantic Resource Integrator Networks & Telecommunications". It displays search results under the heading "Resources:":

- Name:** A Description Logic Primer
Resource Type: Paper
URL: <http://arxiv.org/pdf/1201.4089.pdf>
Topics RyT: Semantic, Description logics, ABox, Ontology, TBox, Axiom, OWL, SROIQ, DL languages,
- Name:** Adrian Vega Blanco
Resource Type: Employee
URL: <http://www.oracle.university.mx/employee/Adrian-Vega-Blanco.asp>
Topics RyT: Ontology, Ontology Reasoning, Knowledge Representation, Web Support Applications, Personal Information Management,

Each result has a blue "More Details" button below it.

(b) Página Dinámica: Resultados de búsqueda

Figura 6.11: Formulario Web para la búsqueda por temas en todos los recursos de información y la página Web mostrando los resultados de esta búsqueda.

por subclases de Persona, 3) Resultados de Búsqueda Avanzada de Personas y 4) Resultados de Búsqueda por tópico(s) de redes y telecomunicaciones.

Si se buscan detalles de un *recurso documento*, entonces, el usuario está en una de estas páginas: 1) Navegación entre documentos, 2) Resultados de Filtrar los recursos por subclases de Documento, 3) Resultados de Búsqueda Avanzada de Documentos y 4) Resultados de Búsqueda por tópico(s) de redes y telecomunicaciones.

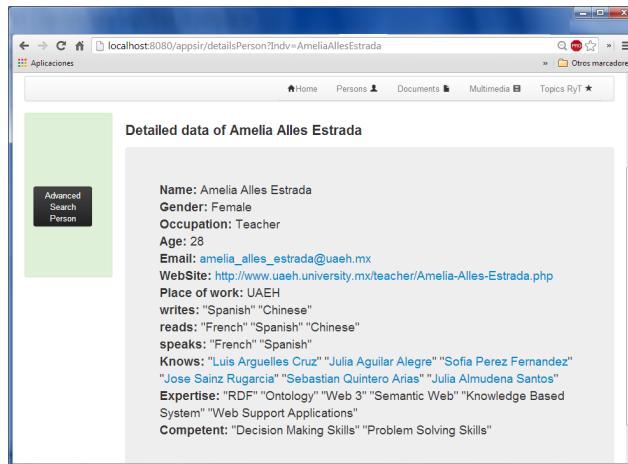
Si se buscan detalles de un *recurso multimedia*, entonces, el usuario está en una de estas páginas: 1) Navegación entre multimedia, 2) Resultados de Filtrar los recursos por subclases de Multimedia, 3) Resultados de Búsqueda Avanzada de Multimedia y 4) Resultados de Búsqueda por tópico(s) de redes y telecomunicaciones.

Post-condición: La interfaz Web muestra toda la información descriptiva del recurso (persona, documento o multimedia) que eligió el usuario.

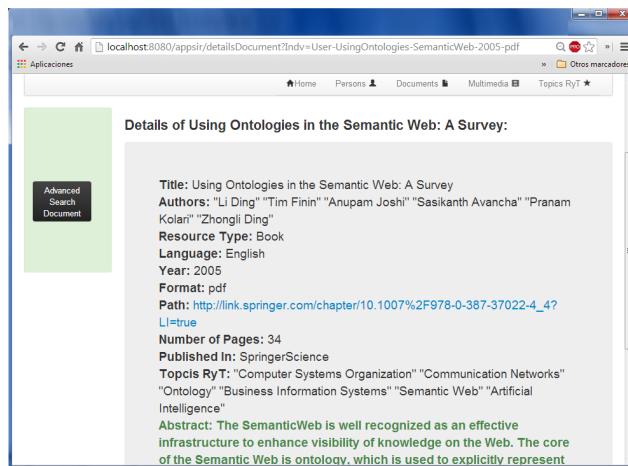
Flujo principal:

1. El usuario da clic en el botón *más detalles* (*More Details*).
2. La interfaz Web muestra todas *características significativas* del *recursos electo*.
 - 2.1. Si es un *recurso persona*, entonces, el sistema muestra el *nombre* (*Name*), *edad* (*Age*), *género* (*Gender*), *ocupación* (*Occupation*), *email* (*Email*), *página personal* (*WebSite*), *lugar de trabajo* (*Place of work*), *temas de investigación* (*Research on*), *idiomas de escritura* (*writes*), *idiomas de lectura* (*reads*), *idiomas de habla* (*speaks*), *relaciones profesionales* (*Knows*), *conocimientos en tópicos de RyT* (*Expertise*) y *competencias profesionales* (*Competent*), para la *persona electa*.
 - 2.2. Si es un *recurso documento*, entonces, el sistema muestra el *título* (*Title*), *autores* (*Authors*), *tipo recurso* (*Resource Type*), *idioma* (*Language*), *Año* (*Year*), *formato o extensión* (*Format*), *ruta* (*Path*), *número de páginas* (*Number of Pages*), *edición* (*Edition*), *organización que lo publico* (*Published In*), *los tópicos de RyT* (*Topcis RyT*) y el *resumen* (*Abstract*), para el *documento electo*.
 - 2.3. Si es un *recurso multimedia*, entonces, el sistema muestra el *título* (*Title*), *autores* (*Authors*), *idioma* (*Language*), *tipo recurso* (*Resource Type*), *Año* (*Year*), *formato o extensión* (*Format*), *ruta* (*Path*), *los tópicos de RyT* (*Topcis RyT*) y la *sinopsis* (*Synopsis*), para el *recurso multimedia electo*.

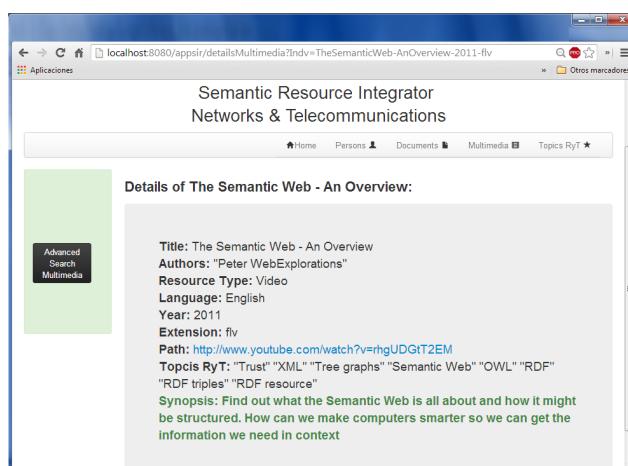
La Figura 6.12 presenta *tres interfaces* que muestran las *fichas descriptivas completas* para una persona, un documento y un recurso multimedia. La interfaz 6.12(a) presenta toda la información descriptiva de la profesora “Amelia Alles Estrada” que es un recurso del tipo persona. La interfaz 6.12(b) muestra la *información descriptiva* de libro “Using Ontologies in the Semantic Web: A Survey”. Mientras, la interfaz 6.12(c) muestra toda la información descriptiva del vídeo “The Semantic Web - An Overview”.



(a) Interfaz Web: mostrar detalles de la profesora *Amelia Alles Estrada*



(b) Interfaz Web: mostrar detalles del libro *Using Ontologies in the Semantic Web: A Survey*



(c) Interfaz Web: mostrar detalles del vídeo *The Semantic Web - An Overview*

Figura 6.12: Interfaces Web: mostrar detalles de un recurso de información.

Capítulo 7

Desarrollo experimental

La *integración semántica de los recursos* es el proceso de búsqueda y recuperación de información en un *grafo de conocimiento (ontología)*. Un *motor de búsqueda de triplets* es el mecanismo encargado de realizar la consulta de información en los grafos de conocimiento, para responder una consulta dada. Este *motor de triplets* generalmente pertenece a un triplestore. En esta tesis, se emplea el *triplestore Jena*. Éste proporciona dos componentes importantes: 1) *un motor de búsqueda* para triplets RDF, denominado **ARQ** y 2) *un motor de inferencia* que soporta los axiomas en nuestras ontologías, los cuales son descritos en la Sección 5.2.

7.1. Observaciones

La *calidad de los resultados* depende del uso o no de inferencia en nuestros modelos semánticos. Si Jena emplea un *modelo sin inferencia* como el ejemplo de la Figura 5.33, entonces el motor ARQ puede proporcionar todos, varios o ningún resultado. Esta variedad en la entrega de resultados, depende de las consultas SPARQL: 1) *consultas sobre las declaraciones de los recursos*, como la consulta de la Figura 5.27, 2) *consultas que agrupan varios patrones para un criterio de búsqueda*, como las consultas de las Figuras 5.28 y 5.30, y 3) *consultas simplificadas*, como en las Figuras 5.29 y 5.31. En contraste, Jena puede entregar mejores resultados cuando emplea un modelo que se obtiene de la inferencia en una ontología. Un ejemplo de este modelo se presenta en la Figura 5.35.

Una característica asociada al uso de inferencia, es el impacto en el *tiempo de procesamiento* de Jena para responder las consultas. Por un lado, se ha observado que el *tiempo de consulta* de Jena es pequeño, menor a medio segundo, cuando usa un *modelo sin inferencia*. Mientras, el *tiempo de consulta* de Jena para un *modelo con inferencia* es mayor en comparación con el tiempo del *modelo sin inferencia*.

7.2. Hipótesis

Con base en estas observaciones, nuestras dos hipótesis de experimentación son éstas:

1. *El triplestore Jena obtiene mejores resultados cuando utiliza nuestros modelos con inferencia.*
2. *El tiempo de consulta de Jena es mayor para nuestros modelos con inferencia en comparación con nuestros modelos sin inferencia.*

7.3. Experimentación

Esta experimentación consiste en la realización de dos actividades para probar nuestras hipótesis de experimentación. *La primera actividad es evaluar la calidad de los resultados de Jena para los modelos con y sin inferencia.* Esta evaluación consiste en estas etapas: 1) establecer una serie de consultas para interrogar nuestros modelos, 2) encontrar manualmente cuántos y cuáles recursos responden las consultas, 3) ejecutar las consultas con el motor ARQ de Jena y 4) comparar los recursos dados por Jena con las respuestas manuales.

La segunda actividad consiste en medir los tiempos promedio de procesamiento de Jena para las consultas de nuestra primera actividad. La finalidad de esta segunda actividad es comparar los *tiempos de consulta* para un modelo con inferencia y otro que no emplea ésta. En ambos modelos se evalúa el tiempo desde *la ejecución de Jena para una consulta a un modelo (con o sin inferencia)* hasta *la presentación de los resultados en pantalla*.

En esta tesis, el proceso de *integración semántica* está asociado a dos *casos de uso* (cartografía de competencias y búsqueda de recursos digitales). Ahora bien, nuestra experimentación consiste en probar la *calidad de los resultados* y el *tiempo de procesamiento* para la *integración semántica de recursos* en la *memoria corporativa* del área de Redes y Telecomunicaciones. Por esta razón, las dos actividades de nuestra experimentación deben ser aplicadas a nuestros dos *casos de uso*.

7.4. Sujetos de experimentación

Los sujetos de nuestra experimentación son un conjunto de personas, documentos y archivos multimedia que son generados artificialmente. Esta *generación artificial* consiste en el uso de scripts para: 1) *asignar un identificador URI para un conjunto de recursos de información ficticios* y 2) *generar triples RDF para estos recursos con base en las propiedades y clases de nuestras ontologías, así como datos aleatorios*.

Un script genera un conjunto de declaraciones para los recursos persona. Mientras, otro script genera las declaraciones para los documentos y archivo multimedia. El algoritmo 1 presenta el funcionamiento general de ambos scripts para la generación y almacenamiento de triples RDF.

El apéndice A presenta los dos algoritmos con el funcionamiento detallado de los scripts. Un algoritmo para los datos simulados de los recursos persona y el otro para los recursos digitales.

```

1  $N \leftarrow$  número de recursos ficticios de información a describir;
2  $modelo_{rdf} \leftarrow$  Crear un modelo rdf;
3 for  $i \leftarrow 1$  to  $N$  do
4    $\sigma_i \leftarrow$  Crear el recurso  $i$  y establecer un identificador URI para éste;
5   Elaborar los valores para cada característica significativa de este recurso ( $\sigma_i$ );
6   Escribir las aserciones, concatenando el URI del recurso ( $\sigma_i$ ), las propiedades de la
      ontología y los valores del paso 5;
7 end
8 Guardar el  $modelo_{rdf}$  en un archivo con extensión “rdf” y sintaxis de serialización
   Turtle;

```

Algorithm 1: Funcionamiento básico de scripts para la generación de triplets artificialmente.

La finalidad del uso de *información simulada* es tener rápidamente un volumen grande de datos en nuestros ABox. La *cantidad de información* en estos ABox debe ser realista con respecto al área de Redes y Telecomunicaciones (RyT). Ya que al tener información realista, nuestra experimentación se ajusta a la cantidad de datos que esperamos manejar en la integración semántica. Otra razón del uso de información simulada es ver si Jena soporta esta escala realista de datos (según los profesores del área RyT).

Las cantidades de *recursos persona* son 60 recursos artificiales y 13 recursos reales, dando un total de 73 personas. Mientras, las cantidades para los *recursos digitales* son 16 recursos reales y 1314 recurso simulados, un total de 1330 recursos digitales.

En nuestra experimentación, algunos *recursos persona* tienen la declaración que los asignan explícitamente a una de estas clases: *Estudiante* (*sirp:Student*), *Empleado* (*sird:Employee*) y *Profesor* (*sirp:Teacher*). Otros recursos persona carecen de esta asignación, pero con inferencia éstos pueden clasificarse en una o varias clases de la *ontología cartografía de competencias*.

En concreto, se tienen estas *cantidades de recursos* por clase: 51 recursos son profesionistas y 23 son *estudiantes*. Los 51 recursos persona mediante inferencia son asignados a la clase *Profesionista* (*sirp:Professional*). De estos 51 profesionistas se tiene que 19 son profesores y 9 son empleados. Por otro lado, de los 23 estudiantes se tiene que 9 recursos están asignados a la clase *Estudiante* y los 14 restantes por inferencia son asignados a la clase *Estudiante*. Existen 13 recursos persona que mediante inferencia se clasifican en la clase *Investigador* (*sirp:Researcher*).

La Tabla 7.1 muestra las cantidades de *recursos persona* por clases de la *ontología cartografía de competencias*. La *primera columna* presenta el nombre de las clases, la *segunda* el número de recursos que tienen la declaración que los asigna explícitamente a una clase y la *tercera columna* el número de recursos que por inferencia tienen la declaración para asignarlos a una clase.

De la misma manera que los *recursos persona*, algunos *recursos digitales* tienen la declaración que los asignan explícitamente a una clase. Otros recursos carecen de esta asignación, pero con inferencia éstos pueden clasificarse en una o varias clases de la *ontología búsqueda*

Clase	Número de Recursos	
	Asignación explícita	Asignación explícita y con inferencia
Persona	0	73
Investigador	0	13
Profesionista	0	51
Estudiante	11	23
Profesor	19	19
Empleado	9	9

Tabla 7.1: Número de recursos persona por clase.

de recursos digitales.

Los 1330 recursos digitales de nuestra experimentación se clasifican en: 156 artículos, 366 libros, 34 reportes técnicos, 146 páginas web, 73 tesis, 42 videos, 42 audios, 77 imágenes y 112 presentaciones. De los 156 artículos, 89 recursos tienen la declaración explícita a la clase *Artículo* (*sird:Paper*) y los restantes 67 mediante inferencia tienen la declaración a esta clase. En los libros, 185 recursos tienen la declaración explícita y 181 recursos mediante inferencia tienen la declaración a la clase *Libro* (*sird:Book*). De la misma manera, 79 recursos tienen la declaración explícita a la clase *Página Web* (*sird:PageWeb*) y 31 recursos a la clase *Tesis* (*sird:Thesis*). Mientras, 67 recursos mediante inferencia pertenecen a la clase *Página Web* y 42 a la clase *Tesis*. Por último, los 1330 recursos se clasifican en 815 *Documentos* (*sird:Document*) y 515 *Multimedia* (*sird:Multimedia*). Las declaraciones a estas dos clases se obtienen mediante inferencia.

La Tabla 7.2 presenta las cantidades de recursos digitales por clases de la *ontología recursos digitales*. Esta Tabla 7.2 presenta la misma estructura de la Tabla 7.1.

7.5. Metodología

Esta experimentación se ha dividido en dos actividades. La primera actividad consiste en la *evaluación de la calidad de los resultados* que proporciona el triplestore Jena. Mientras, la segunda actividad consiste en medir los *tiempos promedios* que toma Jena para: consultar los modelos con/sin inferencia y mostrar los resultados al usuario.

El *primer paso* en la *evaluación de la calidad de resultados* es identificar un conjunto básico de consultas para interrogar las ontologías *cartografía de competencias* y *búsqueda de recursos digitales*. En la Sección 5.3 se hizo un análisis e identificación de preguntas base que posteriormente se transformaron a consultas SPARQL. Estas consultas de la Sección 5.3 son reutilizadas para la primer actividad de esta experimentación.

El *segundo paso* es encontrar cuántos y cuáles recursos son relevantes para responder las

Clase	Número de Recursos	
	Asignación explícita	Asignación explícita y con inferencia
Recurso Digital	0	1330
Documento	0	815
Artículo	89	156
Reporte Técnico	34	34
Página Web	79	146
Tesis	31	73
Libro	185	366
Multimedia	0	515
Presentación	112	112
Audio	42	42
Vídeo	42	42
Imagen	77	77

Tabla 7.2: Número de recursos digitales por clase.

consultas de nuestra experimentación. Para ello, se hace una *búsqueda manual* exhaustiva de los recursos relevantes que responden las consultas de experimentación. Esta búsqueda se hace sobre los recursos de nuestra memoria corporativa que está guiada por los casos de uso. Para cada consulta, se anotan los identificadores URI de los *recursos relevantes* y la cantidad de éstos.

La Tabla 7.3 muestra las preguntas base para la *cartografía de competencias*. En esta Tabla, la primera columna presenta el identificador para cada una de las diecinueve pregunta, la segunda columna enuncia la pregunta y la tercera columna presenta el número de recursos que responden a ésta. De la misma manera, la Tabla 7.4 presenta un identificador, las preguntas y cantidad de recursos que responden a éstas, para la *búsqueda de recursos digitales*. En estas dos Tablas, no se presentan los identificadores URI de los recursos que responden a las preguntas. Porque algunas consultas tienen muchos recursos relevantes.

El *tercer paso* es emplear *Jena* para ejecutar las *consultas SPARQL* a un modelo sin inferencia y otro con inferencia. La finalidad de ésto es ver la calidad de resultados y la cantidad de éstos que al emplear un modelo con inferencia y otro sin inferencia. De esta manera, podemos ver el impacto de la inferencia para la consulta de información en una ontología.

En esta *ejecución de consultas*, se emplea un script para consultar un *modelo semántico* y proporcionar los resultados que están asociados a una consulta dada. El funcionamiento de este script se presenta en el Algoritmo 2. En este algoritmo, los **parámetros de entrada** son: 1) la consulta SPARQL, 2) la ontología (*Personas* o *RecDigi*), 3) inferencia (0 sin inferencia

Id. Consulta	Pregunta	No. de Recursos
Q1.1	¿Cuáles el nombre, correo, sitio web, género y edad de las personas del área de RyT?	73
Q1.2	¿Cuál es el nombre, sitio web y el lugar donde laboran las personas del RyT?	73
Q1.3	¿Quiénes son mayores de 20 años y menores de 45 años?	50
Q1.4	¿Quiénes son profesionistas del área de RyT?	51
Q1.5	¿Quiénes trabajan en la Clark & Parsia y son del sexo Masculino?	3
Q1.6	¿Quiénes son estudiantes y leen en inglés?	8
Q1.7	¿Quienes hablan, leen y escriben en inglés?	16
Q1.8	¿Qué estudiantes saben algo de inglés?	6
Q1.9	¿Qué profesores tienen la capacidad de síntesis?	2
Q1.10	¿Qué profesionistas tienen conocimiento en los temas de Web Semántica?	58
Q1.11	¿Qué profesores tienen conocimientos en Sistemas Distribuidos?	3
Q1.12	¿Quiénes tienen conocimiento en Java, OWL, RDF, Threads, C, OpenMP?	1
Q1.13	¿Qué estudiantes tienen algún conocimiento en los subtemas de Sistemas Operativos?	33
Q1.14	¿Quiénes trabajan en una Universidad?	24
Q1.15	¿Quienes laboran en la UAM y tienen algún conocimiento en Web Semántica?	19
Q1.16	¿Qué personas tienen como asesor a Carolina Medina?	2
Q1.17	¿Quiénes son los colegas de Ricardo Marcelin?	8
Q1.18	¿Quiénes conocen a Carolina Medina Ramirez?	11
Q1.19	¿Qué personas son profesores-investigadores?	9

Tabla 7.3: Preguntas y cantidad de personas que responden a éstas.

Id. Consulta	Pregunta	No. de Recursos
Q2.1	¿Cuáles son los títulos, rutas, extensión, idioma de todos los recursos digitales de RyT?	1330
Q2.2	¿Cuáles libros tratan sobre algunos temas de Sistemas Distribuidos?	103
Q2.3	¿Qué recursos fueron publicados por la UAM?	18
Q2.4	¿Qué documentos son para dar un curso de Sistemas P2P?	31
Q2.5	¿Qué recursos multimedia son mayores al año 2009?	119
Q2.6	¿Cuáles documentos tratan sobre Ontologías?	30
Q2.7	¿Qué recursos fueron publicados en una Revista científica?	156
Q2.8	¿Qué recursos tienen en su contenido las palabras "linked data"?	159
Q2.9	¿Cuáles documentos en inglés y mayores al año 2000 son de autoría de Erik Alarcón Zamora?	2
Q2.10	¿Cuáles la tesis de Samuel Hernández Maza?	4

Tabla 7.4: Preguntas y cantidad de recursos digitales que responden a éstas.

y 1 con inferencia). Mientras, los parámetros de salida son: 1) los identificadores URI que responden la consulta y 2) la cantidad de recursos respuesta. Este script permite ejecutar una consulta a la vez por corrida e imprime los parámetros de salida en pantalla.

Input: Consulta SPARQL: $query$, Nombre del modelo: $model$, Uso de inferencia: $inference$

Output: Resultados: Π_{query} , Número de resultados: $noRes$

```

1  $noRes \leftarrow 0, \Pi_{query} \leftarrow \{\}$ ;
2 if  $model \equiv 'Personas'$  then
3   | Cargar en memoria el ABox y TBox de la ontología cartografía de competencias;
4 else if  $model \equiv 'RecDigi'$  then
5   | Cargar en memoria el ABox y TBox de la ontología búsqueda de recursos digitales;
6 end
7 if  $inference \equiv 0$  then
8   |  $modelo_{semantico} \leftarrow$  modelo ABox y TBox;
9 else if  $inference \equiv 1$  then
10  | Inferir en memoria sobre el ABox y TBox;
11  |  $modelo_{semantico} \leftarrow$  modelo inferido;
12 end
13 Cargar en memoria consulta SPARQL ( $query$ );
14 Ejecutar consulta ( $query$ ) en el modelo ( $modelo_{semantico}$ );
15  $\Pi_{query} \leftarrow$  Recuperar resultados de consulta;           /*  $\Pi_{query} = \{\pi_1, \dots, \pi_n\}$  */
16 forall the elements of  $\Pi_{query}$  do
17   | Imprimir  $\pi_k$  en pantalla;
18   |  $noRes \leftarrow noRes + 1$ ;
19 end
20 Imprimir  $noRes$  en pantalla;
```

Algorithm 2: Algoritmo para la evaluación de la calidad de resultados

El cuarto paso es comparar los resultados que proporciona el script (Algoritmo 2) con los resultados que sabemos responden (búsqueda manual) las consultas de las Tablas 7.3 y 7.4. La finalidad de este paso, es evaluar la calidad de resultados de Jena, para un modelo con inferencia y otro sin el uso de ésta. El cuarto paso, se hace en dos fases: 1) encontrar los recursos relevantes de las respuestas dadas por el script, y 2) usar (métricas de recuperación de la información) para comparar la calidad de los resultados con y sin inferencia.

La fase uno consiste en comparar y anotar las cantidades de recursos relevantes para cada consulta. En esta fase se considera el modelo de uso: con inferencia o sin ésta. Esta comparación y anotación se hizo de forma manual. El primer criterio de comparación es ver la cantidad de recursos de las Tablas 7.3 y 7.4 con el número de respuestas que proporciona el script. Posteriormente, se hace la comparación identificador a identificador de los recursos que son proporcionados por el script con los recursos que se saben responden a la consulta.

La segunda fase, consiste en emplear dos métricas de recuperación de la información (RI):

- La **exhaustividad (recall)** [52] es “la cantidad de elementos relevantes que han sido recuperados, entre la cantidad de elementos relevantes en la base de conocimientos” [53]. La exhaustividad se presenta en la Ecuación 7.1.

$$\text{Exhaustividad} = \frac{\text{Número de recursos relevantes que han sido recuperados}}{\text{Número de recursos relevantes en la base de conocimientos}} * 1,0 \quad (7.1)$$

- La **precisión (Precision)** [52] es la “cantidad de elementos recuperados que son relevantes, entre el total de elementos recuperados” [53]. La precisión se presenta en la Ecuación 7.2.

$$\text{Precisin} = \frac{\text{Número de elementos relevantes que han sido recuperados}}{\text{Total de elementos recuperados en la base de conocimientos}} * 1,0 \quad (7.2)$$

En las ecuaciones de la exhaustividad y precisión, los divisores y dividendos son elementos que se han encontrado en esta primer actividad. En concreto, éstos son los elementos en las dos ecuaciones:

- *Exhaustividad*: La cantidad de elementos relevantes recuperados es el número de recursos relevantes que proporciona el script para una consulta; esta información se obtiene en el etapa uno de este cuarto paso. Mientras, el número de elementos relevantes en la base de conocimiento es la cantidad de resultados de nuestra búsqueda manual que están en las Tablas 7.3 y 7.4.
- *Precisión*: La cantidad de elementos relevantes recuperados es el número de recursos relevantes que proporciona el script para una consulta. Mientras, el total de elementos recuperados en la base de conocimiento es la cantidad de recursos dados por el script, sin el análisis de los recursos relevantes.

El objetivo de emplear estas dos métricas, es evaluar las capacidades de recuperación de la información del triplestadores Jena. En particular, la *exhaustividad* evalúa la habilidad de Jena para encontrar los recursos relevantes en una ontología, mientras la *precisión* evalúa la habilidad de Jena para filtrar los recursos relevantes [54].

En esta experimentación, la **segunda actividad** es tomar los tiempos de procesamiento para la consulta y recuperación de información en un modelo semántico a partir del uso de Jena. El objetivo de esta actividad es comparar estos tiempos de consulta para un modelo con inferencia y otro sin el uso de ésta.

Los tiempos importantes en esta actividad son la medición desde la *consulta de información en un modelo semántico con Jena* hasta la *impresión de los resultados en pantalla*. Esta medición de los tiempos, se hace con base en la modificación del script del Algoritmo 2. Porque este Algoritmo 2 lleva a cabo todo el proceso de búsqueda y recuperación de la

información relevante en los *recursos de información*. También, este Algoritmo permite elegir el modelo: si es cero, entonces no hay inferencia y si es uno, se emplea inferencia.

El funcionamiento del *script modificado* es tomar ‘ n ’ veces el *tiempo de procesamiento* cuando se consulta un modelo semántico (con o sin inferencia). Posteriormente, este script calcula y regresa el *tiempo promedio de consulta* para una *consulta SPARQL* dada. El Algoritmo 3 presenta el funcionamiento de este script modificado.

Input: Consulta SPARQL: $query$, Nombre del modelo: $model$, Uso de inferencia: $inference$, Número de repeticiones: N

Output: Tiempo promedio: $Tiempo_{promedio}$

```

1  $\Pi_{query} \leftarrow \{\}$ ;
2  $t_{ini} \leftarrow 0, t_{fin} \leftarrow 0, Tiempo_{promedio} \leftarrow 0$ ;
3 for  $i \leftarrow 1$  to  $N$  do
4   if  $model \equiv 'Personas'$  then
5     Cargar en memoria el ABox y TBox de la ontología cartografía de
      competencias;
6   else if  $model \equiv 'RecDigi'$  then
7     Cargar en memoria el ABox y TBox de la ontología búsqueda de recursos
      digitales;
8   end
9   if  $inference \equiv 0$  then
10     $modelo_{semantico} \leftarrow$  modelo ABox y TBox;
11   else if  $inference \equiv 1$  then
12    Inferir en memoria sobre el ABox y Tbox;
13     $modelo_{semantico} \leftarrow$  modelo inferido;
14   end
15   Cargar en memoria consulta SPARQL ( $query$ );
16    $t_{ini} \leftarrow \text{time}()$ ;
17   Ejecutar consulta ( $query$ ) en el modelo ( $modelo_{semantico}$ );
18    $\Pi_{query} \leftarrow$  Recuperar resultados de consulta;           /*  $\Pi_{query} = \{\pi_1, \dots, \pi_n\}$  */
19   forall the elements of  $\Pi_{query}$  do
20     Imprimir  $\pi_k$  en pantalla;
21   end
22    $t_{fin} \leftarrow \text{time}()$ ;
23    $Tiempo_{promedio} \leftarrow Tiempo_{promedio} + (t_{fin} - t_{ini})$ ;
24 end
25 return  $Tiempo_{promedio}$ ;
```

Algorithm 3: Algoritmo para medir el tiempo promedio en el proceso de consulta.

Este script (Algoritmo 3) se ejecuta para cada una de las consultas que están asociadas a las preguntas en las Tablas 7.3 y 7.4. Porque, este script permite procesar una consulta a la vez. El número de repeticiones se establece en 20 ($N = 20$). Mientras, las modelos

empleados son las ontologías de *cartografía de competencias y búsqueda de recursos digitales*. El parámetro importante es el uso o no de inferencia para el modelo semántico.

Los scripts asociados a los Algoritmos 2 y 3, se corrieron en una computadora con un procesador Intel Core I7 a 2.3GHz con 8Gb en RAM y 8 núcleos de procesamiento. La versión del Apache Jena es la 2.7.4 para Windows 7 (64bits). Estos scripts se construyeron con el IDE Eclipse y se ejecutaron con la versión 1.7.0_05 de Java.

7.6. Resultados

Los resultados para la *evaluación de la calidad de los resultados* se presentan en dos etapas. Estas dos etapas tienen relación con las dos fases del cuarto paso de esta evaluación. La primera etapa presenta la cantidad de resultados (relevantes y el total) que fueron recuperados por el script del Algoritmo 2. La segunda etapa presenta los valores calculados para la exhaustividad y precisión. Este cálculo se hace para cada consulta de la experimentación.

Id. Consulta	Modelo sin inferencia		Modelo con inferencia	
	Recursos relevantes	Total recursos recuperados	Recursos relevantes	Total recursos recuperados
Q1.1	73	73	73	73
Q1.2	73	73	73	73
Q1.3	50	50	50	50
Q1.4	28	28	51	51
Q1.5	3	3	3	3
Q1.6	4	4	8	8
Q1.7	16	16	16	16
Q1.8	2	2	6	6
Q1.9	10	10	10	10
Q1.10	2	2	58	58
Q1.11	3	3	3	3
Q1.12	1	1	1	1
Q1.13	6	6	20	20
Q1.14	16	16	24	24
Q1.15	2	2	2	2
Q1.16	2	2	2	2
Q1.17	2	2	8	8
Q1.18	7	7	11	11
Q1.19	0	0	13	13

Tabla 7.5: Recursos relevantes y no relevantes asociados a las preguntas de la cartografía de competencias.

En la Tabla 7.5 se presentan las cantidades de recursos relevantes y totales de la primera etapa, para las preguntas de la Tabla 7.3. La primer columna presenta el identificador de consulta. La segunda y tercera columna muestra la cantidad de recursos relevantes y total recursos recuperados cuando se emplea un modelo sin inferencia. La cuarta y quinta presentan los recursos relevantes y total de recursos recuperados para un modelo con inferencia.

De la misma manera, la Tabla 7.6 presenta la cantidad de recursos relevantes y total de recursos recuperados cuando se consulta el modelo sin inferencia y con inferencia de los recursos digitales. En esta Tabla, las consultas están asociadas a las preguntas de la Tabla 7.4. La estructura de la Tabla 7.6 es igual a la Tabla 7.5.

Id. Consulta	Modelo sin inferencia		Modelo con inferencia	
	Recursos relevantes	Total recursos recuperados	Recursos relevantes	Total recursos recuperados
Q2.1	1330	1330	1330	1330
Q2.2	0	0	103	103
Q2.3	18	18	18	18
Q2.4	15	15	31	31
Q2.5	66	66	119	119
Q2.6	15	15	30	30
Q2.7	156	156	156	156
Q2.8	159	159	159	159
Q2.9	0	0	2	2
Q2.10	3	3	4	4

Tabla 7.6: Recursos relevantes y no relevantes asociados a las preguntas de la búsqueda de recursos digitales.

Los resultados para la segunda etapa de la *evaluación de la calidad de resultados*, son los valores que fueron calculados para la exhaustividad (Ecuación 7.1) y precisión (Ecuaciones 7.2) de Jena con nuestras ontologías. El objetivo de esta *segunda etapa* es comparar estas dos métricas de *recuperación de la información* para un modelo con inferencia y otro que no usa ésta, donde estos modelos son procesados por Jena.

La Tabla 7.7 muestra los valores de exhaustividad y precisión para las consultas en la Tabla 7.3. En esta Tabla 7.7, la primer columna muestra el identificador de la consulta. La segunda y tercera columna muestran los valores de exhaustividad y precisión para un modelo sin inferencia. La cuarta y quinta columna presentan los valores de las mismas métricas para un modelo con inferencia.

De la misma manera, la Tabla 7.8 presenta los valores de estas dos métricas de RI, para las preguntas de la Tabla 7.4. Estos valores están asociados al modelo de *búsqueda de recursos digitales*. La estructura de esta Tabla 7.8 es la misma de la Tabla 7.7.

Las Tablas 7.9 y 7.10 presentan los resultados de *evaluar los tiempos promedio de procesamiento*. La Tabla Tablas 7.9 muestra los tiempos promedio de consulta para las preguntas

Id. Consulta	Modelo sin inferencia		Modelo con inferencia	
	Exhaustividad	Precisión	Exhaustividad	Precisión
Q1.1	1	1	1	1
Q1.2	1	1	1	1
Q1.3	1	1	1	1
Q1.4	0.549	1	1	1
Q1.5	1	1	1	1
Q1.6	0.5	1	1	1
Q1.7	1	1	1	1
Q1.8	0.333	1	1	1
Q1.9	1	1	1	1
Q1.10	0.034	1	1	1
Q1.11	1	1	1	1
Q1.12	1	1	1	1
Q1.13	0.3	1	1	1
Q1.14	0.666	1	1	1
Q1.15	1	1	1	1
Q1.16	1	1	1	1
Q1.17	0.25	1	1	1
Q1.18	0.636	1	1	1
Q1.19	0	-	1	1

Tabla 7.7: Exhaustividad y precisión para las consultas de la cartografía de competencias.

Id. Consulta	Modelo sin inferencia		Modelo con inferencia	
	Exhaustividad	Precisión	Exhaustividad	Precisión
Q2.1	1	1	1	1
Q2.2	0	-	1	1
Q2.3	1	1	1	1
Q2.4	0.484	1	1	1
Q2.5	0.555	1	1	1
Q2.6	0.5	1	1	1
Q2.7	1	1	1	1
Q2.8	1	1	1	1
Q2.9	0	-	1	1
Q2.10	0.75	1	1	1

Tabla 7.8: Exhaustividad y precisión para las consultas de la búsqueda de recursos digitales.

en la Tabla 7.3. De la misma manera, la Tabla 7.10 muestra los tiempos para las preguntas en la Tabla 7.4. El objetivo de esta evaluación es comparar los tiempos al consultar un modelo sin inferencia y otro con inferencia.

La estructura es la misma para las Tablas 7.9 y 7.10. Ésta es la estructura de estas Tablas: la primer columna muestra el identificador de la consulta, la segunda el tiempo promedio al consultar un modelo sin inferencia y la tercera columna el tiempo promedio de consulta en un modelo obtenido por inferencia.

Id. Consulta	Tiempo promedio (milisegundos)	
	Modelo sin inferencia	Modelo con inferencia
Q1.1	13	133
Q1.2	17	152
Q1.3	12	138
Q1.4	10	194
Q1.5	7	140
Q1.6	7	125
Q1.7	8	125
Q1.8	9	126
Q1.9	6	123
Q1.10	6	136
Q1.11	7	124
Q1.12	6	125
Q1.13	8	406
Q1.14	8	128
Q1.15	6	126
Q1.16	9	127
Q1.17	25	125
Q1.18	28	129
Q1.19	7	157

Tabla 7.9: Tiempo promedio de procesamiento para las consultas de la cartografía de competencias.

El análisis de las Tablas 7.5, 7.6, 7.9 y 7.10 arroja varios aspectos significativo con relación a los *recursos relevantes recuperados*. En el caso del uso de modelos sin inferencia se tiene los siguientes hechos. 1) En algunos casos, una consulta al conocimiento explícito recupera todos los recursos relevantes y los tiempos de respuesta son pequeños (no pasan del segundo). 2) En otros casos, algunas consultas descartan varios recursos que si responden la consulta. Esto se debe a que algunos recursos carecen un determinado triple.

En el caso de uso de inferencia, Jena permite recuperar todos los recursos esperados, porque el motor de inferencia permite generar algunos triples en *nuestras ontologías*. Estos

Id. Consulta	Tiempo promedio (milisegundos)	
	Modelo sin inferencia	Modelo con inferencia
Q2.1	24	3520
Q2.2	9	4016
Q2.3	12	3520
Q2.4	16	3472
Q2.5	42	3451
Q2.6	14	3392
Q2.7	13	3431
Q2.8	32	3312
Q2.9	34	3570
Q2.10	11	3398

Tabla 7.10: Tiempo promedio de procesamiento para las consultas de la búsqueda de recursos digitales.

triples inferidos son criterios de búsqueda las consultas de nuestra experimentación. Pero, el uso de inferencia consume mayor tiempo, porque el *motor de búsqueda* invierte mayor tiempo en consultar más tripletas de nuestros grafos RDF.

Capítulo 8

Conclusiones y Trabajo Futuro

En este trabajo presentamos un estudio sobre el uso de las *tecnologías semánticas* (TS) para realizar el proceso de *integración semántica de recursos* en una *memoria corporativa*. Esta *integración semántica de recursos* es el proceso de *búsqueda y recuperación* significativa de información existente en los *recursos de información*. Estos *recursos de información*, o bien son documentos, archivos o bases de datos residentes en algún medio de almacenamiento, o bien personas que andan de un lugar a otro.

Este trabajo tiene dos finalidades. Por un lado, construir un *modelo semántico* (ontología) que representa el *conocimiento acerca de los recursos de información en una memoria corporativa* y el *conocimiento sobre el dominio de ésta*. Por otro lado, hacer la *búsqueda y recuperación inteligente de información* para responder las preguntas de los usuarios.

Nuestro trabajo propone un *marco de trabajo* para lograr la *integración semántica de recursos* en una *memoria corporativa*. Esta propuesta comprende tres actividades. La primera actividad es construir un modelo para representar el *conocimiento explícito* sobre los *recursos de información* en una memoria corporativa. La segunda actividad es introducir *reglas de inferencia* para enriquecer al modelo con conocimiento implícito de la memoria corporativa. La tercera actividad es *buscar y recuperar información* acerca de los recursos en el modelo, con el fin de responder las preguntas de los usuarios de la memoria.

Nosotros utilizamos a las tecnologías semánticas, porque éstas proporcionan varias ventajas en la gestión del conocimiento. A continuación se listan las ventajas, las cuales observamos durante y al finalizar el desarrollo e implantación de nuestro *marco de trabajo*.

- El *marco RDF* permite representar cualquier recurso, sea un ser vivo, un archivo digital, un país, ciudad, un edificio, una organización o un concepto abstracto. De esta manera, se supera la integración de información de recursos que son heterogéneos en formato, contenido y estructura.
- El *marco RDF* establece que los recursos tengan un URI, para identificar de manera única a un recurso. De esta forma, se solucionan problemas de homonimia con respecto a los nombres de los recursos.
- En el *marco RDF*, la representación del conocimiento en los recursos no se limita a un pequeño número de características sobre éstos; sino que cualquier información, sea una relación o una característica, puede describirse para un *recurso de información*.

- La introducción de *reglas de inferencia* enriquezcan el conocimiento en un modelo. Porque un razonador a partir de estas reglas puede hacer explícito el conocimiento implícito. Este comportamiento se dio en nuestras ontologías.
- Una *ontología* puede ser extendida con el fin de *incorporara mayor conocimiento o adaptarse a cambios en el conocimiento*. De esta manera, si nuevos recursos se describen, entonces éstos pueden incorporarse a la ontología. De la misma manera, si cambia el conocimiento en algunas ramas de la ontología, entonces éstas pueden ser sustituidas por otras.
- Las *ontologías* están elaboradas en un formato estándar y escritas con vocabularios consensuados. Por esta razón, éstas se pueden intercambiar y reutilizar entre personas o aplicaciones.
- Una *ontología* puede representarse gráficamente como un grafo dirigido. El objetivo es tener una *estructura gráfica* para visualizar las descripciones y reglas de inferencia en forma de tripleta.
- Una *ontología* fomenta el desarrollo de aplicaciones genéricas para aprovechar estos modelos. Ejemplos de estas aplicaciones son editores de ontologías, interfaces gráficas para describir semánticamente los recursos, navegadores sobre ontologías, por mencionar algunas. Estas herramientas genéricas posibilitan que personas expertas en el dominio sean las principales constructoras del grafo de conocimiento. De esta manera, la información en el grafo será confiable, ya que estas personas son las que tienen los conocimientos en el dominio.

La tarea de búsqueda y recuperación de información son actividades que no cualquier usuario puede llevar a cabo. Porque es necesario que un usuario tenga conocimientos en las tecnologías semánticas y los vocabularios en nuestras ontologías. Por esta razón, nosotros proponemos una aplicación que permita la interacción de cualquier usuario con el modelo, para fines de búsqueda de información.

Nuestra interfaz está orientada al usuario, es decir, un ambiente visual que permite a éstos una manera agradable de consultar la información en nuestras ontologías. Esta característica es importante, porque en los trabajos de nuestro *estado del arte* (Ver Sección 4.1), ninguna aplicación está orientada al usuario.

Nosotros llevamos a cabo una evaluación experimental. Los objetivos de esta evaluación son: 1) *evaluar la calidad de los resultados de Jena para los modelos con y sin inferencia*, así como 2) *medir los tiempos promedio de procesamiento de Jena para las consultas de nuestra primera actividad*. En la medición de tiempos, queríamos ver el rendimiento de Jena con la cantidad de datos que esperamos manejar. Mientras, en la evaluación de resultados, queríamos ver si los resultados devueltos por Jena, son los que responden nuestras preguntas.

Los resultados en nuestra experimentación indican lo siguiente. El *tiempo de consulta* es mayor en un *modelo con inferencia* en comparación de un *modelo sin inferencia*. Sin embargo, los resultados de las búsquedas en un modelo sin inferencia son variados. En algunos casos,

todos los *resultados* que se sabe responden la consulta, son recuperados cuando la consulta es sobre el conocimiento explícito. En otros casos, varios recursos que responden la consulta, son descartados, porque algunos recursos carecen de un determinado triple. En cambio, los resultados de búsquedas sobre un *modelo con inferencia* son buenos, porque estas consultas recuperan todos los recursos que de antemano sabemos que responden a las consultas.

En este trabajo presentamos nuestro *marco de trabajo*, un prototipo para la consulta de información, los resultados en nuestra evaluación experimental, así como las ventajas de las tecnologías semánticas, y con base en éstos, nosotros hacemos esta conclusión. **El uso de tecnologías semánticas es una buena aproximación para lograr la integración semántica de recursos en una memoria corporativa.**

En la realización de este trabajo, nosotros identificamos varias actividades que establecimos como trabajos futuros. Estos trabajos se listan a continuación.

- Introducir más *casos de uso*, por ejemplo, administración del inventario en los distintos laboratorios que componen el departamento de IE de la UAM
- Actualizar los *recursos de información* en la *memoria corporativa* del área de RyT y describir a estos recursos con el *marco RDF*.
- Construir una interfaz gráfica de usuario para que éste pueda describir los *recursos de información*. El objetivo es que los usuarios empleen esta aplicación como una aplicación de red social. Algunas funcionalidades de esta interfaz son:
 - Describir un nuevo *recurso de información*.
 - Asignar un URI al nuevo recurso
 - Construir un conjunto de tripletas RDF, asociadas al nuevo recurso.
 - Guardar las tripletas en un archivo o agregar éstas al modelo inmediatamente.
 - Actualizar la información de un *recurso de información*.
 - Ver la información vieja del recurso.
 - Corregir la información en el modelo.
 - Guardar el nuevo modelo.
 - Eliminar las descripciones de un *recurso de información*.
 - Ver la información del recurso a eliminar.
 - Eliminar las tripletas en el modelo, donde el recurso es el sujeto.
 - Guardar el nuevo modelo.
- Evaluar la introducción de nuevos axiomas en las ontologías, los cuales sean más complejos.
- Utiliza el triplestore Stardog.
 - Cargar, inferir y consultar nuestras ontologías con Stardog.

- Hacer nuestra evaluación experimental con Stardog.
- Comparar el desempeño y resultados de experimento de Stardog contra Jena.
- Dar mantenimiento a la ontología del vocabulario de Redes y Telecomunicaciones.
 - Introducir nuevos conceptos, como radios cognitivos, computación en la nube, computación móvil, por mencionar algunos.
 - Introducir tripletas para indicar la traducción de un tema otros idiomas, por ejemplo, semantic web, web semántica, web semântica, semantisch web, anlamsal web, web semantikoa.
- Mejorar el prototipo de integración semántica de recursos.
 - Mejorar la seguridad en el prototipo.
 - Mejorar la visualización de la información en las interfaces.
 - Agregar un cuadro para hacer búsquedas por palabras clave, mediante Lucene Index¹.
 - Optimizar los *tiempos de procesamiento* en la búsqueda y despliegue de información.

¹The Apache Software Foundation, “adding free text searches to SPARQL,” Disponible en: <http://jena.apache.org/documentation/larq/>

Appendices

Apéndice A

Algoritmos para la generación de datos simulados

Algoritmo para generar datos artificialmente de los recursos persona

1. $N \leftarrow$ número de personas a describir;
2. Tópicos de Redes y Telecomunicaciones $TopRyT \leftarrow \{\theta_1 \dots \theta_k\}$;
3. **for** $i \leftarrow 1$ **to** N **do**
 - 3.1. Elegir el nombre, apellido paterno y apellido materno de unas listas predefinidas;
 - 3.2. Concatenar el nombre y apellidos electos, donde cada palabra sea separada por un espacio en blanco;
 - 3.3. Guardar este nombre en una lista tipo cola ($Names$);
4. Lista de Nombres $Names \leftarrow \{nombre_1 \dots nombre_N\}$;
5. $modelo_{rdf} \leftarrow$ Crear un modelo rdf para los recursos persona;
6. **for** $i \leftarrow 1$ **to** N **do**
 - 6.1. $\sigma_i \leftarrow$ crear el recurso i y establecer un identificador URI para éste;
 - 6.1.1. Seleccionar el i-ésimo nombre de la lista de nombres ($nombre_i$);
 - 6.1.2. Quitar los espacios en blanco de esta cadena;
 - 6.1.3. Establecer esta cadena como identificador del recurso;
 - 6.2. Para cada característica significativa de una persona elaborar las literales y elegir los objetos de ésta;
 - 6.2.1. Literal $nombre \leftarrow$ seleccionar el i-ésimo nombre de la lista de nombres ($nombre_i$);
 - 6.2.2. Objeto $Género \leftarrow$ elegir el sexo de la persona a partir del nombre;
 - 6.2.3. Literal $Año \leftarrow$ establecer el año aleatoriamente en un intervalo del 2000 al 2013;

- 6.2.4. Objeto *Ocupación* \leftarrow elegir una ocupación para la persona de las tres posibles (Estudiante, Profesor o Empleado);
 */** Establecer la edad a partir de la ocupación de la persona **/*
 - 6.2.5. **if** (*Ocupación* \equiv ‘Empleado’ || *Ocupación* \equiv ‘Profesor’) **then**
 Literal *Edad* \leftarrow un número aleatorio en el intervalo de 27 a 57;
 - 6.2.6. **else if** (*Ocupación* \equiv ‘Estudiante’) **then**
 Edad \leftarrow un número aleatorio en el intervalo de 18 a 28;
 */** Elegir una organización de un listado a partir de la ocupación de la persona **/*
 - 6.2.7. **if** (*Ocupación* \equiv ‘Estudiante’ || *Ocupación* \equiv ‘Profesor’) **then**
 Literal *Organización* \leftarrow se elige una universidad de un listado preestablecido de universidades;
 - 6.2.8. **else if** (*Ocupación* \equiv ‘Empleado’) **then**
 Organización \leftarrow se elige una organización de una lista preestablecida de organizaciones;
 - 6.2.9. Literal *Email* \leftarrow construir el email a partir del nombre de la persona (cambiando los espacios en blanco por guiones bajos), la organización donde labora y otras palabras especiales;
 - 6.2.10. Literal *Sitio Web* \leftarrow construir el URL concatenando el nombre (cambiando espacios en blanco por guiones), la organización donde labora, la ocupación, una extensión de páginas web y otras palabras especiales;
 */** Elegir otros individuos que se relacionan con esta persona a partir de la ocupación de la persona **/*
 - 6.2.11. **if** (*Ocupación* \equiv ‘Estudiante’) **then**
 Conocidos \leftarrow se eligen dos nombres aleatorios de la lista nombres, cuya ocupación de éstos sea Empleado o Profesor;
 Quitar los espacios en blanco de los nombres;
 Guardar estas cadenas en una lista de conocidos;
 - 6.2.12. **else if** (*Ocupación* \equiv ‘Estudiante’ || *Ocupación* \equiv ‘Profesor’) **then**
 Conocidos \leftarrow se eligen dos nombres a cinco nombre aleatorios de la lista nombres, cuya ocupación de éstos sea Empleado o Profesor;
 Quitar los espacios en blanco de los nombres;
 Guardar estas cadenas en una lista de conocidos;
 - 6.2.13. Objeto *Habilidades lingüísticas* \leftarrow se eligen de forma aleatoria de 1 a 3 idiomas de una lista preestablecida de idiomas;
 - 6.2.14. Objeto *Conocimientos de RyT* \leftarrow se eligen de manera aleatoria de 5 a 7 Tópicos de Redes y Telecomunicaciones (*TopRyT*);
 - 6.2.15. Objeto *Competencias profesionales* \leftarrow se eligen de forma aleatoria de 3 a 4 competencias de un listado preestablecido de competencias;
-

- 6.3. Para cada característica significativa de una persona construir sus aserciones respectivas.

Apéndice B

Transformación de preguntas a consultas SPARQL

```
1 PREFIX sirp: <http://arte.izt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.izt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT ?name ?mail ?ws ?gender ?age
5 WHERE {
6     ?x sirp:has-name ?name;
7     sirp:has-email ?mail;
8     sirp:has-webSite ?ws;
9     sirp:has-gender ?gender.
10    OPTIONAL {?x sirp:has-age ?age}
11 }
```

Algorithm 4: Consulta SPARQL asociada a la pregunta *Q1.1* ¿Cuáles son los nombres, correos, sitios web, géneros y edades de las personas del área de RyT.

```
1 PREFIX sirp: <http://arte.izt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.izt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws ?work
5 WHERE {
6     ?x sirp:has-name ?name;
7     sirp:has-webSite ?ws;
8     sirp:has-workplace ?work.
9 }
```

Algorithm 5: Consulta SPARQL asociada a la pregunta *Q1.2* ¿Cuáles son los nombres, sitios web y los lugares donde laboran las personas del RyT?

```

1 PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT ?name ?ws
5 WHERE {
6     ?x sirp:has-name ?name;
7     sirp:has-webSite ?ws;
8     sirp:has-age ?edad.
9     FILTER (45 >?edad && ?edad >20)
10 }

```

Algorithm 6: Consulta SPARQL asociada a la pregunta *Q1.3* ¿Quiénes son mayores de 20 años y menores de 45 años?

```

1 PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?x rdf:type sirp:Professional;
7     sirp:has-name ?name;
8     sirp:has-webSite ?ws.
9 }

```

Algorithm 7: Consulta SPARQL asociada a la pregunta *Q1.4* ¿Cuáles son los nombres y sitios web de los profesionistas del área de RyT?

```

1 PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?x sirp:has-gender sirp:Male;
7     sirp:has-name ?name;
8     sirp:has-webSite ?ws;
9     sirp:worksIn sirp:ClarkParsia.
10 }

```

Algorithm 8: Consulta SPARQL asociada a la pregunta *Q1.5* ¿Quiénes trabajan en la Clark & Parsia y son del sexo Masculino?

```

1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6   ?x rdf:type sirp:Student;
7   sirp:reads sirp:English;
8   sirp:has-name ?name;
9   sirp:has-webSite ?ws.
10 }
```

Algorithm 9: Consulta SPARQL asociada a la pregunta *Q1.6* ¿Quiénes son estudiantes y leen en inglés?

```

1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6   ?x sirp:reads sirp:English;
7   sirp:speaks sirp:English;
8   sirp:writes sirp:English;
9   sirp:has-name ?name;
10  sirp:has-webSite ?ws.
11 }
```

Algorithm 10: Consulta SPARQL asociada a la pregunta *Q1.7* ¿Quienes hablan, leen y escriben en inglés?

```

1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6   ?x sirp:has-language sirp:German;
7   rdf:type sirp:Student;
8   sirp:has-name ?name;
9   sirp:has-webSite ?ws.
10 }
```

Algorithm 11: Consulta SPARQL asociada a la pregunta *Q1.8* ¿Qué estudiantes saben algo de inglés?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?x sirp:competentIn sirp:Synthesis_Skills;
7     rdf:type sirp:Teacher;
8     sirp:has-name ?name;
9     sirp:has-webSite ?ws.
10 }
```

Algorithm 12: Consulta SPARQL asociada a la pregunta *Q1.9* ¿Qué profesores tienen la capacidad de síntesis?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 SELECT DISTINCT ?name ?ws
6 WHERE {
7     ?k rdfs:subClassOf redes:Semantic_Web.
8     ?x rdf:type sirp:Professional;
9     sirp:expertiseIn ?k;
10    sirp:has-name ?name;
11    sirp:has-webSite ?ws.
12 }
```

Algorithm 13: Consulta SPARQL asociada a la pregunta *Q1.10* ¿Qué profesionistas tienen conocimiento en los temas de Web Semántica?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?x rdf:type sirp:Teacher;
7     sirp:expertiseIn redes:Distributed_Systems;
8     sirp:has-name ?name;
9     sirp:has-webSite ?ws.
10 }
```

Algorithm 14: Consulta SPARQL asociada a la pregunta *Q1.11* ¿Qué profesores tienen conocimientos en Sistemas Distribuidos?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?x sirp:expertiseIn redes:Java, redes:OWL, redes:RDF,
7         redes:Threads, redes:C, redes:OpenMP;
8     sirp:has-name ?name;
9     sirp:has-webSite ?ws.
10 }
```

Algorithm 15: Consulta SPARQL asociada a la pregunta *Q1.12* ¿Quiénes tienen conocimiento en Java, OWL, RDF, Threads, C, OpenMP?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 SELECT DISTINCT ?name ?ws
6 WHERE {
7     ?k rdfs:subClassOf redes:Operating_system.
8     ?x sirp:expertiseIn ?k;
9     sirp:has-name ?name;
10    sirp:has-webSite ?ws.
11 }
```

Algorithm 16: Consulta SPARQL asociada a la pregunta *Q1.13* ¿Qué estudiantes tienen algún conocimiento en los subtemas de Sistemas Operativos?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?k rdf:type sirp:University.
7     ?x sirp:worksIn ?k;
8     sirp:has-name ?name;
9     sirp:has-webSite ?ws.
10 }
```

Algorithm 17: Consulta SPARQL asociada a la pregunta *Q1.14* ¿Quiénes trabajan en una Universidad?

```
1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?x sirp:has-workplace sirp:UAM;
7     sirp:expertiseIn redes:Semantic_Web;
8     sirp:has-name ?name;
9     sirp:has-webSite ?ws.
10 }
```

Algorithm 18: Consulta SPARQL asociada a la pregunta *Q1.15* ¿Quienes laboran en la UAM y tienen algún conocimiento en Web Semántica?

```

1 PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 SELECT DISTINCT ?name ?ws
4 WHERE {
5     ?x sirp:has-advisor ?y.
6     ?y sirp:has-name ?z.
7     FILTER regex(?z, "Carolina Medina", "i")
8     ?x sirp:has-name ?name;
9     sirp:has-webSite ?ws. }

```

Algorithm 19: Consulta SPARQL asociada a la pregunta *Q1.16* ¿Qué personas tienen como asesor a Carolina Medina?

```

1 PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?name ?ws
5 WHERE {
6     ?y sirp:has-name ?z.
7     FILTER regex(?z, "Ricardo Marcellin", "i")
8     ?x sirp:has-colleague ?y;
9     sirp:has-name ?name;
10    sirp:has-webSite ?ws.
11 }

```

Algorithm 20: Consulta SPARQL asociada a la pregunta *Q1.17* ¿Quiénes son los colegas de Ricardo Marcellin?

```

1 PREFIX sirp: <http://arte.itz.uam.mx/ontologies/personRyT.owl#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 SELECT DISTINCT ?name ?ws
4 WHERE {
5     ?y sirp:has-name ?z.
6     FILTER regex(?z, "Carolina Medina Ramirez", "i")
7     ?x sirp:knows ?y;
8     sirp:has-name ?name;
9     sirp:has-webSite ?ws. }

```

Algorithm 21: Consulta SPARQL asociada a la pregunta *Q1.18* ¿Cuáles son los nombres y correos de las personas que conocen a Carolina Medina Ramírez?

```

1 PREFIX sirp: <http://arte.itzt.uam.mx/ontologies/personRyT.owl#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 SELECT DISTINCT ?name ?ws
4 WHERE {
5     ?x rdf:type sirp:Researcher;
6     sirp:has-name ?name;
7     sirp:has-webSite ?ws.}

```

Algorithm 22: Consulta SPARQL asociada a la pregunta *Q1.19* ¿Qué personas son profesores-investigadores?

```

1 PREFIX sird: <http://arte.itzt.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT ?title ?path ?ext ?lang
5 WHERE {
6     ?x sird:has-title ?title;
7     sird:has-filePath ?path;
8     sird:has-fileExtension ?ext;
9     sird:has-languageSource ?lang.
10 }

```

Algorithm 23: Consulta SPARQL asociada a la pregunta *Q2.1* ¿Cuáles son los títulos, rutas, extensión, idioma de todos los recursos digitales de RyT?

```

1 PREFIX sird: <http://arte.itzt.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6     ?x rdf:type sird:Book;
7     sird:has-title ?title;
8     sird:has-filePath ?path;
9     sird:has-topic ?k.
10    ?k rdfs:subClassOf redes:Distributed_Systems.
11 }

```

Algorithm 24: Consulta SPARQL asociada a la pregunta *Q2.2* ¿Cuáles libros tratan sobre algunos temas de Sistemas Distribuidos?

```

1 PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6   ?x sird:publishedIn sirp:UAM;
7   sird:has-filePath ?path;
8   sird:has-title ?title.
9 }
```

Algorithm 25: Consulta SPARQL asociada a la pregunta *Q2.3* ¿Qué recursos fueron publicados por la UAM?

```

1 PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6   ?x rdf:type sird:Document;
7   sird:has-topic redes:Peer_to_Peer_Systems;
8   sird:has-title ?title;
9   sird:has-filePath ?path.
10 }
```

Algorithm 26: Consulta SPARQL asociada a la pregunta *Q2.4* ¿Qué documentos sirven para dar un curso de Sistemas P2P?

```

1 PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6   ?x rdf:type sird:Multimedia;
7   sird:has-year ?year;
8   sird:has-title ?title;
9   sird:has-filePath ?path.
10  FILTER (?year >2009)
11 }
```

Algorithm 27: Consulta SPARQL asociada a la pregunta *Q2.5* ¿Qué recursos multimedia son mayores al año 2009?

```

1 PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6     ?x rdf:type sird:Document;
7     sird:has-topic redes:Ontology;
8     sird:has-title ?title;
9     sird:has-filePath ?path.
10 }
```

Algorithm 28: Consulta SPARQL asociada a la pregunta *Q2.6* ¿Cuáles documentos tratan sobre Ontologías?

```

1 PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6     ?x sird:publishedInJournal ?k;
7     sird:has-title ?title;
8     sird:has-filePath ?path.
9 }
```

Algorithm 29: Consulta SPARQL asociada a la pregunta *Q2.7* ¿Qué recursos fueron publicados en una Revista científica?

```

1 PREFIX sird: <http://arte.itz.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itz.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6     ?x sird:has-compendium ?txt
7     FILTER regex(?txt,“state of the art”,“i”)
8     ?x sird:has-title ?title;
9     sird:has-filePath ?path.
10 }
```

Algorithm 30: Consulta SPARQL asociada a la pregunta *Q2.8* ¿Qué recursos tienen en su descripción las palabras “linked data”?

```

1 PREFIX sird: <http://arte.itzt.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6   ?x rdf:type sird:Document;
7   sird:has-year ?year.
8   FILTER (?year >2000)
9   ?x sird:has-languageSource sirp:English;
10  sird:has-author ?z;
11  sird:has-title ?title;
12  sird:has-filePath ?path.
13  FILTER regex(?z, "Erik Alarcon Zamora", "i")
14 }

```

Algorithm 31: Consulta SPARQL asociada a la pregunta *Q2.9* ¿Cuáles documentos en Inglés y mayores al año 2000 son de autoría de Erik Alarcon Zamora?

```

1 PREFIX sird: <http://arte.itzt.uam.mx/ontologies/digiResourceRyT.owl#>
2 PREFIX redes: <http://mcyti.itzt.uam.mx/arios/odaryt.owl#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT DISTINCT ?title ?path
5 WHERE {
6   ?x rdf:type sird:Thesis;
7   sird:has-author ?y;
8   sird:has-title ?title;
9   sird:has-filePath ?path.
10  FILTER regex(?y, "Samuel Hernandez Maza", "i")
11 }

```

Algorithm 32: Consulta SPARQL asociada a la pregunta *Q2.10* ¿Cuál son las tesis de Samuel Hernandez Maza?

Bibliografía

- [1] L. Gandon, Fabien. Ontology Engineering: a Survey and a Return on Experience. Technical Report RR-4396, INRIA, March 2002.
- [2] James G. March, Herbert A. Simon, and Harold S. Guetzkow. *Teoría de la Organización*. Ariel, 1987.
- [3] Richard L. Daft. *Teoría Y Diseño Organizacional*. Cengage Learning, 09 edition, 2007.
- [4] Reinaldo O. Silva. *Teorías de la administración*. Thomson, 01 edition, 2002.
- [5] Juan J. Gilli. *Diseño organizativo: estructura y procesos*. Granica, 2007.
- [6] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on World Wide Web 7*, WWW7, pages 107–117, Amsterdam, The Netherlands, The Netherlands, 1998. Elsevier Science Publishers B. V.
- [7] Peter Rob and Carlos Coronel. *Sistemas de bases de datos: diseño, implementación y administración*. Thomson, 05 edition, 2004.
- [8] 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.
- [9] Rose Dieng, Olivier Corby, Alain Giboin, and Myriam Ribiére. Methods and Tools for Corporate Knowledge Management. Technical Report RR-3485, INRIA, September 1998.
- [10] John Lyons. *Semántica Lingüística: Una Introducción*. Paidos Iberica, 1997.
- [11] Torcoroma Velásquez Pérez, Andrés Puentes Velásquez, and Jaime Guzmán Luna. Ontologias: una tecnica de representacion de conocimiento. *Avances en Sistemas e Informática*, 8(2), 2011.
- [12] S. Bouzid, C. Cauvet, and J. Pinaton. A survey of semantic web standards to representing knowledge in problem solving situations. In *Information Retrieval Knowledge Management (CAMP), 2012 International Conference on*, pages 121–125, 2012.
- [13] C. Gueret, S. Schlobach, K. Dentler, M. Schut, and G. Eiben. Evolutionary and swarm computing for the semantic web. *Computational Intelligence Magazine, IEEE*, 7(2):16–31, 2012.

- [14] Tim Berners-Lee, Roy T. Fielding, and Larry Masinter. Uniform resource identifier (URI): Generic syntax. RFC 3986, RFC Editor, January 2005.
 - [15] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, May 2006.
 - [16] T. Fujino and N. Fukuta. A sparql query rewriting approach on heterogeneous ontologies with mapping reliability. In *Advanced Applied Informatics (IIAAAI), 2012 IIAI International Conference on*, pages 230–235, 2012.
 - [17] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220, June 1993.
 - [18] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 1–17. Springer Berlin Heidelberg, 2009.
 - [19] Markus Krötzsch, František Simančík, and Ian Horrocks. A description logic primer. *Computing Research Repository (CoRR)*, abs/1201.4089, 2012.
 - [20] Matthew Horridge, Holger Knublauch, Alan Rector, Robert Stevens, and Chris Wroe. A Practical Guide To Building OWL Ontologies Using The Protege-OWL Plugin and CO-ODE Tools Edition 1.0. August 2004.
 - [21] Magdalena Ortiz. Introducción a las Lógicas Descriptivas. Technical report, Vienna University of Technology, 2009.
 - [22] Yun Lin and John Krogstie. Semantic annotation of process models for facilitating process knowledge management. *Int. J. Inf. Syst. Model. Des.*, 1(3):45–67, July 2010.
 - [23] Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, Bijan Parsia, Peter Patel-Schneider, and Ulrike Sattler. Owl 2: The next step for owl. *Web Semant.*, 6(4):309–322, November 2008.
 - [24] R. B. Mishra and Sandeep Kumar. Semantic web reasoners and languages. *Artif. Intell. Rev.*, 35(4):339–368, April 2011.
 - [25] A.Q. Al-Namiy and F.S. Majeed. Towards automatic extracted semantic annotation (esa) for web documents. In *Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on*, volume 2, pages 614–617, 2009.
 - [26] A. Norta, R. Yangarber, and L. Carlson. Utility evaluation of tools for collaborative development and maintenance of ontologies. In *Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2010 14th IEEE International*, pages 207–214, 2010.
 - [27] Li Ding, Pranam Kolari, Zhongli Ding, and Sasikanth Avancha. Using ontologies in the semantic web: A survey. In Raj Sharman, Rajiv Kishore, and Ram Ramesh, editors, *Ontologies*, volume 14 of *Integrated Series in Information Systems*, pages 79–113. Springer US, 2007.
 - [28] T. Aruna, K. Saranya, and C. Bhandari. A survey on ontology evaluation tools. In *Process Automation, Control and Computing (PACC), 2011 International Conference on*, pages 1–5, 2011.
-

- [29] David Moner, J.A. Maldonado, D. Bosca, J.T. Fernandez, C. Angulo, P. Crespo, P.J. Vivancos, and M. Robles. Archetype-based semantic integration and standardization of clinical data. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 5141–5144, 2006.
- [30] Jun Zhai, Jianfeng Li, and Qinglian Wang. Using ontology and xml for semantic integration of electricity information systems. In *Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on*, pages 2197–2201, 2008.
- [31] Kai Yang and R. Steele. A semantic integration solution for online accommodation information integration. In *Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on*, pages 1105–1110, 2011.
- [32] Tuan-Dung Cao, Thanh-Hien Phan, and Anh-Duc Nguyen. An ontology based approach to data representation and information search in smart tourist guide system. In *Knowledge and Systems Engineering (KSE), 2011 Third International Conference on*, pages 171–175, 2011.
- [33] Inay Ha, Kyeong-Jin Oh, and Geun-Sik Jo. Ontology-driven visualization system for semantic search. In *Information Science and Applications (ICISA), 2011 International Conference on*, pages 1–6, 2011.
- [34] R. Suganyakala and R.R. Rajalaxmi. Movie related information retrieval using ontology based semantic search. In *Information Communication and Embedded Systems (ICICES), 2013 International Conference on*, pages 421–424, 2013.
- [35] F. Salam. New semantic indexing and search system based on ontology. In *Emerging Intelligent Data and Web Technologies (EIDWT), 2013 Fourth International Conference on*, pages 313–318, 2013.
- [36] Wang Xin and Xiong Guangleng. Design rationale as part of corporate technical memory. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 3, pages 1904–1908 vol.3, 2001.
- [37] R. Chakhmoune, H. Behja, and A. Marzak. Building corporate memories in collaborative way using ontologies: Case study of a ssii. In *Next Generation Networks and Services (NGNS), 2011 3rd International Conference on*, pages 23–28, 2011.
- [38] Victoria Uren, Philipp Cimiano, José Iria, Siegfried Handschuh, Maria Vargas-Vera, Enrico Motta, and Fabio Ciravegna. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1):14–28, January 2006.
- [39] Oscar Corcho. Ontology based document annotation: trends and open research problems. *Int. J. Metadata Semant. Ontologies*, 1(1):47–57, 2006.
- [40] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Niraj Aswani, Ian Roberts, Genevieve Gorrell, Adam Funk, Angus Roberts, Danica Damljanovic, Thomas Heitz, Mark A. Greenwood, Horacio Saggion, Johann Petrak, Yaoyong Li, and Wim Peters. *Text Processing with GATE (Version 6)*. 2011.

- [41] N. Islam, M.S. Siddiqui, and Z.A. Shaikh. Tode : A dot net based tool for ontology development and editing. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, volume 6, pages V6–229–V6–233, 2010.
 - [42] Holger Knublauch, Ray W. Fergerson, Natalya F. Noy, and Mark A. Musen. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In Sheila .. McIlraith, Dimitris Plexousakis, and r. a. n. k. van Harmelen, editors, *The Semantic Web - ISWC 2004*, volume 3298 of *Lecture Notes in Computer Science*, chapter 17, pages 229–243. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2004.
 - [43] Sören Auer. Powl - a web based platform for collaborative semantic web development. In *Proceeding of 1st Workshop Scripting for the Semantic Web (SFSW'05), Hersonissos, Greece, May 30*. CEUR Workshop Proceedings, May 2005.
 - [44] Walter Waterfeld, Moritz Weiten, and Peter Haase. Ontology management infrastructures. In Martin Hepp, Pieter Leenheer, Aldo Moor, and York Sure, editors, *Ontology Management*, volume 7 of *Computing for Human Experience*, pages 59–87. Springer US, 2008.
 - [45] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, Bernardo Cuenca Grau, and James Hendler. Swoop: A web ontology editing browser. *Web Semant.*, 4(2):144–153, June 2006.
 - [46] Mohamed Morsey, Jens Lehmann, Sören Auer, Claus Stadler, and Sebastian Hellmann. DBpedia and the Live Extraction of Structured Data from Wikipedia. *Program: electronic library and information systems*, 46:27, 2012.
 - [47] B. McBride. Jena: a semantic web toolkit. *Internet Computing, IEEE*, 6(6):55–59, 2002.
 - [48] Karlis Cerans, Guntis Barzdins, Renars Liepins, Julija Ovcinnikova, Sergejs Rikacovs, and Arturns Sprogis. Graphical schema editing for stardog owl/rdf databases using owlgred/s. In Pavel Klinov and Matthew Horridge, editors, *OWLED*, volume 849 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.
 - [49] M. Salvadores, G. Correndo, T. Omitola, N. Gibbins, S. Harris, and N. Shadbolt. 4s-reasoner: Rdfs backward chained reasoning support in 4store. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 3, pages 261–264, 2010.
 - [50] Jeen Broekstra, Arjohn Kampman, and Frank van Harmelen. Sesame: A generic architecture for storing and querying rdf and rdf schema. In *Proceedings of the First International Semantic Web Conference on The Semantic Web, ISWC '02*, pages 54–68, London, UK, UK, 2002. Springer-Verlag.
 - [51] Ivar Jacobson, James Rumbaugh, and Grady Booch. *El lenguaje unificado de modelado*. ADDISON-WESLEY, 2001.
 - [52] O. Nasraoui and L. Zhuhadar. Improving recall and precision of a personalized semantic search engine for e-learning. In *Digital Society, 2010. ICDS '10. Fourth International Conference on*, pages 216–221, 2010.
-

- [53] J. Beall. Measuring duplicate metadata records in library databases. In *Library Hi Tech News*, volume 27, pages 10–12, 2010.
- [54] Cyril Cleverdon and Michael Keen. Methods and Tools for Corporate Knowledge Management. Technical report, 1968.