

PA153-02

**SISTEMA DE SOFTWARE SEMÁNTICO PARA INCLUIR Y COMPARAR OBJETOS
PATRIMONIALES**

JORGE DANILO RUEDA SANABRIA

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
BOGOTÁ, D.C.
2015**

PA153-02

**SISTEMA DE SOFTWARE SEMÁNTICO PARA INCLUIR Y
COMPARAR OBJETOS PATRIMONIALES**

Autor:

JORGE DANILO RUEDA SANABRIA

MEMORIA DEL TRABAJO DE GRADO REALIZADO PARA CUMPLIR UNO
DE LOS REQUISITOS PARA OPTAR AL TÍTULO DE
MAGÍSTER EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

Director

RAFAEL ANDRÉS GONZÁLEZ RIVERA

Asesor

LUIS MANUEL VILCHES BLÁZQUEZ

Comité de Evaluación del Trabajo de Grado

ALEXANDRA POMARES QUIMBAYA

LINA CONSTANZA BELTRÁN BELTRÁN

Página web del Trabajo de Grado

<http://pegasus.javeriana.edu.co/~PA153-02-SEMANPHI>

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
MAESTRÍA EN INGENIERIA DE SISTEMAS Y COMPUTACIÓN
BOGOTÁ, D.C.
Noviembre, 2015

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

Rector Magnífico

Jorge Humberto Peláez Piedrahita, S.J.

Decano Académico Facultad de Ingeniería

Ingeniero Jorge Luís Sánchez Téllez

Director Maestría en Ingeniería de Sistemas y Computación

Ingeniera Ángela Cristina Carrillo Ramos

Director Departamento de Ingeniería de Sistemas

Ingeniero Rafael Andrés González Rivera

Artículo 23 de la Resolución No. 1 de Junio de 1946

“La Universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus proyectos de grado. Sólo velará porque no se publique nada contrario al dogma y la moral católica y porque no contengan ataques o polémicas puramente personales. Antes bien, que se vean en ellos el anhelo de buscar la verdad y la Justicia”

AGRADECIMIENTOS

A Dios por haberme permitido llegar hasta este punto y haberme dado bendición y fuerza para lograr mis objetivos.

A mi esposa Claudia Patricia y mis hijos Dany y Lucy, quienes fueron mi motivación para no desfallecer en los momentos difíciles. Gracias por su amor y entendimiento.

A mis padres por los ejemplos de perseverancia y constancia que los caracteriza y que siempre me enseñaron.

A mi asesor del trabajo de grado, el ingeniero Luis Manuel Vilches quien fue mi mentor, motivador y amigo permanente en este proceso. Sin él no hubiera sido posible este trabajo.

A mi director del trabajo de grado, el ingeniero Rafael Andrés González porque gracias a sus consejos, paciencia y dedicación se pudo terminar de manera exitosa este trabajo.

A mi compañero el ingeniero Néstor Armando Nova quien gracias a sus estudios y casos de estudio en el Doctorado pude tener este tema tan interesante para mi trabajo de grado.

A la comunidad de la red PHI y los docentes de la facultad de arquitectura de la universidad, Carlos y Lina por brindarme toda la información requerida para desarrollar el proyecto.

A la directora de la maestría, la ingeniera Ángela Cristina Carrillo por su gestión y asesoría para poder culminar con mis estudios.

A todos los docentes de la maestría que de manera directa o indirecta ayudaron al desarrollo de este trabajo.

Debo agradecer también de manera especial a muchos compañeros y amigos por su colaboración adicional.

Gracias a todos por su apoyo.

CONTENIDO

CONTENIDO	V
LISTA DE FIGURAS.....	VII
LISTA DE TABLAS	IX
INTRODUCCIÓN.....	1
I – PRESENTACION DEL TRABAJO DE GRADO.....	2
1. PROBLEMÁTICA U OPORTUNIDAD DE MEJORA	2
2. OBJETIVOS	3
2.1 OBJETIVO GENERAL.....	3
2.2 OBJETIVOS ESPECÍFICOS.....	3
3. METODOLOGÍA	4
4. POTENCIAL DE INNOVACIÓN	6
II – MARCO TEÓRICO.....	8
2.1 PATRIMONIO HITÓRICO CULTURAL IBEROAMERICANA.....	8
2.2 GESTIÓN DEL CONOCIMIENTO	9
2.3 TECNOLOGÍAS PARA LA GESTIÓN DEL CONOCIMIENTO	10
2.3.1 WEB SEMÁNTICA	10
2.3.2 LENGUAJES DE LA WEB SEMÁNTICA.....	12
2.3.3 ONTOLOGIAS.....	15
2.3.4 METODOLOGÍAS PARA EL DESARROLLO DE ONTOLOGIAS.....	15
2.3.5 LINKED (OPEN) DATA.....	16
2.4 TRABAJOS RELACIONADOS.....	17
2.4.1 MODELOS DE LA WEB SEMÁNTICA SOBRE OBJETOS PATRIMONIALES.....	17
2.4.2 INICIATIVAS DE LINKED DATA RELACIONADAS CON PATRIMONIO	20
2.5 METODOLOGÍAS ÁGILES	21
2.5.1 Extreme Programming (XP).....	23
2.6 HERRAMIENTAS TECNOLÓGICAS	23
2.6.1 Servidores Base de Datos.....	23
2.6.2 Servidores Web.....	24
2.6.3 Frameworks.....	24
2.6.4 Triplestore	24
III – DISEÑO Y DESARROLLO	25

3.1 PROCESO DE LA SOLUCIÓN	25
3.1.1 MÉTODO	25
3.1.2 PASOS DE LA SOLUCIÓN.....	26
3.2 DISEÑO DE LA SOLUCIÓN	32
3.1.1 ARQUITECTURA.....	33
3.1.2 PATRONES DE DISEÑO IMPLEMENTADOS	35
3.1.3 ALGORITMOS	35
3.1.4 HERRAMIENTAS UTILIZADAS	39
3.1.3 CRITERIOS DE EVALUACIÓN E ITERACIÓN.....	39
IV – PROTOTIPO	46
4.1 TRANSFORMADOR.....	46
4.2 COMPARADOR	49
V – EVALUACION Y VALIDACIÓN DEL PROTOTIPO	52
VI – TRABAJOS FUTUROS	57
VII – CONCLUSIONES Y APORTES.....	58
VIII – BIBLIOGRAFÍA	59

LISTA DE FIGURAS

Figura 1. Arquitectura de la Web semántica.....	11
Figura 2. Un triple con dos nodos (Sujeto y Objeto) y conectados (Predicado).....	13
Figura 3. Estructura general de un documento OWL	14
Figura 4. SPARQL consultas y manipulación de grafos	14
Figura 5. Diseño y desarrollo de la solución.....	30
Figura 6. Diagrama de contexto.....	33
Figura 7. Vista lógica.....	34
Figura 8. Modelo de componentes del sistema.....	34
Figura 9. Detalle de la generación de triples directos	41
Figura 10. Comprobación del grafo generado por medio de herramientas semánticas	43
Figura 11. Comprobación de la publicación del grafo y resultado de consulta	44
Figura 12. Evaluación de funcionalidad e integridad de los resultados	44
Figura 13. Evaluación de la integración con otros sistemas	45
Figura 14. Evaluación del resultado de la comparación del lenguaje natural	45
Figura 15. Pasos de la transformación en un modelo RDF.....	46
Figura 16. Paso de conectividad a BD y selección del esquema a utilizar	47
Figura 17. Mapeo directo o manual.	47
Figura 18. Conciliación entre ontología y mapeo inicial	48
Figura 19. Transformación adicional.	48
Figura 20. Generación del archivo grafo resultante en formato RDF.....	49
Figura 21. Opción de búsqueda	49
Figura 22. Ejemplo de iglesias ubicada en Bogotá	50
Figura 23. Ejemplo de identificación de ruta entre dos puntos.....	50

Figura 24. Ejemplo de vista callejera..... 51

Figura 25. Ejemplo del detalle de comparación..... 51

Figura 26. Resultado promedio para cada variable TAM..... 55

LISTA DE TABLAS

Tabla 1. Alineación de objetivos específicos y fases de la metodología	5
Tabla 2. Comparación de las características entre tipos de conocimiento	10
Tabla 3. Buenas prácticas para la elaboración de casos de uso	22
Tabla 4. Ranking de “Agilidad”	22
Tabla 5. Características del método utilizado.	26
Tabla 6. Requerimientos funcionales de la ontología.	28
Tabla 7. Identificación de Requerimientos	30
Tabla 8. Algoritmos de conciliación ontología y esquema.	37
Tabla 9. Algoritmo Relacionar clases con campos del Esquema.....	38
Tabla 10. Algoritmo Comparación semántica.	39
Tabla 11. Lista de herramientas utilizadas.....	39
Tabla 12. Ejemplo de aplicación de criterios de primer nivel.....	40
Tabla 13. Porcentaje de elementos utilizados en la generación del mapeo directo	40
Tabla 14. Comparación de técnicas string matching	41
Tabla 15. Identificación de equivalencia y cantidad de elementos resultantes.....	42
Tabla 16. Grado de eficiencia en la aplicación de las transformaciones	43
Tabla 17. Evaluación característica de funcionalidad.....	52
Tabla 18. Evaluación característica de fiabilidad	53
Tabla 19. Evaluación característica de usabilidad	53
Tabla 20. Evaluación característica de eficiencia	53
Tabla 21. Evaluación característica de mantenibilidad.....	54
Tabla 22. Evaluación característica de portabilidad	54
Tabla 23. Variables utilizadas en el modelo de aceptación tecnológica.	55

Tabla 24. Resultado promedio total de la validación TAM 56

ABSTRACT

The project involved the creation of a system support network Patrimonio Histórico-Cultural Iberoamericano (PHI) for inclusion and comparison of heritage objects to semantic level to share this knowledge as a strategic value within the network. The principal goal is to share knowledge semantically enriched to be exploitable in terms of usefulness for members of the heritage community and likewise can be implemented in other different areas. This document allows understanding the process carried in the project, from problem identification through each stage of development to evaluation and validation.

RESUMEN

El proyecto consistió en la creación de un sistema de apoyo a la Red de Patrimonio Histórico-Cultural Iberoamericano (PHI) para la inclusión y comparación de objetos patrimoniales a nivel semántico con el fin de compartir este conocimiento como valor estratégico dentro de la Red. Compartir conocimiento enriquecido semánticamente es el objetivo primordial del proyecto, para que sea explotable en términos de utilidad por los miembros de la comunidad de patrimonio y que de igual manera pueda implementarse en otras áreas diferentes. El presente documento permite entender el proceso llevado a cabo en el desarrollo del proyecto, desde la identificación del problema pasando por cada uno de las etapas de desarrollo hasta la evaluación y validación del mismo.

RESUMEN EJECUTIVO

El avance de la Web y la manera como se publican los datos ha expuesto de manera general las problemáticas en el intercambio del conocimiento a través de Internet. La utilización de técnicas heterogéneas para documentar los grandes volúmenes de información que son trasladados a la web, las diferentes estructuras utilizadas para la representación de los datos y la baja calidad de la información que se encuentra publicada hace que los procesos de búsqueda, análisis y compresión se dificulten aún más por parte de las personas que quieren acceder a este conocimiento.

La red de Patrimonio Histórico-Cultural Iberoamericano no es ajena a esta problemática por lo que su objetivo principal es el de contar con instrumentos innovadores que sirvan para conocer mejor el valor estratégico del patrimonio y que permitan gestionar de modo más eficiente este legado común.

La implementación y mejoras en las tecnologías de gestión de conocimiento, puede conducir a cambios sustanciales en la manera de compartir información de casos, eventos y métodos registrándolos semánticamente para realizar un análisis más sistemático y una colaboración más efectiva en el desarrollo de proyectos de patrimonio.

Se propone como solución la creación de un sistema que permita a la Red de Patrimonio Histórico-Cultural Iberoamericano incluir y comparar objetos patrimoniales haciendo uso de técnicas semánticas apoyadas en las tecnologías de la información y comunicación con el fin de intercambiar conocimiento.

El proyecto se desarrolla teniendo en cuenta la información de interés y las características de los objetos patrimoniales tangibles que requieren los investigadores de la red PHI. Haciendo uso de la ingeniería de software se diseña y construye el sistema de software semántico, identificando los principales componentes y sus relaciones. Se construye una red de ontologías de patrimonio histórico y cultural recogiendo las características iberoamericanas del dominio. Se transforma la información existente de la red PHI para enriquecer su contenido semánticamente y explotar su utilización en el entorno de la red a través de búsquedas, comparaciones y actividades de georreferenciación.

Para la implementación de la solución se utilizan tecnologías para el manejo de bases de datos, servidores web, diferentes plataformas de sistemas operativos, marcos de trabajo para los lenguajes de desarrollo, almacenes de triples y librerías externas que permiten desarrollar el proyecto de software.

Para la formalización de la evaluación y la validación se realizan experimentos con el sistema resultante y se presenta a la comunidad de la red PHI. Se realizan pruebas funcionales del sistema y a través del modelo de aceptación tecnológica se realiza una prueba de concepto incluyendo a expertos del dominio tecnológico y de patrimonio.

La realización del proyecto aporta tanto para la red PHI como para otros dominios del conocimiento un cambio en la manera de publicar y compartir su información enriquecida semánticamente permitiendo realizar análisis más sistemáticos y una colaboración más efectiva en el desarrollo de sus proyectos. Otro aporte importante es que durante el proceso de desarrollo se diseñaron algoritmos y artefactos que pueden ser utilizados para trabajos futuros o ser implementados en otros procesos de enriquecimiento semántico.

Este proyecto sirve como base para trabajos futuros con el fin de implementar mejoras en su diseño, en el soporte de múltiples plataformas, en la utilización de diferentes técnicas de pareo como la minería de datos y el alignment server, en una mayor explotación semántica, en una transformación o enriquecimiento semántico más automatizado y en la aplicabilidad que otros investigadores puedan proponer para este trabajo.

INTRODUCCIÓN

Las bases de conocimiento están desempeñando un papel cada vez más importante en la mejora de la inteligencia de la Web, de la búsqueda de información y como base a la integración de la información. Actualmente, la mayoría de las bases de conocimientos cubren sólo dominios específicos, son creados por grupos relativamente pequeños de ingenieros del conocimiento, y cuesta mucho el mantenimiento a medida que evolucionan los dominios.

Este proyecto hace uso de las redes ontológicas y de la integración con la web semántica para fortalecer las bases y la compartición de conocimiento inicialmente en la red de Patrimonio Histórico-Cultural Iberoamericano.

El desarrollo del documento inicia con una presentación sobre la problemática y la motivación que llevaron a la definición de los objetivos, la metodología y el potencial de innovación del presente proyecto (capítulo I). Posteriormente en el capítulo II, se define el marco teórico como base para delimitar el problema, formular definiciones, fundamentar afirmaciones y presentar trabajos relacionados. El capítulo III es el diseño y desarrollo de la solución que incluye el proceso, el diseño y los criterios de evaluación. El prototipo se presenta en el capítulo IV. El método y resultados de la evaluación y validación se presentan en el capítulo V. Los capítulos finales presentan los trabajos futuros, las conclusiones y aportes.

I – PRESENTACION DEL TRABAJO DE GRADO

En el presente capítulo se realiza la presentación del trabajo de grado desde la identificación de la problemática, pasando por la definición de los objetivos de la solución, la metodología a usar con sus respectivas fases y el potencial de innovación que se tiene en el desarrollo del actual proyecto.

1. PROBLEMÁTICA U OPORTUNIDAD DE MEJORA

El principal objetivo que persigue la red de Patrimonio Histórico y Cultural Iberoamericano (PHI) es el de contar con instrumentos innovadores que sirvan para conocer mejor el valor estratégico del patrimonio y que permitan gestionar de modo más eficiente este legado común para activar su capacidad de ordenación del espacio habitado [1].

Dentro de las problemáticas que más afectan el intercambio de conocimiento sobre la conservación de los objetos patrimoniales se encuentra la utilización de técnicas para documentar grandes volúmenes de información [2], frecuentemente heterogéneos, de baja calidad y con diferentes estructuras de datos, impidiendo comprender y analizar mejor los procesos por parte de diferentes especialistas involucrados (restauradores, historiadores, arqueólogos, arquitectos, ingenieros, etc.), haciendo difícil la comunicación y comparación de buenas prácticas aplicadas sobre los objetos patrimoniales [3].

Se han propuesto diferentes herramientas y técnicas para la gestión del conocimiento [4] [5] que ayudan a solventar el problema anteriormente mencionado, entre las que se encuentra la utilización de Tecnologías de la Información y las Comunicaciones (TIC) para el intercambio y compartición de datos a través del uso de aplicaciones o plataformas sobre Internet, intranets o redes privadas virtuales. Estas pueden ser utilizadas en sistemas como soporte para la toma de decisiones, nuevos procesos para el desarrollo de productos, aprendizaje organizacional, transferencia de conocimiento, integración del conocimiento, ontologías, reingeniería, intercambio de información y simulación [6].

Con la utilización de tecnologías semánticas y redes de ontologías se puede dotar de mayor significado la información que se publica en Internet permitiendo que las personas pueden encontrar respuestas a sus preguntas de manera más fácil y rápida. Cuando se cuenta con más significado y semántica en los datos se pueden crear soluciones que puedan compartir, procesar y transferir información de una mejor manera gracias a esa estructura común [7].

En consecuencia, se propone un sistema con el objetivo de incluir y comparar objetos patrimoniales haciendo uso de técnicas semánticas [8] apoyadas en las tecnologías de la información y comunicación, con el fin de apoyar la compartición de conocimiento dentro de la red de patrimonio-cultural histórico iberoamericano (PHI).

Utilizar tecnologías semánticas y redes de ontologías permite la eliminación de fronteras institucionales haciendo uso de información de sucesos históricos, de conservación y demás información relevante de los objetos patrimoniales en diferentes bibliotecas y otras fuentes diferentes del sistema utilizado [9] [10]. Publicar los vocabularios como Datos Enlazados Abiertos

(Linked Open Data) faculta a otras personas relacionar esta información con la suya, creando una red mundial de información sobre el patrimonio histórico cultural de manera automática [11].

Sin embargo, la implementación y mejoras en las tecnologías de gestión de conocimiento, puede conducir a cambios sustanciales en la manera de compartir información de casos, eventos y métodos registrándolos semánticamente para realizar un análisis más sistemático y una colaboración más efectiva en el desarrollo de proyectos de patrimonio [12].

Adicionalmente, este proyecto hace parte de un marco de trabajo mucho más amplio conformado por un grupo de investigadores del área de arquitectura y el desarrollo de un estudio de caso aplicado a la Red PHI de una tesis doctoral en Ingeniería de la Pontificia Universidad Javeriana.

2. OBJETIVOS

La identificación del problema sobre el intercambio de conocimiento propone como solución la creación de un sistema que permita a la Red de Patrimonio Histórico-Cultural Iberoamericano incluir y comparar objetos patrimoniales haciendo uso de técnicas semánticas apoyadas en las tecnologías de la información y comunicación. Teniendo en cuenta lo anterior se definieron los siguientes objetivos del proyecto.

2.1 OBJETIVO GENERAL

Realizar un sistema de software semántico que permita introducir conocimiento de proyectos de patrimonio en el contexto de la Web Semántica incluyendo y comparando objetos patrimoniales de la Red de Patrimonio Histórico-cultural Iberoamericano (PHI) para compartir conocimiento codificado semánticamente.

2.2 OBJETIVOS ESPECÍFICOS

- Identificar la información de interés y las características de los objetos patrimoniales tangibles, con base en la información que requieren los investigadores de la red PHI.
- Construir o ajustar la red de ontologías de patrimonio histórico y cultural recogiendo las características iberoamericanas del dominio.
- Incluir y publicar datos semánticos de patrimonio histórico y cultural conforme a los principios de Linked Data.
- Diseñar la arquitectura del sistema de software semántico, identificando los principales componentes y sus relaciones.
- Construir el sistema de software para realizar la comparación semántica de objetos patrimoniales tangibles.
- Validar el sistema de software semántico mediante su implementación como una prueba de concepto en la Pontificia Universidad Javeriana, para la inclusión y comparación de objetos patrimoniales.

3. METODOLOGÍA

La estrategia que subyace la metodología de investigación utilizada es la ciencia basada en el diseño, donde se busca realizar el proceso de investigación desde la lógica del diseño como solución de problemas. El trabajo de Hevner y otros [13], justifica que la construcción de un artefacto es una forma de contribuir al conocimiento en ciencias de lo artificial o de diseño (particularmente en informática y sistemas de información), al tiempo que se soluciona un problema real; esto con el objeto de equilibrar el rigor y la relevancia en una investigación.

La estrategia antes mencionada se instancia en una metodología concreta, propuesta por Peffers [14] para operacionalizar la ciencia basada en el diseño en sistemas de información. En tanto que Hevner incluye una filosofía y lógica de investigación centrada en el diseño para investigar en sistemas de información, Peffers aterriza esto en una serie de pasos derivados de la lógica de Hevner y de investigaciones ya publicadas. Esto resulta en una guía concreta del paso a paso de la investigación que sigue la lógica de un proceso de diseño ingenieril típico, incorporando elementos de validación y comunicación (divulgación) que posibiliten una contribución novedosa al conocimiento.

Se identifica la motivación y el problema en una primera fase, en la segunda fase se definen los objetivos de la solución, en la tercera fase se procede con la realización del diseño y desarrollo; para finalizar, el artefacto resultante se demuestra, evalúa y los resultados se comunican. La aplicación de la ciencia del diseño permite utilizar ciclos o iteraciones entre las fases definidas hasta que los criterios de evaluación se cumplan.

Ni Hevner ni Peffers prescriben la manera en que se debe realizar el diseño, pues este depende del tipo de artefacto y del conocimiento aplicable que puede sugerir métodos y técnicas diversas. En este caso, la fase de diseño y desarrollo se orienta a la implementación de los aspectos semánticos de los datos de patrimonio. Se seleccionará una metodología de desarrollo de software para la construcción del sistema que permita dividir el desarrollo en fases para modelarlo y diseñarlo de manera ágil, iterativa e incremental [15] [16]. Por último para el desarrollo de la red de ontologías se utilizará la metodología NeOn [17] y para la generación de datos semánticos conforme a los principios de Linked Data se utilizará la metodología descrita en [8].

La evaluación y validación se formalizarán mediante la realización de experimentos con el sistema resultante del diseño y desarrollo. De esta manera la evaluación se realizará por medio de pruebas funcionales del sistema y la validación a través de una prueba de concepto [18].

Cada uno de los objetivos específicos se alinea a las fases de la metodología propuesta (Tabla 1):

OBJETIVOS ESPECÍFICOS	FASES DE LA METODOLOGÍA
1. Identificar la información de interés y las características de los objetos patrimoniales tangibles, con base en la información que requieren los investigadores de la red PHI.	1. Identificación del problema y la motivación. 2. Objetivos de la solución

2. Construir o ajustar la red de ontologías de patrimonio histórico y cultural recogiendo las características iberoamericanas del dominio.	3. Diseño y desarrollo
3. Incluir y publicar datos semánticos de patrimonio histórico y cultural conforme a los principios de Linked Data.	
4. Diseñar la arquitectura del sistema de software semántico, identificando los principales componentes y sus relaciones.	
5. Construir el sistema de software semántico que contenga un comparador de objetos patrimoniales tangibles.	
6. Validar el sistema de software semántico mediante su implementación como una prueba de concepto en la Pontificia Universidad Javeriana, para la comparación de objetos patrimoniales.	4. Demostración, evaluación y comunicación

Tabla 1. Alineación de objetivos específicos y fases de la metodología

A continuación se describen cada una de las fases a desarrollar y cómo se relacionan con el proceso de investigación científica basada en el diseño:

Fase 1. Identificación del problema y la motivación: En esta fase se definirá el problema específico de investigación y se justificará el valor de una solución. Al detallar conceptualmente el problema se planteará una solución que pueda capturar la complejidad del problema. Justificar el valor de la solución motiva la investigación para buscar la solución y aceptar los resultados [14]. Para el desarrollo de esta fase se realizaron las siguientes actividades:

- a. Identificación del grupo de usuarios del sistema.
- b. Toma de requerimientos específicos.
- c. Identificación de debilidades y oportunidades de la Red PHI para compartir conocimiento de los objetos patrimoniales.
- d. Formulación del problema a resolver.

Fase 2. Objetivos de la solución: En esta fase se definirán los objetivos del proyecto que permitan en términos cualitativos, apoyar la solución del problema a través de un nuevo artefacto. Los objetivos deben ser lógicamente inferidos de la especificación problema [14]. Se realizará una revisión del estado del arte sobre gestión del conocimiento y semántica de datos en particular. Para el desarrollo de esta fase se realizaron las siguientes actividades:

- a. Buscar en las bases bibliográficas material relacionado.
- b. Realizar análisis de citas.
- c. Identificación de aplicaciones y/o plataformas de las Tecnologías de la Información y las Comunicaciones (TIC) que puedan ser usadas en el desarrollo del sistema de software semántico.

- d. Formulación de los objetivos de la solución.

Fase 3. Diseño y desarrollo: Esta fase incluye la determinación de la funcionalidad deseada del artefacto y su arquitectura y luego crear el artefacto real. Se modelaran los datos semánticos [HORR2004] teniendo en cuenta el marco de descripción de recursos (RDF) [19]. En términos generales, se deben plasmar los objetivos en el diseño y desarrollo para que se puedan aplicar como una solución [14]. Para el desarrollo de esta fase se realizaron las siguientes actividades:

- a. Seleccionar y aplicar una metodología general ágil para el desarrollo del sistema.
- b. Identificar y crear del modelo de componentes y sus relaciones.
- c. Construir la red de ontologías que identifican los objetos patrimoniales tangibles con base en recursos no ontológicos y ontológicos existentes.
- d. Modelar los datos semánticos de patrimonio histórico y cultural conforme a las características de Linked Data
- e. Crear los modelos de datos y sus relaciones con las redes de ontologías utilizando Resource Description Framework (RDF).
- f. Desarrollar los elementos del sistema de inclusión y comparación.
- g. Verificar el funcionamiento del sistema.

Fase 4. Demostración, evaluación y comunicación: En esta fase se implementará el sistema de software semántico en un ambiente de pruebas para la Red de Patrimonio Histórico y Cultural Iberoamericano. Implicará realizar una experimentación o prueba apropiada para demostrar cómo usar el artefacto para resolver el problema. Adicionalmente se debe evaluar el cumplimiento de los objetivos con los resultados reales observados por el uso del artefacto [14]. Para el desarrollo de esta fase se realizaron las siguientes actividades:

- a. Preparar el hardware y software requerido para un ambiente de pruebas.
- b. Instalar componentes en el ambiente de pruebas.
- c. Realizar migración de datos de algunos casos o proyectos existentes sobre objetos patrimoniales.
- d. Validar la arquitectura de desarrollo del sistema por parte de un experto en gestión de conocimiento y tecnologías utilizadas.
- e. Realizar una prueba de completitud y coherencia de los objetos patrimoniales representados de manera semántica.
- f. Utilizar pruebas de aceptación del modelo tecnológico (TAM) como instrumento de validación.
- g. Validar la credibilidad y plausibilidad en la comparación de objetos patrimoniales por parte de un experto del dominio de patrimonio.
- h. Consolidar y analizar los resultados de la validación.

4. POTENCIAL DE INNOVACIÓN

El desarrollo del sistema de software semántico permitirá la representación e inclusión de los datos de patrimonio a través de Linked Data y comparará recursos semánticos de objetos patri-

moniales. Este es el aporte potencial de innovación de uso del conocimiento porque su implementación a través de estas tecnologías de gestión de conocimiento generarán cambios en la manera de compartir información de casos, eventos y métodos a nivel patrimonial, permitiendo registrar la información de manera semántica y de esta manera poder realizar análisis más sistemáticos y una colaboración más efectiva en el desarrollo de proyectos de patrimonio dentro de la Red de Patrimonio Histórico-Cultural Iberoamericana (PHI).

Otro aporte potencial de innovación es que el resultado del caso específico en la red PHI puede ser trasladado a otros contextos donde la tecnología de comparación y la arquitectura que lo soporta puedan ser utilizadas de manera útil en la inclusión y comparación de objetos, trayendo mejoras en los procesos de gestión de su propio conocimiento.

II – MARCO TEÓRICO

En este capítulo se describen los referentes teóricos y conceptuales que soportan el rigor sobre la base del conocimiento. Se presentan apartados sobre el patrimonio histórico-cultural iberoamericano, la contextualización de la gestión del conocimiento, las tecnologías para gestionar conocimiento, los trabajos realizados y las herramientas tecnológicas usadas.

2.1 PATRIMONIO HISTÓRICO CULTURAL IBEROAMERICANA

Según la UNESCO, el patrimonio cultural considera los monumentos, conjuntos y lugares que tengan un valor universal excepcional desde el punto de vista de la historia, el arte o la ciencia. Valor universal excepcional significa una importancia cultural y/o natural tan extraordinaria que trasciende las fronteras nacionales y cobra importancia para las generaciones presentes y venideras de toda la humanidad. Por lo tanto, la protección permanente de este patrimonio es de suma importancia para la comunidad internacional [20].

En el dominio cultural, los sistemas de información han servido de apoyo en el manejo del diverso y amplio volumen de datos que emergen de un proyecto patrimonial, y que requieren sistemas de información especializados en actividades como la gestión de colecciones, conservación, investigación y presentación [21]. La necesidad específica de apoyo se refleja en actividades como el registro, la catalogación, la representación visual y la publicación de información de objetos patrimoniales, en una tendencia a la disposición de información al público en formato digital, muchas veces exigida por los gobiernos para atender el interés público de conocimiento cultural. Estas tareas requieren describir y controlar detalladamente los objetos patrimoniales [22]; sin embargo, aunque existen varios recursos tecnológicos para la catalogación de los recursos patrimoniales, estos no reflejan un acuerdo entre expertos [21].

Un esfuerzo conjunto entre universidades expertas en el ámbito patrimonial intenta consensuar una metodología para la recopilación de metadatos y recuperación de información que facilite la catalogación y la gestión de la innumerable información asociada con el manejo patrimonial. Este es el caso de la Red de Patrimonio Histórico y Cultural Iberoamericano – Red PHI, compuesta por siete universidades expertas en el tema patrimonial, que actualmente se encuentran desarrollando en conjunto un sistema de gestión del conocimiento patrimonial cuyo objetivo es crear una plataforma que sirva para conocer mejor el valor estratégico del patrimonio y que permita gestionar de modo más eficiente este legado común para activar su capacidad de ordenación del espacio habitado. En el ámbito tecnológico, el cumplimiento de este objetivo requiere que el sistema desarrollado permita la interoperabilidad con otras bases de datos y facilite la búsqueda de información dentro del conjunto de casos de estudio que se hayan registrado y catalogado en el sistema [23].

Teniendo en cuenta que el ámbito del patrimonio cultural implica un conjunto de disciplinas diversas que son altamente especializadas, y que los datos que emergen de estas actividades generalmente no están estructurados, se hace necesaria la aplicación de tecnologías avanzadas de representación y relacionamiento del conocimiento como la ingeniería ontológica y Linked

Data, que facilitan la descripción del dominio cultural en sus términos fundamentales o conceptos, y la definición de relaciones entre estos en un sentido semántico que conserva el significado de la información y a su vez facilita el almacenamiento, recuperación y explotación de la información [24].

2.2 GESTIÓN DEL CONOCIMIENTO

El conocimiento es un activo dentro de las organizaciones; debe verse como un recurso explotable [25] permitiendo la generación de nuevas oportunidades. Es importante la implementación activa de la gestión del conocimiento con el fin de mejorar las capacidades en la toma de decisiones a través de la construcción de comunidades de aprendizaje y de esta manera motivar el cambio cultural y la innovación [26].

Si bien muchas veces no existe una diferenciación clara entre el conocimiento, la información y los datos, algunos autores ponen de manifiesto que “el conocimiento no es datos ni información, aunque se relaciona con ambos y a menudo las diferencias entre estos términos es una cuestión de grado” [27] por lo que es importante definir claramente esos conceptos [26].

- **Datos.** Una definición sencilla es que los datos simplemente existen y no tienen importancia más allá de su existencia [28]. Desde el punto de vista organizacional se refieren a hechos discretos sobre acontecimientos de actividades realizadas. Los datos no suministran explicaciones ni dan razones [29]. Con frecuencia los datos se almacenan a través de algún medio tecnológico. Por ejemplo, un texto o una fecha serían datos.
- **Información.** La información son los datos relacionados que permiten tener un significado [28]. A diferencia de los datos, la información tiene significado, debe ser importante para algún propósito y estar organizada. Se dice que los datos se convierten en información cuando se les agrega significado [30]. Por ejemplo, el conjunto de datos que representan un objeto patrimonial (su título, su descripción, su ubicación, su autor, etc.) sería la información propia de ese elemento.
- **Conocimiento.** El conocimiento es la apropiación adecuada de la información con la intención de ser útil. El conocimiento es un proceso determinista [28]. Los datos se pueden convertir en información así como la información en conocimiento. Podemos entonces aplicar para el conocimiento una definición formal indicando que “es una mezcla de experiencias estructuradas, valores, información contextual e internalización experta, que proporciona un marco para la evaluación e incorporación de nuevas experiencias e información. Se origina y se aplica en la mente de los conocedores” [27]. Por ejemplo, las técnicas y métodos que se utilizan dentro de la red PHI para identificar y representar el valor histórico de los objetos patrimoniales sería una representación clara de conocimiento.

La gestión del conocimiento organizacional se basa en la teoría de la generación del conocimiento, que es un proceso de comunicación del conocimiento, teniendo en cuenta las transformaciones del conocimiento tácito y el explícito, de una manera dinámica y continua [31]. Las características de cada uno de estos tipos de conocimiento se muestran en la tabla 2.

Fuente. Nonaka y Takehuchi

Conocimiento Tácito	Conocimiento Explícito
<ul style="list-style-type: none"> • Representa la experiencia y es la forma más poderosa de conocimiento • Dificultad para articular formalmente • Dificultad para comunicarlo y compartirlo • Incluye ideas personales, sentimientos, cultura y valores • Compartido únicamente de manera voluntaria a través de la interacción social 	<ul style="list-style-type: none"> • Puede llegar a ser obsoleto rápidamente • Es posible articularlo formalmente, puede ser procesado y almacenado por medios automatizados u otros medios de comunicación • Fácil de comunicarlo y compartirlo • Formalmente relacionado y público • Puede ser copiado e imitado fácilmente • Puede ser transmitido.

Tabla 2. Comparación de las características entre tipos de conocimiento

En el caso de la red de PHI se pueden identificar el conocimiento tácito en cada uno de los investigadores y estudiantes que a lo largo de los años han ido identificando la manera de mejorar los procesos de análisis y percepción del valor patrimonial.

Un ejemplo del conocimiento explícito dentro de la red es la documentación formal de sus procesos y procedimientos de cómo, cuándo y quién debe realizar actividades propias de la red PHI.

2.3 TECNOLOGÍAS PARA LA GESTIÓN DEL CONOCIMIENTO

Se han propuesto diferentes herramientas y técnicas que apoyan la implementación de la gestión del conocimiento, entre las que se encuentra la utilización de Tecnologías de la Información y las Comunicaciones (TIC) para el intercambio y compartición de datos a través del uso de aplicaciones o plataformas sobre Internet, intranets o redes privadas virtuales [32]. Las TIC pueden ser utilizadas en sistemas como soporte para la toma de decisiones, nuevos procesos para el desarrollo de productos, aprendizaje organizacional, transferencia de conocimiento, integración del conocimiento, ontologías, reingeniería, intercambio de información y simulación [33].

2.3.1 WEB SEMÁNTICA

La Web semántica se presenta como una opción que tienen las organizaciones para clasificar y modelar correctamente el conocimiento asociado a sus procesos de negocio o cualquier dominio específico [34]. Tim Berners-Lee creador de la Web establece que "La Web Semántica es una extensión de la Web actual donde la información tiene un significado bien definido, lo que

permite una mejor cooperación del trabajo entre los ordenadores y las personas" [35]. Teniendo en cuenta lo anterior podemos decir la Web Semántica puede ser un marco de trabajo colaborativo donde las personas y organizaciones pueden incluir y compartir conocimiento.

En la práctica, la idea es enriquecer la información existente mediante la adición de la semántica que representan los recursos de una manera comprensible tanto para los seres humanos, como para las máquinas [36]. Esto hace que la información se pueda descubrir, clasificar y ordenar, a diferencia de los enfoques tradicionales. Además, al tener la información "comprensible" por la máquina, hace que se puedan utilizar agentes de software capaces de consumir y entender la información de manera que permita la inferencia [37].

La arquitectura de la web semántica se puede definir por capas (Figura 1). Fuente. Tim Berners-Lee (2001)

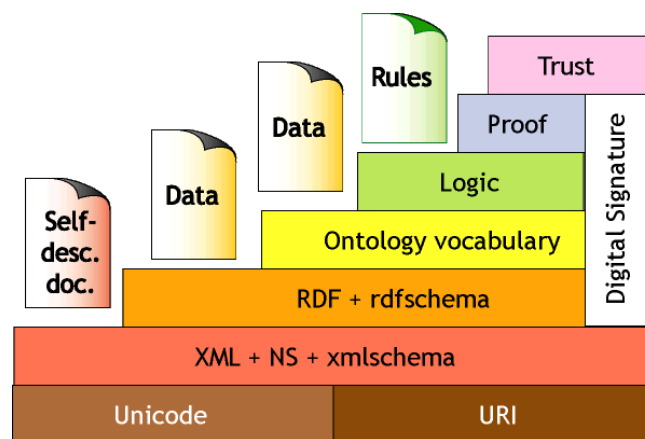


Figura 1. Arquitectura de la Web semántica

- **Unicode:** Codificación para permitir símbolos de diferentes idiomas. De esta manera, se puede expresar información en la Web Semántica en cualquier idioma.
- **URI:** "Uniform Resource Identifier" o Identificador Uniforme de Recursos, permite la localización de un recurso de manera unívoca que puede ser accedido a través de Internet.
- **XML + NS + xmlschema:** En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan interactuar. XML es un lenguaje de marcado que ofrece un formato común para intercambio de documentos, NS (namespaces) sirve asociar elementos de XML con espacios de nombre y XML Schema ofrece un esquema para elaborar documentos estándar.
- **RDF + rdfschemata:** Esta capa define el lenguaje universal con el cual podemos expresar los diferentes conceptos en la Web Semántica. RDF es un lenguaje simple mediante el cual definimos sentencias en el formato de una 3-upla o triple (sujeto: el recurso al que nos referimos; predicado: el recurso que indica qué es lo que estamos definiendo;

y objeto: puede ser el recurso o un literal que podría considerarse el valor de lo que acabamos de definir). RDF Schema provee un esquema de metadatos definido sobre RDF que permite el modelado de objetos con una semántica claramente definida.

- **Lenguaje de Ontologías:** El uso de ontologías permite clasificar objetos y sus relaciones con otros objetos pertenecientes a un dominio concreto del conocimiento. Esta capa permite extender la funcionalidad de la Web Semántica, describiendo recursos a través de nuevas clases y propiedades.
- **Lógica:** Se requieren reglas de inferencia para aplicar “razonadores” sobre las ontologías para que de esta manera las máquinas puedan consultar y manipular términos de manera eficiente haciendo uso de la inteligibilidad humana.
- **Pruebas:** Intercambio de "pruebas" escritas en el lenguaje unificador de la Web Semántica con el fin de posibilitar las inferencias lógicas.
- **Confianza:** Suministran comprobación sobre el uso de fuentes de información en la Web Semántica. (Web Of Trust RDF Ontology -WOT- <http://xmlns.com/wot/0.1/> y FOAF <http://xmlns.com/foaf/0.1/>)
- **Firma digital:** Método de encriptación para verificar que la información adjunta ha sido ofrecida por una fuente específica confiable. (XML Signature WG: <http://www.w3.org/Signature/>)

El desarrollo de la Web semántica necesita utilizar lenguajes estructurados que puedan dotar a los recursos o contenidos, de una lógica y un significado con el fin que puedan visualizados, integrados y reutilizados por parte de los computadores [38].

2.3.2 LENGUAJES DE LA WEB SEMÁNTICA

La mejor manera de aprovechar el uso de la Web Semántica es por medio de lenguajes semánticos capaces de representar el conocimiento a través de metadatos y ontologías. A continuación se presentan algunos de estos lenguajes.

- **Resource Description Framework (RDF).** Es un marco para representar información en la Web [39]. La estructura base de la sintaxis es un conjunto de triples o triplete, cada uno compuesto de un sujeto, un predicado y un objeto. Un conjunto de este tipo de triples se llama grafo RDF. Un gráfico RDF se puede visualizar como un nodo y el diagrama de arco dirigido, en el que cada Triple se representa como un nodo de enlace de arco-nodo (Figura 2). Fuente: <http://www.w3.org>

Pueden haber tres tipos de nodos en un grafo RDF: IRIs, literales y nodos en blanco. Un triple de RDF muestra algún tipo de relación indicada por el predicado, y sostenida entre los recursos indicados por el sujeto y el objeto. Esta declaración correspondiente a un triple de RDF se conoce como una declaración RDF. El predicado en sí es un IRI y denota una propiedad, es decir, un recurso que puede ser pensado como una relación binaria.



Figura 2. Un triplete con dos nodos (Sujeto y Objeto) y conectados (Predicado)

- **Web Ontology Language (OWL)**. El primer nivel por encima de RDF requerido para la Web Semántica es un lenguaje de ontologías lo que se puede describir formalmente el significado de la terminología utilizada en los documentos Web. Si se espera que las máquinas para realizar tareas de razonamiento útiles en estos documentos, el lenguaje debe ir más allá de la semántica básica del RDF Schema [40].

OWL ha sido diseñado para satisfacer esta necesidad de un Lenguaje de Ontologías Web. OWL hace parte de las recomendaciones del W3C relacionadas con la Web Semántica y proporciona tres sublenguajes en diferentes niveles de expresión para el uso específico de los usuarios [41].

- *OWL Lite* es compatible con aquellos usuarios que necesitan sobre todo una jerarquía de clasificación y restricciones simples.
- *OWL DL* apoya aquellos usuarios que desean la máxima expresividad conservando completitud computacional (todas las conclusiones se garantiza que sea computable) y decidibilidad (todos los cálculos terminarán en tiempo finito).
- *OWL Full* está pensado para usuarios que quieren máxima expresividad y la libertad sintáctica de RDF sin garantías computacionales.

La estructura general de un documento OWL se muestra en la Figura 3. Clases, tipos de datos, propiedades de los objetos, propiedades de datos, propiedades de anotación, e instancias de las entidades. Todos los elementos se identifican por medio de una IRI. Las clases representan grupos de individuos; los tipos de datos identifican el dominio de los datos como cadenas o enteros, decimales, etc.; las propiedades de objetos y datos se pueden utilizar para representar relaciones en el dominio; las anotaciones se pueden utilizar para asociar información no lógica con ontologías, axiomas y entidades; y las instancias o individuos pueden usarse para representar objetos reales del dominio [42].

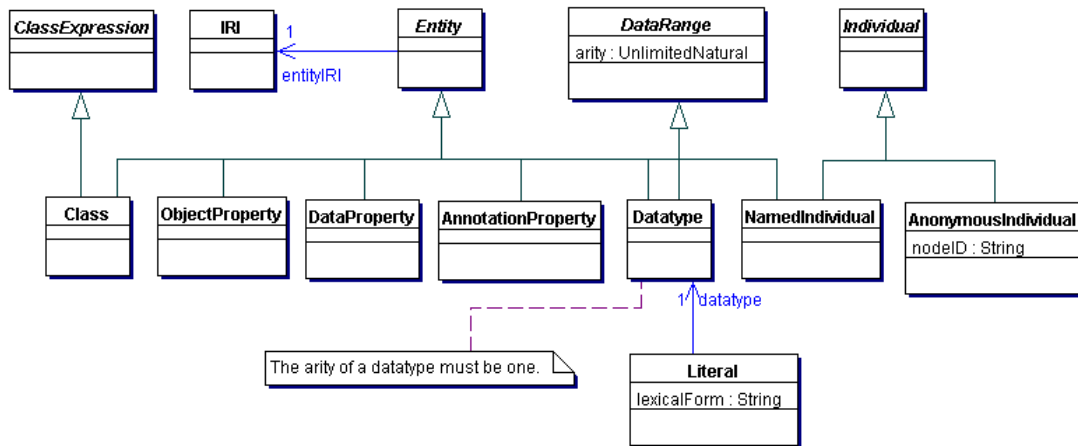


Figura 3. Estructura general de un documento OWL

- **SPARQL Protocol and RDF Query Language (SPARQL)**. Es un lenguaje de consulta que incluye protocolos para consultar y manipular contenido de grafos RDF en la Web o en un almacén de RDF [43].

El lenguaje también se puede utilizar para expresar consultas a través de diversas fuentes de datos, si los datos se almacenan de forma nativa como RDF o como un middleware para ver RDF. Contiene capacidades para consultar los patrones obligatorios y opcionales de los grafos, junto con sus conjunciones y disyunciones. SPARQL también soporta la agregación, subconsultas, la negación, la creación de valores por las expresiones, las pruebas de valor extensible, y restringir las consultas por fuente (Figura 4). Los resultados de consultas pueden ser un conjunto de resultados o grafos RDF [44].

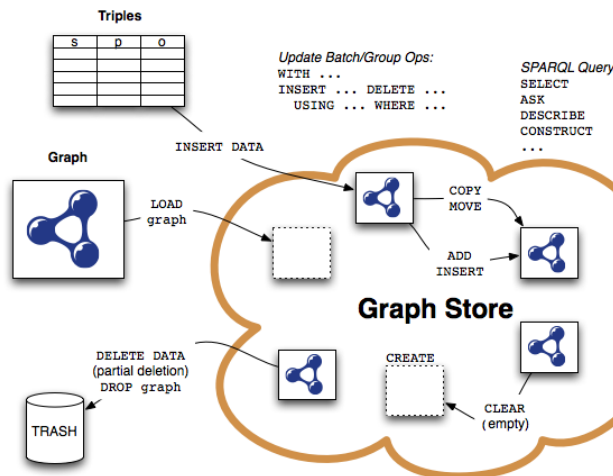


Figura 4. SPARQL consultas y manipulación de grafos

2.3.3 ONTOLOGIAS

La Web semántica se ocupa de dar significado a los contenidos con el fin de mejorar la búsqueda y recuperación de la información mientras que las ontologías proveen las estructuras conceptuales con contenido semántico acerca de un dominio específico que ayudan al desarrollo y diseño de sistemas de información colaborativos [45].

Para la Web Semántica uno de los aspectos más importantes es la estandarización en la representación de los dominios, es por esto que las ontologías surgen como una herramienta para esa representación. La ontología resulta del conjunto de clases o conceptos propios de un área del conocimiento llamado dominio. Por lo tanto, una ontología es el modelo conceptual del dominio modelado [46].

La ontología es la base de la Web Semántica. Adicionalmente a la tarea de identificar los conceptos del dominio se deben determinar las relaciones entre ellos y sus jerarquías. También es importante evitar aplicar malas prácticas al momento de diseñar una ontología. Se deben aplicar las fases de creación de una ontología que incluye [47]:

- Determinar el dominio y alcance de la ontología
- Reutilizar ontologías existentes.
- Enumerar los términos importantes de la ontología.
- Definir las clases y las jerarquías.
- Definir las relaciones entre las clases.
- Definir las propiedades de las clases.
- Crear las instancias.

2.3.4 METODOLOGÍAS PARA EL DESARROLLO DE ONTOLOGIAS

La Metodología NeON para la construcción de redes de ontologías es una metodología que utiliza escenarios que apoyan el desarrollo de ontologías, la reutilización y la evolución dinámica de las redes de ontologías en entornos distribuidos [17].

Las características más importantes son:

- Un conjunto de nueve escenarios para la construcción de ontologías y redes de ontologías donde se valora la reutilización de los recursos ontológicos y no ontológicos, la reingeniería y la fusión, y teniendo en cuenta la colaboración y el dinamismo.
- Un glosario de procesos y actividades que identifica y define aquellos que están involucrados en el desarrollo de redes de ontologías.
- Directrices metodológicas para diferentes procesos y actividades del proceso de desarrollo de la ontología de la red.

Los escenarios permiten utilizar o no las pautas del desarrollo de la red de ontologías:

- Escenario 1: Desde la especificación de la aplicación.

- Escenario 2: La reutilización y reingeniería de los recursos no ontológicos (NOR).
- Escenario 3: La reutilización de los recursos ontológicos.
- Escenario 4: La reutilización y re-ingeniería de los recursos ontológicos.
- Escenario 5: La reutilización y la fusión de los recursos ontológicos.
- Escenario 6: Reutilización, la fusión y re-ingeniería de los recursos ontológicos.
- Escenario 7: Reutilización de los patrones de diseño de ontologías (ODPs).
- Escenario 8: Reestructuración de recursos ontológicos.
- Escenario 9: Localización de recursos ontológicos.

2.3.5 LINKED (OPEN) DATA

La madurez de la tecnología ha permitido la publicación y explotación de bases de datos anotadas semánticamente, dando lugar a la creación de un movimiento más amplio hacia Linked Open Data (LOD).

El paradigma de Linked Data ha evolucionado como un habilitador poderoso para la transición de la Web actual orientada a documentos hacia la Web de datos relacionados entre sí y, en última instancia, hacia la Web Semántica. Linked Data se refiere al conjunto de buenas prácticas para la publicación y la interrelación de datos estructurados en la Web [48].

Actualmente se están creando continuamente un gran número de repositorios anotados semánticamente. En el dominio de los datos de gobierno, cada vez más gobiernos facilitan esta información de manera abierta, con el fin de promulgar la transparencia, la eficacia, la precisión y la participación ciudadana en el ámbito de un gobierno abierto y conectado [49].

Linked Open Data (LOD) significa publicación estructurada de datos en un formato abierto y disponible para ser usada por cualquiera. Esto tiene que ver con la evolución de la conexión y publicación en la Web, de las tecnologías Web y de los datos relacionados que no están siendo vinculados, o que eran vinculados mediante otros métodos. El uso de identificadores únicos de recursos (URI) y del marco para la descripción de recursos (RDF) son importantes debido a que los datos pueden ser relacionados entre sí, creando un gran número de datos, ofreciendo las capacidades de búsqueda, combinación y explotación del conocimiento. Los usuarios pueden incluso navegar entre las diferentes fuentes de datos, siguiendo enlaces RDF, y navegar por una potencialmente interminable Web de fuentes de datos conectados [50].

Los principios básicos de Linked Data son:

- a. Utilizar el modelo de datos RDF para publicar datos estructurados en la Web.
- b. Utilizar enlaces RDF para interconectar los datos de diferentes fuentes de datos.

La aplicación de los principios conlleva a la creación de un activo común de datos en la Web, un espacio donde las personas y las organizaciones pueden publicar y la utilización o consumo de datos por parte de cualquiera. Este patrimonio común es conocido la Web de Datos o Web Semántica. La Web de datos se puede acceder mediante los navegadores para Linked Data, al

igual que se accede a la Web tradicional de documentos utilizando navegadores HTML. En lugar de seguir los vínculos entre las páginas HTML, los navegadores de Linked Data permiten a los usuarios navegar entre las diferentes fuentes de datos siguiendo los enlaces RDF [51].

Así como la Web documento tradicional puede ser rastreado siguiendo los enlaces de hipertexto, la web de los datos puede ser rastreado por los siguientes enlaces RDF. Trabajando en los datos rastreados, los motores de búsqueda pueden proporcionar capacidades de consulta sofisticada, similar a los proporcionados por las bases de datos relacionales convencionales. Debido a que la consulta se traduce en sí son datos estructurados, no sólo vincula a las páginas HTML, que se pueden procesar de inmediato, lo que permite una nueva clase de aplicaciones basadas en la Web de Datos [52].

2.4 TRABAJOS RELACIONADOS

2.4.1 MODELOS DE LA WEB SEMÁNTICA SOBRE OBJETOS PATRIMONIALES

- *Vocabularios y tesauros.* Para ayudar a la síntesis de la información sobre activos patrimoniales, distintas instituciones nacionales e internacionales han definido guías, estructuras e indicadores, adoptando tesauros y vocabularios controlados para la estandarización de términos [53], los cuales pueden mejorar la gestión de información cultural [54]. Los tesauros juegan un rol importante en los sistemas de almacenamiento y recuperación de información.

Un tesauro es un directorio de palabras, organizado alfabética o sistemáticamente, que muestra las relaciones terminológicas entre las palabras (por ejemplo, homónimos, sinónimos, antónimos) dentro de un determinado dominio. Un tesauro puede ser utilizado ya sea intencionalmente o automáticamente para mejorar los términos de búsqueda en un sistema de información [33]. El propósito de los tesauros es evitar la ambigüedad y subjetividad de la documentación sobre los objetos patrimoniales, así como la pérdida de información importante.

Muchos de estos tesauros están disponibles conforme a SKOS [55]. SKOS proporciona estándares para el uso de los sistemas de organización del conocimiento en el marco de la Web Semántica. Así, por ejemplo, SKOS provee una forma estándar para representar los KOS utilizando el Marco de Descripción de Recursos (RDF). La codificación de esta información en RDF permite que esta sea transmitida entre aplicaciones informáticas de una forma interoperable [56]. La interoperabilidad semántica se facilita por el uso de vocabularios comunes en la terminología utilizada en los datos [57].

Existe amplia diversidad de tesauros nacionales como INSCRIPTION-Inglaterra, RCAHMS-Escocia, MiBACT-Italia, etc., internacionales como el AAT, CONA, ULAN, DCMI, ELSST, PACTOLS, entre otros, y relacionados con el espacio geográfico como el TGN, Pleiades, entre otros [58].

Teniendo en cuenta las especificaciones del proyecto de la red PHI, son de interés los vocabularios internacionales del instituto GETYE como son: AAT, CONA y ULAN, y del espacio geográfico es de interés el TGN.

Los cuatro vocabularios Getty: The Art & Architecture Thesaurus (AAT), the Getty Thesaurus of Geographic Names (TGN), the Union List of Artist Names (ULAN), and the Cultural Objects Name Authority (CONA) son vocabularios estructurados que están destinados a proporcionar terminología y otra información sobre objetos, artistas, conceptos y lugares importantes para las diversas disciplinas que se especializan en arte, arquitectura y cultura material [59].

Los dominios de aplicación de los vocabularios Getty son tres: catalogación, recuperación e investigación. En catalogación, los vocabularios Getty pueden ser utilizados como estándares para metadatos, en recuperación pueden ser usados como asistentes de búsqueda en sistemas de recuperación en bases de datos, como bases de conocimiento que incluyen redes semánticas que muestran links y caminos entre conceptos las cuales hacen la recuperación más fácil. Por su parte, los vocabularios Getty pueden emplearse como herramientas de investigación para aprovechar la información y conocimiento contextual que estos contienen [59]. Los vocabularios Getty son recursos compilados, no son comprehensivos, y su construcción es colaborativa, enriqueciéndose de contribuciones de expertos. Los vocabularios Getty son construidos para permitir su uso con Linked Open Data.

- The Art & Architecture Thesaurus (AAT): El enfoque de cada registro de AAT es un concepto. AAT contiene términos genéricos, es decir clases o conceptos pero no sus instancias. Un registro mínimo en AAT contiene una identificación numérica, un término, y una posición en la jerarquía. La mayoría de los campos en los registros de AAT están escritos en inglés, sin embargo, la estructura de la AAT soporta el multilingüismo para los términos y notas descriptivas. El AAT cumple de las normas ISO y NISO, [60].
- The Cultural Objects Name Authority (CONA): El enfoque de cada registro CONA es una obra de arte o de arquitectura. CONA contiene títulos y otra información sobre la arquitectura y obras móviles que incluyen pinturas, esculturas y una amplia gama de objetos. Un registro mínimo en CONA contiene una identificación numérica, título o nombre, creador, objeto / tipo de trabajo, y otros campos. CONA registra obras físicas únicas. Sin embargo, CONA puede incluir obras que nunca se construyeron o que ya no existen, por ejemplo diseños de edificios que no se construyeron o que han sido destruidos. Toda la información está escrita en alfabeto romano, pero la mayoría de los campos están escritos en inglés y soporta multilingüismo en títulos / nombres y notas descriptivas [61].
- The Union List of Artist Names (ULAN): El enfoque de cada registro ULAN es el artista. ULAN incluye nombres propios o denominaciones anónimas, biografías, personas relacionadas o jurídicas, y otra información acerca de artistas y otras personas y grupos involucrados en la creación, distribución, recolección, mantenimiento y estudio del arte y la arquitectura. Los artistas pueden ser tanto particulares (personas) o grupos de individuos que trabajan juntos (entidades corporativas). Un registro mínimo en ULAN contiene una identificación numérica, un nombre, un rol, nacionalidad y fechas. Los nombres pueden estar en varios idiomas, y pueden cambiar con el tiempo [62].
- The Getty Thesaurus of Geographic Names (TGN): El enfoque de cada registro TGN es el lugar. TGN incluye nombres e información asociada sobre los lugares. Los lugares en TGN incluyen entidades políticas administrativas (por ejemplo, ciudades, nacio-

nes) y las características físicas (por ejemplo, montañas, ríos). Se incluyen lugares históricos y actuales. Un registro mínimo en TGN contiene una identificación numérica, un nombre, un lugar en la jerarquía, y un tipo de lugar. La estructura de la TGN soporta el multilingüismo [63].

- **Ontologías.** Existen varios recursos ontológicos elaborados para la integración de información sobre patrimonio cultural, entre las cuales resaltan CIDOC CRM [64] y Europea Data Model – EDM [65]. Las ventajas de utilizar ontologías para la descripción y catalogación de recursos del patrimonio cultural es que permiten modelar el conocimiento heterogéneo sobre el dominio cultural, ayuda en la estandarización de la terminología, facilita la publicación de datos relacionados para su explotación a través de motores de búsqueda y finalmente pueden ser incorporadas a sistemas basados en conocimiento que permitan almacenar conocimiento experto y usarlo para inferencias generación de nuevo conocimiento.

Profundizando sobre CIDOC CRM podemos decir que es una ontología formal diseñada para facilitar la integración, mediación e intercambio de información heterogénea sobre patrimonio cultural. Su objetivo es proporcionar las definiciones semánticas y aclaraciones necesarias para transformar las fuentes de información dispares y localizadas en un recurso mundial coherente, ya sea dentro de una institución, en intranets o en Internet. Para facilitar la integración de recursos de información, provee amplia flexibilidad para lograr compatibilidad con otros sistemas y no impone una solución particular [64].

CRM proporciona un medio de interpretación de la información estructurada de manera que grandes cantidades de datos pueden ser transformados o mediados automáticamente. No proporciona más estructura a la ya definida previamente por los usuarios. La CRM, sin embargo, permite que la información de texto libre para ser integrada con información estructurada [64].

CRM está disponible en idioma inglés como RDF (Resource Description Framework) y contiene las siguientes extensiones de libre acceso (FORTH - Institute of Computer Science, 2011):

- CRMdig: es una ontología y esquema RDF para codificar metadatos sobre los pasos y métodos de producción (procedencia) de los productos de digitalización y representaciones digitales sintéticas como 2D, 3D o incluso modelos animados creados por diversas tecnologías.
- CRMsci: es una ontología formal destinada a ser utilizada como un esquema global para la integración de metadatos sobre la observación científica, las mediciones y procesamiento de datos en ciencias descriptivas y empíricas como biodiversidad, geología, geografía, arqueología, conservación del patrimonio cultural y otros en ambientes de investigación en IT y en librerías de datos.
- CRMinf: es una ontología con propósito similar a CRMsci, sin embargo, ésta se enfoca a facilitar la gestión, integración, mediación, intercambio y acceso a datos acerca de razonamientos sobre de una descripción de las relaciones semánticas entre las premisas, conclusiones y actividades de razonamiento.
- CRMgeo: proporciona una articulación o enlace entre las normas del geo espacio y la comunidad del patrimonio cultural, en particular entre GeoSPARQL y CIDOC CRM.

- CRMarcheo: es una ontología formal que codifica metadatos sobre procesos de excavación arqueológica.

2.4.2 INICIATIVAS DE LINKED DATA RELACIONADAS CON PATRIMONIO

En el contexto de la preservación del patrimonio cultural, múltiples museos y bibliotecas se han puesto en la tarea de almacenar información apoyándose en esquemas de metadatos vinculados a través de Linked Data como los definidos por Dublin Core y POLIS o para la agregación de datos como LIDO y CARARE [66].

La elección de un buen esquema de datos enlazados es un factor crítico en la especificación de una red de ontologías. Por lo tanto, es importante definir el esquema de codificación que se va a utilizar para la identificación de objetos patrimoniales, sus propiedades y posibles valores. Teniendo en cuenta lo anterior, a continuación se revisan los esquemas de metadatos más utilizados para catalogar sitios, monumentos y áreas de carácter patrimonial.

- **Monument Inventory Data Standard (MIDAS)**. Se crea por medio de la comisión de construcciones históricas del Reino Unido con el fin de ofrecer un estándar de datos para la gestión de monumentos. MIDAS es una aplicación de la CIDOC-CRM para el campo particular del patrimonio que permite definir esquemas para monumentos y el inventario de sus elementos, establece conocimientos sobre su pasado, eventos o actividades ocurridos para su conservación o destrucción, también se pueden incluir recursos como publicaciones, mapas, planos, fotografías, entre otros. Además de la gestión de actividades para su mantenimiento se maneja información sobre personas, organizaciones y roles que han intervenido en los objetos patrimoniales [67].

- **Europeana Data Model (EDM)**. Fue diseñado para reemplazar el Europeana Semantic Elements (ESE). EDM ofrece una mayor expresividad y flexibilidad y permite representación más rica y verdaderamente semántica de los millones de objetos de todo tipo de las comunidades de patrimonio cultural en Europeana. Es la generalización de propiedades de metadatos más radical hasta el momento en el área del patrimonio cultural y no constriñe la representación de los metadatos introducidos a un esquema común. EDM integra ontologías bien establecidas como SKOS, Dublin Core y FOAF con el fin de permitir descripciones ricas e interoperables de objetos de Europeana. Como ontología de alto nivel permite integrar las distintas perspectivas de información y las necesidades de las diferentes comunidades que aportan datos a Europeana y conserva la riqueza original de los estándares de dichas comunidades, como LIDO, CIDOC CRM, MARC o EAD [21].

- **Connecting ARchaeology and ARchitecture in Europeana (CARARE Schema)**. Es un esquema basado en estándares existentes en los dominios de arqueología y arquitectura. La fortaleza del esquema es su habilidad para soportar el amplio rango de información descriptiva y detallada de monumentos, edificaciones, áreas paisajísticas y sus representaciones, pero no soporta actividades de gestión y ni protección. Carare está basado en MIDAS Heritage y POLIS DTD para inventario de monumentos. La versión 2.0 incluye elementos de CIDOC CRMdig y 3D ICONS y fue modificada para incluir desarrollos de Europeana. Gestiona información y

documentación base del objeto patrimonial, así como contenido 3D y Realidad Virtual basada en X3DOM [68].

- ***Cultural Heritage Identity Cards (EU CHIC CHICEBERG)***. Es un esquema de gestión enfocado en la recolección y organización unificada de datos sobre activos de patrimonio cultural con el fin de aportar a la efectividad en la toma de decisiones relativa a su preservación, incluyendo los procesos de deterioro natural y las intervenciones humanas. Está subdividido en tres niveles de datos: datos generales, conjunto de conocimientos y datos de apoyo a las decisiones [54]. Constituye también un protocolo para la documentación integrada de patrimonio cultural tangible. El Protocolo ICEBERG para la documentación integrada del patrimonio, está basado en una taxonomía de los edificios históricos desarrollado por el proyecto de EU Perpetuate [69].

- ***Science and Technology in Archaeology Research Centre (STARC)***. Este esquema tiene como objetivo permitir la interoperabilidad de datos y el acceso a recursos digitales almacenados en un repositorio. El conjunto de datos almacenados en STARC incluye datos arqueológicos en 2D y 3D, sitios arqueológicos, objetos de museo y elementos arquitectónicos. Tiene tres niveles: adquisición, procesamiento y publicación [70]. Permite la recuperación de modelos, actividades, decisiones y la interpretación de datos. El esquema se basa en LIDO y CARARE y es compatible con CIDOC CRM. Gestiona información de los proyectos, activos de patrimonio cultural, procedencia de recursos digitales y actividades [71].

- ***Central Institute for Catalogue and Documentation (ICCD)***. El ICCD define las normas para la catalogar y documentar los patrimonios arqueológicos, arquitectura, arte, historia y patrimonio etno-antropológico. Estas normas y herramientas están orientadas exclusivamente a las regiones y patrimonio italiano. Se definen estándares y mejores prácticas para la representación esquematizada de documentación fotográfica, digitalización de las imágenes fotográficas e imágenes gráficas, documentación multimedia y transferencia de datos [72].

2.5 METODOLOGÍAS ÁGILES

Para apoyar el diseño del sistema, se requieren los conceptos, modelos, tecnologías y trabajos relacionados antes mencionados. Adicionalmente, se requiere de un soporte metodológico para la fase de construcción de software. Dado el alcance de este proyecto, se contemplan entonces metodologías ágiles. Como consecuencia de la creación del manifiesto ágil para el desarrollo de software se han establecido una serie de normas y estándares para mejorar el proceso. Dicho documento pone en evidencia la importancia que tienen los individuos y sus relaciones, el funcionamiento del software, la interacción con el cliente y la respuesta al cambio [73].

En estas metodologías la identificación de requerimientos es de suma importancia, teniendo en cuenta al usuario como principal fuente de información para la administración de requisitos, funcionalidades e interacción con el sistema. Los casos de uso son independientes de la interacción con el usuario, por lo que se escriben en términos de la relación que tiene el usuario con el sistema, donde se tienen una entrada por parte del usuario y una respuesta por parte del sistema, lo anterior permite que el usuario se centre en la utilización del sistema [74].

Los casos de uso tienen dos componentes principales: Los diagramas UML y la descripción del caso de uso. Los diagramas especifican gráficamente la interacción entre el usuario y el sistema mientras que la descripción del caso de uso permite detallar qué se requiere y cómo se resolverá [75].

La aplicación de buenas prácticas permitirá tener claridad sobre los casos de uso para su futuro desarrollo [76] (Tabla 3):

BUENAS PRÁCTICAS – CASOS DE USO

La descripción debe hacerse en términos de la interacción de los actores y no en funcionalidades.
El objetivo debe plasmarse en el nombre de caso de uso.
Se debe tener claridad cuáles casos de uso son requisitos previos
Se recomienda no utilizar condicionales o flujos en las interacciones porque muchas veces no se tiene claridad por parte del usuario.
Se debe normalizar la elaboración de los casos de uso y tener un mismo nivel de detalle.
No solapar elementos en el diseño de diagramas con el fin de que sean lo más claro posibles.

Tabla 3. Buenas prácticas para la elaboración de casos de uso

A través del nivel de madurez de los modelos y la comparación de las distintas aproximaciones ágiles con base en tres parámetros: vista del sistema como algo cambiante, tener en cuenta la colaboración entre los miembros del equipo y características más específicas de la propia metodología como son simplicidad, excelencia técnica, resultados, adaptabilidad, se puede obtener la siguiente tabla ilustrativa (Tabla 4) [77].

	<i>CMM</i>	<i>ASD</i>	<i>Crystal</i>	<i>DSDM</i>	<i>FDD</i>	<i>LD</i>	<i>Scrum</i>	<i>XP</i>
<i>Sistema como algo cambiante</i>	1	5	4	3	3	4	5	5
<i>Colaboración</i>	2	5	5	4	4	4	5	5
<i>Características Metodología (CM)</i>								
Resultados	2	5	5	4	4	4	5	5
Simplicidad	1	4	4	3	5	3	5	5
Adaptabilidad	2	5	5	3	3	4	4	3
Excelencia técnica	4	3	3	4	4	4	3	4
Prácticas de colaboración	2	5	5	4	3	3	4	5
Media CM	2.2	4.4	4.4	3.6	3.8	3.6	4.2	4.4
Media Total	1.7	4.8	4.5	3.6	3.6	3.9	4.7	4.8

Tabla 4. Ranking de “Agilidad”

La media total nos indica que existen metodologías más ágiles que otras y su nivel de madurez; entre ellas destacan ASD, Scrum y XP como las más ágiles.

La escogencia de la metodología depende de varios factores entre los que se pueden enumerar cantidad de personas en el equipo, relevancia de la documentación, tamaño del proyecto, variabilidad de requerimientos, etc. Para este proyecto se tendrá en cuenta eXtreme Programming (XP) por la evaluación de cada uno de los ítems y la facilidad con que puede ser efectuada, desde el punto de vista del desarrollo, por un solo individuo [78].

2.5.1 *Extreme Programming (XP)*

Es una metodología de desarrollo ágil de software que tiene la característica de adaptarse a los cambios en requerimientos de los proyectos. Dentro de esta metodología se tienen en cuenta 5 valores: simplicidad, comunicación, retroalimentación, coraje, y respeto, como base para la mejora continua dentro de los proyectos [79]. Los ciclos de desarrollo son muy cortos, permitiendo que se realicen retroalimentaciones prontas, continuas y concretas. Por ser iterativa, también es incremental y ofrece flexibilidad integrándose con las necesidades del negocio obteniendo como resultado mejoras en la comunicación entre todos los stakeholders [80].

2.6 HERRAMIENTAS TECNOLÓGICAS

2.6.1 *Servidores Base de Datos*

- *Microsoft® SQL Server™*. “Es un sistema de gestión de bases de datos relacionales de Microsoft para soluciones de comercio electrónico, línea de negocio y almacenamiento de datos. Proporciona rendimiento, disponibilidad y facilidad de uso innovadores para las integrarse con otras aplicaciones. Microsoft SQL Server ofrece nuevas capacidades en memoria en la base de datos principal para el procesamiento de transacciones en línea (OLTP) y el almacenamiento de datos, que complementan nuestras capacidades de almacenamiento de datos en memoria e inteligencia de negocio” [81].
- *PostgreSQL*. “Es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras bases de datos comerciales. PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando” [82].
- *MySQL*. “Es la base de datos de código abierto de mayor aceptación mundial y permite la oferta económica de aplicaciones de bases de datos fiables, de alto rendimiento y fácilmente ampliables basadas en la web e integradas. Además de la facilidad de uso, el alto rendimiento y la fiabilidad de MySQL, se puede beneficiar de funciones avanzadas, herramientas de gestión y soporte técnico para desarrollar, implementar y gestionar sus aplicaciones MySQL” [83].

2.6.2 Servidores Web

Un servidor web, también llamado servidor HTTP es una aplicación que permite estar instalada del lado del servidor, recibir peticiones HTTP y emitir respuestas hacia el lado del cliente. El código recibido por el cliente suele ser compilado y ejecutado por un navegador web o browser [84].

- Internet Information Server (IIS). Es un potente servidor Web que ofrece una infraestructura de gran fiabilidad, capacidad de manejo y escalabilidad para aplicaciones Web. IIS hace posible que las organizaciones aumenten la disponibilidad de sus sitios y aplicaciones Web [85].
- Apache Tomcat. Es un contenedor web basado en el lenguaje Java que actúa como motor de servlets y JSPs. Se ha convertido en la implementación de referencia para las especificaciones de aplicaciones Web [86].

2.6.3 Frameworks

Es una estructura de software que tiene componentes personalizables e intercambiables para el desarrollo de una aplicación informática. Los objetivos principales de los framework son los de agilizar el proceso de desarrollo, reutilizar código y fomentar buenas prácticas como la implementación de patrones de desarrollo [87].

- .NET. Es un entorno de ejecución administrado que proporciona diversos servicios a las aplicaciones en ejecución. Consta de dos componentes principales: Common Language Runtime (CLR), que es el motor de ejecución que controla las aplicaciones en ejecución, y la biblioteca de clases de .NET Framework, que proporciona una biblioteca de código probado y reutilizable al que pueden llamar los desarrolladores desde sus propias aplicaciones [88].

2.6.4 Triplestore

Un triplestore es una base de datos especialmente diseñada para el almacenamiento y recuperación de triples a través de consultas semánticas. Al igual que una base de datos relacional, la información se almacena en el triplestore y se recupera a través de un lenguaje de consulta. A diferencia de una base de datos relacional, un triplestore está optimizado para el almacenamiento y recuperación de triples. Además de las consultas, generalmente se pueden importar/exportar triples utilizando Resource Description Framework (RDF) y otros formatos [89].

- OpenLink VIRTUOSO. Virtuoso Universal Server es un híbrido de motor middleware y base de datos que combina la funcionalidad de un RDBMS tradicional, ORDBMS, base de datos virtual, RDF, XML, texto libre, servidor de aplicaciones Web y la funcionalidad de servidor de archivos en un solo sistema. En lugar de tener servidores dedicados para cada uno de los reinos de funcionalidad antes mencionados, Virtuoso es un "servidor universal"; que permite a un único proceso de servidor multiproceso que implementa múltiples protocolos. La edición de código abierto de Virtuoso Universal Server y también se conoce como OpenLink Virtuoso [90].

III – DISEÑO Y DESARROLLO

Después de haber definido los objetivos de la solución y el marco conceptual se continua con el paso del diseño y desarrollo de la solución, en el cual se evidencian un diseño del proceso y un diseño de la solución empleado para llevar a cabo el desarrollo del proyecto [91].

3.1 PROCESO DE LA SOLUCIÓN

Se describe el método establecido, los pasos seguidos en el desarrollo, los criterios de evaluación y la iteración que llevaron a la construcción del prototipo final.

3.1.1 MÉTODO

Se hace necesario establecer un esquema metodológico de trabajo para el desarrollo de las aplicaciones, para contar con un instrumento útil con el fin de realizar las actividades que dan soporte al ciclo de vida del software de una manera sistemática. Es así como se establecen unos requerimientos para la metodología con las siguientes premisas:

- Ser ágil, con iteraciones incrementales.
- Permitir permanentemente ajustar el diseño del sistema al cambio en el marco conceptual.
- Construir sobre lo construido (partir del levantamiento de requerimientos y casos de uso previamente realizados)
- Garantizar una documentación mínima que permita lograr la obtención de resultados tempranos, así como la transferencia de conocimiento.
- Ser aplicada para el desarrollo por uno o varios individuos
- Aprovechar las plataformas tecnológicas transversales disponibles para facilitar del desarrollo.

Para el desarrollo de software se ha tenido en cuenta XP como metodología ágil para la aplicación de estándares de la Ingeniería de Software y Calidad, además de referencias en cuanto a seguridad y gestión de proyectos (Tabla 5).

CARÁCTERÍSTICAS	eXtreme Programming
<i>Fases</i>	No existen fases. Se programa la funcionalidad requerida y se prueba.
<i>Iteraciones</i>	Permanentes entre el desarrollo y las pruebas.
<i>Control de cambios</i>	No existe. Se pueden añadir o quitar requerimientos en cualquier momento.

<i>Control del proyecto</i>	Aunque no existe en XP, el proyecto de desarrollo se controla con avances semanales con el director y el asesor del trabajo de grado.
<i>Pruebas</i>	A medida que se va codificando se van realizando las respectivas pruebas, esto permite comprobar la calidad del desarrollo.
<i>Código fuente</i>	La programación se realiza por el responsable del trabajo de grado.
<i>Control del Software</i>	No aplica.
<i>Carga de Trabajo</i>	El programador tiene menos tareas de documentación.
<i>Dedicación al proyecto por parte del usuario</i>	El usuario interactúa permanentemente

Tabla 5. Características del método utilizado.

3.1.2 PASOS DE LA SOLUCIÓN

Teniendo en cuenta la ciencia del diseño y la concepción establecida para la construcción de la herramienta informática que soporte el modelo conceptual se estableció un plan de creación y evolución del sistema, respondiendo a los ciclos de relevancia, diseño y rigor.

Mediante esta premisa se abordaron de forma sistemática y cíclica cada una de las etapas implícitas en el desarrollo e implementación de un sistema de información, buscando iniciar con una conceptualización a largo plazo sobre qué debe ser el sistema (pensar en grande), segmentar la solución propuesta en módulos funcionales (construir en pequeño) de tal forma que la solución se vaya generando a partir de resultados operativos funcionales muy rápidamente. A través de las iteraciones se hicieron ajustes en la medida que el modelo conceptual evolucionaba, se cambiaban las restricciones y se generaban nuevos requerimientos de los usuarios producto de la experiencia [92].

Se definieron los requerimientos a nivel macro y se desarrollaron como se presenta a continuación los siguientes pasos:

- Investigación Marco Teórico. La investigación requirió actividades de búsqueda y recopilación de información a través de bases de datos en línea sobre libros, documentos, artículos e investigaciones realizadas sobre diferentes temas del dominio del problema como base teórica, tecnologías, aplicabilidad, enfoques, limitaciones y herramientas.

Posteriormente se hizo una lectura, interpretación y clasificación de acuerdo a la importancia dentro del proyecto de estas fuentes de datos. Este proceso de depuración implicó nuevamente actividades de búsqueda y recopilación hasta poder consolidar una fuente relevante de información.

Se utilizó Cmap como herramienta para la representación de mapas mentales con el fin de desarrollar de una manera organizada las fuentes de datos e ideas de generación del marco teórico.

Con el estado del arte definido se procedió a realizar una redacción sistemática del marco de referencia incluyendo una contextualización, bases teóricas de la gestión del conocimiento, tecnologías y trabajos relacionados con el fin de utilizarlos en el análisis del problema y que son la base del presente trabajo de grado [93].

- **Análisis Modelo Ontológico.** Para el proceso de crear un modelo para representar el conocimiento se realizaron varios pasos, entre los cuales tenemos:

- *Selección de la fuente de datos.* Como fuente de datos se toma la información suministrada por los expertos en el dominio del patrimonio cultural de la facultad de arquitectura de la Pontificia Universidad Javeriana. Esta información contiene las fichas de metadatos utilizadas para el registro de los casos de estudio según las políticas de gestión información definidas por la Red PHI. Estas fichas son la base para el desarrollo de la ontología, sin embargo en este punto del proceso únicamente se tenían únicamente cuatro casos de estudios registrados en estos formatos, por lo cual fue necesario obtener un mayor número de registros.
- *Obtención de acceso a la fuente de datos.* Para obtener acceso a las fuentes de datos fue necesario acudir de manera formal a la Universidad Politécnica de Madrid, que es la encargada de coordinar el desarrollo del Sistema de Gestión del Conocimiento de la Red PHI. Respondieron la solicitud con la remisión de 54 casos de estudio. Se realizó una conversión de los registros en formato Excel para facilitar su análisis. Se asume que la información contenida en la base de datos es real y está ajustada al esquema de metadatos definido para el desarrollo de la plataforma información.
- *Análisis de las licencias de la fuente de datos.* Fue necesario determinar quién es el titular de los derechos y qué licencias han sido declaradas para su uso con el fin de evitar inconvenientes legales. Se solicitó autorización al coordinador de la Red PHI para el uso de información porque esas políticas de uso se encuentran en construcción.
- *Análisis de la fuente de datos.* Se realizó un análisis de las características cuantitativas de los datos y de sus valores con el fin de determinar que la estructura y organización de los datos estuvieran ajustadas al esquema de metadatos definidos por la Red PHI. También se identificó la completitud de la información y se pudo observar que existen muchos campos que pueden ser importantes pero son opcionales en el modelo haciendo que la información no se llene de manera heterogénea en los registros.

- **Diseño y desarrollo de la ontología.** Se utilizó la metodología NeON para el desarrollo de la ontología. Se definieron los requisitos de la ontología a través de los documentos plantilla para la especificación de requerimientos de ontologías (ORSO) y se procedió a la aplicación de algunas etapas que se identificaron como válidas en el desarrollo del modelo ontológico.

Dentro de la plantilla de requerimientos se establecieron los parámetros necesarios para la realización de la ontología.

- *Propósito.* El propósito de la construcción de la ontología de referencia es proporcionar un modelo de conocimiento consensual del dominio de los objetos patrimoniales (bienes inmuebles).
- *Alcance.* La ontología tiene que centrarse en los casos estudios que se realice a los objetos patrimoniales desarrollados dentro de la red PHI, así como el análisis de la información sobre su uso, material, historia, gestión, sociedad, cultura, ubicación, patología, conservación y medio ambiente. El nivel de granularidad está directamente relacionado con las cuestiones de competencia y términos identificados.
- *Lenguaje de Implementación.* La ontología será implementada en lenguaje OWL.
- *Usuarios Previstos.* Personas, organizaciones nacionales o internacionales que realizan búsquedas de la información sobre bienes patrimoniales. Personas encargadas de la publicación de la información.
- *Usos previstos.* Realizar búsquedas, publicaciones y complementaciones, revisiones, correcciones, aprobaciones y comparaciones.
- *Requerimientos No funcionales.* La ontología debe ser genérica para caracterizar casos de estudio de bienes patrimoniales.
- *Requerimientos Funcionales:* (Tabla 6)

Grupos de preguntas competentes
<p>USOS: FICHA</p> <p>Con cuál información se registra el caso de estudio?</p> <p>Con cuál información se identifica el caso de estudio?</p> <p>Como se clasifica el caso de estudio?</p> <p>Como se geo localiza el caso de estudio?</p> <p>Cuál es la historia del caso de estudio?</p> <p>Con cuál información se identifica la propiedad del caso de estudio?</p> <p>USOS: DIAGNOSTICO</p> <p>Con cuál información se registra el diagnóstico del caso de estudio?</p> <p>Cuál información contiene el diagnóstico del caso de estudio?</p> <p>Como se resume el diagnóstico del caso de estudio?</p> <p>USOS: PROPUESTA</p> <p>Con cuál información se registra la propuesta de intervención del caso de estudio?</p> <p>Cuál información contiene la propuesta de intervención del caso de estudio?</p> <p>Cuáles son los detalles que contiene la propuesta de intervención del caso de estudio?</p>

Tabla 6. Requerimientos funcionales de la ontología.

Dentro de la lista de escenarios planteados por la metodología se seleccionaron y adoptaron los siguientes:

- Escenario 2. Reutilización de recursos no ontológicos. Se realizó una exploración de los recursos no ontológicos y se propuso una clasificación inicial teniendo en cuenta que la unidad de análisis para el proyecto PHI es el caso de estudio. Donde el caso de estudio se compone de al menos un objeto de estudio y de un diagnóstico asociado. El objeto de estudio puede definirse como un área geográfica o urbana, un objeto o conjunto de objetos arqueológicos, o una construcción arquitectónica.

Un caso de estudio completo estará conformado por tres modelos de fichas que se usan para capturar datos de: El objeto de estudio, el diagnóstico asociado al objeto de estudio y la propuesta asociada al diagnóstico.

- Escenario 3. Reutilización de recursos ontológicos. Se realizó una búsqueda de recursos ontológicos existentes más importantes y se encontraron las siguientes ontologías en formato OWL: Getty, DBpedia, EDM, Protonu, Erlangen, Foaf y Ausnc_md_model
- Escenario 6. Reutilización, la fusión y re-ingeniería de los recursos ontológicos. Se reutilizaron ontologías existentes, se combinaron y reorganizaron estos recursos-ontológicos. Getty, DBpedia, EDM, Erlangen y Foaf
- Escenario 8. Reestructuración de recursos ontológicos. Se reestructuraron los recursos ontológicos, puesto que se realizaron podas a las ontologías reutilizadas y se integraron a la red de ontologías.

- **Identificación de requerimientos para la solución informática.** Se realizó un análisis con base en las expectativas de los miembros de la red PHI y los alcances versus limitaciones tecnológicas para establecer los requerimientos no funcionales, los funcionales y de calidad. El documento de requerimientos del sistema se encuentra publicado con el presente trabajo de grado (Tabla 7).

TIPO DE REQUERIMIENTOS	ELEMENTOS
NO FUNCIONALES	RNFU_01_Ayuda RNFU_02_AccesoMenu RNFU_03_GuiaDeDiseño RNFU_04_Concurrencia RNFU_05_AlmacenamientoDatos RNFU_06_MostrarInformacion
FUNCIONALES	RF_01_Recibir Notificaciones RF_02_Desplega_Menú RF_03_Diligenciar_Información RF_04_Conexión_BaseDatos RF_05_Obtener_Esquema_BaseDatos RF_06_RealizarMapeoIndirecto

	RF_07_RealizarMapeoDirecto RF_08_ConciliarOntologia RF_09_EfectuarConfiguracionAdicional RF_10_CrearArchivosRDF RF_11_PublicarArchivoRDF RF_12_BuscarInformacion RF_13_GeolocalizarInformacion RF_14_CompararInformacion
DE CALIDAD	QA_01_Escalabilidad QA_02_SistemaOperativo QA_03_Funcionamiento QA_04_TrazaDeErrores QA_05_Concurrencia QA_06_ConfiabilidadSolicitudes QA_07_Interoperabilidad

Tabla 7. Identificación de Requerimientos

- *Diseño y desarrollo la solución informática.* Con base en los requerimientos definidos se desarrollan actividades de extracción sobre bases de datos, conciliación con modelos ontológicos, generación y publicación de grafos y una explotación de esta información como se muestra en la figura 5.

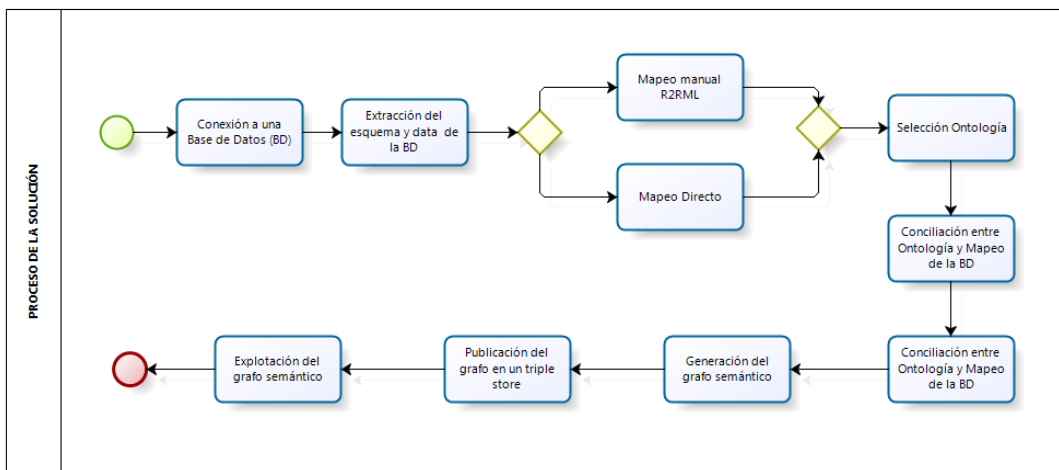


Figura 5. Diseño y desarrollo de la solución

- Extracción esquema y datos de Bases de Datos en RDF y/o OWL. La información de los casos de estudio de la Red PHI fue remitida en una base de datos por lo que para obtener el esquema se hace necesario realizar una extracción a nivel lógico de las relaciones de las tablas y sus elementos. La generación en RDF implica tomar cada uno de los registros existente en la base de datos y generar una correspondencia con el sistema de triples (sujeto, predicado, objeto). Se identificó que el proceso podría hacerse a través de 2 métodos, uno manual definiendo la estructura por parte del usuario u otro directo haciendo uno del esquema de la base de datos.

Cualquiera de las dos maneras implicó la definición de una IRI que fue tomada como base para la identificación de cada uno de los elementos resultantes. Como este algoritmo ya existe y está definido por la W3C se hizo reutilización de librerías para este proceso. Sin embargo, este paso tiene el inconveniente que el modelo de la ontología o el RDF es sobre la base del esquema de la base de datos.

- Conciliación entre Modelo Ontológico y el RDF inicial. Después de la generación del RDF inicial se debía conciliar los elementos en los que se basa su modelo que son elementos propios de la base de datos y del modelo ontológico que representa el conocimiento del dominio. Este paso implicó determinar e identificar relaciones entre los dos modelos, se relacionaron clases con tablas y propiedades de datos con campos.

A un nivel más de detalle se identificó que se debe establecer cuál elemento tabla tiene su correspondiente elemento clase y cuál elemento campo tiene su correspondiente propiedad de datos. Este proceso inicialmente se diseñó de manera manual porque se trabajó con un modelo de base de datos muy reducido para las pruebas.

Cuando se contó con el modelo de base de datos real este proceso se volvió ineficiente. A partir de este momento se determinó que el conocimiento modelado en la ontología podía servir para relacionar de manera automática los elementos.

Se escogieron técnicas de pareo de cadenas donde se pudo determinar un grado de simetría entre dos cadenas comparadas. Teniendo los nombres de todos los elementos se pudo hacer uso de esta técnica. Sin embargo, el pareo de los nombres no fue suficiente porque en la realización de las pruebas se identificó que no se generaban pareos entre elementos existentes y que con conocimiento del dominio y del esquema eran equivalentes.

Teniendo en cuenta lo anterior se hace uso de los elementos de anotación que tiene la ontología para cada uno de sus elementos y también se le aplican las técnicas de pareo para mejorar el resultado de las equivalencias. Este paso implicó también ajustar el paso de la construcción de la ontología enriqueciendo su anotación.

- Generación Grafo Semántico. Una vez conciliado el modelo ontológico y el mapeo directo, se procedió a la generación del grafo semántico en formato RDF para posteriormente ser publicado en un almacén de triples. Para este proceso se tomó el archivo

de la ontología como base y se incluyeron los triples del grafo generado con el fin de mantener los elementos correspondientes instanciados.

- Publicación Grafo Semántico. La publicación del grafo resultante se realizó en un almacén de triples, previamente instalado y configurado. Se adicionó la IRI general que trabajará el grafo y que fue definida en la generación del mapeo directo. Este almacén se recomienda que se encuentre disponible en la red de Internet y que no se encuentre en máquinas locales.
- Explotación Grafo Semántico. Para la explotación de los datos se utilizaron interfaces de búsqueda, geolocalización y comparación para cada uno de los elementos del dominio de la red PHI como son casos de estudio de objetos patrimoniales, sus diagnósticos y propuestas asociadas.

Se realizó una interfaz que permitiera a los usuarios encontrar relaciones a través de cualquier elemento o de información contenida en los mismos. Los resultados se integraron con la Web de despliegue que tiene la Red PHI. Las consultas se evaluaron a través del almacén de triples que contiene el archivo, realizando conexiones desde la interfaz a los endpoint de consulta.

Los datos geográficos son aprovechados por medio de una segunda interfaz que implementó la visualización de la posición de los elementos en el mapa y utilización de funcionalidades externas para integrar los datos geográficos con las funcionalidades propias de los servicios ofrecidos por proveedores de mapas como son rutas, Street View, sitios de interés entre otros.

La tercera interfaz que se desarrolló para comparar información en lenguaje natural utilizando técnicas de anotación y enriquecimiento semántico entre dos elementos del esquema de la red PHI. Esta comparación se realizó entre dos elementos seleccionados previamente e información que contienen. A la información que se va a comparar y se le aplicaron técnicas de eliminación de palabras no relevantes, traducción para realizar la comparación en un lenguaje unificado y una anotación semántica con el fin de comparar la información entre los elementos seleccionados.

El resultado de la comparación se diseñó con el objetivo de navegar a través de recursos encontrados en Internet. Por último se realizó una presentación del detalle de los resultados de la comparación de los elementos.

3.2 DISEÑO DE LA SOLUCIÓN

El diseño de la solución integra elementos de arquitectura, requerimientos, diagramas, componentes, patrones de diseño, algoritmos y la tecnología utilizada.

3.1.1 ARQUITECTURA

La arquitectura y el modelo de diseño del Sistema están definidos por las vistas arquitectónicas Conceptual, Lógica y Física, su representación se realiza a través de los diferentes diagramas UML.

- **Vista Conceptual.** La vista conceptual se utiliza con el fin de los requerimientos funcionales y la visión que los usuarios del negocio tienen de la aplicación y se describe el modelo de negocio [94].

El diagrama Contextual del sistema permite identificar algunos artefactos claves requeridos para la construcción de la solución completa. El flujo de información entre el sistema a ser construido y cada elemento proporciona insumos claves para el modelo de información. Las características de los elementos determinan la necesidad de adaptadores que facilitarían la integración de la tecnología. Los flujos de información también representan actividades importantes que se remontan a los modelos de procesos de negocio (Figura 6).

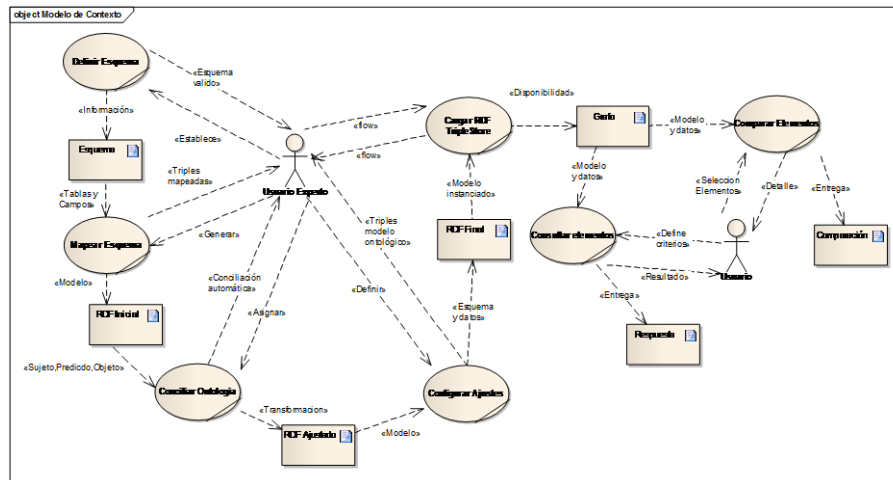


Figura 6. Diagrama de contexto

- **Vista Lógica.** Para dar mayor claridad a la vista lógica se utiliza el Modelo Vista controlador (MVC). Los principales componentes que conforman las capas de presentación, de negocios y de datos; además, permite identificar la interrelación entre los componentes de cada capa y cómo ocurre el flujo de información entre ellas (Figura 7).

La arquitectura de 3 capas tiene las siguientes ventajas:

- La interfaz del cliente no es requerida para comprender o comunicarse con el receptor de los datos. Por lo tanto, esa estructura de los datos puede ser modificada sin cambiar la interfaz del usuario.
- El código de la capa intermedia puede ser reutilizado por múltiples aplicaciones si está diseñado en formato modular.

- La separación de roles en tres capas, hace más fácil reemplazar o modificar una capa sin afectar a los módulos restantes.

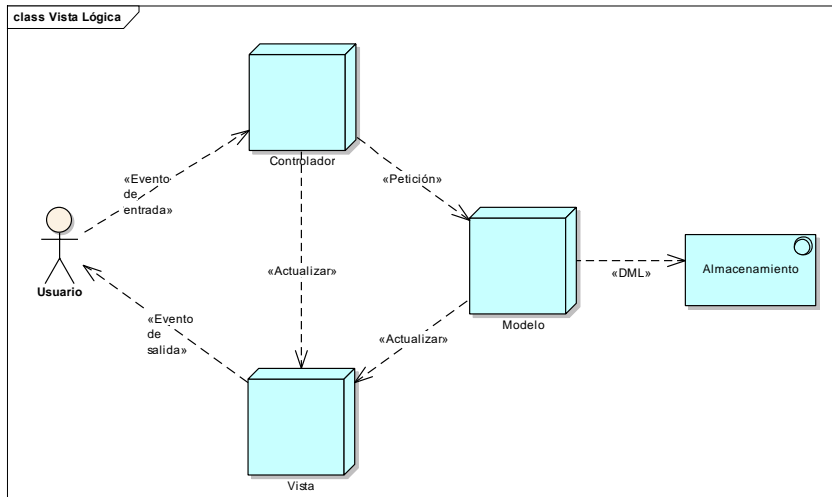


Figura 7. Vista lógica

- **Vista Física.** La vista física, permite identificar las relaciones físicas de los distintos nodos que componen el sistema, la distribución y configuración de hardware en dichos nodos. También se muestran los protocolos de comunicaciones y los métodos de acceso que utilizan sus componentes (Figura 8).

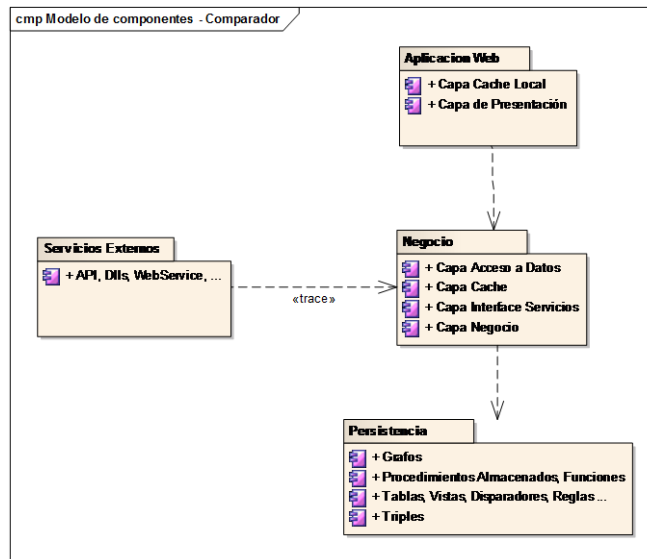


Figura 8. Modelo de componentes del sistema

3.1.2 PATRONES DE DISEÑO IMPLEMENTADOS

Para realizar el proceso de desarrollo de software es necesario implementar patrones de diseño que nos permitan estandarizar métodos y unificar la manera de resolver los problemas. Estos elementos realmente son modelos que son pautas para que los desarrolladores los utilicen [95].

Dentro de los objetivos básicos que debe cumplir un patrón tenemos la estandarización del lenguaje entre programadores, aprovechamiento del tiempo en soluciones conocidas y la creación de código reutilizable.

Se utilizaron los siguientes patrones de diseño que hace parte del anexo Patrones de Diseño adjunto en la documentación del presente trabajo.

- Modelo vista controlador (MVC)
- Front Controller
- Composite
- Cadena de responsabilidad (Chain of Responsibility)
- Fabrica abstracta (Abstract factory).
- Método plantilla (Template Method).
- Estrategia (Strategy).

El documento de los patrones de diseño implementados en el sistema se encuentra publicado con el presente trabajo de grado.

3.1.3 ALGORITMOS

A continuación se listan los algoritmos más importantes utilizados para resolver diversos problemas que se presentaron en el desarrollo del proyecto [96]. La mayoría de algoritmos se reutilizaron de propuestas existentes y en otros casos se diseñaron instrucciones propias que hacen parte de las contribuciones del proyecto.

Los algoritmos diseñados se presentan en pseudocódigo por las ventajas en cuanto a que se presentan en un lenguaje natural, es ordenado, fácil de entender y puede codificarse posteriormente en cualquier lenguaje de programación [97], para los demás algoritmos se presenta la fuente oficial.

- ***RDB to RDF Mapping Language (R2RML)***. Algoritmo que permite expresar asignaciones personalizadas de bases de datos relacionales en conjuntos de datos RDF. Las asignaciones permiten mapear el modelo de una base de datos relacional en una estructura y vocabulario definido por el autor. Las asignaciones R2RML en sí mismas también son grafos RDF y se escriben utilizando la sintaxis del lenguaje turtle [98].

- ***A Direct Mapping of Relational Data to RDF***. Algoritmo que permite exponer datos de bases de datos relacionales (RDB) en la Web Semántica. El mapeo directo consiste en transformar datos relacionales en RDF. El mapeo directo toma como entrada una base de datos relacional

(datos y el esquema) y genera un grafo RDF que se llama Direct Mapping. El algoritmo convierte tablas en clases, campos en data properties y las relaciones entre las tablas en object properties [99].

- **Conciliación Ontología.** Este algoritmo permite conciliar y asociar los elementos de la ontología con los elementos del esquema de la base de datos. La conciliación consiste en verificar coincidencia de los nombres de los elementos y sus descripciones para asociarlos como posibles elementos compatibles y de esta manera tener cada elemento de la ontología con su correspondiente elemento en el esquema de la base de datos. El algoritmo aplica para asociar tipos de elementos así:

- Clases con Tablas
- Propiedades de datos con Campos

Posteriormente a la ejecución del algoritmo se reemplazan las asociaciones o coincidencias en el Direct Mapping realizado (Tabla 8).

ALGORITMO 1. SECUENCIAL

<u>CONCILIAR ONTOLOGIA</u>	
1	ABRIR ARCHIVO ONTOLOGIA
2	PARA CADA ELEMENTO DE LA RAIZ
2.1	IDENTIFICAR EL TIPO DE ELEMENTO DE LA ONTOLOGIA
2.2	LLAMAR ALGORITMO <u>ASIGNAR ELEMENTO A LISTA</u>
2.3	LLAMAR ALGORITMO <u>CARGAR ELEMENTO ESQUEMA BD</u>
3	FIN PARA

ALGORITMO 2. RECURSIVO

<u>ASIGNAR ELEMENTO A LISTA</u>	
1	SI ELEMENTO TIENE ELEMENTOS HIJOS
1.1	PARA CADA ELEMENTO DE LA RAIZ
1.1.1	LLAMAR ALGORITMO <u>ASIGNAR ELEMENTO A LISTA</u>
1.2	FIN PARA
2	SINO
2.1	EXTRAER INFORMACION DEL ELEMENTO (ETIQUETAS Y NOMBRE RECURSO)
2.2	ADICIONAR INFORMACION A LA LISTA DEL TIPO DE ELEMENTO IDENTIFICADO
3	FIN SI

ALGORITMO 3. SECUENCIAL		
<u>CARGAR ELEMENTO ESQUEMA BD</u>		
1	PARA CADA ELEMENTO DEL ESQUEMA	
1.1	ASOCIAR ELEMENTO DEL ESQUEMA CON ELEMENTO DE LA ONTOLOGIA	
1.2	IDENTIFICAR CORRELACION DE ELEMENTOS	
1.3	SI LOS ELEMENTOS SE CORRELACIONAN	
	1.3.1	APLICAR TECNICAS STRING MATCHING (ETIQUETA Y NOMBRE) <ul style="list-style-type: none"> - Levenshtein Distance - Jaccard Distance - Ratcliff Obershelp Similarity - Longest Common Substring
	1.3.2	SI LA TÉCNICA INDICA MATCH CON EL ELEMENTO
		1.3.2.1 ASOCIAR ELEMENTOS
	1.3.3	FIN SI
	1.4	FIN SI
2	FIN PARA	

Tabla 8. Algoritmos de conciliación ontología y esquema.

- Relacionar clases con campos del Esquema. Cuando en la conciliación de la ontología no se encuentran correlaciones (automáticas o semiautomáticas) entre clases y alguna de las tablas del esquema de la base de datos y el experto no la relaciona manualmente es posible que la clase se encuentre dentro de uno o varios campos de una tabla, ya sea por el diseño de la ontología o por el diseño del modelo de la base de datos.

De esta manera es importante asociar ese conjunto de campos con una clase y determinar cuál es la relación que se tiene con la clase “padre” relacionada con la tabla que tiene estos campos. Inicialmente se aplica la el algoritmo de la conciliación de la ontología pero entre clases y campos del esquema que no han sido asociados previamente. Posteriormente de la conciliación entre estos dos tipos de elementos se hace necesario identificar la relación (object property) y transformar el Direct Mapping para que modifique los conjuntos de campos por una nueva clase y la relación que tiene con su clase padre (Tabla 9).

ALGORITMO 1. SECUENCIAL		
<u>TRANSFORMACIÓN DE CAMPOS POR CLASES</u>		
1	PARA CADA CLASE MAPEADA CON UN CAMPO	
	1.1	BUSCAR LA CLASE ORIGINAL QUE ESTA MAPEADA CON LA TABLA DEL CAMPO
	1.2	ASIGNAR A UNA LISTA LAS TRIPLES DE ESA CLASE ORIGINAL QUE CONTENGA EL CAMPO MAPEADO
	1.3	PARA CADA TRIPLE DE LA LISTA
	1.3.1	ELIMINAR LA TRIPLE ENCONTRADA
	1.3.2	CREAR UNA TRIPLE PARA LA RELACION

	1.3.3	CREAR UNA TRIPLE PARA LA INSTANCIA
	1.3.4	PARA CADA DATAPROPERTY DE LA CLASE MAPEADA
	1.3.4.1	ASOCIARLO CON EL NUEVO RECURSO
	1.3.5	FIN PARA
1.4		FIN PARA
2		FIN PARA
3		ABRIR EL GRAFO BASE
3		PARA CADA TRIPLE DE TODO EL CONJUNTO
	3.1	ASIGNAR EL TRIPLE AL GRAFO
4		FIN PARA
5		CREAR ARCHIVO RDF

Tabla 9. Algoritmo Relacionar clases con campos del Esquema.

- **Comparación semántica.** La comparación semántica también implicó el diseño de un algoritmo que pudiera tomar los elementos a comparar y aplicarles una serie de ajustes al lenguaje natural como la eliminación de StopWords, la traducción al mismo idioma y por último un enriquecimiento semántico de las palabras o frases de los textos comparados con el fin de verificar coincidencias (Tabla 10).

ALGORITMO 1. SECUENCIAL

COMPARACIÓN SEMÁNTICA

1		SELECCIONAR DOS ELEMENTOS A COMPARAR
2		ELIMINAR LOS STOPWORDS DE LOS DOS ELEMENTOS
3		TRADUCIR LOS TEXTOS SIN LOS STOPWORDS DE LOS DOS ELEMENTOS
4		PARA CADA TEXTO DEPURADO Y TRADUCIDO
	4.1	OBTENER LA LISTA DE PALABRAS
	4.2	PARA CADA PALABRA DE LA LISTA
	4.2.1	BUSCAR LA PALABRA COMO RECURSO EN DBPEDIA
	4.2.2	SI ENCUENTRA RECURSO
	4.2.2.1	OBTIENE SUB ELEMENTOS DEL RECURSO
	4.2.2.2	ENRIQUECE ELEMENTO ACTUAL CON EL RECURSO
	4.2.3	FIN DE SI
4.3		FIN PARA
5		FIN PARA
6		PARA CADA RECURSO ENCONTRADO DEL ELEMENTO A
	6.1	COMPARAR EL RECURSO CON CADA UNO DE LOS RECURSOS DEL ELEMENTO B
	6.2	SI ENCUENTRA
	6.2.1	INCLUYE ESTADÍSTICA
6.3		FIN SI

8 | FIN PARA

9 | DETALLA NIVEL DE CONCORDANCIA

Tabla 10. Algoritmo Comparación semántica.**3.1.4 HERRAMIENTAS UTILIZADAS**

Se listan las herramientas tecnológicas utilizadas para cumplir con los objetivos del proyecto. (Tabla 11)

TECNOLOGÍAS	HERRAMIENTAS
Servidores Base de Datos	<ul style="list-style-type: none"> • MS Sql Server • PostgreSQL • MySQL
Servidores Web	<ul style="list-style-type: none"> • Internet Information Server (IIS) • Apache Tomcat
Plataformas	<ul style="list-style-type: none"> • Azure • Linux • Windows
Framework	<ul style="list-style-type: none"> • .NET
Lenguajes de Programacion	<ul style="list-style-type: none"> • Javascript • C# .NET
Triplestore	<ul style="list-style-type: none"> • Vistutoso Openlink Server
Libererías Externas	<ul style="list-style-type: none"> • DataBase Schema Reader. • DotNetRDF. • Fuzzy String. • API Google Maps. • API DBpedia. • API MS Translator.

Tabla 11. Lista de herramientas utilizadas**3.1.3 CRITERIOS DE EVALUACIÓN E ITERACIÓN**

- **Marco Teórico.** Para el caso del estado del arte se utilizaron criterios de evaluación como matrices de oportunidades hasta un tercer nivel. De esta manera se iteró 3 veces y en cada iteración se seleccionaban los elementos con un puntaje superior a la media, los demás elementos de desechaban (Tabla 12).

- **Modelo Ontológico.** El proceso de construcción de la ontología fue iterativo donde se analizaban los elementos con el experto del dominio de la Red PHI y un experto del dominio ontológico para ir depurando el modelo constantemente a través de revisiones. El modelo inició con un nivel de abstracción alto que hizo la ontología tuviera clases y propiedades que a medida

que se avanzaba en el trabajo se identificaban como innecesarias o parte de otros elementos. El modelo final es genérico y permite representar el conocimiento del dominio fácilmente y de manera concreta.

TECNOLOGÍAS / U. ANÁLISIS	Individual	Grupo Comunidad	Organizacional	Inter- Organizacional
Marcos gestión del conocimiento	✘ 0,11	⚠ 0,53	✔ 0,86	✔ 0,93
Sistemas basados conocimiento	✘ 0,26	⚠ 0,53	✔ 0,86	✔ 0,93
Minería de datos	✘ 0,11	⚠ 0,53	✔ 0,86	✔ 0,79
La información y la comunicación	✘ 0,26	⚠ 0,53	✔ 0,86	✔ 0,93
Inteligencia Artificial / S. Expertos	✘ 0,11	⚠ 0,53	✔ 0,86	✔ 0,93
Bases de datos	✘ 0,26	⚠ 0,53	✔ 0,86	✔ 0,93
Modelado	✘ 0,26	⚠ 0,53	✔ 0,79	✔ 0,71

Tabla 12. Ejemplo de aplicación de criterios de primer nivel.

Se utilizó el documento plantilla de especificación de requerimientos para la creación de ontologías de la metodología utilizada. Dentro de los escenarios planteados por la metodología se aplicaron y adoptaron los siguientes:

- Escenario 2. Reutilización de recursos no ontológicos.
- Escenario 3. Reutilización de recursos ontológicos
- Escenario 6. Reutilización, la fusión y re-ingeniería de los recursos ontológicos.
- Escenario 8. Reestructuración de recursos ontológicos.

- **Extracción esquema y datos de Bases de Datos en RDF.** Los criterios utilizados en este paso se verificaron contra la completitud tanto de los elementos del esquema como de los registros de datos encontrados. De esta manera, se obtiene una tabla comparativa con los porcentajes de evaluación y cumplimiento (Tabla 13).

TIPO DE ELEMENTO	TOTAL	UTILIZADOS
Tablas	10	100%
Campos	184	100%
Registros	347	100%
Datos con información	581	100%
Datos sin información	539	100%

Tabla 13. Porcentaje de elementos utilizados en la generación del mapeo directo

Dentro de la aplicación del transformador se presenta también un detalle de los elementos generados, los que tienen y no tienen información en la base de datos. Estos elementos que no tienen valores son porque son opcionales y no fueron diligenciados pero en la generación el valor del objeto en la tripleta quedan en blanco (Figura 9).

Detalle de la Generación		
No. Total: 1120	Con Información: 581	Sin Información: 539
<input type="button" value="Cerrar y Continuar"/>		
Triples		
Sujeto	Predicado	Objeto
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://wgvubuntu01.c
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	34
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	Universidad Politécnica
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	Laura Gilibert Sansalva
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	gilibert.laura@gmail.com
http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	http://wgvubuntu01.cloudapp.net:8080/redphi/catalogo_propuesta...	2014-10-16

Figura 9. Detalle de la generación de triples directos

- **Selección de técnicas String Matching.** De las técnicas existentes y permitidas por la librería FuzzyString se deben seleccionar las más eficientes por lo que se decide realizar una prueba individual y de los resultados obtenidos se seleccionan las técnicas con mayor efectividad y que no presenten inconvenientes en su utilización. Se comparó un palabra contra 20 palabras (19 diferentes y una igual a la comparada) haciendo uso de las técnicas, el resultado de la comparación arrojó una lista de posibles palabras parecidas o iguales. Dentro del resultado de todas las técnicas se encontró la palabra igual comparada, pero para la selección de las técnicas se eligieron las que menos palabras devolvieron (Tabla 14).

La efectividad se calculó por el menor número de palabras seleccionadas sobre la cantidad de palabras y la formula aplicada fue $1 - \left(\frac{\#palabras\ encontradas}{\#total\ de\ palabras} \right) * 100$.

TECNICA	PALABRAS	EFFECTIVIDAD
<i>Hamming Distance</i>	9	55%
<i>Jaccard Distance</i>	4	80%
<i>Jaro Distance</i>	5	75%
<i>Jaro Winkler Distance</i>	4	80%
<i>Levenshtein Distance</i>	3	85%
<i>Longest Common Subsequence</i>	4	80%
<i>Longest Common Substring</i>	4	80%
<i>Normalized Levenshtein Distance</i>	3	85%
<i>Overlap Coefficient</i>	6	70%
<i>Ratcliff Obershelp Similarity</i>	4	80%
<i>Sorensen Dice Distance</i>	8	60%
<i>Tanimoto Coefficient</i>	8	60%

Tabla 14. Comparación de técnicas string matching

Para las técnicas con una eficiencia del 80% o superior se hicieron pruebas para aplicarlas en conjunto y la combinación de las técnicas Jaccard Distance, Levenshtein Distance, Longest Common substring y Ratcliff Obershelp Similarity no produjo lentitud en el procesamiento ni

bloqueos en su ejecución por lo que fueron las técnicas seleccionadas para la realización del pareo entre elementos.

- **Conciliación entre Modelo Ontológico y el RDF inicial.** El uso de las técnicas de pareo de cadenas en la conciliación hace necesario crear un criterio de selección en la visualización de los resultados obtenidos. De esta manera se presentan los criterios utilizados en este paso del proceso (Tabla 15).

Se debe aclarar que el criterio la selección de las técnicas de pareo se hizo con base en la recomendación de expertos y estudios realizados [100] [101] [102].

CRITERIO	ASIGNACIÓN	CANTIDAD
Pareo Exacto	Verde	Un único elemento seleccionado del total de elementos.
Algunos Pareos	Naranja	Más de un elemento encontrado pero menos del 30% del total de elementos totales
Múltiples pareos o No existe Pareo	Rojo	Más del 30% de elementos encontrados o ningún elemento encontrado por lo que se muestra el 100% de los elementos.

Tabla 15. Identificación de equivalencia y cantidad de elementos resultantes.

Adicionalmente en las conciliaciones iniciales y pruebas se identificaron que elementos del modelo ontológico y elementos del grafo inicial no encontraban su correspondencia. Esto se debía a que la correlación creada entre clases – tablas y propiedades de datos – campos no fuera suficiente debido a que el modelo de la base de datos no necesariamente va a generar un mapeo exacto con cada uno de los niveles del modelo de la ontología.

Se hizo necesario iterar a un segundo nivel de transformación para relacionar elementos entre niveles y encontrar su posible correspondencia. En este paso el sistema hace recomendaciones pero es el experto del dominio quien aprueba las asignaciones, las modifica, resetea o elimina como crea conveniente.

La transformación de segundo nivel también presenta la dificultad para la identificación de la correlación de los subelementos del elemento asignado con los demás elemento restantes sin una correlación. Por eso se realiza un tercer nivel de transformación donde nuevamente se aplican técnicas de pareo de cadenas para identificar y proponer correlaciones automáticas que el experto edita si es necesario.

Todas las transformaciones se evalúan en sí mismas para comprobar el grado de eficiencia en la selección de pareos automáticos y como se genera una evolución en la conciliación entre los dos modelos (Tabla 16).

GRADO DE EFICIENCIA	DIRECT MAPPING	CONCILIACIÓN DE LA ONTOLOGIA		
		NIVEL 1	NIVEL 2	NIVEL 3
Alto	13,22%	21,49%	25,62%	46,28%
Medio	9,92%	39,67%	45,45%	35,54%
Bajo	76,86%	19,83%	8,26%	7,44%
Incorrecto	0,00%	10,74%	5,79%	3,31%
Falso positivo	0,00%	8,26%	14,88%	7,44%
TOTAL	100,00%	100,00%	100,00%	100,00%

Tabla 16. Grado de eficiencia en la aplicación de las transformaciones

El grado de eficiencia se mide teniendo las siguientes premisas de evaluación:

- *Alto*. El valor mostrado concuerda perfectamente.
- *Medio*. Dentro de la lista de valores filtrados se encuentra el valor correcto.
- *Bajo*. Dentro de la lista completa se muestra el valor correcto.
- *Incorrecto*. Dentro de la lista de valores filtrados NO se encuentra el valor correcto.
- *Falso Positivo*. El único valor mostrado NO es el correcto.

- **Grafo Semántico**. La evaluación de este proceso se realiza a través de herramientas semánticas existentes, en donde el archivo generado debe abrir sin errores, poder mostrar la ontología y la instanciación de los elementos de la ontología (Figura 10).

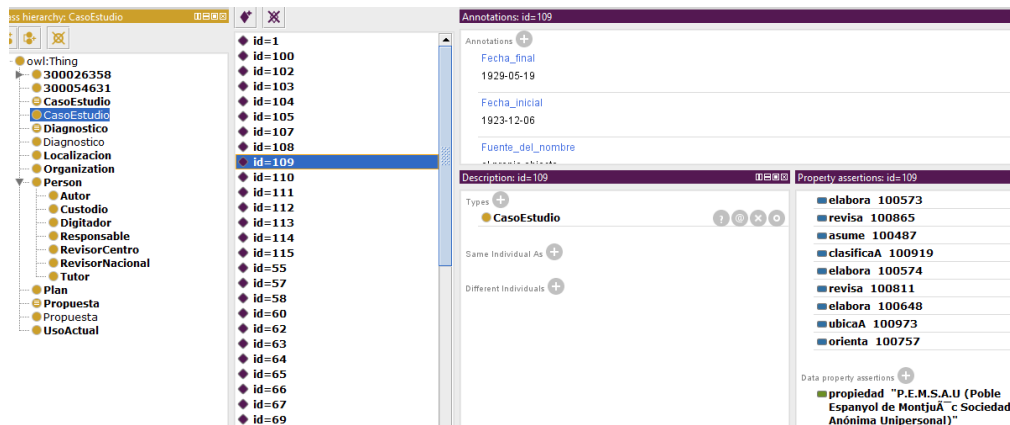


Figura 10. Comprobación del grafo generado por medio de herramientas semánticas

También se hacen pruebas con el archivo generado modificando su contenido y estructura con el fin de verificar su correcto formato.

- **Publicación del Grafo Semántico**. Para este paso se realizó una evaluación a través de consultas a través del endpoint de consultas Sparql del almacén de datos utilizado. Se comprobó

que la carga del archivo se hubiese realizado correctamente y con la IRI establecida. Adicionalmente para las consultas que se realicen se evalúa que su resultado este acorde a la integridad del grafo y no generarse errores (Figura 11).

The screenshot shows the Virtuoso SPARQL Query Editor interface. At the top, it displays the title 'Virtuoso SPARQL Query Editor' and navigation links for 'About', 'Namespace Prefixes', 'Inference rules', and 'SPARQL'. Below this, the 'Default Data Set Name (Graph IRI)' is set to 'http://wgvubuntu01.cloudapp.net:8080/redphi'. The 'Query Text' field contains the SPARQL query: 'select distinct ?Concept where {[a ?Concept] LIMIT 100'. Below the query field, there are settings for 'Results Format' (HTML), 'Execution timeout' (0 milliseconds), and 'Options' (Strict checking of void variables checked). At the bottom, there are 'Run Query' and 'Reset' buttons. The results are displayed in a table with the following content:

Concept
http://www.w3.org/2002/07/owl#Class
http://www.w3.org/2002/07/owl#Ontology
http://www.w3.org/2002/07/owl#AnnotationProperty
http://www.w3.org/2002/07/owl#Restriction
http://www.w3.org/2002/07/owl#ObjectProperty
http://www.w3.org/2002/07/owl#DatatypeProperty
http://www.w3.org/2002/07/owl#NamedIndividual
http://wgvubuntu01.cloudapp.net:8080/redphi#Autor
http://wgvubuntu01.cloudapp.net:8080/redphi#Localizacion

Figura 11. Comprobación de la publicación del grafo y resultado de consulta

- *Explotación Grafo Semántico.* La evaluación y los criterios de evaluación de la explotación del grafo semántico se realizaron de manera lógica con base en los requerimientos definidos y de manera técnica a través de la conectividad de las interfaces de búsqueda, geolocalización y comparación con el almacén de datos que contiene el grafo semántico y de los servicios externos utilizados. De igual manera se analizaron las respuestas de cada interfaz para comprobar la integridad de los resultados teniendo en cuenta los datos almacenados (Figura 12).

The screenshot shows a search interface with the following elements:

- Criterios:**
 - En: Caso de Estudio (dropdown)
 - Campo: Titulo (dropdown)
 - Valor: fábrica (text input)
 - Buscar (button)
- Resultado:**
 - Resultado de la Búsqueda**
 - 110 - Rehabilitación y reconversión para uso sociales de la antigua fábrica de aceite
 - 82 - Ex fábrica de chocolates "La Cubana"
 - 8 - Fábrica Mirat

Figura 12. Evaluación de funcionalidad e integridad de los resultados

Se comprobó la integración de la aplicación con otros sistemas externos (Figura 13).

PHI Patrimonio Histórico + Cultural Iberoamericano

Ficha Diagnósticos Propuestas INICIO CATÁLOGO VISOR CAR

Inicio

Datos básicos

Datos del tutor

Datos de revisión

Datos del caso de estudio

España · **Fábrica Mirat**
Autor: Marta Bueno Martín

Resumen del caso de estudio

1535 fundación- 1812 Guerra de la Independencia. Fabrica formada por edificios de varias épocas desde el XIX de su fundación hasta la actualidad. Construida sobre los restos del Convento de Nuestra Señora de la Victoria del que se conserva solo las trazas del claustro y bóvedas semienterradas.

Figura 13. Evaluación de la integración con otros sistemas

Por último, se evaluaron los resultados de la comparación de lenguaje natural mostrando un detalle del ejercicio de la funcionalidad y comprobando que los datos estuvieran acordes con los requerimientos y respuestas (Figura 14).

Resultado		
Years Deficit Maintenance Privatization Service Operated A '90 Last Century Meant Notable Deterioration Building Attempted Against Architectural Integrity Spaces Function Mail Been Relegated Surface Much Less Original Much Federal Tax Administration Truly Predatory Plant Floor Destroying Substantial Character Building Situation Explain / Finished This Serious		
COMPARACIÓN		
DETALLE	Label	Label
<i>No. Palabras</i>	52	54
<i>No. URIS</i>	45	49
<i>No. Match</i>	4	4
<i>% Confiabilidad</i>	86.54	90.74
<i>% Coincidencia</i>	8.89	8.16

Figura 14. Evaluación del resultado de la comparación del lenguaje natural

IV – PROTOTIPO

En el presente apartado se explica el funcionamiento del prototipo del sistema del cual hacen parte dos artefactos. El transformador encargado de incluir información de la Red PHI en un contexto semántico y el comparador delegado para la explotación del contenido semántico a través de funcionalidades de búsqueda, geolocalización y comparación.

4.1 TRANSFORMADOR

El transformador es un artefacto que tiene como objetivo la aplicación de anotaciones y enriquecimiento semántico a elementos de una base de datos convirtiéndolos en triples de un grafo apoyándose en la definición de una ontología del dominio.

Los pasos establecidos para cumplir con los objetivos del transformador son conectividad a una base de datos, generación inicial de un mapeo directo de la base de datos, una conciliación con una ontología del dominio trabajado, una configuración de ajustes al modelo de elementos no conciliados y por último la generación del grafo en un archivo RDF (Figura 15).



Figura 15. Pasos de la transformación en un modelo RDF

- El primer paso consiste en utilizar una base de datos que contenga el esquema a utilizar. La conectividad permite sistemas de gestión de base de datos como MS Sql Server, PostgreSQL y MySQL. En este punto también se selecciona todo el esquema de la base de datos o parte de él por medio de la selección de tablas y campos de la base de datos (Figura 16).

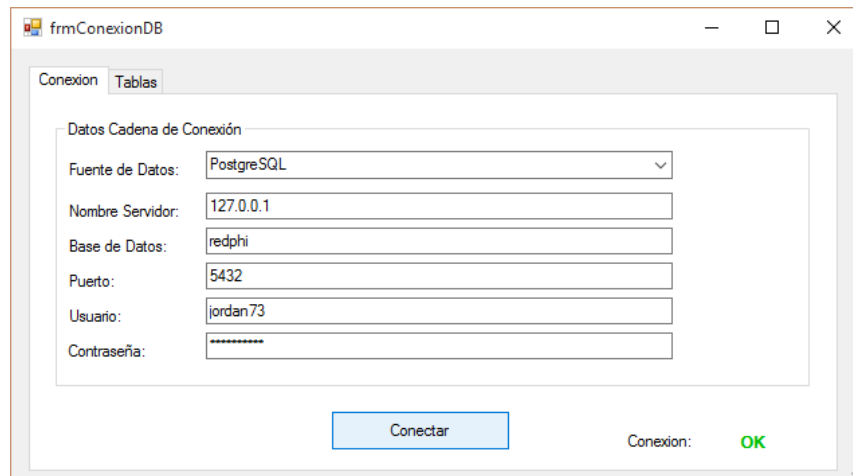


Figura 16. Paso de conectividad a BD y selección del esquema a utilizar

- El siguiente paso permite que el usuario genere un grafo inicial en memoria o en archivo físico por medio de una configuración manual o de un mapeo directo a la base de datos (Figura 17).

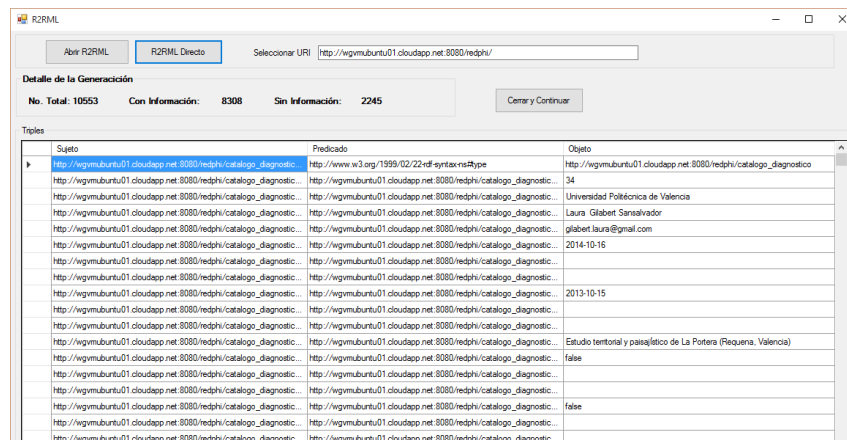


Figura 17. Mapeo directo o manual.

- Posteriormente el usuario debe conciliar su ontología con el proceso de mapeo del punto anterior. Esto permitirá alinear con los conceptos y propiedades con el esquema de la base de datos (Figura 18).

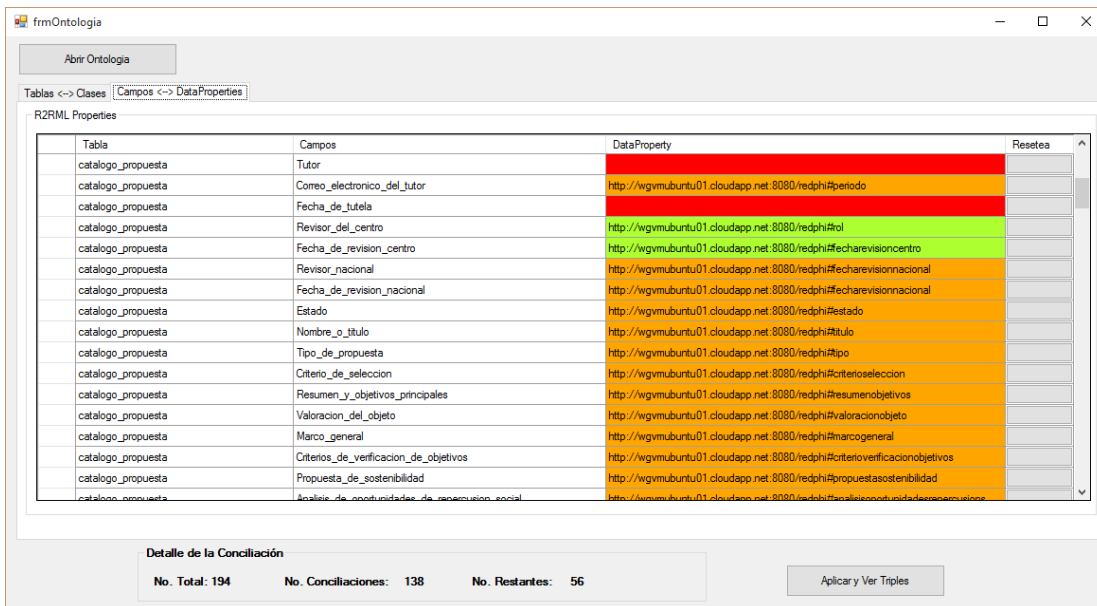


Figura 18. Conciliación entre ontología y mapeo inicial

- El proceso de conciliación implica una segunda y tercera iteración para conciliar elementos entre diferentes niveles y no relacionados (Figura 19).

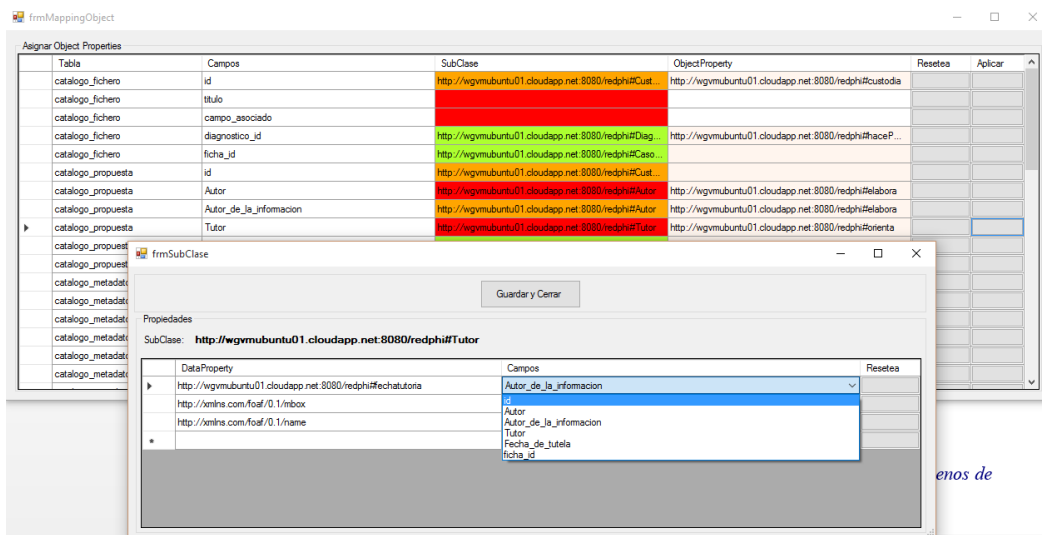


Figura 19. Transformación adicional.

- El resultado de la transformación se evidencia a través de la creación de un grafo RDF que contienen los triples instanciados de los elementos de la ontología. Este archivo se almacena por parte del usuario para posteriormente ser cargado en un triple store (Figura 20).

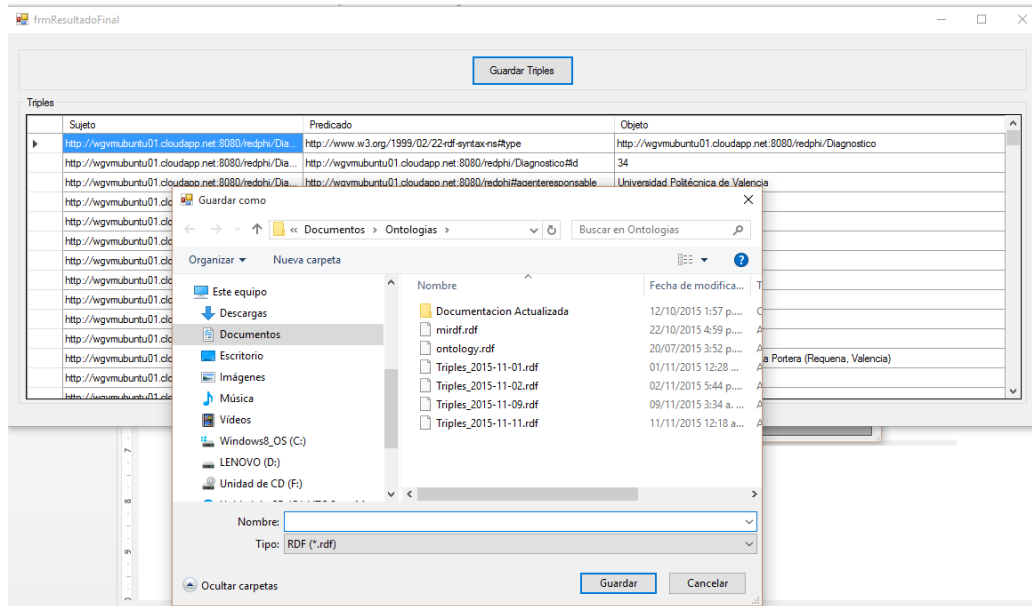


Figura 20. Generación del archivo grafo resultante en formato RDF

4.2 COMPARADOR

El comparador es un artefacto que tiene como objetivo explotar el contenido semántico de las instancias generadas y que se encuentran ubicada en un grafo de un almacén de datos.

Se utilizan tres funcionalidades para realizar la explotación que son: búsqueda, geolocalización y comparación. Esto aplica para los elementos del dominio de la Red PHI como son casos de estudio de objetos patrimoniales, sus diagnósticos y propuestas asociadas.

La opción de búsqueda permite consultar información sobre cualquier elemento de los casos de estudio, diagnósticos y propuestas. El resultado de la búsqueda se integrará con el sistema de la red PHI (Figura 21).

Criterios

En:

Campo:

Valor:

Resultado

Resultado de la Búsqueda
110 - Rehabilitación y re conversión para uso sociales de la antigua fábrica de aceite
82 - Ex fábrica de chocolates "La Cubana"
8 - Fábrica Mirat

Figura 21. Opción de búsqueda

La opción de herramientas de geolocalización permite consultar información sobre los mapas como sitios de interés, rutas, posicionamiento de los casos de estudio y vistas de las calles u objetos patrimoniales.

Los sitios de interés en el mapa permiten al usuario consultar en los mapas elementos del mismo tipo en lugares diferentes (áreas, ciudades, países) (Figura 22).

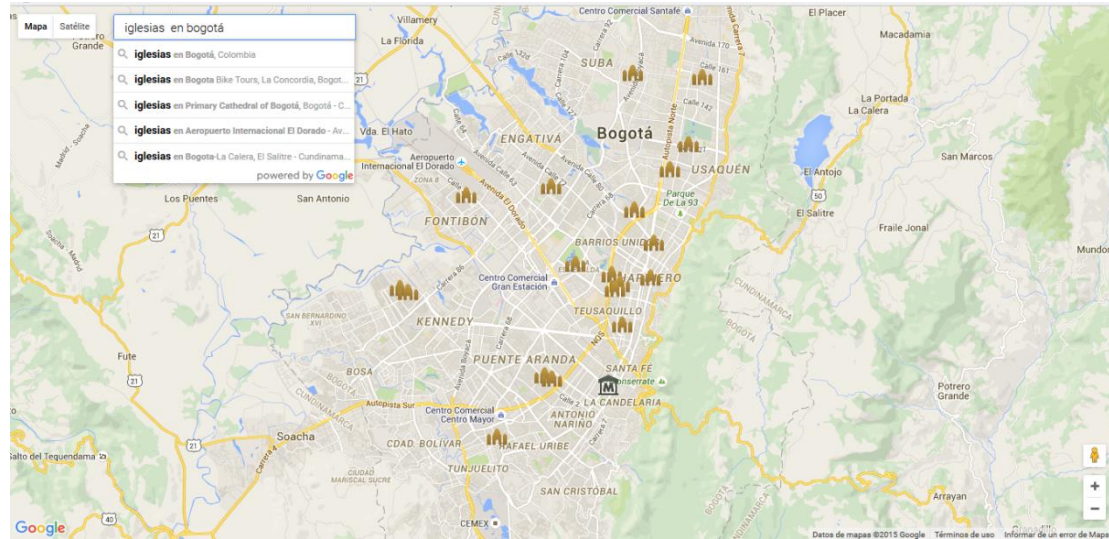


Figura 22. Ejemplo de iglesias ubicada en Bogotá

Las rutas permiten al usuario visualizar los caminos para desplazarse de un lugar de origen a un destino (Figura 23).

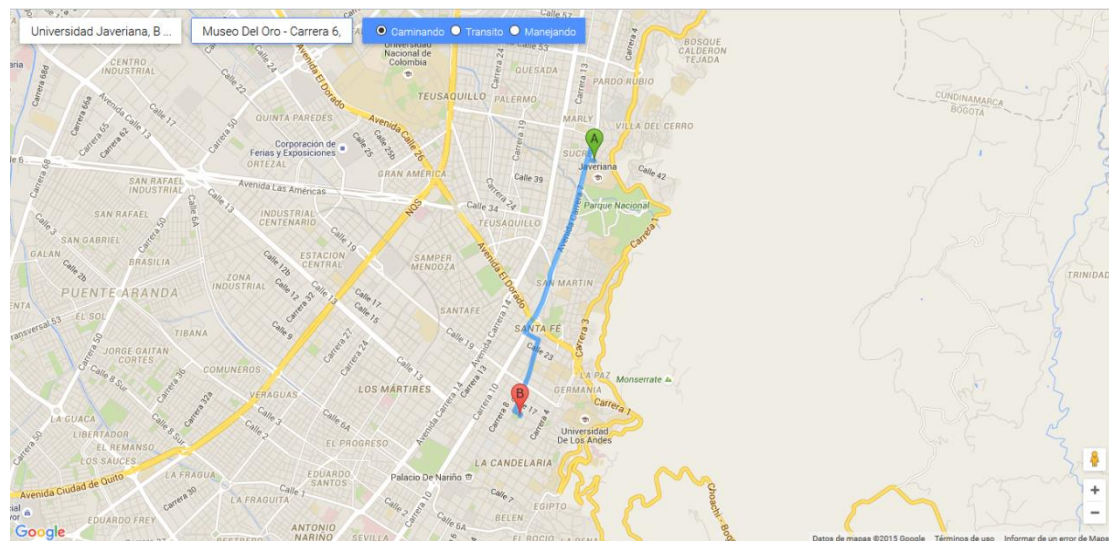


Figura 23. Ejemplo de identificación de ruta entre dos puntos.

La visualización de calles o lugares de interés permiten interactuar o visitar virtualmente lugares de patrimonio (Figura 24).



Figura 24. Ejemplo de vista callejera

La opción de comparación permite seleccionar dos elementos y realizar una comparación de lenguaje natural de los campos que el investigador o usuario crea convenientes comparar. La comparación puede realizarse entre elementos del mismo tipo, elementos diferentes y hasta en idiomas diferentes (Figura 25).

Muchos años de deficitario mantenimiento y la privatización del servicio operada en los años 90 del siglo pasado, significaron un deterioro notable en los aspectos edilicios y en las modalidades de ocupación que atentaron contra la integridad arquitectónica de los espacios. La función de correo ha quedado relegada en una superficie mucho menor que la original, buena parte de la placa de oficinas se encuentra desocupada y la Administración Federal de Impuestos ha hecho uso verdaderamente depredatorio de la planta baja, destruyendo aspectos sustanciales al carácter del edificio. La situación que se explicará en el anexo de Patologías da acabada cuenta de este grave deterioro.

Resultado

Years Deficit Maintenance Privatization Service Operated A 90 Last Century Meant Notable Deterioration Building Aspects Modalities Occupation Attempted Against Architectural Integrity Spaces Function Mail Been Relegated Surface Much Less Original Much Plate Offices Located Unemployed Federal Tax Administration Truly Predatory Plant Floor Destroying Substantial Character Building Situation Explain Annex de Patologías Realizes Finished This Serious

COMPARACIÓN		
DETALLE	Label	Label
<i>No. Palabras</i>	52	54
<i>No. URIS</i>	45	49
<i>No. Match</i>	4	4
<i>% Confabilidad</i>	86.54	90.74
<i>% Coincidencia</i>	8.89	8.16

Eliminar StopWords Traducir Comparar

Figura 25. Ejemplo del detalle de comparación.

V – EVALUACION Y VALIDACIÓN DEL PROTOTIPO

Al culminar de manera formal con la fase diseño y desarrollo se hace necesario realizar una evaluación y validación de los resultados.

5.1 EVALUACIÓN TÉCNICA

Para la evaluación técnica del prototipo se utiliza el estándar internacional ISO/IEC 9126 para la evaluación del Software, fue desarrollado para proporcionar un esquema para la evaluación de calidad del software. La normativa define seis características (funcionalidad, fiabilidad, usabilidad, mantenibilidad, eficiencia, portabilidad), las cuales representan un modelo detallado para la evaluación de cualquier sistema informático [103].

La evaluación técnica fue realizada por el ingeniero Néstor Armando Nova para cada una de las seis características. A continuación se presentan los resultados (Tablas 17, 18, 19, 20, 21 y 22).

CARACTERÍSTICA FUNCIONALIDAD		
SUBCARACTERÍSTICA	PREGUNTA DE EVALUACIÓN	SI/NO
<i>Idoneidad</i>	<i>Al someter al programa a diversas tareas básicas, cumple con los requerimientos establecidos?</i>	SI
<i>Exactitud</i>	<i>Realiza las tareas de manera tal que el resultado de las mismas sea correcto?</i>	SI
<i>Interoperabilidad</i>	<i>El software puede interactuar correctamente con otros sistemas previamente especificados?</i>	SI
<i>Seguridad</i>	<i>Están protegidos los datos que manipula el sistema, ya sea en su tiempo de proceso y tránsito, como así también en su estado de almacenamiento?</i>	SI
	<i>Es seguro ante al ataque de hackers?</i>	SI
	<i>Tiene contemplado un sistema de recuperación, ante pérdida de datos?</i>	NO

Tabla 17. Evaluación característica de funcionalidad

CARACTERÍSTICA FIABILIDAD		
SUBCARACTERÍSTICA	PREGUNTA DE EVALUACIÓN	SI/NO
<i>Madurez</i>	<i>Es confiable el software para el usuario final?</i>	SI
<i>Tolerancia a fallos</i>	<i>Después de un buen periodo de uso: sucede a veces que el usuario “desconfía” porque en ocasiones anteriores ha perdido datos importantes que le ha elevado tiempo cargar?</i>	NO
	<i>Cuándo el sistema falla, son fallas graves o leves, según las consecuencias que provocan?</i>	NO
<i>Facilidad de recuperación</i>	<i>El sistema, presenta fallas muy a menudo?</i>	NO

	<i>Es preciso inhabilitarlo por mucho tiempo cada vez que hay que hacer tareas de mantenimiento?</i>	<i>NO</i>
--	--	-----------

Tabla 18. Evaluación característica de fiabilidad**CARACTERÍSTICA USABILIDAD**

SUBCARACTERÍSTICA	PREGUNTA DE EVALUACIÓN	SI/NO
<i>Facilidad de comprensión</i>	<i>Es amigable el software para los desarrolladores?</i>	<i>SI</i>
	<i>Pueden comprender su estructura lógica, sus funciones de ejecución y procesamiento, su código fuente es fácilmente legible y comprensible?</i>	<i>SI</i>
	<i>Son cómodos los botones, las opciones, las ventanas de interfaces, los cuadros de diálogo, los formularios, etc.? Las jerarquías visuales son correctas?</i>	<i>SI</i>
<i>Facilidad de aprendizaje</i>	<i>El usuario puede aprender a manejar y utilizar el prototipo en un tiempo corto?</i>	<i>SI</i>
	<i>Es intuitivo, y posee la información y ayudas adecuadas como para que el usuario no tenga que depender de alguien que explique cómo utilizar cada función?</i>	<i>NO</i>
<i>Operatividad</i>	<i>Es sencillo de entender y manejar el software para los usuarios a los cuales está destinado su uso?</i>	<i>SI</i>
	<i>Es sencillo buscar y filtrar información dentro del programa?</i>	<i>SI</i>

Tabla 19. Evaluación característica de usabilidad**CARACTERÍSTICA EFICIENCIA**

SUBCARACTERÍSTICA	PREGUNTA DE EVALUACIÓN	SI/NO
<i>Tiempo de uso</i>	<i>Se ve afectada la productividad de los usuarios por esta lentitud?</i>	<i>NO</i>
<i>Recursos utilizados</i>	<i>Cuando el volumen de datos crece dentro de lo contemplado, ¿el software se vuelve lento?</i>	<i>SI</i>
	<i>Es capaz el software de procesar/almacenar datos de manera eficiente?</i>	<i>SI</i>

Tabla 20. Evaluación característica de eficiencia**CARACTERÍSTICA MANTENIBILIDAD**

SUBCARACTERÍSTICA	PREGUNTA DE EVALUACIÓN	SI/NO
<i>Facilidad de análisis</i>	<i>Es sencillo corregir errores del software (bugs o funcionalidades mal definidas)? (esto depende del grado de modularidad del software) si el software es modular se aísla problemas fácilmente y se gana tiempo encontrando y corrigiendo errores.</i>	<i>SI</i>
<i>Facilidad de cambio</i>	<i>Es sencillo hacer adaptaciones cuando se alteran levemente los requerimientos iniciales?</i>	<i>SI</i>
<i>Estabilidad</i>	<i>Reacciona bien el sistema ante situaciones o casos no previstos o no contemplados en los requerimientos?</i>	<i>NO</i>
<i>Facilidad de prueba</i>	<i>Se realiza un esfuerzo necesario para validar las modificaciones realizadas al sistema?</i>	<i>NO</i>

Tabla 21. Evaluación característica de mantenibilidad

<i>CARACTERÍSTICA PORTABILIDAD</i>		
<i>SUBCARACTERÍSTICA</i>	<i>PREGUNTA DE EVALUACIÓN</i>	<i>SI/NO</i>
<i>Facilidad de instalación</i>	<i>El software es fácil de instalar</i>	<i>NO</i>
<i>Facilidad de ajuste</i>	<i>El software es portable a diferentes sistemas operativos y plataformas?</i>	<i>NO</i>
<i>Facilidad de adaptación al cambio</i>	<i>Es sencillo "trasladar" el software de una intranet a otra, o de un dominio/servidor a otro sin mayores problemas, y configurando tan solo unos pocos parámetros?</i>	<i>NO</i>

Tabla 22. Evaluación característica de portabilidad

El comentario por parte del ingeniero evaluador posterior a la aplicación del estándar ISO/IEC 9126 es el siguiente: “Los artefactos de cumplen con los requisitos establecidos en el documento de requerimientos. Su funcionalidad es correcta teniendo en cuenta las pruebas realizadas, el transformador tiene opciones de mejora puesto que el usuario debe ser especializado y aun así no resulta fácil aprender su uso rápidamente. En el comparador por otro lado se presenta una interfaz mucho más amigable y sencilla. Se recomienda utilizar una misma plantilla para ambos artefactos.”

5.2 VALIDACIÓN DE LA UTILIDAD POTENCIAL

Para la validación se utiliza un modelo de aceptación tecnológica por parte de uno o varios individuos. El Modelo de Aceptación Tecnológica (TAM) es uno de los modelos más frecuentemente empleado para la investigación en la aceptación de nuevas tecnologías de la información. Básicamente, se verifica la percepción de la facilidad de uso y la utilidad de la tecnología que determinan las actitudes positivas o negativas que se tengan para su adopción [104].

Las variables tenidas en cuenta para la validación del modelo son percepción de la utilidad (PU), percepción de la facilidad de uso (PEOU), la curiosidad y satisfacción (C/E), la confianza, el altruismo, la experiencia de usuario, la reciprocidad esperada (ER), los costos operativos (EC) y la intención de uso (ITU).

La validación fue realizada por el arquitecto Carlos Eduardo Nieto González y un grupo de estudiantes a los cuales se le compartió acceso al link del sistema. Aprovechando la utilización de la aplicación se incluyó una opción para realizar la validación de una manera intuitiva y fácil en vez de hacerlo manualmente.

El usuario puede ingresar a la opción y el sistema presenta cada una de las variables del modelo de validación y sencillamente responde de manera positiva o negativa las preguntas marcando o desmarcando según sea el caso. Esta información queda almacenada de manera anónima para ser posteriormente utilizada estadísticamente (Tabla 23).

Preguntas

Perceived usefulness (PU)		
No.	Pregunta	Si/No?
1	El comparador es útil para mí.	<input checked="" type="checkbox"/>
2	El comparador mejora el rendimiento de mi trabajo y el aprendizaje.	<input checked="" type="checkbox"/>
3	El comparador mejora la eficiencia de mi vida.	<input checked="" type="checkbox"/>
4	El comparador aumenta la comodidad y la productividad en mi vida.	<input checked="" type="checkbox"/>

Perceived ease of use (PEOU)		
No.	Pregunta	Si/No?
1	Es fácil para mí hacer uso del comparador.	<input checked="" type="checkbox"/>
2	Aprender a utilizar el comparador es fácil para mí.	<input checked="" type="checkbox"/>
3	En general, creo que es fácil usar el comparador.	<input checked="" type="checkbox"/>

Curiosity and enjoyment (C/E)		
No.	Pregunta	Si/No?
1	Es interesante participar en el uso del comparador.	<input checked="" type="checkbox"/>
2	Es satisfactorio utilizar el comparador.	<input checked="" type="checkbox"/>
3	Me divierto utilizando el comparador y sus funciones.	<input checked="" type="checkbox"/>

Trust		
No.	Pregunta	Si/No?
1	Creo que la información del comparador es veraz.	<input checked="" type="checkbox"/>
2	Creo que el uso del comparador no revelará mi privacidad.	<input checked="" type="checkbox"/>
3	La gente que utiliza el comparador es de confianza, sobre todo aquellos que tienen contacto conmigo.	<input checked="" type="checkbox"/>
4	Confío en usuarios que hace parte de la red que usa	<input checked="" type="checkbox"/>

Tabla 23. Variables utilizadas en el modelo de aceptación tecnológica.

Para poder realizar los cálculos y obtener un resultado cuantitativo para las variables se asumieron las respuestas positivas con un valor de 5 y las respuestas negativas con un valor de 1. Se tomaron todas las respuestas del mismo tipo y se obtuvo la media. En el ejercicio de la validación en línea se obtuvieron 12 registros entre docentes y estudiantes (Figura 26).

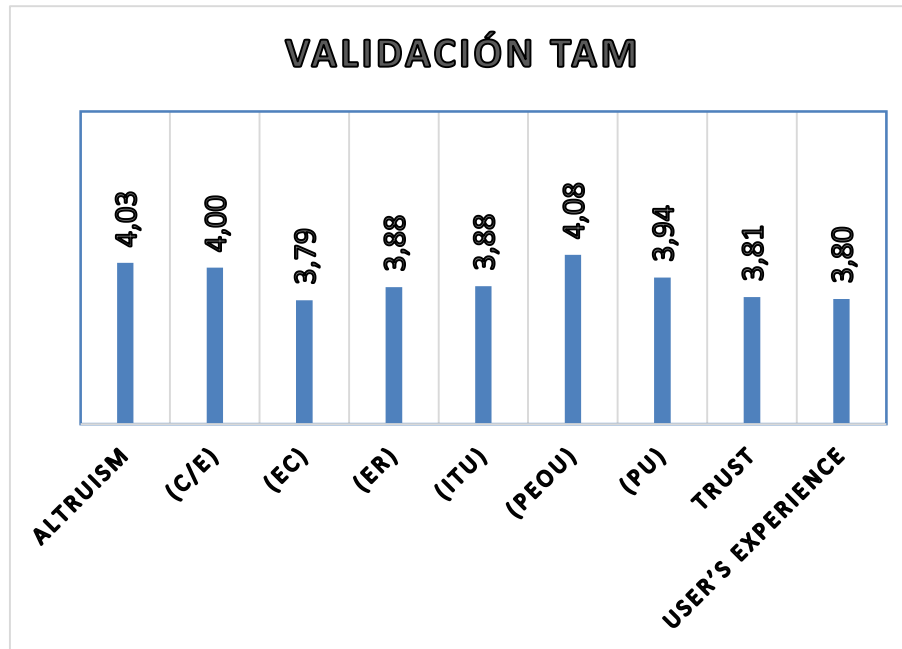


Figura 26. Resultado promedio para cada variable TAM

En la validación los usuarios que realizaron la prueba tenían una opción de recomendaciones para mejorar los artefactos. De esta manera también se pudo hacer una valoración cualitativa. Algunas de sus recomendaciones fueron las siguientes:

- “La utilización del aplicativo ayuda a los estudiantes a comparar información que crean relevante contra las características de otros objetos patrimoniales”.
- “Se recomienda para futuras versiones una comparación de varias o todas las características contra la información de todos los casos de estudio”.
- “El uso de mapas para la representación de información me parece que le da un valor agregado al sistema”.
- “Se cuenta con una base de datos reducida de objetos patrimoniales, sería importante que a medida que la información llegase a la red de patrimonio también se viera reflejada inmediatamente en el sistema”.

El resultado promedio de todas las variables arrojó un valor de 3.91 sobre un máximo de 5 por lo que se puede concluir que el modelo de aceptación tecnológica es positivo en términos generales (Tabla 24).

VARIABLE	TOTAL
Altruism	4,03
(C/E)	4,00
(EC)	3,79
(ER)	3,88
(ITU)	3,88
(PEOU)	4,08
(PU)	3,94
Trust	3,81
User's experience	3,80
PROMEDIO	3,91

Tabla 24. Resultado promedio total de la validación TAM

VI – TRABAJOS FUTUROS

Algunos de los posibles trabajos futuros para continuar desarrollando como resultado de la investigación o profundización del presente trabajo se presentan a continuación.

Los artefactos resultantes deben poder soportar múltiples plataformas por lo que en este punto se podría pensar en un transformador para la web y no como una herramienta de escritorio. De esta manera este artefacto podría utilizarse desde cualquier plataforma con acceso a la web y no requeriría instalación en los equipos donde se vaya a utilizar.

La utilización de técnicas de pareo sobre conjunto de cadenas es importante en la realización de la conciliación para encontrar correspondencias, sin embargo se podrían aplicar otras técnicas más efectivas como minería de datos, alignment server u otros algoritmos que permitan incrementar el grado de coincidencia entre los elementos pareados.

La explotación a través de la web semántica es el objetivo final que buscan los sistemas con enriquecimiento semántico por lo que se puede ampliar esta característica por medio de relacionamiento con otras bases semánticas y no solo con DBpedia. También se puede hacer uso de otras herramientas para verificar funcionalidades que bien podrían adaptarse al transformador con el fin de generar más enriquecimiento semántico.

Contar con la característica de una transformación más automática y con menos intervención de un experto podría ser una oportunidad para un trabajo futuro de manera que las correlaciones se generan nivel a nivel y entre niveles de tipos de elementos diferentes con el fin de gestionar de manera completa la adquisición de un modelo más robusto en los primeros pasos.

VII – CONCLUSIONES Y APORTES

Se presentan de manera concreta las conclusiones y los aportes resultantes de la elaboración de este trabajo de grado.

La utilidad de la web semántica solo se alcanza a vislumbrar cuando se empieza a trabajar con las tecnologías asociadas a esta web y dejando por un momento a un lado las tecnologías tradicionales. La gran potencialidad que se tiene para gestionar información y conocimiento entre comunidades permite trabajar de manera colaborativa con el fin de buscar, seleccionar, organizar, evaluar, comparar, compartir y publicar de manera eficiente la información y generar constantemente nuevo conocimiento es el mejor aporte para los profesionales que no han mirado hacia esta web del futuro.

El desarrollo de la red de ontologías facilita la representación formal del conocimiento entre los actores relacionados con el patrimonio cultural de Iberoamérica y representa un paso importante hacia la integración de la Red PHI como una red colaborativa y de gestión del conocimiento entre universidades de varios países en el ámbito del patrimonio cultural.

Otro aporte importante es que se identifica claramente durante el proceso de desarrollo es el diseño y formalización de algoritmos que proponen para la comparación semántica de elementos, otro para la conciliación entre elementos de una ontología y un esquema RDF resultado de un mapeo directo y un último algoritmo para correlacionar elementos entre niveles diferentes como son las clases y los campos.

El desarrollo e implementación del artefacto transformador permite ser utilizado en diferentes dominios donde se requiere enriquecer semánticamente información y relacionarla con un modelo de conocimiento (ontología). La generación de manera automática y asistida permite reducir tiempos, relacionar todos los elementos e identificar elementos que no han sido correlacionados.

El enriquecimiento y aprovechamiento de las características de la web semántica podría darse en cualquier área del conocimiento por lo que aunque la red PHI fue el caso de implementación de este trabajo, este proceso podrá trasladarse a otras áreas con el fin de gestionar conocimiento y cambiar la forma de cómo se deberían publicar la información de manera eficiente en la Web.

Este proyecto se logró gracias a la dirección de la Facultad de Ingeniería, los docentes de arquitectura de la Universidad Javeriana adscritos a la red PHI y a los miembros de la red pertenecientes a la Universidad Politécnica de Madrid, quienes en todo momento apoyaron el proceso para cumplir el compromiso de la Universidad Javeriana para aportar una herramienta de comparación semántica a la red PHI.

VIII – BIBLIOGRAFÍA

- [1] Red PHI. http://rede-phi.net/?page_id=6&lang=es
- [2] Newman, A., & McLean, F. (1998). Heritage builds communities: the application of heritage resources to the problems of social exclusion. *International Journal of Heritage Studies*, 4(3-4), 143-153.
- [3] Ronzino, P., Amico, N., & Niccolucci, F. (2011). Assessment and comparison of meta-data schemas for architectural heritage. *Proc. of CIPA*.
- [4] Liao, S. H. (2003). Knowledge management technologies and applications—literature review from 1995 to 2002. *Expert systems with applications*, 25(2), 155-164.
- [5] Tyndale, P. (2002). A taxonomy of knowledge management software tools: origins and applications. *Evaluation and Program Planning*. 25(2), pp. 183-190.
- [6] Maier, R. (2002). Knowledge management systems: Information and communication technologies for knowledge management. Springer-Verlag New York, Inc.
- [7] BOUQUET, Paolo, et al. C-owl: Contextualizing ontologies. En *The Semantic Web-ISWC 2003*. Springer Berlin Heidelberg, 2003. p. 164-179.
- [8] Villazón, B., Vilches, L. M., Corcho, O. & Gómez, A. (2011). Methodological guidelines for publishing government linked data. In *Linking government data* (pp. 27-49). Springer New York.
- [9] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28-37.
- [10] Lynch, C. (2002). Digital collections, digital libraries & the digitization of cultural heritage information. *Microform & imaging review*, 31(4), 131-145.
- [11] Jain, P., Hitzler, P., Sheth, A. P., Verma, K., & Yeh, P. Z. (2010). Ontology alignment for linked open data. In *The Semantic Web-ISWC 2010* (pp. 402-417). Springer Berlin Heidelberg.
- [12] Johannessen, J. A., Olsen, B., & Olaisen, (1999). Aspects of innovation theory based on knowledge-management. *International Journal of Information Management*, 19, pp. 121-139.
- [13] Hevner, A.R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.
- [14] Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: a model for producing and presenting

- information systems research. In Proceedings of the first international conference on design science research in information systems and technology. pp. 83-106.
- [15] Zhang, X., & Dorn, B. (2011, April). Agile Practices in a Small-Scale, Time-Intensive Web Development Project. In Information Technology: New Generations (ITNG), 2011 Eighth International Conference on (pp. 1060-1061). IEEE.
- [16] Schwaber, K., & Sutherland, J. (2013). The definitive guide to scrum: The rules of the game. Improving the Profession of Software development.
- [17] Suárez, M. C., Gómez, A., Motta, E., & Gangemi, A. (Eds.). (2012). Ontology engineering in a networked world. Springer Science & Business Media.
- [18] WALLS, Joseph G.; WIDMEYER, George R.; EL SAWY, Omar A. Building an information system design theory for vigilant EIS. Information systems research, 1992, vol. 3, no 1, p. 36-59.
- [19] Lassila, O., & Swick, R. R. (1999). Resource description framework (RDF) model and syntax specification.
- [20] UNESCO. Gestión del patrimonio mundial cultural. UNESCO. 2014.
- [21] DOERR, Martin, et al. El Modelo de Datos de Europeana (EDM). En Meeting: 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management. 2010.
- [22] D'ANDREA, Andrea; FERNIE, Kate. CARARE 2.0: a metadata schema for 3D Cultural Objects. Digital Heritage, 2013.
- [23] Red PHI. Proyecto [PHI] Patrimonio Histórico+Cultural Iberoamericano. 2013.
- [24] CACCIOTTI, Riccardo; BLAŠKO, Miroslav; VALACH, Jaroslav. A diagnostic ontological model for damages to historical constructions. Journal of Cultural Heritage, 2014.
- [25] DRUCKER, Peter F.; DRUCKER, Peter Ferdinand. Post-capitalist society. Routledge, 1994.
- [26] CHOO, Chun Wei. The knowing organization: How organizations use information to construct meaning, create knowledge, and make decisions. Oxford University Press, USA, 2006.
- [27] DAVENPORT, T. H. i Prusak, L. (2000). Conocimiento en acción (Cómo las organizaciones manejan lo que saben).
- [28] BELLINGER, Gene; CASTRO, Durval; MILLS, Anthony. Data, information, knowledge, and wisdom. 2004.

- [29] PONJUÁN DANTE, Gloria. Introducción a la gestión del conocimiento. La Habana: Universidad de La Habana. Facultad de Comunicación. Dpto. de Bibliotecología y Ciencia de la Información, 2006, p. 6.
- [30] DRUCKER, Peter Ferdinand. Technology, management, and society. Harvard Business Press, 2011.
- [31] NONAKA, Ikujiro. A dynamic theory of organizational knowledge creation. Organization science, 1994, vol. 5, no 1, p. 14-37.
- [32] LIAO, Shu-hsien. Knowledge management technologies and applications—literature review from 1995 to 2002. Expert systems with applications, 2003, vol. 25, no 2, p. 155-164.
- [33] MAIER, Ronald. Knowledge management systems: Information and communication technologies for knowledge management. Springer-Verlag New York, Inc., 2002.
- [34] DACONTA, Michael C.; OBRST, Leo J.; SMITH, Kevin T. The semantic web: a guide to the future of XML, web services, and knowledge management. John Wiley & Sons, 2003.
- [35] BERNERS-LEE, Tim, et al. The semantic web. Scientific american, 2001, vol. 284, no 5, p. 28-37.
- [36] SOLTERO, Alonso Pérez; VALENZUELA, Mario Barceló; SCHMITZ, Guzmán Gerardo Alfonso Sánchez. La web semántica como apoyo a la gestión del conocimiento y al modelo organizacional. Ingeniería informática, 2006, no 12, p. 4.
- [37] KONSTANTINOU, Nikolaos; SPANOS, Dimitrios-Emmanuel. Materializing the Web of Linked Data. 2015.
- [38] BERNERS-LEE, Tim, et al. Semantic web road map. 1998.
- [39] KLYNE, Graham; CARROLL, Jeremy J. Resource description framework (RDF): Concepts and abstract syntax. 2006.
- [40] MCGUINNESS, Deborah L., et al. OWL web ontology language overview. W3C recommendation, 2004, vol. 10, no 10, p. 2004.
- [41] OWL Web Ontology Language [W3C Recommendation 10 February 2004] <<http://www.w3.org/TR/2004/REC-owl-features-20040210>>
- [42] OWL 2 Web Ontology Language. Structural Specification and Functional-Style Syntax. <http://www.w3.org/TR/owl2-syntax/>

- [43] PÉREZ, Jorge; ARENAS, Marcelo; GUTIERREZ, Claudio. Semantics and Complexity of SPARQL. En *The Semantic Web-ISWC 2006*. Springer Berlin Heidelberg, 2006. p. 30-43.
- [44] OWL Web Ontology Language [W3C Recommendation 10 February 2004] <<http://www.w3.org/TR/2004/REC-owl-features-20040210>>
- [45] FLÓREZ, Alex Giovanni Amaya; GONZÁLEZ, Rafael Andrés. Construcción de un Catálogo de Capacidades de Investigación, Usando Ontologías.
- [46] POVEDA VILLALON, Maria; SUÁREZ-FIGUEROA, Mari Carmen; GÓMEZ-PÉREZ, Asunción. *Malas prácticas en ontologías*. 2009.
- [47] CORCHO, Oscar; GÓMEZ-PÉREZ, Asunción. A roadmap to ontology specification languages. En *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*. Springer Berlin Heidelberg, 2000. p. 80-96.
- [48] AUER, Sören; BRYL, Volha; TRAMP, Sebastian. *Linked Open Data--Creating Knowledge Out of Interlinked Data*. Springer, 2014.
- [49] UBALDI, Barbara. *Open Government Data*. 2013
- [50] BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. *Linked data-the story so far*. 2009.
- [51] BIZER, Chris, et al. *How to publish linked data on the web*. 2007.
- [52] HALLO, Maria Asunción; MARTÍNEZ-GONZÁLEZ, Maria Mercedes; DE LA FUENTE REDONDO, Pablo. Las tecnologías de Linked Data y sus aplicaciones en el gobierno electrónico. *Scire: representación y organización del conocimiento*, 2012, vol. 18, no 1, p. 49-61.
- [53] RONZINO, Paola; AMICO, Nicola; NICCOLUCCI, Franco. Assessment and comparison of metadata schemas for architectural heritage. *Proc. of CIPA*, 2011.
- [54] KIOUSSI, Anastasia, et al. Recommendations and strategies for the establishment of a guideline for monument documentation harmonized with existing European standards and codes. En *Proceedings of the XXIIIrd Symposium on CIPA*, Czech Technical University, Prague. 2011.
- [55] D'ANDREA, Andrea; FERNIE, Kate. Centro Interdipartimentale di Servizi di Archeologia, Università degli Studi di Napoli "L'Orientale", Naples, Italy. En *Digital Heritage International Congress (DigitalHeritage)*, 2013. IEEE, 2013. p. 137-143.
- [56] World Wide Web Consortium -W3C. (2014). Simple Knowledge Organization System - SKOS. Retrieved from <http://www.w3.org/2004/02/skos/intro>

- [57] McKenna, G., & Carolien Fokke. (2012). EUROPEANA INSIDE - Recommendations for Technical Standards. Retrieved from <http://www.europeana-newspapers.eu/public-materials/deliverables/>
- [58] Papatheodorou, C., Gavrilis, D., Fernie, K., Wright, H., Richards, J., Ronzino, P., & Meghini, C. (2012). ARIADNE. Retrieved from <http://www.ariadne-infrastructure.eu/Resources/D3.1-Initial-Report-on-the-project-registry>
- [59] Getty Research Institute. (2014b). Getty Vocabularies. Retrieved from <http://www.getty.edu/research/tools/vocabularies/>
- [60] Getty Research Institute. (2014c). The Art & Architecture Thesaurus-AAT. Retrieved from <http://www.getty.edu/research/tools/vocabularies/aat/about.html>
- [61] Getty Research Institute. (2010). Cultural Objects Name Authority-CONA. Retrieved from <http://www.getty.edu/research/tools/vocabularies/cona/about.html>
- [62] Getty Research Institute. (2015). The Union List of Artist Names-ULAN. Retrieved from <http://www.getty.edu/research/tools/vocabularies/ulan/about.html>
- [63] Getty Research Institute. (2014a). Getty Thesaurus of Geographic Names-TGN. Retrieved from <http://www.getty.edu/research/tools/vocabularies/tgn/about.html>
- [64] International Council of Museums. (2011). CIDOC CRM. Retrieved from <http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html>
- [65] Europeana. (2011). Europeana - EDM. Retrieved April 24, 2015, from <http://labs.europeana.eu/api/linked-open-data/data-structure/>
- [66] RONZINO, Paola, et al. European standards for the documentation of historic buildings and their relationship with CIDOC-CRM.
- [67] GOERZ, Günther; SCHOLZ, Martin. Content analysis of museum documentation with a transdisciplinary perspective. En Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education. Association for Computational Linguistics, 2009. p. 1-9.
- [68] MASCI, Maria Emilia, et al. 3D in the CARARE project: Providing Europeana with 3D content for the archaeological and architectural heritage: The Pompeii case study. En Virtual Systems and Multimedia (VSMM), 2012 18th International Conference on. IEEE, 2012. p. 227-234.
- [69] PAWŁOWSKI, Adam. Polish Monolingual Task within Cultural Heritage in CLEF (CHiC) 2013. Wrocław Runs.

-
- [70] RONZINO, P.; HERMON, S.; NICCOLUCCI, F. A metadata schema for cultural heritage documentation. V., CApellini (ed.), *Electronic Imaging & the Visual Arts: EVA*, 2012, p. 36-41.
- [71] VASSALLO, V., et al. The Cyprus Institute-Science and Technology in Archaeology Research Center, Cyl-STARC, Nicosia, Cyprus. En *Digital Heritage International Congress (DigitalHeritage)*, 2013. IEEE, 2013. p. 667-670.
- [72] FELICETTI, A., et al. Mapping ICCD Archaeological Data to CIDOC-CRM: the RA Schema.
- [73] Kent Beck, "Manifesto for Agile Software Development." 2001 [Online]. Available: <http://agilemanifesto.org/>
- [74] JACOBSON, Ivar. *Object oriented software engineering: a use case driven approach*. 1992.
- [75] ADOLPH, Steve; COCKBURN, Alistair; BRAMBLE, Paul. *Patterns for effective use cases*. Addison-Wesley Longman Publishing Co., Inc., 2002.
- [76] WIEGERS, Karl; BEATTY, Joy. *Software requirements*. Pearson Education, 2013.
- [77] Highsmith, J. "Agile Software Development Ecosystems". Addison-Wesley. 2002.
- [78] K. S. Rubin, *Essential Scrum: a practical guide to the most popular agile process*. Upper Saddle River, NJ: Addison-Wesley, 2013.
- [79] Don Wells, "Extreme Programming Project." [Online]. Available: <http://www.extreme-programming.org/map/project.html>.
- [80] K. Beck, *Extreme Programming Explained: Embrace Change*, US ed. Addison-Wesley Professional, 1999.
- [81] Microsoft SQL Server. <https://msdn.microsoft.com/es-es/library/bb545450.aspx>. Microsoft 2014.
- [82] PostgreSQL. http://www.postgresql.org.es/sobre_postgresql. Portal en Español sobre PostgreSQL 2010.
- [83] MySQL. <http://www.oracle.com/es/products/mysql/overview/index.html>. ORACLE 2014.
- [84] CARDELLINI, Valeria; COLAJANNI, Michele; PHILIP, S. Yu. Dynamic load balancing on web-server systems. *IEEE Internet computing*, 1999, no 3, p. 28-39.

- [85] SHELDON, Thomas; SÁNCHEZ, Antonio Vaquero; YÁÑEZ, Luis Hernández. Manual de Microsoft Internet information server. McGraw-Hill, Interamericana de España, 1997.
- [86] TOURIS, Alberto Molpeceres; MARIÑÁN, Martín Pérez. Arquitectura empresarial y software libre, J2EE. 2002.
- [87] MONTILVA, Jonás; ARAPÉ, Nelson; COLMENARES, J. Desarrollo de software basado en componentes. En Actas del IV. Congreso de Automatización y Control. Mérida, Venezuela. 2003.
- [88] Introducción a .NET Framework. <https://msdn.microsoft.com/es-es/library/hh425099%28v=vs.110%29.aspx>. Microsoft 2014.
- [89] ROHLOFF, Kurt, et al. An evaluation of triple-store technologies for large data stores. En On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. Springer Berlin Heidelberg, 2007. p. 1105-1114.
- [90] OpenLink Virtuoso. https://www.w3.org/2001/sw/wiki/OpenLink_Virtuoso. W3C Semantic Web. 2012
- [91] PRESSMAN, Roger S. Software engineering: a practitioner's approach. Palgrave Macmillan, 2005.
- [92] González, R.A. & Pomares, A. "La investigación científica basada en el diseño como eje de proyectos de investigación en ingeniería". Reunión Nacional ACOFI, Sep. 12-14, Medellín, 2012.
- [93] GENTNER, Dedre. Structure-mapping: A theoretical framework for analogy. Cognitive science, 1983, vol. 7, no 2, p. 155-170.
- [94] CLEMENTS, Paul, et al. Documenting software architectures: views and beyond. En ICSE. 2003. p. 740-741.
- [95] O'REILLY, Tim. Qué es Web 2.0. Patrones del diseño y modelos del negocio para la siguiente generación del software. Boletín de la Sociedad de la Información: Tecnología e Innovación, 2006, p. 177-201.
- [96] AHO, Alfred V.; HOPCROFT, John E.; ULLMAN, Jeffrey D. Estructura de datos y algoritmos. México, DF: Addison-Wesley Iberoamericana, 1988.
- [97] OLSEN, Anne L. Using pseudocode to teach problem solving. *Journal of Computing Sciences in Colleges*, 2005, vol. 21, no 2, p. 231-236.
- [98] R2RML: RDB to RDF Mapping Language. <http://www.w3.org/TR/r2rml/>. W3C Recommendation 27 September 2012.

- [99] A Direct Mapping of Relational Data to RDF. <http://www.w3.org/TR/rdb-direct-mapping/>. W3C Recommendation 27 September 2012
- [100] UKKONEN, Esko. Approximate string-matching with q-grams and maximal matches. *Theoretical computer science*, 1992, vol. 92, no 1, p. 191-211.
- [101] COHEN, William; RAVIKUMAR, Pradeep; FIENBERG, Stephen. A comparison of string metrics for matching names and records. En *Kdd workshop on data cleaning and object consolidation*. 2003. p. 73-78.
- [102] STOILOS, Giorgos; STAMOU, Giorgos; KOLLIAS, Stefanos. A string metric for ontology alignment. En *The Semantic Web–ISWC 2005*. Springer Berlin Heidelberg, 2005. p. 624-637.
- [103] ZEISS, Benjamin, et al. Applying the iso 9126 quality model to test specifications. *Software Engineering*, 2007, vol. 15, no 6, p. 231-242.
- [104] VENKATESH, Viswanath; DAVIS, Fred D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 2000, vol. 46, no 2, p. 186-204.