

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS E INGENIERÍA**



**PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DEL PERÚ**

**PUBLICACIÓN Y CONSUMO DE INFORMACIÓN DE ATRACTIVOS  
TURÍSTICOS Y CULTURALES LOCALES, UTILIZANDO LOS  
PRINCIPIOS Y TECNOLOGÍAS DE LINKED DATA**

Tesis para optar el Título de Ingeniero Informático, que presenta el bachiller:

**Cristhian Roberto Yesquen Lihim**

**ASESOR: Héctor Andrés Melgar Sasieta**

Lima, Junio de 2016

## RESUMEN

En la actualidad, la Web se ha convertido en la plataforma social e interactiva más utilizada. Las personas ahora tienen la posibilidad de interactuar unos con otros y aportar mayor contenido que permita enriquecer la experiencia de navegar en Internet. Sin embargo, esto ha sido el causante de una serie de problemas relacionados a la gestión y organización de los recursos que son publicados en la Web.

En Internet podemos encontrar una gran cantidad de contenido, pero muchas veces la información que encontramos trata sobre el mismo tema o elemento en particular, sobrecargando la Web con información ya existente. Muchas veces, estos recursos no están relacionados y, la forma en que son publicados imposibilita la existencia de alguna forma de conectarlos unos con otros, de manera que se pueda, por un lado, evitar la duplicidad de información, y, por otro lado, promover la reutilización de información.

El presente proyecto se enfocará en el ámbito del turismo, en específico, en relación a la información publicada respecto a los atractivos turísticos y culturales presentes en el Perú. Existen gran cantidad de sitios Web, tanto nacionales como internacionales, en donde se puede buscar y obtener información y datos de interés de atractivos turísticos locales. Esta información puede ser exactamente igual o muy similar, con lo cual se incurre en una sobrecarga y duplicidad de información en la Web, o inclusive muy distinta, pero sin posibilidad de complementarse.

Con el objetivo de poder reducir la sobrecarga de información en la Web, y aumentar las posibilidades de complementarse entre distintas fuentes, surge un conjunto de principios, buenas prácticas y tecnologías bajo el concepto de *Linked Data* o Datos Enlazados. Este conjunto de elementos describe un método de publicación de recursos en la Web, estructurado y llevado a cabo de tal manera que los datos e información puedan ser reutilizados por fuentes y ordenadores de orígenes distintos.

De acuerdo a lo expuesto anteriormente, se plantea brindar una alternativa de solución al problema de la publicación de recursos en la Web, en específico en el dominio de atractivos turísticos y culturales del Perú. Para ello, se hará uso de los principios, buenas prácticas y tecnologías de *Linked Data* para la publicación y consumo de dichos recursos. En el presente proyecto se diseñará e implementará una estructura que permita publicar en la Web datos de interés relacionados a atractivos turísticos y culturales locales, siguiendo los lineamientos de *Linked Data*, y además de la construcción de una herramienta que permita consultar y obtener información de dichos atractivos.

## Índice de contenidos

CAPÍTULO 1.....	1
<b>1. Problemática.....</b>	<b>1</b>
<b>1.1. Objetivo general.....</b>	<b>3</b>
<b>1.2. Objetivos específicos.....</b>	<b>3</b>
<b>1.3. Resultados esperados.....</b>	<b>3</b>
<b>2. Herramientas, métodos, metodologías y procedimientos.....</b>	<b>4</b>
<b>2.1. Herramientas.....</b>	<b>5</b>
<b>2.2. Métodos y Procedimientos.....</b>	<b>9</b>
<b>2.3. Metodologías.....</b>	<b>9</b>
<b>3. Alcance y Limitaciones.....</b>	<b>10</b>
<b>4. Justificación.....</b>	<b>11</b>
CAPÍTULO 2.....	12
<b>1. Marco conceptual.....</b>	<b>12</b>
<b>2. Estado del arte.....</b>	<b>16</b>
CAPÍTULO 3.....	26
<b>1. Introducción.....</b>	<b>26</b>
<b>2. Resultados Esperados del Objetivo Especifico 1 - Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales.....</b>	<b>26</b>
<b>2.1. Análisis y diseño conceptual del conocimiento en el dominio de atractivos turísticos y culturales locales.....</b>	<b>26</b>
<b>2.2. Modelo de Ontología que represente el conocimiento en el dominio de atractivos turísticos y culturales locales.....</b>	<b>30</b>
CAPÍTULO 4.....	31
<b>1. Introducción.....</b>	<b>31</b>
<b>2. Resultados Esperados del Objetivo Especifico 2 - Desarrollar un repositorio con información respecto a atractivos turísticos y culturales locales aplicando los principios y tecnologías de <i>Linked Data</i>.....</b>	<b>32</b>
<b>2.1. Base de datos relacional con información de atractivos turísticos y culturales locales .....</b>	<b>32</b>
<b>2.2. Mapeo entre la base de datos relacional y la Ontología.....</b>	<b>33</b>
<b>2.3. Publicación de la base de datos de atractivos turísticos y culturales locales de acuerdo a los principios de <i>Linked Data</i>.....</b>	<b>35</b>

2.4. Conjunto de pruebas de consulta a la información publicada en la Web.....	36
CAPÍTULO 5.....	38
1. Introducción .....	38
2. Resultados Esperados del Objetivo Especifico 3 - Desarrollar una herramienta informática que permita informar acerca de atractivos turístico y culturales locales ....	38
2.1. Arquitectura de software.....	38
2.2. Prototipo de aplicación Web .....	39
CAPÍTULO 6.....	42
1. Conclusiones.....	42
1.1. Conclusiones del Objetivo Específico 1 - Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales.....	42
1.2. Conclusiones del Objetivo Específico 2 - Desarrollar un repositorio con información respecto a atractivos turísticos y culturales locales aplicando los principios y tecnologías de <i>Linked Data</i> .....	42
1.3. Conclusiones del Objetivo Específico 3 - Desarrollar una herramienta informática que permita informar acerca de atractivos turístico y culturales locales	42
1.4. Conclusión final.....	43
2. Recomendaciones .....	43
3. Observaciones.....	44
Referencias bibliográficas.....	45

## Índice de Figuras

<b>Figura 1:</b> Comparación de información respecto a la Catedral de Lima de los portales Web minube.pe y de PromPerú. [Portal minube.pe, 2015], [PromPerú, 2015] .....	3
<b>Figura 2:</b> Ejemplo de metadato de la Fortaleza del Real Felipe. ....	13
<b>Figura 3:</b> CONTUR, ontología de turismo creada por la Universidad de Standford.....	14
<b>Figura 4:</b> Ejemplo de esquema grafico básico del uso de Linked Data. ....	15
<b>Figura 5:</b> Arquitectura empleada para el desarrollo de la aplicación SmartTourism. [Fermoso, Mateos, Beato, Berjon 2015] .....	23
<b>Figura 6:</b> Arquitectura empleada para el desarrollo de la aplicación eZaragoza. [Tejo, Fernández, Barrueta, Polo 2013].....	24
<b>Figura 7:</b> Arquitectura empleada para el desarrollo de la aplicación LinkedQR.....	25
<b>Figura 8:</b> Ontología desarrollada en el dominio de atractivos turísticos y culturales locales. Imagen de autoría propia.....	30
<b>Figura 9:</b> Patrones de publicación de información en Linked Data, según origen, tipo y forma de almacenamiento. Imagen recuperada de [Heath, Bizer, 2011] .....	31
<b>Figura 10:</b> Modelo de la base de datos relacional para el almacenamiento de información de atractivos turísticos y culturales locales. Imagen de autoría propia .....	33
<b>Figura 11:</b> Mapeo de la base de datos “tourismperu” a la entidad “database”. Imagen de autoría propia .....	34
<b>Figura 12:</b> Mapeo de la tabla “toupe_atractivo” de la base de datos con la clase “Attraction” de la ontología. Imagen de autoría propia .....	35
<b>Figura 13:</b> Mapeo de la columna “toupe_atractivo.nombre” de la base de datos con la propiedad “attraction_name” de la ontología.....	35
<b>Figura 14:</b> Ejemplo de consulta SPARQL realizada a la información de atractivos turísticos publicada de acuerdo a Linked Data. Imagen de autoría propia .....	36
<b>Figura 15:</b> Diagrama general de los componentes involucrados en el desarrollo de la solución propuesta. Imagen de autoría propia.....	39
<b>Figura 16:</b> Vista de Despliegue de la aplicación Web. Imagen de autoría propia.....	39
<b>Figura 17:</b> Menú de búsqueda de la aplicación Web Turismo Perú. Imagen de autoría propia.....	40
<b>Figura 18:</b> Vista detallada de un atractivo turístico en la aplicación Web Turismo Perú...	41

## Índice de Tablas

<b>Tabla 1:</b> Tabla de Resultados Esperados por objetivo. ....	4
<b>Tabla 2:</b> Tabla de herramientas, métodos y procedimientos a usarse por resultado esperado y objetivo. ....	4
<b>Tabla 3:</b> Cadenas generales de búsqueda utilizadas para llevar a cabo la revisión de la literatura .....	17
<b>Tabla 4:</b> Resumen de investigaciones .....	18
<b>Tabla 5:</b> Resumen de aplicaciones revisadas .....	19
<b>Tabla 6:</b> Consultas y respectivas medidas de Precision y Recall.....	37



# CAPÍTULO 1

## 1. Problemática

En la actualidad, la Web se ha convertido en el más grande repositorio y fuente de información en relación a cualquier tema o dominio que sea del conocimiento de las personas que la usan. Con el paso de los años, la Web dejó de ser una simple plataforma informativa, en la que el usuario sólo era un agente pasivo que recibía o publicaba información sin posibilidad alguna de interacción, para convertirse en una plataforma mucho más social e interactiva. Los usuarios tienen la posibilidad de interactuar con otros y aportar mucho mayor contenido que permita enriquecer la experiencia de navegar en Internet [Pastor Sánchez, 2011]. A esta nueva forma de ver y utilizar la Web se le denominó como la Web Social [Pastor Sánchez, 2011] o la llamada Web 2.0 [O'Reilly, 2005]. Se caracteriza principalmente por la creación de plataformas para publicar y compartir contenidos como blogs, o redes sociales como Facebook, o páginas de alojamiento de contenido multimedia como Youtube; todas con el claro objetivo de dotar al usuario de un rol mucho más activo en la Web de lo que tenía anteriormente.

Si bien este paradigma de la Web brinda a las personas mayores capacidades y participación en ella, también causa serie de problemas. Al ser una Web centrada en el usuario, se pensó en ella de manera tal que sea entendible simplemente por las personas, convirtiéndola en una Web conformada por documentos que puedan ser visualizados por las personas, y en formatos que sólo éstas entendieran [Berners-Lee, Hendler, Lassila, 2001]. Esto trae como consecuencias problemas en la gestión y organización de los recursos publicados en la Web. Por ejemplo, existe una gran cantidad de recursos que han sido publicados por entidades y/o personas distintas, pero que tratan sobre el mismo tema o algún elemento en particular similar, sobrecargando la Web con información ya existente [Pastor Sánchez, 2011]. Muchas veces, estos recursos no están relacionados de alguna manera, ya que forma en que son publicados en la Web imposibilita la existencia de algún mecanismo para conectarlos unos con otros, de manera que pueda evitarse la duplicidad de información, y promover la reutilización de datos ya publicados previamente [Bizer, Heath, Berners-Lee, 2009]. Por último, los formatos empleados para publicar recursos en la Web son diversos, dependen de quién los publique y sus necesidades específicas, generalmente no se sigue un formato estándar, complicando más su reutilización y acceso por parte de las personas y programas de computadora [Pastor Sánchez, 2011].

Debido a los problemas mencionados previamente, se vio en la necesidad de redefinir la forma en que los recursos son publicados, almacenados y accedidos en la Web, para que ahora no sólo el usuario final sea el foco de atención, sino que también las máquinas y los programas de computadoras pudieran procesar estos recursos con algún propósito u objetivo en específico [Shadbolt, Hall, 2006]. Es por ello que surgió la Web Semántica, o la denominada Web de los Datos, que propone el uso de un conjunto de métodos, tecnología y estándares para redefinir la forma en que los recursos son publicados en la Web, y así brindarles una estructura y significado, dando solución a los problemas que



están presentes en la Web tradicional. Algunos de estos son, por ejemplo, el uso de metadatos para describir recursos, el uso de ontologías para la representación formal del conocimiento, o el uso *Linked Data* o datos enlazados.

A pesar de la introducción de la Web Semántica, el uso de los métodos y tecnologías que propone aun no es muy frecuente. Por ejemplo, la información publicada respecto a los atractivos turísticos y culturales presentes en la ciudad de Lima. En la actualidad, existen diversas páginas Web que informan acerca de los mismos atractivos turísticos locales, pudiendo ser esta información exactamente igual o muy similar, con lo cual se incurre en una sobrecarga y duplicidad de información en la Web, o inclusive muy distinta, pero sin posibilidad de complementarse. Por ejemplo, si visitamos el portal web de PromPerú dedicado a informar acerca de lugares turísticos de interés en la ciudad de Lima, podemos encontrar una lista de sitios a los que una persona puede visitar. Tomando el caso de la Catedral de Lima, en este portal podemos obtener información acerca de dónde se encuentra, cuáles son los horarios de atención y una pequeña reseña informativa. Por otro lado, si accedemos al portal web [www.minube.pe](http://www.minube.pe), el cual está orientado a personas que viajan constantemente brindando información sobre lugares turísticos, también podremos encontrar información sobre la Catedral de Lima, pero en este caso obtenemos información acerca de dónde se encuentra, un teléfono de referencia, la página Web oficial, y otros enlaces de interés. La comparación de ambas páginas Web, tomando como ejemplo información de la Catedral de Lima, se puede visualizar en la Figura 1. Como se puede comprobar, los datos e información no son los mismos, bien podrían complementarse, pero la información es propia de cada entidad u organismo que la ha publicado, dificultando la reutilización de los datos de una manera conveniente y sencilla. Cabe precisar que, además de páginas del tipo de las que fueron mencionadas previamente, en la Web también existen otras un poco más informales que también podrían contener información similar o distinta sobre el mismo tema, y estos son los blogs personales, principalmente de personas o comunidades dedicadas al turismo y viaje.

Como se pudo apreciar, a pesar de las ventajas que conllevaría el uso de los métodos y tecnologías propuestos por la Web Semántica, aún existe mucho trabajo que realizar respecto a ello. Un claro ejemplo de eso está compuesto por la información acerca de atractivos y sitios turísticos de la ciudad de Lima que se encuentra en la Web. Existe gran cantidad de información relacionada a ese tema; sin embargo, ya que es publicada, de manera individual y específica a cada entidad o persona que es dueña de una página Web, genera una sobrecarga de información y aumenta la dificultad de poder reutilizarla. Además, la arquitectura empleada por estas entidades o personas en el desarrollo de sus páginas Web, así como también el uso de tecnologías y estándares, como por ejemplo RDF, propuesto principalmente por la World Wide Web Consortium (W3C), juegan un papel muy importante en la consecución del objetivo de publicar los recursos en la Web de una manera más estructurada y con posibilidad de ser reutilizada.





**Figura 1:** Comparación de información respecto a la Catedral de Lima de los portales Web minube.pe y de PromPerú. [Portal minube.pe, 2015], [PromPerú, 2015]

Por esta razón, el presente proyecto de fin de carrera brinda una alternativa de solución al problema de la publicación de recursos en la Web, en específico en el dominio de atractivos turísticos y culturales del Perú. Se aplicarán los principios y tecnologías de *Linked Data*, o Datos enlazados, con la finalidad de publicar información en el dominio mencionado, para su posterior uso en una herramienta informática.

### 1.1. Objetivo general

Aplicar los principios y tecnologías de *Linked Data* para la publicación y consumo de información en el dominio de atractivos turísticos y culturales del Perú.

### 1.2. Objetivos específicos

Objetivo 1: Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales del Perú mediante la implementación de una ontología.

Objetivo 2: Desarrollar un repositorio con información respecto a atractivos turísticos y culturales del Perú aplicando los principios y tecnologías de *Linked Data*.

Objetivo 3: Desarrollar una herramienta informática que permita informar acerca de atractivos turístico y culturales del Perú, mediante el uso del repositorio de información.

### 1.3. Resultados esperados

En la Tabla 1 se presentan los resultados esperados por objetivo.

**Tabla 1:** Tabla de Resultados Esperados por objetivo.

Objetivo	Resultado esperado
Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales del Perú mediante la implementación de una ontología.	1) Análisis y diseño conceptual del conocimiento en el dominio de atractivos turísticos y culturales del Perú.
	2) Modelo de Ontología que represente el conocimiento en el dominio de atractivos turísticos y culturales del Perú.
Desarrollar un repositorio con información respecto a atractivos turísticos y culturales del Perú aplicando los principios y tecnologías de <i>Linked Data</i> .	3) Esquema de base de datos para el almacenamiento de información relacionada a atractivos turísticos y culturales del Perú, y su posterior implementación con datos reales.
	4) Documento con la especificación del mapeo entre la base de datos relacional y la ontología.
	5) Publicación de la base de datos de atractivos turísticos y culturales del Perú de acuerdo a los principios de <i>Linked Data</i> en un servidor y SPARQL Endpoint.
	6) Conjunto de pruebas de consulta y validación a la información publicada en la Web.
Desarrollar una herramienta informática que permita informar acerca de atractivos turísticos y culturales del Perú, mediante el uso del repositorio de información.	7) Modelo de arquitectura de software para el desarrollo de la herramienta.
	8) Prototipo de aplicación web que permita usar la información publicada en la Web respecto a atractivos turísticos y culturales del Perú.

## 2. Herramientas, métodos, metodologías y procedimientos

En esta sección se presentarán las herramientas, métodos y procedimientos que se emplearán para desarrollar los resultados esperados indicados previamente. En la Tabla 2 se muestra un mapeo entre las herramientas a usar por cada resultado esperado.

**Tabla 2:** Tabla de herramientas, métodos y procedimientos a usarse por resultado esperado y objetivo.

Objetivo	Resultado esperado	Herramientas a usarse
Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales del Perú mediante la implementación de una ontología.	1) Análisis y diseño conceptual del conocimiento en el dominio de atractivos turísticos y culturales del Perú	<i>Ontology Development 101</i> .
	2) Modelo de Ontología que represente el conocimiento en el dominio de atractivos turísticos y culturales del Perú	Protégé, RDF, OWL.

Objetivo	Resultado esperado	Herramientas a usarse
Desarrollar un repositorio con información respecto a atractivos turísticos y culturales del Perú aplicando los principios y tecnologías de <i>Linked Data</i> .	3) Esquema de base de datos para el almacenamiento de información relacionada a atractivos turísticos y culturales del Perú, y su posterior implementación con datos reales.	MySQL
	4) Documento con la especificación del mapeo entre la base de datos relacional y la ontología.	<i>D2RQ Mapping Language</i>
	5) Publicación de la base de datos de atractivos turísticos y culturales del Perú de acuerdo a los principios de <i>Linked Data</i> en un servidor y SPARQL Endpoint.	<i>D2RQ Platform, Linked Data, SPARQL</i>
	6) Conjunto de pruebas de consulta y validación a la información publicada en la Web.	SPARQL, Precision and Recall
Desarrollar una herramienta informática que permita informar acerca de atractivos turísticos y culturales del Perú, mediante el uso del repositorio de información.	7) Modelo de arquitectura de software para el desarrollo de la herramienta.	UML.
	8) Prototipo de aplicación web que permita usar la información publicada en la Web respecto a atractivos turísticos y culturales del Perú.	MySQL, Visual Studio, ASP.NET, dotNetRDF.

## 2.1. Herramientas

### 2.1.1. Protégé

Protégé es un framework gratuito que provee una interfaz de usuario gráfica para la creación y edición de ontologías, y que es desarrollado y mantenido por la Universidad de Stanford bajo el lenguaje de programación Java [Stanford Protégé]. En su núcleo, Protégé usa un conjunto de estructuras para el modelado del conocimiento y acciones para soportar la creación, visualización y manipulación de ontologías en distintos formatos de representación, como por ejemplo un listado de clases y sus propiedades, o un esquema en forma de árbol y sus nodos [Sivakumar, Arivoli, 2011].

Se decidió emplear esta herramienta ya que permitirá implementar la ontología que represente el conocimiento en el dominio de los atractivos turísticos y culturales del Perú. Además, el hecho de que sea gratuito y cuente con la documentación necesaria en la Web para su aprendizaje y uso son factores que también influyen en su selección.

### **2.1.2. RDF**

RDF es un modelo estándar para representar recursos en la Web mediante una serie de tripletas, en donde cada una de las tres partes es llamada sujeto, predicado y objeto respectivamente [W3C RDF]. El sujeto de una tripleta es un identificador URI que identifica al recurso que se describe; el objeto puede ser un valor literal como una cadena de caracteres, número o fecha, pero también puede ser otro identificador URI que identifica a otro recurso con el que el sujeto tiene algún tipo de relación; por último, el predicado indica que tipo de relación existe entre el sujeto y el objeto, y también está representado por un identificador URI [Bizer, Cyganiak, Heath, 2007].

Se decidió emplear este estándar ya que es el modelo con el que se representarán los recursos que serán publicados en la Web, y que además es soportado completamente por la herramienta Protégé y por el lenguaje OWL para la implementación de una ontología.

### **2.1.3. OWL**

OWL es un lenguaje desarrollado bajo el contexto de la Web Semántica, con el propósito de representar conocimiento complejo acerca de las cosas, grupo de cosas, y sus relaciones entre ellas, es decir, para desarrollar ontologías, de manera que el conocimiento puede ser representado de tal forma que pueda ser procesado por los programas de computadora [W3C OWL]. Fue diseñado para ser compatible con el lenguaje de marcado XML, así como también con otros estándares de W3C, y en particular RDF [Heflin, 2006]. Sintácticamente, una ontología desarrollada bajo OWL es un documento RDF válido, y como tal, está conformado por etiquetas XML, lo cual permite que pueda ser procesado por una gran variedad de herramientas de XML y RDF [Heflin, 2006].

Se decidió emplear OWL ya que será el lenguaje que se utilizará para la creación y edición de la ontología que represente el conocimiento en el dominio de los atractivos turísticos y culturales del Perú, y que además es soportado completamente por la herramienta Protégé y complementado con el uso del estándar RDF.

### **2.1.4. MySQL**

MySQL es un sistema de gestión de bases de datos relacional, multihilo y multiusuario, que cuenta con un tipo de licenciamiento dual, es decir, tiene una versión que es de código libre y otra privada [MySQL PROD]. La versión a utilizar para el presente proyecto de carrera será la MySQL Community Edition, que tiene licenciamiento GPL de código libre.

Se decidió emplear el motor de base de datos MySQL ya que permitirá implementar una base de datos relacional en donde se almacenará la información respecto a los atractivos turísticos y culturales del Perú. Además, MySQL es segura y fácil de usar, y cuenta con extensa documentación oficial sobre su aprendizaje y uso en la Web.

### **2.1.5.D2RQ Platform**

La plataforma D2RQ es un sistema para acceder bases de datos relacionales como grafos virtuales RDF de solo lectura, de manera que no sea necesario replicar la información almacenada en la base de datos en un repositorio de tripletas RDF [D2RQ Plat]. Gracias al uso de esta plataforma se pueden hacer consultas SPARQL a base de datos relacionales, acceder al contenido de una base de datos en forma de *Linked Data* y crear archivos de texto en formato RDF a partir de la información de una base de datos relacional [D2RQ Plat].

Se decidió emplear esta herramienta ya que permitirá realizar el proceso de publicación en la Web, de acuerdo a los principios de *Linked Data*, de la información de atractivos turísticos y culturales del Perú que estén almacenados en una base de datos relacional.

### **2.1.6.D2RQ Mapping Language**

D2RQ es un lenguaje de mapeo declarativo, por medio del cual se puede describir la relación entre el esquema de una base de datos relacional y un vocabulario RDF o una ontología OWL. Gracias a este mapeo, se puede definir un grafo virtual basado en RDF que contiene la información de la base de datos relacional, y puede ser accedido de diversas maneras, como por ejemplo utilizando el lenguaje SPARQL para consultar la información almacenada [D2RQ Map].

Se decidió emplear esta herramienta ya que permitirá realizar el mapeo entre la base de datos relacional con la ontología a desarrollar, de manera que la información almacenada en la base de datos pueda ser posteriormente publicada en la Web en formato RDF y según los principios de *Linked Data*, y también pueda ser accedida por medio del lenguaje SPARQL.

### **2.1.7.SPARQL**

SPARQL es un lenguaje estandarizado para la consulta de información representada en grados RDF [Pastor Sánchez, 2011]. Es reconocido por el W3C como una tecnología clave en el ámbito de la Web Semántica. SPARQL puede ser usado para expresar consultas que involucren diversas fuentes de datos, en las cuales los datos están almacenados nativamente mediante RDF, o que pueden ser expresados mediante RDF por medio de alguna herramienta intermedia [W3C SPARQL].

Se decidió emplear SPARQL ya que es un estándar propuesto por el W3C, y será el lenguaje empleado para poder acceder y consultar a la información que será publicada en la Web de acuerdo a los principios de *Linked Data*. Además, el hecho de que cuente con la documentación necesaria en la Web para su aprendizaje y uso es un factor que también influye en su selección.



### **2.1.8. UML**

UML es un lenguaje que permite modelar, construir y documentar los elementos que conforman un software antes que estos sean codificados. El uso de estos modelos ofrece beneficios a las personas que van a implementar el software ya que brinda una documentación y sustento fáciles de ser comprendidos que especifican las características y funcionalidades que tendrá el producto software [UML Model].

Se escogió emplear este lenguaje de modelado ya que permitirá especificar de manera conveniente y entendible la arquitectura a utilizar para el desarrollo de la herramienta de software propuesto en el presente proyecto de fin de carrera. Además, su facilidad y amplio uso en el entorno académico es un factor que también influye en su selección.

### **2.1.9. Visual Studio**

Visual Studio es un entorno de desarrollo para sistemas operativos Windows que brinda un completo conjunto de herramientas y servicios que permiten desarrollar una gran variedad de aplicaciones, tanto como para plataformas de Microsoft como para otras plataformas. Soporta múltiples lenguajes de programación como C++, C#, Visual Basic .NET, entre otros; al igual que entornos de desarrollo web como ASP.NET [VISUAL].

Se escogió emplear esta herramienta ya que permitirá realizar el prototipo de la aplicación Web por medio del cual se podrá acceder a la información de atractivos turísticos y culturales publicada de acuerdo a *Linked Data*. Además, su facilidad de uso, amplio uso en el ámbito académico, son factores que también influyen en su selección

### **2.1.10. ASP.NET**

En un framework para desarrollar aplicaciones web, creado y comercializado por Microsoft. Permite implementar páginas y aplicaciones web usando HTML, CSS y JavaScript [ASPNET].

Se escogió emplear este framework ya que forma parte íntegramente de la plataforma Microsoft para desarrollar aplicaciones, y se encuentra incorporado por defecto al entorno de desarrollo Visual Studio. Además, su facilidad de uso, amplio uso en el ámbito académico, son factores que también influyen en su selección

### **2.1.11. dotNetRDF**

Es una librería creada para trabajar con el framework de Microsoft .NET que provee un conjunto de funcionalidades compatibles con el uso de RDF, SPARQL y otras tecnologías de la Web Semántica [NETRDF].

Se escogió emplear esta librería ya que ofrece las características necesarias para poder integrar las tecnologías de la Web Semántica y *Linked Data* en la creación de una aplicación Web.



## 2.2. Métodos y Procedimientos

### 2.2.1. *Linked Data*

Para poder llevar a cabo la publicación de la base de datos de atractivos turísticos y culturales a implementar, se seguirán los principios propuestos por *Linked Data*, de manera que esta información esté publicada en forma de datos públicos y enlazados en la Web. A continuación se mencionan estos principios:

- 1) Utilizar URIs para identificar los recursos almacenados en la Web.
- 2) Utilizar HTTP URIs de manera que las personas puedan localizar a los recursos en la Web.
- 3) Ofrecer información sobre los recursos de la Web utilizando estándares como RDF.
- 4) Incluir, en los recursos Web, enlaces URIs a otros recursos, de manera que se pueda descubrir más información.

### 2.2.2. *Precisión y Recall*

Para poder determinar que la información publicada de acuerdo a los principios de *Linked Data* sea la misma que aquella almacenada en la base de datos relacional, se emplearán dos medidas: precisión y *recall*. La precisión y *Recall* son las medidas básicas usadas en el ámbito de la recuperación de la información y estrategias de búsqueda; para ello se debe considerar las siguientes asunciones [Jizba, 2007]:

- Existe un conjunto de registros en una base de datos que son relevantes para el tema de búsqueda.
- Todos los registros almacenados pueden ser, o bien relevantes, o no relevantes.
- El conjunto de registros que den como resultado de una búsqueda pueden no coincidir perfectamente con el conjunto de registros relevantes.

La precisión es el ratio del número de registros relevantes encontrados en una búsqueda, entre el número total de registros relevantes y no relevantes encontrados. El *Recall* es el ratio del número de registros relevantes encontrados en una búsqueda, entre el número total de registros relevantes que se encuentran almacenados en la base de datos [Jizba, 2007].

## 2.3. Metodologías

### 2.3.1. *Ontology Development 101*

Esta metodología es una propuesta para abarcar el proceso iterativo que comprende desarrollar una ontología. Para ello, define un conjunto de tres reglas y siete pasos a seguir [Noy, McGuinness, 2001]:

- No hay forma correcta para modelar un dominio, siempre existen otras alternativas viables. La mejor solución dependerá de la aplicación que se tiene en mente desarrollar, y del alcance que se puede anticipar.
- El desarrollo de una ontología es necesariamente un proceso iterativo.
- Los conceptos en una ontología deben ser cercanos a los objetos (tanto físicos como lógicos) y a las relaciones en el dominio de interés. En su mayoría, estos pueden ser sustantivos (objetos) o verbos (relaciones) en oraciones que describen el dominio.

Los siete pasos propuestos son los siguientes:

- Determinar el dominio y alcance de la ontología.
- Considerar reutilizar ontologías existentes.
- Enumerar términos importantes en la ontología.
- Definir las clases y la jerarquía de clases.
- Definir las propiedades de las clases (ranuras).
- Definir los aspectos de las propiedades.
- Crear instancias.

Las reglas y pasos propuestos en esta metodología representan una forma adecuada y fácil de entender y aplicar, para el proceso de desarrollo de la ontología; además, se cuenta con una breve documentación respecto a ella que facilita aún más su aprendizaje. Es por ello, que esta metodología será aplicada para realizar el análisis y diseño de la ontología propuesta como parte de los objetivos del presente proyecto.

### **3. Alcance y Limitaciones**

En el presente proyecto de fin de carrera se desarrollará un prototipo de aplicación Web que use como fuente de información repositorios que han sido publicados en la Web siguiendo los principios y tecnologías de *Linked Data* para informar acerca de atractivos turísticos y culturales del Perú. Para ello, primero se desarrollará una ontología para representar el conocimiento en el dominio de atractivos turísticos y culturales del Perú. Luego, se implementará un repositorio de información bajo un motor de base de datos relacional, con una estructura similar a la desarrollada en la ontología. Como siguiente paso, se realizará un mapeo entre la base de datos relacional y la ontología, lo cual servirá para poder publicar la información almacenada en la base de datos en formato RDF, y además de aplicar los principios de *Linked Data* al proceso de publicación. Finalmente, se construirá una aplicación Web que permita mostrar al usuario final información de interés en relación a los atractivos turísticos y culturales del Perú.

Para el presente proyecto, se considerará información de una selección de atractivos turísticos y culturales del Perú, y será recabada a través de datos que se puedan obtener de páginas Web del estado, de municipalidades, u otra organización.

#### 4. Justificación

En la actualidad, la información que se encuentra publicada en la Web sobre atractivos turísticos y culturales del Perú, es muy dispersa y variada. Existen páginas dedicadas a dicho tema que son oficiales, es decir, desarrolladas y mantenidas por el estado u otras entidades como lo son municipalidades, y muchas otras que son páginas o blogs personales. Esto conlleva a una serie de problemas, como por ejemplo la duplicidad de información que existe en la Web, o que la información sobre el mismo atractivo turístico no sea coherente en distintas páginas sin ninguna relación o mecanismo que pueda relacionarlos. Es en este punto que la aplicación de los principios y tecnologías propuestos por *Linked Data* surge como un adecuada posibilidad para poder dar inicio a una solución al escenario antes planteado.

Por otra parte, cuando una persona desea consultar información respecto a atractivos turísticos y culturales, las páginas Web no son la única fuente de información, si no que existen otros medios como lo son por ejemplo periódicos, revistas, anuncios televisivos o afiches en la calle, es decir, existe más de una forma como dar cuenta respecto a ellos. Sin embargo, esto puede constituir un problema a algunas personas en términos de comodidad y facilidad de acceso a la información. Es por ello que el proponer una aplicación Web como herramienta que permita mostrar información de interés acerca de atractivos turísticos y culturales del Perú, y que pueda ser accedida por distintos dispositivos, se afirma como una solución adecuada.

Por último, en el Anexo 3 se detalla la viabilidad para llevar a cabo el proyecto teniendo en cuenta tres aspectos: técnico, económico y tiempo.

## CAPÍTULO 2

### Marco conceptual y Estado del arte

#### 1. Marco conceptual

En el presente marco conceptual se presentarán y definirán conceptos y términos, de manera que permita un adecuado entendimiento del problema presentado en el Capítulo 1 del presente proyecto, es decir, en relación al estado actual de la Web, los problemas de gestión y organización de los recursos que son publicados en ella, y cómo la Web Semántica y *Linked Data* surgen como iniciativas para dar solución a dicho problema. El objetivo que se busca alcanzar con esta sección es, en primer lugar, comprender los principios, tanto conceptuales como tecnológicos, relacionados a la Web Semántica, para posteriormente presentar el concepto de *Linked Data* y cómo se relacionan, para que se pueda entender de mejor manera el beneficio que conlleva su aplicación en el ámbito del turismo peruano.

##### 1.1. Web Semántica

La Web actual, o también llamada Web 2.0 o Web Social, está orientada a facilitar la conexión entre las personas, la idea principal se centra en el establecimiento de redes y comunidades de usuarios que trabajan con una serie de servicios Web como los blogs y redes sociales, fomentando la colaboración, difusión e intercambio de información de manera rápida y sencilla [Pastor Sánchez, 2011]. Sin embargo, a lo largo del tiempo han surgido problemas debido a la propia naturaleza de esta Web en la que los usuarios son los agentes principales. Entre estos problemas podemos mencionar: la sobrecarga de información, la existencia de sistemas aislados sin posibilidad de compartir información a otros sistemas externos, y una pobre capacidad de agregación de contenido [Fensel, 2003]. Es en este punto donde surge el concepto de la Web semántica, proponiendo el uso de un conjunto de métodos y tecnologías con el objetivo de dar solución a los problemas mencionados previamente.

Se puede definir a la Web semántica como una extensión de la Web actual, en la cual se le brinda un significado bien definido a toda la información que se encuentre almacenada en ella, con la finalidad de que tanto los ordenadores como las personas puedan trabajar conjuntamente de manera más eficaz. Para ello, se debe proporcionar a la Web datos que estén bien definidos siguiendo algún tipo de estructura, y que además estén enlazados permitiendo que aplicaciones de distintos orígenes localicen, integren y reutilicen la información presente en ella [Berners-Lee, Hendler, Lassila 2001].

##### 1.2. Metadatos

Los metadatos es la primera idea fundamental sobre la cual se basa la representación de la información en la Web semántica. Limitándonos sólo a la definición etimológica del término, éste significa “datos sobre datos” [Pastor Sánchez, 2011]. Sin embargo, tomando un punto de vista más general, se puede definir a los metadatos como los

elementos que describen un determinado objeto siguiendo algún modelo o conjunto de reglas. Dentro del escenario de la Web semántica, el uso de metadatos representa un rol muy importante, ya que permitirá estructurar y codificar las características y propiedades que describen a los recursos, de manera que facilite las actividades de localización, recuperación, valoración, administración, persistencia e interoperabilidad [Pastor Sánchez, 2011].

Gracias al uso de los metadatos, cada elemento o recurso de la Web puede ser definido de manera más precisa y legible para un ordenador, esto debido principalmente al uso de tecnologías como por ejemplo Resource Description Framework (RDF). El uso de RDF permite un esquema de mapeo de las características de cada elemento, en donde la semántica de ellos es formulada de manera más explícita. Desde el punto de vista de la gestión del conocimiento, RDF representa un esquema descentralizado para representar el conocimiento. Está desarrollado en base al lenguaje de marcado eXtensible Markup Language (XML), de manera que permite crear descripciones complejas de los recursos, además de ofrecer un conjunto de reglas para crear relaciones semánticas. [Wittenburg, 2002]. En la Figura 2 se muestra un ejemplo de metadato de la Fortaleza del Real Felipe, el cual es un atractivo turístico localizado en la ciudad de Lima. En él se puede apreciar, de manera formal, datos como por ejemplo: cuándo y por quién fue construido, qué tipo de atractivo turístico es, el material con el que está construido, quién es su actual dueño, entre otros.

```
<metadata>
  <rdf:RDF
    xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
    xmlns:dc = "http://purl.org/dc/elements/1.1/" >
  </rdf:Description>
</rdf:RDF>
<rdf:Description
  rdf:about="https://es.wikipedia.org/wiki/Fortaleza_del_Real_Felipe">
  <dc:abstract>La Fortaleza del Real Felipe es una edificación militar de estilo Vauban
  construida en el siglo XVIII en la bahía del Callao</dc:abstract>
  <dc:constructedIn>Siglo XVIII</dc:constructedIn>
  <dc:constructedBy>José Antonio Manso de Velasco</dc:constructedBy>
  <dc:reConstructedIn>1925</dc:reConstructedIn>
  <dc:reConstructedBy>Agustín Tovar de Albertis en</dc:reConstructedBy>
  <dc:type>Fortaleza</dc:type>
  <dc:material>Piedra</dc:material>
  <dc:owner>Ejército del Perú</dc:owner>
  <dc:openToPublic>Si</dc:openToPublic>
  <dc:location>En la parte central de la bahía del Callao</dc:location>
  <dc:condition>Restaurada</dc:condition>
</rdf:Description>
</rdf:RDF>
</metadata>
```

**Figura 2:** Ejemplo de metadato de la Fortaleza del Real Felipe.  
Imagen de autoría propia.



## Ontologías

Una ontología es una especificación explícita de una conceptualización [Gruber, 1993]. Es decir, una ontología es una abstracción simple y esquemática de algún objeto o hecho del mundo real, el cual se desea representar con algún fin en específico.

Las ontologías es la segunda idea fundamental sobre la cual se basa la representación de la información en la Web semántica. El uso de metadatos permite únicamente la descripción de los recursos Web en forma de propiedades y sus valores. Sin embargo, es necesario el uso de ontologías con la finalidad de introducir en los recursos un conocimiento estructurado que permite realizar deducciones e inferencias para obtener información que no ha sido declarada explícitamente en los datos. Una ontología, aplicada en el ámbito de la Web semántica, permite realizar descripciones de los recursos, basados en principios de interoperabilidad semántica, gracias a la definición de clases, propiedades, relaciones y axiomas [Pastor Sánchez, 2011]. En la Figura 3 se muestra parte de CONTUR, una ontología desarrollada en el dominio del turismo por la Universidad de Standford.



**Figura 3:** CONTUR, ontología de turismo creada por la Universidad de Standford. [Fermoso, Mateos, Beato, Berjon 2015]

### 1.3. *Linked Data*

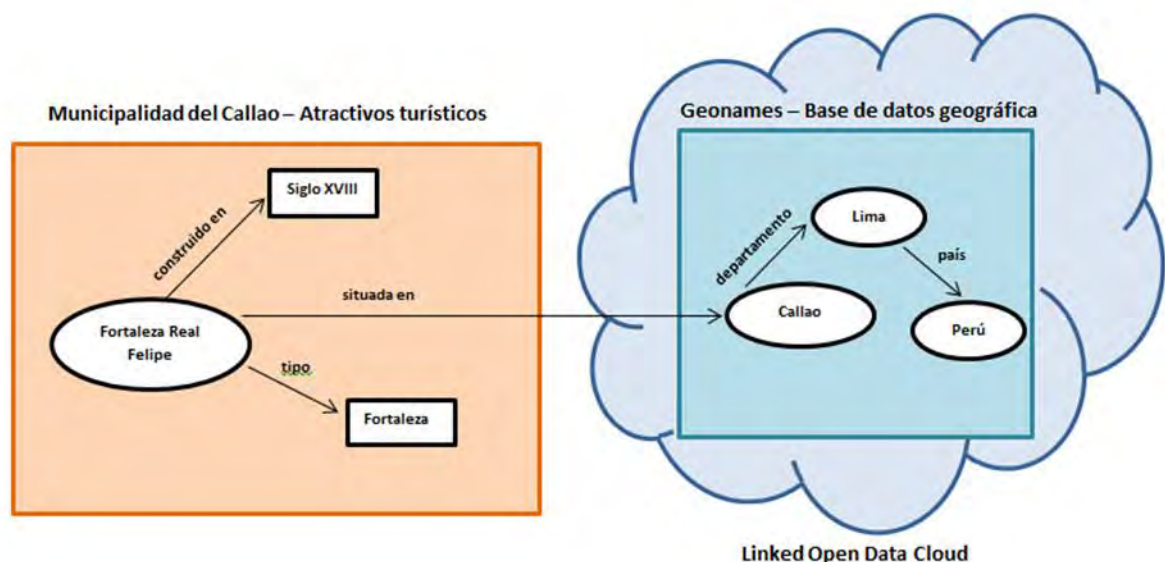
El término *Linked Data* se refiere al conjunto de buenas prácticas que se utilizan para publicar y conectar datos estructurados en la web [Bizer, Heath, Berners-Lee, 2009]. Uno de los principales objetivos de la Web semántica, además de la publicación estructurada de datos en la Web, es la publicación de datos de tal forma que estos puedan vincularse a otros, de manera que una persona o un ordenador pueda explorar la Web llegando a otros datos, externos y de distintas fuentes, que están relacionados a los



datos iniciales [Berners-Lee, 2006]. Es éste último punto el que *Linked Data* trata de alcanzar.

El uso de *Linked Data* permite construir la denominada “Web de los Datos”, la cual consiste en una gran base de datos que están interconectados y distribuidos a lo largo de toda la web. Estos datos se vinculan y exploran de manera similar a la utilizada para vincular documentos HTML [Berners-Lee, 2006]. Técnicamente hablando, *Linked Data* hace referencia a datos publicados en la Web de una manera tal que sea legible y posible de leer por un ordenador, que su significado este definido explícitamente, que esté vinculado a otros datos externos, y que además pueda ser vinculado desde otros datos externos [Bizer, Heath, Berners-Lee, 2009].

Por ejemplo, suponiendo que una municipalidad distrital de la ciudad de Lima publica información específica respecto a cada atractivo turístico que están localizados dentro de su distrito, como por ejemplo información histórica, horarios de atención, precios, entre otros, es posible que también desee publicar información relacionada a su ubicación geográfica. Pero, teniendo en cuenta que en la Web existen sitios que poseen bases de datos geográficas con información detallada, la municipalidad puede hacer referencia a esta fuente externa, de manera que la información inicial de cada atractivo turístico se enriquece con datos adicionales geográficos obtenidos de otro sitio (ver Figura 4).



**Figura 4:** Ejemplo de esquema grafico básico del uso de *Linked Data*.  
Imagen de autoría propia (2015)

### 1.3.1. Principios de *Linked Data*

Con el objetivo fomentar el crecimiento de la denominada Web de los Datos, tanto a nivel de documentos HTML, como a nivel de datos estructurados semánticamente, existen cuatro principios o reglas para publicar información en la Web [Berners-Lee, 2006]:

- 5) Utilizar URIs para identificar los recursos almacenados en la Web.
- 6) Utilizar HTTP URIs de manera que las personas puedan localizar a los recursos en la Web.
- 7) Ofrecer información sobre los recursos de la Web utilizando estándares como RDF.
- 8) Incluir, en los recursos Web, enlaces URIs a otros recursos, de manera que se pueda descubrir más información.

### **1.3.2. Tecnologías de *Linked Data***

*Linked Data* se basa principalmente en el uso de tres tecnologías: Uniform Resource Identifiers (URIs), HyperText Transfer Protocol (HTTP) y Resource Description Framework (RDF). Los URIs son cadenas de caracteres estandarizadas que conforman espacios de identificadores asociados a recursos abstractos o localizables en internet [Pastor Sánchez, 2011]. Al identificar o nombrar los recursos mediante URIs, se ofrece una abstracción del lenguaje natural, con lo cual se puede hacer referencia a cualquier recurso de la Web de manera estándar y sin ambigüedades. El protocolo HTTP provee un mecanismo sencillo pero universal para localizar y obtener recursos que han sido publicados en la Web. El uso de URIs y HTTP esta complementado por el uso de RDF, el cual es una tecnología fundamental para la Web semántica. RDF provee un modelo de datos genérico por medio del cual se puede estructurar y enlazar datos que describen objetos del mundo real [Bizer, Heath, Berners-Lee, 2009]. Una vez que se accede a un recurso en la Web mediante el uso de URIs y el protocolo HTTP, se debe poder obtener información útil de dicho recurso, para ello, esta información debe estar representada mediante descripciones estándares en RDF.

### **1.4. *Linked Data Mashup***

La aplicación de los principios de *Linked Data*, y la evolución de la denominada Web de los Datos, aumenta las posibilidades para el desarrollo de aplicaciones basadas en dominios y temas en específicos. A comparación de la Web tradicional, en la cual las aplicaciones trabajan sobre un conjunto fijo de fuentes de datos, las aplicaciones basadas en *Linked Data* pueden trabajar sobre un espacio global de fuentes de datos que son públicas y enlazadas unas con otras, permitiendo cada vez más ofrecer información más completa [Bizer, Heath, Berners-Lee, 2009]. Estas aplicaciones, denominadas *Linked Data Mashups*, son posibles principalmente al uso del estándar RDF, ya que gracias a él se puede combinar, agregar y transformar datos que provienen de fuentes de datos heterogéneos, abstrayendo los datos de la lógica de las aplicaciones [Hoang, Cung, Truong, Hwang, 2014].

## **2. Estado del arte**

Debido al gran volumen de datos publicación en la Web, siguiendo el conjunto de principios y buenas prácticas que propone *Linked Data*, existen varias formas y manera de poder utilizar y explotar dichos datos. Una de las principales, y que tiene gran uso en la actualidad, es la creación de aplicaciones que utilizan datos e información de un

dominio o tema en específico provenientes de fuentes publicadas según los principios de *Linked Data*, los denominados *Linked Data Mashups* [Bizer, Heath, Berners-Lee, 2009].

En esta sección se presentarán estudios relacionados tanto al uso de *Linked Data* para la publicación de datos como a la creación de aplicaciones orientadas a dominios y temas en específico basadas en el uso de datos enlazados en la Web. Los estudios que se presentarán más adelante fueron seleccionados a partir de una revisión que se realizó del estado del arte en base a una revisión sistemática. El realizar una revisión sistemática, en comparación con una revisión tradicional, tiene la ventaja de ofrecer un método formal y estructurado para llevar a cabo una revisión de la literatura relacionada al tema de investigación, con la finalidad de identificar aquellas que son las más relevantes, minimizando el grado de parcialidad [Kitchenham, 2004].

## 2.1. Formulación de las preguntas de investigación

Para poder proseguir con la revisión sistemática, se planteó las siguientes preguntas de investigación:

- ¿De qué maneras se han estado utilizando los datos enlazados en la Web para la creación de aplicaciones en cualquier dominio, y en particular, que informen y promuevan el turismo?
- ¿Qué fuentes *Linked Data* son las más empleadas en el ámbito del turismo?
- ¿Qué tipo de arquitecturas han sido utilizadas en aplicaciones que hacen uso de datos enlazados en la Web relacionadas al turismo?
- ¿De qué manera se han utilizado las ontologías en aplicaciones que hacen uso de datos enlazados en la Web relacionadas al turismo?

## 2.2. Selección de las fuentes

En base a las preguntas de investigación planteadas anteriormente, se definieron las siguientes palabras claves: “*Linked Data*”, “*tourism*”, “*semantic web*” y “*ontology*”. En la Tabla 3 se muestra las cadenas generales básicas de búsqueda, que se forman a partir de la combinación de las palabras claves definidas previamente con operadores lógicos, y que se utilizaron para llevar a cabo la revisión.

**Tabla 3:** Cadenas generales de búsqueda utilizadas para llevar a cabo la revisión de la literatura

Cadenas generales de búsqueda	
1	“ <i>Linked Data</i> ” AND “ <i>semantic web</i> ” AND “ <i>tourism</i> ”
2	“ <i>Linked Data</i> ” AND “ <i>tourism</i> ”
3	“ <i>semantic web</i> ” AND “ <i>tourism</i> ”
4	“ <i>tourism</i> ” AND “ <i>ontology</i> ”

Para llevar a cabo la revisión, las cadenas generales de búsqueda fueron usadas directamente en páginas de las librerías digitales de ACM, IEEE, Science Direct y Scopus. Se limitó la búsqueda a documentos publicados desde el año 2012 hasta la

fecha, ya que teniendo en cuenta que *Linked Data* se popularizó y oficializó desde el año 2007 con la creación del Linked Open Data Cloud, se desea considerar sólo los estudios que se han desarrollado en los últimos años en este ámbito; esto vendría a ser el criterio de exclusión empleado. Luego de encontrar una gran cantidad de artículos, la forma para determinar su inclusión dentro del estudio, estuvo basado en el título, palabras claves y resumen, teniendo en cuenta que el objetivo de dichos artículos debía ser el uso de datos enlazados en la Web para la creación de aplicaciones e investigaciones en el ámbito del turismo. Para determinar la calidad de los estudios, se plantearon las siguientes preguntas:

- ¿Se especifica con claridad las fuentes *Linked Data* que fueron usados para llevar a cabo el estudio?
- ¿Las investigaciones dieron como producto final una herramienta informática que es usada en la actualidad?
- ¿Está claramente definido y especificado la metodología o procedimiento empleado para llevar a cabo el estudio?

Este criterio de calidad, sumado al criterio de exclusión e inclusión definidos previamente, redujeron el número de artículos considerablemente. A continuación se procederá a presentar una selección de las aplicaciones e investigaciones obtenidas en la revisión del estado del arte.

### 2.3. Tablas Resumen del estado del arte

#### 2.3.1. Investigaciones en distintos dominios

En la Tabla 4 se puede observar el resumen de las investigaciones realizadas en el ámbito de *Linked Data* y su aplicación en diferentes dominios.

**Tabla 4:** Resumen de investigaciones

	Dominio del estudio	¿Publica información en la Web según <i>Linked Data</i> ?	Fuentes <i>Linked Data</i> empleadas	Uso de ontologías	Se diferencia en
[Curé, 2013]	Salud y automedicación de personas.	No	DrugBank DailyMed Sider DBPedia	No	Estudio sobre repositorios <i>Linked Data</i> relacionados a fármacos y automedicación. Concluye en el desarrollo de una aplicación web para ayudar a las personas que se auto medican para que estén mejor informadas.

	Dominio del estudio	¿Publica información en la Web según <i>Linked Data</i> ?	Fuentes <i>Linked Data</i> empleadas	Uso de ontologías	Se diferencia en
[Galiotou, Fragkou, 2013]	Información de administración pública del gobierno de Grecia.	Sí	No	Sí	Propone una metodología para publicar, de acuerdo a los principios de <i>Linked Data</i> , información de administración pública del gobierno de Grecia.

### 2.3.2. Aplicaciones en el ámbito del turismo

En la Tabla 5 se puede contemplar el resumen de las aplicaciones que han sido desarrolladas en el ámbito de *Linked Data* que pertenecen al ámbito del turismo.

**Tabla 5:** Resumen de aplicaciones revisadas

Nombre de la herramienta	Dominio de aplicación	Fuentes <i>Linked Data</i> empleadas	Uso de ontologías	Resumen
SmartTourism	Atractivos turísticos y patrimonios culturales de la región de Castilla y León en España.	DBPedia.	Ontología propia con nombre SmartTourism, que se basó en dos ontologías ya existentes: Contur y Cruzar.	Aplicación móvil como herramienta para acceder a la información de atractivos turísticos y patrimonio cultural de la región de Castilla y León en España. Emplea datos publicados por entidades gubernamentales.
eZaragoza	Atractivos turísticos de la ciudad de Zaragoza en España.	Proyecto Cruzar. (Proyecto de Web Semántica realizado por el gobierno de Zaragoza, que además posee una base de datos publicada en la Web mediante <i>Linked Data</i> ).	No menciona.	Aplicación Web con la finalidad de promover el turismo de la ciudad, que se basa principalmente en el uso y explotación de datos publicados en la web siguiendo los principios de <i>Linked Data</i> . Utiliza otras fuentes como Flickr, YouTube y Google Maps.



Nombre de la herramienta	Dominio de aplicación	Fuentes <i>Linked Data</i> empleadas	Uso de ontologías	Resumen
LinkedQR	Piezas de arte en una galería localizada en la ciudad de San Sebastián en España.	DBPedia.	No menciona.	Aplicación móvil basada en el uso de <i>Linked Data</i> con los códigos QR, específicamente para su uso en una galería de arte.

## 2.4. Investigaciones en distintos dominios

A continuación se presentaran en detalle las investigaciones que han sido consideradas como relevantes de acuerdo a la revisión realizada, y que fueron desarrolladas en los últimos años, basándose principalmente en el uso de *Linked Data* en distintos dominios.

### 2.4.1. Aplicación Web orientada a la automedicación de personas

Debido a la creciente práctica de la automedicación por parte de las personas, muchas entidades gubernamentales están solicitando la creación de herramientas educativas que permitan instruir a las personas en temas de salud y automedicación. Como resultado de ello, existe el Open Self-Medication Project, que consiste en una aplicación Web basada en datos abiertos enlazados, o mejor llamados *Linked Open Data*, especializados en fármacos y medicamentos. Esta aplicación presenta información detallada a los usuarios respecto a fármacos y medicamentos en base a algún síntoma que puedan tener.

Las fuentes de datos usados para desarrollar esta aplicación son DrugBank, DailyMed y DBPedia. Estas fuentes de datos forman parte del llamado *Linked Open Data Cloud*, el cual es un conjunto de varias bases de datos publicadas en la Web siguiendo los principios de *Linked Data*, pudiendo ser de temas y dominios distintos. Debido a que en estas fuentes de datos aún no existe información del todo precisa y consistente respecto a productos farmacéuticos complejos, este proyecto sólo considera aquellos productos que solo tengan un compuesto químico. Además, para poder determinar si un producto puede ser catalogado como apto para el autoconsumo por parte de las personas, fue necesario el trabajo en conjunto de profesionales expertos en salud, debido a que estos productos carecen de una advertencia o propiedad explícita que los indique como aptos para el autoconsumo. Por medio de esta aplicación Web, las personas pueden seleccionar entre un conjunto de síntomas más comunes, como por ejemplo tos, para que luego se muestre una lista de los compuestos que pueden dar solución a ese síntoma. Para poder obtener mayor información respecto a algún compuesto, el usuario puede seleccionarlo, con lo cual se abrirá otra pantalla con información específica y detallada del compuesto seleccionado.



## 2.4.2. Aplicación de tecnologías de *Linked Data* a datos libres del gobierno griego

Este caso de estudio tiene como objetivo definir y proponer una metodología para la aplicación de las tecnologías de *Linked Data* en la información que tiene el gobierno de Grecia en su portal ERMIS, dedicado a la administración pública. En dicho portal, se almacena información que es de interés público para los ciudadanos y empresas privadas que operan en Grecia, en específico información relacionada a los tramites y transacciones que ellos, tanto los ciudadanos como las empresas, pudieran tener con alguna entidad pública del estado [Portal ERMIS, 2015]. Este caso de estudio se enfoca sólo en aquella información relacionada a las actividades y requerimientos necesarios para la provisión de cualquier tipo de servicio, y que está regulada bajo la directiva 123/2006/EC. Esta directiva se enfoca en simplificar el procedimiento administrativo necesario por el que un ciudadano europeo debe pasar para poder practicar alguna profesión en algún país miembro de la Unión Europea. Para ello, cada país miembro posee un portal denominado PSC o Point of Single Contact el cual contiene información relacionada a la documentación necesaria para la provisión de algún servicio o profesión.

La manera en que esta información esta almacenada en el portal PSC de Grecia no ofrece ningún tipo de conexión o relación explícita entre servicios o profesiones similares o que en realidad sí debieran tener algún tipo de relación, como por ejemplo, un servicio ser pre requisito de otra. Ante esta situación se propone una metodología de cuatro pasos, los cuales se describirán a continuación:

### 1) Recuperación y almacenamiento de páginas Web

Debido a que las páginas Web del portal ERMIS, y los datos almacenados en ellas, no pueden ser modificados, como por ejemplo añadirle etiquetas RDF o anotaciones, es necesario realizar una copia de cada una de las paginas, así como también definir mecanismos que permitan verificar cambios en las páginas originales periódicamente.

### 2) Uso de una ontología.

Es necesario el uso de una ontología para la definición e identificación de conceptos básicos, y sus atributos, así como también las relaciones entre ellos, existentes en la información que se manejará en relación a los servicios o profesiones. Para ello, ya existe una ontología usada por el portal ERMIS llamada e-GIF, como herramienta básica para la organización de su información.

### 3) Definición de conexiones potenciales entre las fuentes

Se plantea la posibilidad de definir conexiones entre la información almacenada en los portales de cada uno de los países miembros que posean un PSC. Para poder llevar a cabo esta tarea, una ontología en común debe ser usada como medio de interconexión, es por ello que se propone a la ontología empleada DBPedia como el candidato.

#### 4) Resultados estadísticos

De las conexiones realizadas a los datos, se puede obtener estadísticas que sirvan de fuente a autoridades públicas con la finalidad de que ellos obtengan más información acerca de si un procedimiento administrativo necesita ser mejorado o no. Por ejemplo, se podría determinar ¿Cuántos servicios o profesiones son regulados por el sector público X?, ¿En qué servicios o profesiones la documentación A y B es requerida?, entre otros.

#### 2.5. Aplicaciones en el ámbito del turismo

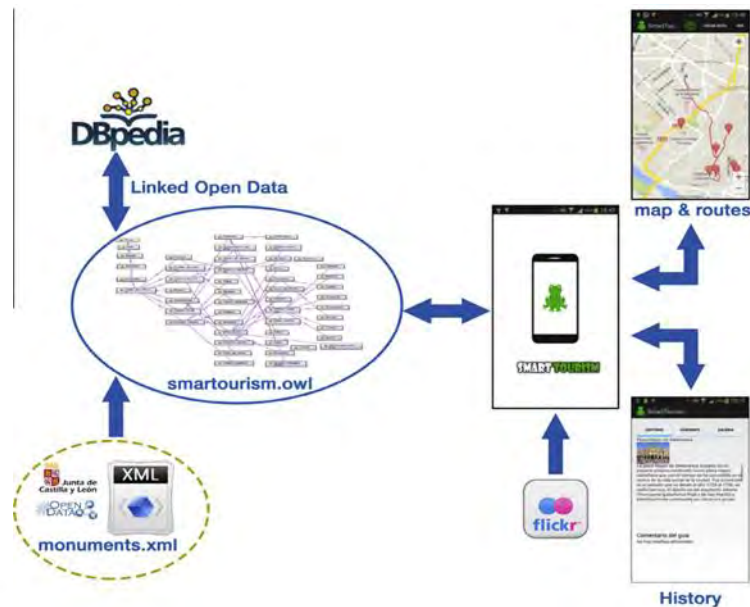
A continuación se presentaran en detalle las aplicaciones que han sido desarrolladas en los últimos años, basándose en el uso de *Linked Data* en el ámbito del turismo. Se hará énfasis en el tipo de arquitectura que emplean, así como también en el uso de ontologías o algún otro tipo de tecnología adicional.

##### 2.5.1. Aplicación que hace uso de ontologías y *Linked Data* – SmartTourism

SmartTourism es un sistema cuyo objetivo es promover información del patrimonio cultural por medio de la colaboración entre los datos publicados por entidades gubernamentales regionales y enriquecidas con datos de otras fuentes provenientes del Linked Open Data Cloud. Esto en la práctica dio como resultado una aplicación móvil como herramienta para acceder a la información mencionada.

La información que provee la aplicación proviene principalmente de tres fuentes de datos. En primer lugar, se obtiene datos que son publicados por el gobierno regional de Castilla y León en España siguiendo los principios de *Linked Data*. Esta información está relacionada al patrimonio cultural de dicha región, en específico se seleccionó la información acerca de monumentos. Entre los datos seleccionados para cada monumento se puede mencionar su nombre, dirección, estilo arquitectónico, horario de atención, precios, entre otros. En segundo lugar se usó DBPedia, que proviene del Linked Open Data Cloud, para obtener información detallada respecto a la historia de los monumentos. Por último, se obtiene fotografías de cada uno de los monumentos, provenientes del sitio Web Flickr.

Para poder compartir e integrar la información proveniente de las distintas fuentes, en especial aquella de DBPedia, se creó una ontología. Los datos provenientes de DBPedia son almacenados de acuerdo a las especificaciones y formato diseñado en la ontología creada, de manera que posteriormente puedan ser consultadas. Esta ontología creada tiene el nombre de smarttourism.owl, y está basada en dos ontologías ya existentes relacionadas al área de cultura y turismo: Cruzar, el cual es una ontología creada por el municipio de Zaragoza, y Contur, que es una ontología de la Universidad de Standford. En la Figura 5 se puede apreciar un esquema básico de la arquitectura empleada para el desarrollo de esta aplicación móvil.

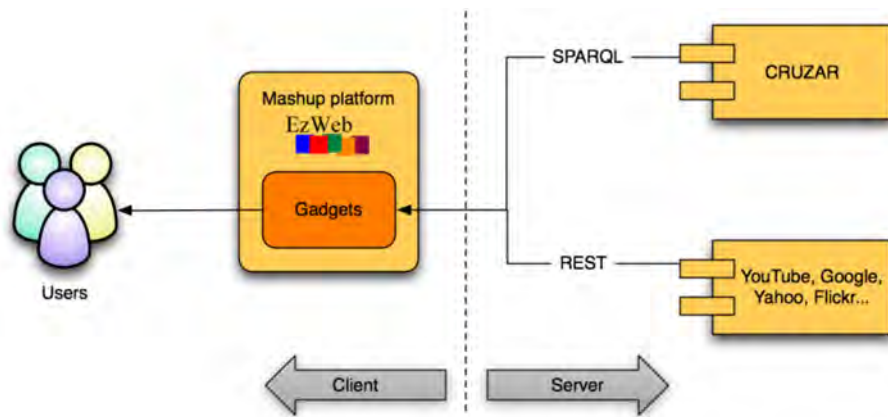


**Figura 5:** Arquitectura empleada para el desarrollo de la aplicación SmartTourism.  
[Fermoso, Mateos, Beato, Berjon 2015]

### 2.5.2. Aplicación que hace uso de una plataforma *Linked Data Mashup* para su desarrollo – eZaragoza

El municipio de la ciudad de Zaragoza creó una aplicación Web, llamada eZaragoza, con la finalidad de promover el turismo, que se basa principalmente en el uso y explotación de datos publicados en la web siguiendo los principios de *Linked Data*. eZaragoza fue desarrollado bajo EzWeb, el cual es una plataforma del tipo *Linked Data Mashup* en donde los componentes, más conocidos como gadgets, intercambian información entre ellos, y es lo suficientemente flexible que permite a los usuarios crear y desplegar nuevos componentes o gadgets [Portal EzWeb, 2015]. El municipio de Zaragoza posee una gran cantidad de bases de datos con información relacionada al patrimonio cultural de dicha ciudad, y desde el año 2008 están disponibles en la Web como *Linked Data*.

La arquitectura empleada para esta aplicación es del tipo cliente/servidor. Los componentes o gadgets vienen a ser los clientes, y son desplegados dentro de la plataforma EzWeb que es ejecutado desde cualquier navegador Web. El servidor es el que provee la información, y teniendo en cuenta que la mayoría de la información se obtiene desde fuentes del tipo *Linked Data*, la información se obtiene a partir del protocolo y lenguaje de consulta SPARQL. SPARQL es un lenguaje estandarizado para la consulta de información representada en grados RDF [Pastor Sánchez, 2011]. También es posible obtener datos a partir de fuentes que no son del tipo *Linked Data*, como por ejemplo Flickr (imágenes), Youtube (videos) y Google Maps (mapas), para ello, la comunicación con estas fuentes de datos se hace a partir de las APIs propietarias de cada uno de ellos, básicamente a partir de servicios Web. En la Figura 6 se puede apreciar un esquema básico de la arquitectura mencionada previamente.



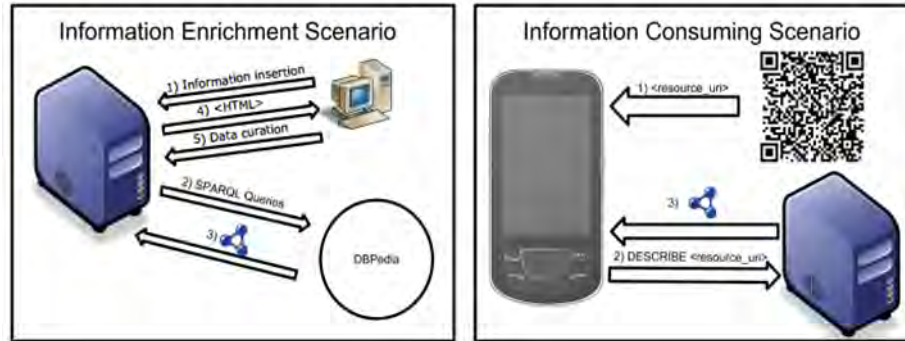
**Figura 6:** Arquitectura empleada para el desarrollo de la aplicación eZaragoza. [Tejo, Fernández, Barrueta, Polo 2013]

### 2.5.3. Aplicación que hace uso de *Linked Data* y la tecnología QR – LinkedQR

LinkedQR es una aplicación móvil basada en la colaboración del uso de *Linked Data* con los códigos QR, aplicado al sector de turismo, específicamente para su uso en una galería de arte. Se aplicó como caso de estudio en una galería de arte en la ciudad de San Sebastián, España. Esta aplicación permite enriquecer la información respecto a piezas de arte que se encuentran en una galería, con datos provenientes de repositorios localizados en la Web siguiendo los principios de *Linked Data*. La información proveniente de estas fuentes externas es almacenada en formato RDF en un repositorio para su posterior consulta a través del lenguaje SPARQL. Esta información almacenada es consumida por una aplicación móvil, que a través de códigos QR logra identificar cada pieza de arte mediante una HTTP URI única.

Como se puede apreciar en la Figura 7, el sistema propuesto tiene dos escenarios distintos, una denominada “Enriquecimiento de información” y la otra “Consumo de información”. En el primer escenario, se obtiene información relacionada a piezas de arte almacenada en una base de datos relacional para posteriormente ser adaptada de acuerdo a la Web Semántica. Luego, con esta información adaptada, se hacen consultas mediante el uso del lenguaje SPARQL a DBPedia para obtener más datos respecto a las piezas de arte. Finalmente, esta información es almacenada en un servidor propietario de la galería o museo en formato RDF, con la posibilidad de que los encargados de puedan modificarla de acuerdo a sus necesidades. En el segundo escenario, los códigos QR son asociados a cada pieza de arte que se encuentre en la galería o museo. El visitante deberá descargar la aplicación desde el portal Google Play Store. Una vez con la aplicación instalada en el dispositivo móvil, el usuario podrá seleccionar la opción “QR” del menú principal, e inmediatamente la aplicación usará un lector de códigos QR para obtener la imagen del respectivo código que estará localizado junto a la pieza de arte, de esta manera se obtiene el HTTP URI de dicha pieza de arte, con lo cual se accederá al servidor RDF para obtener la información respectiva. Finalmente, toda la información respecto a la pieza de arte es presentada al usuario en su dispositivo móvil.





**Figura 7:** Arquitectura empleada para el desarrollo de la aplicación LinkedQR.  
[Emaldi, Lázaro, Laiseca, Lopez 2012]

## 2.6. Conclusiones sobre el estado del arte

Una vez presentado el estado del arte, se tiene una idea más clara de cómo abarcar el proceso de publicación de información relacionada a atractivos turísticos y culturales del Perú, de manera que siga los principios de *Linked Data*. Las herramientas presentadas son específicas a contextos, culturas y sociedades distintas a la nuestra, es por ello que es necesario desarrollar una herramienta que tome como base información relacionada al contexto turístico del país, pero que ofrezca las mismas, o similares, características que presentan estas herramientas. De acuerdo a lo que se investigó, queda muy en claro que existen dos factores que son los más determinantes: la información que ya existe en la Web, y la arquitectura a emplear para el desarrollo de la herramienta. La complejidad de la herramienta, dependerá directamente de que tanta información ya exista en la Web respecto a atractivos turísticos del Perú, y de qué forma esta publicada dicha información, es decir, si es que sigue los principios de *Linked Data* o no.

## CAPÍTULO 3

### Análisis y desarrollo de la representación del conocimiento en el dominio

#### 1. Introducción

En el presente capítulo se presentarán y discutirán los resultados esperados relacionados al primer objetivo del presente proyecto. Como se recuerda, lo que se busca lograr con este objetivo es el análisis, diseño e implementación de la representación del conocimiento en el dominio de los atractivos turísticos y culturales del Perú. Para ello, se seguirá la metodología *Ontology Development 101*, que servirá de apoyo para la fase de análisis y diseño de la ontología. Finalmente, se procederá a implementar la ontología diseñada en la herramienta Protégé.

#### 2. Resultados Esperados del Objetivo Especifico 1 - Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales

Como primer resultado esperado se tiene el análisis y diseño conceptual en el dominio de los atractivos turísticos y culturales, el cual se llevará a cabo teniendo como referencia la metodología *Ontology Development 101*, que permitirá identificar cada elemento o término que se desea conceptualizar, además de cómo se relacionarán entre ellos. Posteriormente, se procederá a implementar una ontología en la herramienta Protégé teniendo como base el análisis y diseño realizado previamente, además de hacer uso del motor razonador que ofrece esta herramienta para verificar su consistencia. Una vez alcanzados los resultados esperados del presente objetivo, en específico la ontología, permitirá ofrecer una estructura semántica, basada en el estándar RDF, para la representación y posterior publicación de la información, respecto a atractivos turísticos y culturales locales, de acuerdo a los principios de *Linked Data*.

##### 2.1. Análisis y diseño conceptual del conocimiento en el dominio de atractivos turísticos y culturales locales

Para el desarrollo de este resultado esperado se empleó la metodología *Ontology Development 101* [Noy, McGuinness, 2001], la cual propone un conjunto de siete pasos a seguir para el análisis y diseño de una ontología.

###### 2.1.1. Paso 1: Determinar el dominio y alcance de la ontología

Para apoyar la consecución de este paso, la presente metodología sugiere contestar las siguientes preguntas [Noy, McGuinness, 2001]:

- ¿Cuál es el dominio que abarcará la ontología?  
El dominio que abarcará esta ontología será los atractivos turísticos y culturales locales, es decir que se encuentran localizados o se desarrollen en el Perú.



- ¿Para qué se utilizará la ontología?  
Se planea desarrollar esta ontología para su uso en aplicaciones que informen respecto a atractivos y eventos turísticos y culturales que tomen lugar en territorio nacional.

- ¿A qué tipo de preguntas la información en la ontología proveerá respuesta?  
Preguntas de carácter informativo acerca de cada atractivo turístico y cultural. Ejemplos de preguntas son las siguientes:

¿Qué lugares puedo visitar en un distrito, o ciudad, determinado?

¿Qué tipos de atractivos puedo visitar en un distrito, o ciudad, determinado?

¿Se cobra algún tipo de tarifa o pago, por visitar o asistir, a un lugar determinado?

¿Cuáles son los horarios de atención de un lugar determinado?

¿Cuál es la dirección exacta de determinado atractivo turístico o cultural?

¿Existe alguna referencia que me pueda guiar para llegar a un lugar determinado?

- ¿Quién usará y dará mantenimiento a la ontología?  
El autor de esta ontología será el encargado de su uso y mantenimiento.

### **2.1.2. Paso 2: Considerar reutilizar ontologías existentes**

Es importante averiguar si es que existen ontologías que traten sobre el mismo dominio, para determinar si pueden ser reutilizadas o servir como base para crear una nueva [Noy, McGuinness, 2001]. Es por ello que se buscó en la Web, en específico en la biblioteca de ontologías Ontolingua, ontologías sobre el dominio del turismo. Del proceso de búsqueda, se seleccionaron dos ontologías, que se presentarán a continuación:

- ETP-tourism: es una ontología desarrollada por la Universidad de Quebec en Montreal. Se utilizará como base debido a que ofrece un modelo fácil de entender, con clases que pueden ser tomadas como base, como por ejemplo: Actividad, Destino y Evento. Se decidió no reutilizarla por completo debido a que contempla clases que no son del interés del presente proyecto, lo cual aumenta su complejidad, como por ejemplo las clases Contacto o Habitación, usadas para brindar información de sitios que brindan alojamiento, o la clase Servicio y Transporte, usadas para brindar información respecto a servicios básicos y de transporte ofrecidos en una ciudad.
- Contur: es una ontología desarrollada por la Universidad de Standford. Similar al caso anterior, se puede utilizar como modelo de referencia debido a que existen clases que pueden ser aplicables, en específico las clases Actividad y Evento. Sin embargo, su alcance es mucho mayor, y su complejidad es muy alta, ya que no se centra en un carácter informativo sobre atractivos turísticos, si no que abarca temas del turismo

más específicos, como medios de transporte, calificaciones de usuarios, estadísticas, lugares de hospedaje, posibilidad de más de un idioma, entre otros.

### 2.1.3. Paso 3: Enumerar términos importantes en la ontología

Previo a definir una estructura jerarquizada, es útil realizar una lista con términos relacionados al dominio de interés. ¿Cuáles son los términos de los que se quiere hablar?, ¿Qué propiedades tienen estos términos?, ¿Qué nos gustaría decir acerca de dichos términos? [Noy, McGuinness, 2001]. Para ello, se realizó la siguiente lista de términos:

- |                      |                         |
|----------------------|-------------------------|
| • Actividad          | • Festival              |
| • Agrupación vegetal | • Horario de visita     |
| • Atractivo Cultural | • Litorales             |
| • Atractivo Natural  | • Mercado Tradicional   |
| • Aventura           | • Museo                 |
| • Biblioteca         | • Parque                |
| • Celebración típica | • Patrimonio Religioso  |
| • Comida y bebida    | • Precio o tarifa       |
| • Departamento       | • Provincia             |
| • Deporte            | • Referencia            |
| • Depósito de agua   | • Relieve               |
| • Dirección          | • Reserva               |
| • Distrito           | • Restricción           |
| • Evento             | • Sightseeing           |
| • Evento Deportivo   | • Vestigio Arqueológico |
| • Exposición         | • Zoológico y Acuario   |
| • Feria              |                         |

### 2.1.4. Paso 4: Definir las clases y jerarquía de clases

Para el desarrollo de la jerarquía de clases, se utilizará el método descendente, o también llamado top-down, es decir, primero se definen los conceptos más generales del dominio, para luego definir aquellos más especializados. Se seleccionan los términos que describen objetos reales y que no dependan de otros; estos términos serán las clases en la ontología [Noy, McGuinness, 2001]. Se organizarán los términos definidos en el paso anterior para definir una jerarquía.

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>1) Actividad <ul style="list-style-type: none"> <li>a. Aventura</li> <li>b. Comida y bebida</li> <li>c. Deporte</li> <li>d. Sightseeing</li> </ul> </li> <li>2) Atractivo Natural <ul style="list-style-type: none"> <li>a. Vegetación</li> <li>b. Depósito de agua</li> <li>c. Litoral</li> <li>d. Accidente geográfico</li> <li>e. Reserva natural</li> </ul> </li> <li>3) Atractivo Cultural <ul style="list-style-type: none"> <li>a. Biblioteca</li> <li>b. Mercado Tradicional</li> <li>c. Museo</li> <li>d. Parque</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>e. Patrimonio Religioso</li> <li>f. Patrimonio Histórico</li> <li>g. Vestigio Arqueológico</li> <li>h. Zoológico y Acuario</li> <li>4) Evento <ul style="list-style-type: none"> <li>a. Celebración típica</li> <li>b. Evento Deportivo</li> <li>c. Exposición</li> <li>d. Feria</li> <li>e. Festival</li> </ul> </li> <li>5) Departamento</li> <li>6) Distrito</li> <li>7) Provincia</li> </ul> |
|---|---|

### 2.1.5. Paso 5 y 6: Definir las propiedades de las clases y Definir los aspectos de las propiedades

Los términos que no fueron seleccionados como clases de la lista inicial, son candidatos a ser propiedades de las clases, es decir, elementos que brindan información de cada clase [Noy, McGuinness, 2001]. En el caso de la presente ontología, se ha definido por ejemplo que los términos “Dirección”, “Referencia”, “Horario de visita”, “Precio o tarifa” y “Restricción” son propiedades de las clases de primer nivel “Actividad”, “Atractivo Natural”, “Atractivo Cultural” y “Evento”. Además, existen otro tipo de propiedades que surgen a partir de relaciones entre clases. En este caso, se puede mencionar, por ejemplo, que la clase “Distrito” tiene una propiedad, que se denominará “Pertenece a una provincia” que la relacionará con la clase “Provincia”; o que la clase “Distrito” tiene la propiedad “Posee atractivos naturales” que la relacionará con la clase “Atractivo Natural”.

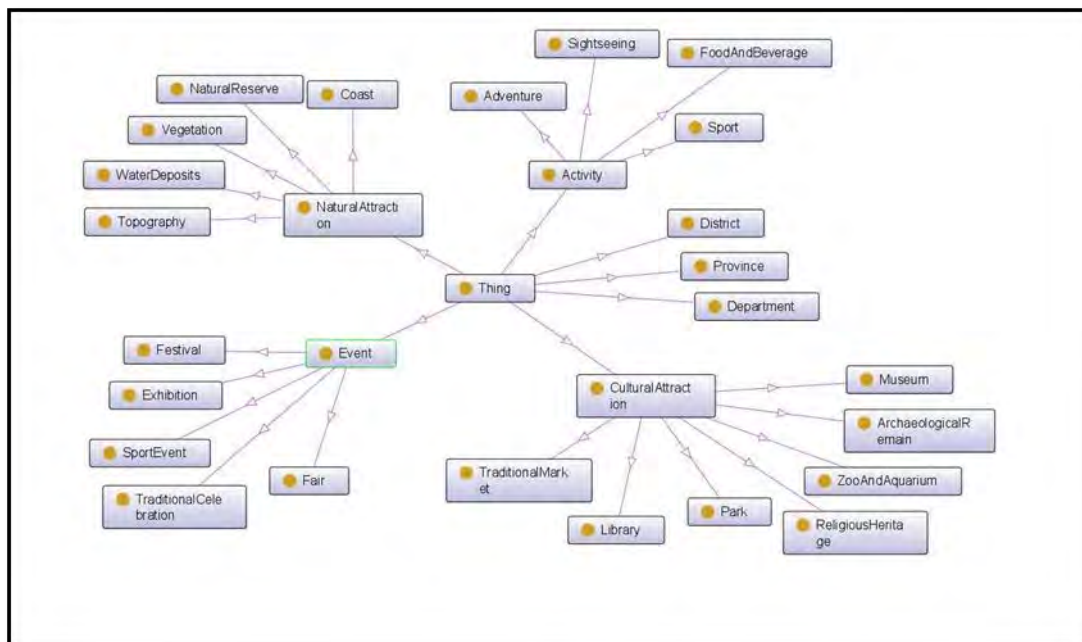
Por último, se definieron aspectos de cada una de las propiedades, es decir, su cardinalidad, tipo de valor, dominio y rango. Por ejemplo, la propiedad “Dirección”, de la clase “Evento”, tendrá la cardinalidad de uno, debido a que un evento tendrá una única dirección en donde se llevará a cabo. El tipo de valor que puede tener la propiedad “Dirección” será una cadena de caracteres.

### 2.1.6. Paso 7: Crear instancias

El último paso consiste en crear instancias de las clases de la jerarquía. Para ello se requiere (1) escoger una clase, (2) crear una instancia de esa clase, (3) llenar sus propiedades [Noy, McGuinness, 2001]. Para este paso, se escogió por ejemplo la clase “Museo”, y se instanció con el valor de “Museo de Arte de Lima – MALI”. Siguiendo el mismo ejemplo, se llenaron sus propiedades “Dirección” y “Referencia” con las cadenas de caracteres “Paseo Colón 125, Lima” y “Dentro del Parque de la Exposición de Lima, a unas cuadras del Centro Cívico” respectivamente.

## 2.2. Modelo de Ontología que represente el conocimiento en el dominio de atractivos turísticos y culturales locales

Para el desarrollo de la ontología se hizo uso de la herramienta Protégé, y se procedió a seguir la estructura y jerarquía de clases definida en el punto anterior. Al momento de implementar la ontología en la herramienta, se decidió utilizar los términos definidos previamente, pero en el idioma inglés, de manera que sea más estándar y pueda ser reutilizado y entendible por otras personas. El nombre seleccionado para la ontología es “tourismPeru “. En la Figura 8 se puede observar la ontología desarrollada en Protégé.



**Figura 8:** Ontología desarrollada en el dominio de atractivos turísticos y culturales locales. Imagen de autoría propia

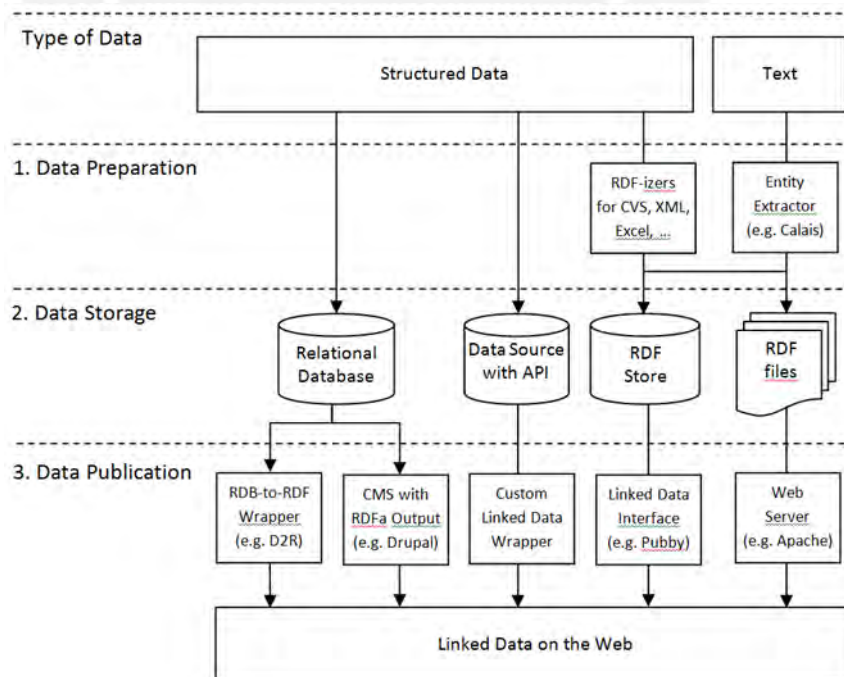
## CAPÍTULO 4

### Publicación de información según *Linked Data*

#### 1. Introducción

En el presente capítulo se presentarán y discutirán los resultados esperados correspondientes al segundo objetivo específico del presente proyecto. Lo que se busca es publicar la información relacionada al dominio de atractivos turísticos y culturales locales en la Web de acuerdo a los principios de *Linked Data*. Para ello, es preciso indicar que existen diversas formas o patrones para poder llevar a cabo la publicación, y la elección de cuál usar dependerá principalmente del tipo de información que se tenga, como esta almacenada y su volumen.

Principalmente, existen dos tipos de información, estructurada y no estructurada [Heath, Bizer, 2011]. Un ejemplo de información estructurada puede ser aquella que está almacenada en una base de datos relacional, de acuerdo a una estructura definida previamente. Un ejemplo de información no estructurada es aquella que puede provenir de archivos de texto plano sin ningún tipo de formato determinado. En la figura 9 se puede apreciar un diagrama indicando algunos de las maneras en que se puede publicar información según *Linked Data*. Como se puede ver, como primer factor se considera el tipo de dato, principalmente si está estructurado o no. Posteriormente, puede ser necesario llevar a cabo procesos de preparación de los datos, para luego almacenarlos en un repositorio. De acuerdo al tipo de repositorio, por ejemplo si es una base de datos relacional o un repositorio RDF, se procede a la publicación de la información.



**Figura 9:** Patrones de publicación de información en *Linked Data*, según origen, tipo y forma de almacenamiento. Imagen recuperada de [Heath, Bizer, 2011]



Para el presente proyecto, se considerará información de atractivos turísticos y culturales del Perú que estará almacenada en una base de datos relacional, debido a que su mantenimiento y actualización resultará mucho más sencillo, en contraste a manejar simples archivos de texto.

## **2. Resultados Esperados del Objetivo Especifico 2 - Desarrollar un repositorio con información respecto a atractivos turísticos y culturales locales aplicando los principios y tecnologías de *Linked Data***

Cada uno de los resultados que se presentarán conforma un paso en todo el proceso de publicación de la información en la Web de acuerdo a *Linked Data*. Como se mencionó previamente, la información se almacenará en una base de datos relacional, por lo cual se tendrá como primer resultado la estructura de esta base de datos y su implementación con una selección de datos reales. Luego se mapeará esta estructural relacional con la ontología desarrollada de modo que la información pueda ser publicada posteriormente según el estándar RDF. Como último paso se llevará a cabo la publicación de la información, para lo cual se hará uso de la plataforma D2RQ, que ofrece la característica de publicar información que proviene de una base de datos relacional y mapeada a una ontología, en un equipo o servidor para que luego pueda ser accedida.

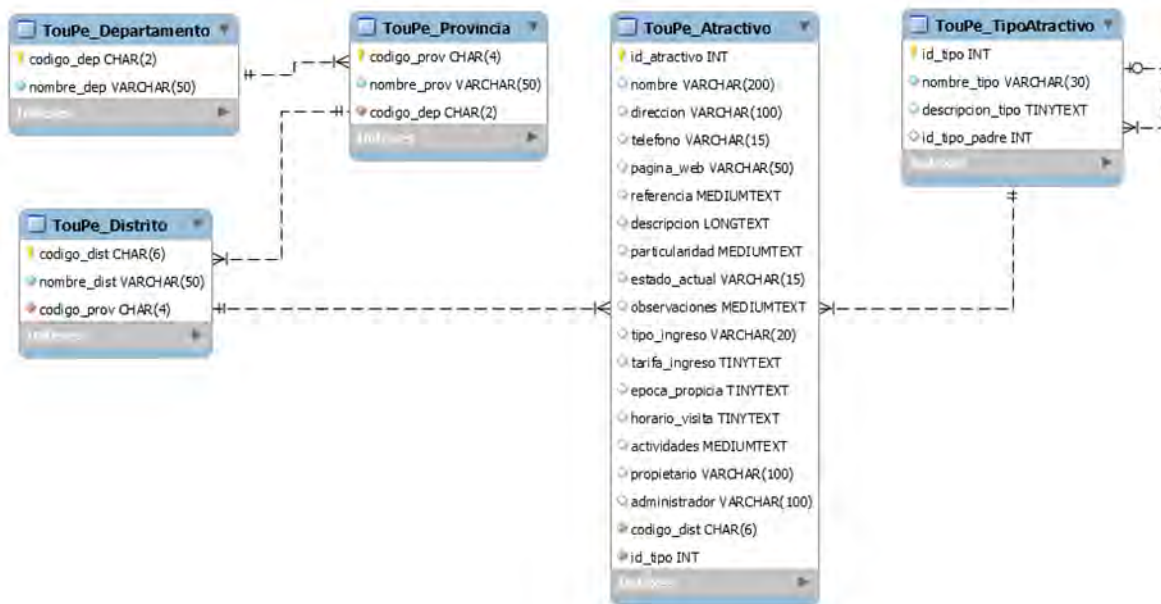
### **2.1. Base de datos relacional con información de atractivos turísticos y culturales locales**

De acuerdo a los principios de *Linked Data*, cuando se accede a un recurso en la Web, este debe ofrecer su información utilizando ciertos estándares como RDF. Sin embargo, teniendo en cuenta que la información a publicar en el presente proyecto se encontrará almacenada en una base de datos relacional, la forma en que está estructurada es totalmente distinta. Una forma para lograr la publicación de información almacenada en una base de datos relacional es mapeando dicha estructura relacional con un esquema o estructura semántica como lo son los vocabularios RDF o las ontologías [Heath, Bizer, 2011]. Es por ello que, como parte del primer objetivo, se desarrolló una ontología que represente el conocimiento en el dominio de los atractivos turísticos y culturales locales en un enfoque informativo, y en esta sección se desarrollará una base de datos relacional para almacenar la información requerida, y posteriormente mapear su estructura con la ontología.

La información que se almacenará respecto a atractivos turísticos y culturales locales, tendrá un enfoque solo informativo, es decir, datos que brinden información a personas que tengan el interés de visitar dichos lugares. Además, para cumplir el principio de *Linked Data* que refiere sobre la necesidad de enlazar con otros repositorios, se agregó a cada atractivo almacenado en la base de datos, un campo cuyo valor será el enlace URI de DBPedia, en donde se encuentra información de cada atractivo, o en su defecto, un enlace URI que haga referencia al departamento o país en donde está localizado. Este mismo enfoque se tuvo en cuenta al momento del análisis, diseño e implementación de la ontología. Es por ello que es necesario aclarar que todos los datos que se desean considerar deben estar definidos en ambos modelos de representación, tanto en el modelo semántico representado por la ontología, como en el modelo relacional, representado por la base de datos, cada uno con su forma y lógica distinta de estructurar el conocimiento.

Para el presente proyecto, los datos a almacenar en la base de datos se obtendrán del Ministerio de Comercio Exterior y Turismo, específicamente de su Inventario de Recursos Turísticos del Perú. Si bien esta información puede ser accedida de manera pública mediante un portal Web, no presenta alguna opción que permita su exportación de manera masiva, por lo que la información será obtenida de dicha fuente de manera manual, para posteriormente ser importados a la base de datos relacional que se desarrollará. Debido a este proceso manual, y a la gran cantidad de tiempo que requeriría obtener la información de todos los atractivos turísticos del país, se optará por utilizar sólo una selección de ellos.

Teniendo en cuenta lo explicado y definido previamente, y a la información disponible a la que se puede tener acceso, se puede apreciar en la Figura 10 el modelo de la base de datos. El nombre que se le dio a la base de datos es “tourismperu”. Como se puede ver, la entidad o tabla principal del modelo es la denominada “TouPe\_Atractivo”, en la cual se almacenará los atractivos turísticos y sus principales atributos como nombre, dirección, pagina web, entre otros. También se cuenta con una entidad, denominada “TouPe\_TipoAtractivo” para almacenar el tipo de atractivo turístico. Además se cuenta con tablas específicas para almacenar los departamentos, provincias y distritos del país, que estará asociado a los atractivos turísticos. Para la implementación de la base de datos con datos reales, se empleó MySQL, que además ofrece la funcionalidad de poder realizar el modelado previo dentro del su mismo entorno. El detalle y descripción de todas las tablas y columnas de la base de datos se encuentra en el Anexo 1.



**Figura 10:** Modelo de la base de datos relacional para el almacenamiento de información de atractivos turísticos y culturales locales. Imagen de autoría propia

## 2.2. Mapeo entre la base de datos relacional y la Ontología

Para poder publicar la información en la Web, es necesario realizar un proceso previo, debido a que la forma en la que la información está estructurada en la base de datos no

cumple con los principios de *Linked Data*, específicamente con aquel principio relacionado al uso del estándar RDF para la representación de la información. La solución para este escenario es realizar la publicación de una vista, de la información almacenada en la base de datos, que siga los principios de *Linked Data*, para lo cual es necesario el uso de ciertas herramientas y técnicas que apoyen a dicho propósito [Heath, Bizer, 2011].

Una de las herramientas diseñadas para dicho fin, y que es una de las más usadas, es la plataforma D2RQ, que se basa principalmente en un mapeo declarativo entre el esquema de una base de datos y estructuras que emplean RDF, como por ejemplo vocabularios RDF u ontologías [Heath, Bizer, 2011]. Es por ello que se decidió emplear dicha herramienta, junto al lenguaje declarativo que emplea para hacer el mapeo denominado *D2RQ Mapping Language*, con la finalidad de poder mapear la ontología desarrollada con la estructura de la base de datos relacional para la publicación de la información almacenada en él de acuerdo a *Linked Data*. Como resultado de esta actividad se tiene un archivo en el que estará definido, mediante sentencias en el lenguaje D2RQ, el mapeo entre ambas estructuras; el nombre de este archivo es “touPeru-mapping.ttl”

En la Figura 11 se puede apreciar la sección del archivo en donde se realiza la conexión a la base de datos y la creación de una entidad denominada “database” para mapearla e identificarla en el archivo, esto gracias a la sentencia “map:database a d2rq:Database;”. Por otro lado, un ejemplo de mapeo entre una tabla de la base de datos con una clase de la ontología se puede apreciar en la Figura 12. Como se puede ver se hace uso de la sentencia “map:toupe\_atractivo a d2rq:ClassMap;” para indicar que se quiere asociar la tabla “toupe\_atractivo” y para indicar la clase de la ontología se utilizó la sentencia “d2rq:class tourismPeru:Attraction;”. Por último, para mapear las columnas de una tabla de la base de datos a las propiedades de una clase de la ontología, se hace uso de la sentencia “d2rq:PropertyBridge”; un ejemplo de ello se puede apreciar en la Figura 13, en donde se asocia la columna “nombre” de la tabla “toupe\_atractivo” a la propiedad “attraction\_name” de la ontología.

```
map:database a d2rq:Database;  
d2rq:jdbcDriver "com.mysql.jdbc.Driver";  
d2rq:jdbcDSN "jdbc:mysql:///tourismperu";  
d2rq:username "root";  
jdbc:autoReconnect "true";  
jdbc:zeroDateTimeBehavior "convertToNull";
```

**Figura 11:** Mapeo de la base de datos “tourismperu” a la entidad “database”. Imagen de autoría propia

```
# Table toupe_atractivo
map:toupe_atractivo a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "toupe_atractivo/@@toupe_atractivo.id_atractivo@";
  d2rq:class tourismPeru:Attraction;
  d2rq:classDefinitionLabel "toupe_atractivo";
```

**Figura 12:** Mapeo de la tabla “toupe\_atractivo” de la base de datos con la clase “Attraction” de la ontología. Imagen de autoría propia

```
map:toupe_atractivo_nombre a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:toupe_atractivo;
  d2rq:property tourismPeru:attraction_name;
  d2rq:propertyDefinitionLabel "toupe_atractivo nombre";
  d2rq:column "toupe_atractivo.nombre";
```

**Figura 13:** Mapeo de la columna “toupe\_atractivo.nombre” de la base de datos con la propiedad “attraction\_name” de la ontología. Imagen de autoría propia

### 2.3. Publicación de la base de datos de atractivos turísticos y culturales locales de acuerdo a los principios de *Linked Data*

Una vez realizado el mapeo entre la estructura de la base de datos relacional con la ontología, se procederá a publicar la información de acuerdo a los principios de *Linked Data*. Para ello, es necesario el uso de un componente de la plataforma D2RQ denominado D2R Server, el cual tomará como entrada la información de la base de datos relacional, y haciendo uso del documento de mapeo “touPeru-mapping.ttl”, publicará la información, en una URI en específica, en formato de tripletas RDF.

El uso de una URI para la publicación se da en concordancia con los principios de *Linked Data*, de manera que el recurso, en este caso la información, sea identificado y accedido gracias a esta URI. Para efectos prácticos del presente proyecto, la información sólo será publicada y accedida localmente en un equipo para dicho fin.

Por último, se usó el entorno web que ofrece la plataforma D2RQ para acceder y consultar la información publicada, es decir, de un SPARQL Endpoint. En la Figura 14, en la parte posterior, se puede apreciar un ejemplo de consulta SPARQL que se ejecutó en el SPARQL Endpoint para obtener información de los atractivos turísticos como su nombre y dirección, de aquellos que están situados en el distrito del Callao, así como también su respectivo resultado en la parte inferior de la figura.



```

SELECT ?atractivo
      ?nombreAtractivo
      ?direccion
      ?distrito
      ?nombDistrito
WHERE
{
  ?atractivo <http://localhost:8080/resource/vocab/toupe_atractivo_direccion> ?direccion.
  ?atractivo <http://localhost:8080/resource/vocab/toupe_atractivo_nombre> ?nombreAtractivo.
  ?atractivo <http://localhost:8080/resource/vocab/toupe_atractivo_codigo_dist> ?distrito.
  ?distrito <http://localhost:8080/resource/vocab/toupe_distrito_nombre_dist> ?nombDistrito.
  FILTER (?nombDistrito = "CALLAO")
}

```

atractivo	nombreAtractivo	direccion	distrito	nombDistrito
db.toupe_atractivo/14	"Isla San Lorenzo"	"Situada a una distancia de 4 Km. al frente de la bahía del Callao"	db.toupe_distrito/070101	"CALLAO"

**Figura 14:** Ejemplo de consulta SPARQL realizada a la información de atractivos turísticos publicada de acuerdo a *Linked Data*. Imagen de autoría propia

#### 2.4. Conjunto de pruebas de consulta a la información publicada en la Web

Para poder validar la correcta publicación de la información de acuerdo al estándar RDF propuesto por *Linked Data*, se aplicará las medidas de Precisión y *Recall*. La Precisión es el ratio entre el número de registros relevantes obtenidos y el número de registros relevantes e irrelevantes obtenidos en una búsqueda. Por otro lado, el *Recall* es el ratio entre el número de registros relevantes obtenidos y el número de registros relevantes totales almacenados en el repositorio para una búsqueda en específico.

Para estas pruebas se consideró una selección de 50 atractivos turísticos y culturales, los cuales fueron almacenados en la base de datos. Además, las tablas correspondientes para el almacenamiento de los departamentos, provincias, distritos y tipos de atractivos turísticos también fueron poblados con datos reales y en su totalidad. En las pruebas se realizaron consultas para determinar el número de registros totales de cada una de las tablas, así como también consultas más elaboradas empleando filtros.

Se ejecutaron una serie de consultas, en el SPARQL Endpoint, sobre la información publicada y se comparó con la información almacenada originalmente en la base de datos relacional. En el proceso de la obtención de las mediciones, se vio necesario ajustar el documento de mapeo, de manera que tanto las medidas de Precisión y *Recall* alcancen el 100% ambas, ya que esto significa que el mapeo fue realizado correctamente. Cabe precisar que, si bien desde un principio se sabía que estas medidas deberían llegar al 100%, su uso fue necesario para poder validar que el documento de mapeo fue realizado correctamente. En la Tabla 5 se encuentra las pruebas de consulta realizadas y las medidas obtenidas. En el Anexo 2 se especifica más a detalle el proceso para obtener las medidas de Precisión y *Recall*.



**Tabla 6:** Consultas y respectivas medidas de Precision y *Recall*

N°	Consulta	BD Relacional Número registros (Registros relevantes)	Linked Data Número registros relevantes obtenidos	Linked Data Número registros irrelevantes obtenidos	Precisión	Recall
1	Todos los atractivos	50	50	0	100%	100%
2	Todos los departamentos	25	25	0	100%	100%
3	Todas las provincias	195	195	0	100%	100%
4	Todos los distritos	1834	1834	0	100%	100%
5	Todos los tipos de atractivos	26	26	0	100%	100%
6	Atractivos que se encuentran en el distrito "CALLAO"	4	4	0	100%	100%
7	Atractivos que tienen tipo de acceso "LIBRE"	20	20	0	100%	100%

## CAPÍTULO 5

### Construcción de prototipo de herramienta informática

#### 1. Introducción

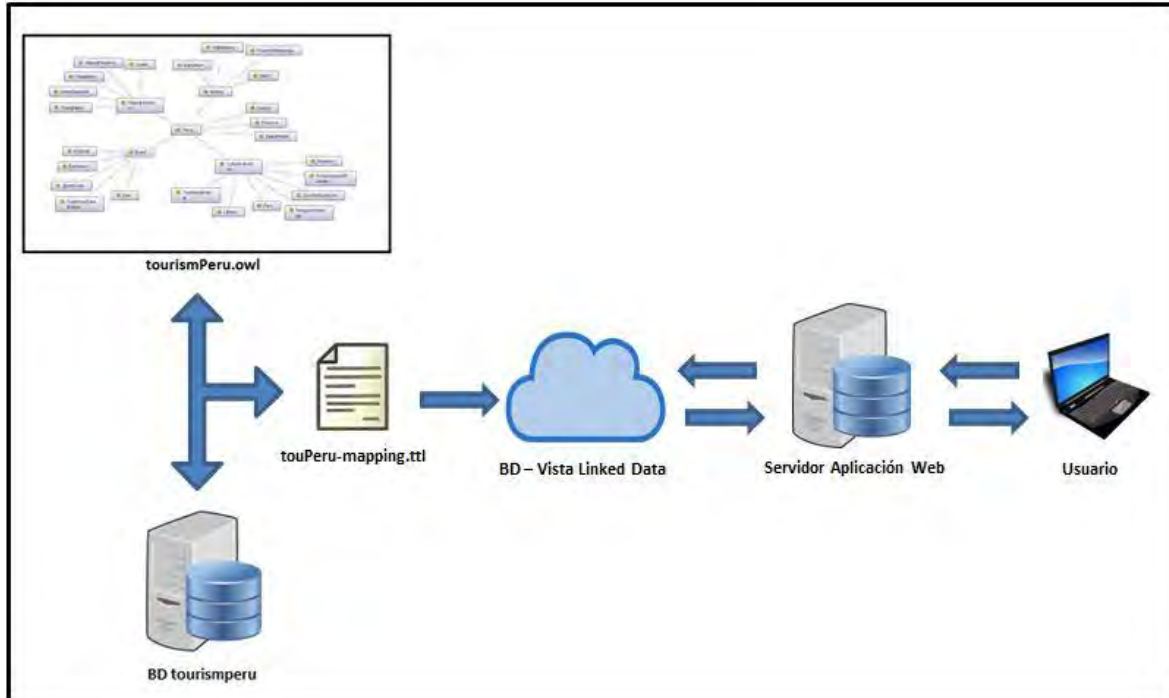
En el presente capítulo se presentarán y discutirán los resultados esperados relacionados al tercer y último objetivo del presente proyecto. Lo que se busca lograr es implementar un prototipo funcional de una aplicación Web que, consumiendo la información publicada que dio como resultado el objetivo anterior, permita informar acerca de atractivos turísticos y culturales locales. Para ello, es importante diseñar un modelo de arquitectura sobre la que se desarrollará la solución, debido a que existen elementos diversos que se relacionan entre sí y conforman la arquitectura, como lo son la base de datos relacional, la ontología, el documento de mapeo, y los componentes propios a usar para desarrollar la aplicación Web. Por último, se procederá a desarrollar una aplicación Web que permita mostrar a los usuarios la información de atractivos turísticos y culturales, publicada de acuerdo a *Linked Data*.

#### 2. Resultados Esperados del Objetivo Especifico 3 - Desarrollar una herramienta informática que permita informar acerca de atractivos turístico y culturales locales

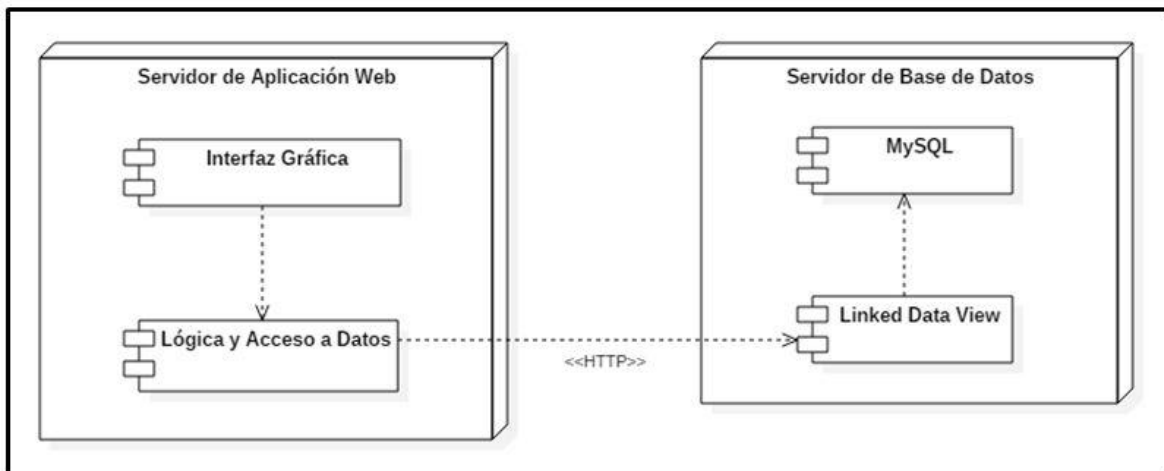
##### 2.1. Arquitectura de software

Dado que existen varios componentes, y de naturaleza distinta, que son empleados para poder llevar a cabo tanto la publicación como el consumo de información de acuerdo a los principios de *Linked Data*, es preciso identificarlos y determinar cómo interactúan entre ellos. En la Figura 15 se aprecia una vista general de todos los elementos involucrados y cómo se relacionan entre ellos. Como se puede observar, la ontología desarrollada “tourismPeru.owl” es mapeada a la estructura de la base de datos relacional “tourismperu”, dando como resultado el documento de mapeo “touPeru-mapping.ttl”. Gracias a este documento de mapeo, se puede publicar una vista de la información almacenada en la base de datos de acuerdo a los principios de *Linked Data*, en especial en relación al uso del estándar RDF para la representación de los datos. Es a esta vista a la que la aplicación Web, con nombre Turismo Perú accederá, tanto para hacer consultas y obtener información de ella, sin necesidad de interactuar directamente con la base de datos.

En la Figura 16 se puede apreciar la Vista de Despliegue de la aplicación Web. Los componentes de la Interfaz Gráfica y Lógica y Acceso a Datos, estarán instalados en un equipo y servidor Web. Por medio del componente de Lógica y Acceso a Datos, se accederá a la vista *Linked Data* de la información publicada



**Figura 15:** Diagrama general de los componentes involucrados en el desarrollo de la solución propuesta. Imagen de autoría propia



**Figura 16:** Vista de Despliegue de la aplicación Web. Imagen de autoría propia

## 2.2. Prototipo de aplicación Web

Como último resultado esperado, se tiene el desarrollo de un prototipo de una aplicación Web denominada Turismo Perú, que consuma la información publicada de atractivos turísticos y culturales, de manera que pueda ser accedida por una persona que desee obtener información sobre algún sitio, lugar turístico, o actividad cultural en específico desde un navegador Web. Debido a que esta información está publicada según el estándar RDF, se

empleó el entorno de desarrollo ASP.NET en conjunto con la librería dotNetRDF, el cual ofrece las características necesarias para poder integrar las tecnologías de la Web Semántica y *Linked Data* en la creación de una aplicación Web [NETRDF].

Como se puede apreciar en la Figura 17, en la aplicación Web el usuario puede buscar atractivos turísticos en base a donde estén localizados, es decir, teniendo la posibilidad de filtrar su búsqueda en base a algún departamento, provincia o distrito del país en específico. En base a estos parámetros, se realizará una consulta SPARQL hacia la vista *Linked Data* de la información de atractivos turísticos publicada, para luego mostrar los resultados al usuario. Posteriormente, el usuario podrá seleccionar un atractivo turístico en particular para visualizar su información, como por ejemplo su dirección, teléfono, tipo de ingreso, tarifa, entre otros; un ejemplo de la visualización de estos datos se puede apreciar en la Figura 18.

BIENVENIDO A TURISMO PERÚ

Busqueda Contacto

Ingrese los siguientes datos para la búsqueda

Nombre:

Departamento:

Provincia:

Distrito:

Tipo de ingreso:

Tipo de atractivo:

Buscar Cancelar

**Figura 17:** Menú de búsqueda de la aplicación Web Turismo Perú. Imagen de autoría propia

**BIENVENIDO A TURISMO PERÚ**

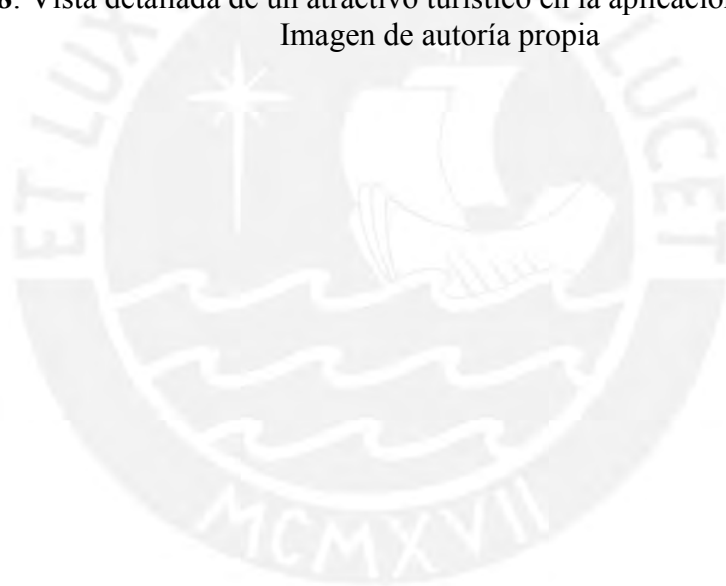
[Busqueda](#) [Contacto](#)

**Museo de Arte de Lima**

Dirección: Paseo Colón N° 125 - 128 -Cercado de Lima  
Teléfono: (51-1) 204-0000  
Página Web: www.malipe  
Referencia: A 1.2 km. / 10 min de la Plaza de Armas.  
Descripción: El Museo de Arte de Lima se encuentra en el Palacio de la Exposición, edificio republicano de arquitectura ecléctica, construido en hierro, con decoración renacentista italiana, y fue diseñado por el arquitecto italiano Antonio Leonardi.  
Estado actual: Bueno.  
Tipo de ingreso: Libre  
Tarifa: Presentación de Documento de Identidad  
Epoca propicia: Todo el Año  
Horario de visita: Todos los días de 9:00 - 5:00 p.m.  
Actividades: Actividades Culturales, Estudios e Investigación  
Propietario: ESTADO PERUANO  
Administrado por: Ministerio de Cultura

[Regresar](#)

**Figura 18:** Vista detallada de un atractivo turístico en la aplicación Web Turismo Perú.  
Imagen de autoría propia





## CAPÍTULO 6

### Conclusiones, recomendaciones y observaciones

#### 1. Conclusiones

Una vez llevado a cabo la aplicación de los conceptos y tecnologías propuestos por *Linked Data* para la publicación de recursos en la Web, se presentarán a continuación las conclusiones obtenidas por cada uno de los objetivos específicos que se plantearon al inicio del presente proyecto de fin de carrera.

##### 1.1. Conclusiones del Objetivo Específico 1 - Analizar y desarrollar la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales

El análisis llevado a cabo permitió diseñar y modelar una estructura que represente semánticamente el conocimiento en el dominio de los atractivos turísticos y culturales del Perú. En base a este diseño, se pudo implementar una ontología en la herramienta Protégé. La ontología desarrollada como resultado del primer objetivo específico, permitió definir una estructura por la cual se pudo representar semánticamente la información de los atractivos turísticos y culturales del Perú. Esta estructura permitió que la información publicada, como parte de los resultados del objetivo siguiente, se amolde al estándar RDF previo a su publicación de acuerdo a los principios de *Linked Data*.

##### 1.2. Conclusiones del Objetivo Específico 2 - Desarrollar un repositorio con información respecto a atractivos turísticos y culturales locales aplicando los principios y tecnologías de *Linked Data*

Para poder almacenar y manipular la información de atractivos turísticos y culturales se desarrolló una base de datos, usando el motor MySQL, en base a una estructura con un enfoque relacional tradicional. Además de posibilitar la implementación de la base de datos, la estructura definida sirvió como elemento de entrada para la siguiente actividad de este objetivo, es decir, para llevar a cabo el mapeo entre la estructura de la ontología y la base de datos relacional implementada. Es gracias a este mapeo, que la información en la base de datos se estructura de acuerdo a la ontología previa a ser publicada. Una vez producido el mapeo, y su respectivo documento de mapeo, se pudo llevar a cabo el proceso de publicación de una vista *Linked Data* de la información. Por último, gracias a la ejecución de consultas SPARQL, y en base a las medidas obtenidas de Precisión y *Recall* de dichas consultas, se pudo validar que toda la información que está almacenada en la base de datos relacional, esté también publicada en el SPARQL Endpoint.

##### 1.3. Conclusiones del Objetivo Específico 3 - Desarrollar una herramienta informática que permita informar acerca de atractivos turístico y culturales locales

Como parte de este objetivo, se definió una arquitectura mediante la cual se pudo determinar y especificar los elementos que están involucrados en el desarrollo del presente proyecto, y cómo es que se relacionan entre ellos, como por ejemplo la ontología, la base

de datos relacional y el documento de mapeo. Además, se definió una arquitectura básica de software para la posterior implementación de una herramienta que permita consultar y mostrar la información publicada en la Web de acuerdo a *Linked Data*. Por último, se desarrolló un prototipo de una aplicación Web, que consuma la información de los atractivos turísticos y culturales publicada. Esta aplicación Web permite realizar búsquedas de atractivos turísticos en base a ciertos parámetros. En base al valor que tomen estos parámetros, la aplicación Web realiza una consulta SPARQL sobre la información publicada en el SPARQL Endpoint, y obtiene un conjunto de resultados que son mostrados al usuario. Finalmente, el usuario puede seleccionar un atractivo turístico en particular para visualizar información más detallada sobre él, como su dirección, teléfono, tipo de ingreso, tarifa, entre otros.

#### **1.4. Conclusión final**

Como resultado general del presente proyecto de fin de carrera, se aplicaron los principios de *Linked Data* para la publicación de información en el dominio de los atractivos turísticos y culturales del Perú. Para ello, se plantearon y desarrollaron los objetivos y resultados esperados presentados en las secciones anteriores. Como se pudo ver, cada uno de los objetivos específicos se logró satisfactoriamente, de manera que, en su conjunto, permitió, por una parte, la publicación de información en la Web siguiendo los principios de *Linked Data*, y por otra parte, consumir y utilizar esta información con el fin de poder brindar a las personas datos e información de interés, respecto a atractivos turísticos y culturales que se encuentran en nuestro país.

#### **2. Recomendaciones**

En el transcurso del desarrollo de los objetivos planteados, y sus respectivos resultados esperados, se identificaron posibilidades de mejora que valen la pena mencionar. En el diseño de la representación semántica del conocimiento en el dominio de los atractivos turísticos y culturales locales, y su posterior implementación en la ontología, pueden considerarse otros aspectos de los estrictamente informativos que han sido planteados en el presente proyecto. Por ejemplo, pueden considerarse clases y propiedades para poder representar los servicios que ofrece una ciudad y que están relacionados directamente con algún atractivo turístico o actividad cultural, como por ejemplo rutas y líneas de transporte público. De esta manera, se puede ampliar el enfoque y el alcance de la solución planteada, y por consiguiente ofrecer mucha mayor información relacionada a las personas interesadas.

Por otra parte, teniendo en cuenta que en el ámbito del turismo las opiniones y críticas de las personas son muy valiosas, se puede añadir ciertas características a la solución planteada de manera que sea mucho más colaborativa y se nutra con información proveniente de las mismas personas que visitan los atractivos turísticos. En este sentido, se podría almacenar las opiniones y calificaciones que puedan brindar las personas, turistas, que visitan estos lugares, y que esta información esté almacenada en la base de datos relacional para posteriormente publicarla en la Web de acuerdo a los principios de *Linked Data*, y pueda ser accedida de manera pública

### 3. Observaciones

Como parte del desarrollo del primer objetivo, se comprobó que en la actualidad ya existen ontologías en el ámbito del turismo, pero que tienen un enfoque y alcance distintos. Las ontologías que se revisaron tienen un enfoque más estadístico y detallado, y aplicados bajo contextos distintos, es decir, culturas y países distintos al Perú. Sin embargo, se pudo tomar como base algunas clases y atributos que eran comunes a todas para estructurarlas y añadirlas de acuerdo a la necesidad del presente proyecto.

Cabe además mencionar que la información obtenida del Inventario de Recursos Turísticos del Perú, del Ministerio de Comercio Exterior y Turismo, si bien es cierto es abierta al público, no es publicada de manera que pueda ser accedida fácil, rápida y automáticamente por las personas o alguna herramienta informática. Es por ello que, debido al extenso tiempo que consume el extraer esta información de manera manual, se decidió considerar para el presente proyecto sólo una selección de atractivos turísticos y culturales.

Por último, durante el desarrollo de la aplicación Web planteada como último resultado esperado, se pudo comprobar que las tecnologías y herramientas que existen para el desarrollo de aplicaciones en el ámbito de *Linked Data* aun no son muy maduras, y muy pocas tienen soporte y mantenimiento oficial. Es por ello que se decidió emplear la librería dotNetRDF para la creación de la aplicación Web, pues es una de las pocas que cuenta con un soporte continuo por parte de sus desarrolladores.

**Papers y Libros**

1. **[Heflin, 2006]**  
Heflin, J. (2006).  
An Introduction to the OWL Web Ontology Language. Chapter 2.  
Lehigh University. Pensilvania, USA
2. **[Tejo, Fernández, Barrueta, Polo 2013]**  
Tejo-Alonso, C., Fernández, S., Barrueta, D., Polo, L. (2013)  
eZaragoza, a tourist promotional mashup.  
Fundacion CTIC. Gijon, España
3. **[Bizer, Heath, Berners-Lee, 2009]**  
Bizer, C., Heath, T., Berners-Lee, T. (2009)  
Linked Data – The Story so far. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems.
4. **[Hoang, Cung, Truong, Hwang, 2014]**  
Hoang, H., Cung, T., Truong, D., Hwang D. (2014)  
Semantic Information Integration with Linked Data Mashups Approaches.
5. **[Gruber, 1993]**  
Gruber, T. (1993)  
Toward Principles for the Design of Ontologies Used for Knowledge Sharing.  
Stanford Knowledge Systems Laboratory. California, USA.
6. **[Sivakumar, Arivoli, 2011]**  
Sivakumar, R., Arivoli P. (2011)  
ONTOLOGY VISUALIZATION PROTÉGÉ TOOLS – A REVIEW.  
Bharathidasan University. Trichirappalli, India
7. **[O’Reilly, 2005]**  
O’Reilly, T. (2005)  
What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software.
8. **[Berners-Lee, Hendler, Lassila, 2001]**  
Berners-Lee, T., Hendler, J., Lassila, O. (2001)  
Scientific American: Feature Article: The Semantic Web.
9. **[Shadbolt, Hall, 2006]**  
Shadbolt, N., Hall, W. (2006)  
The Semantic Web Revisited.  
IEEE Computer Society

10. **[Fermoso, Mateos, Beato, Berjon 2015]**  
Fermoso, A., Mateos, M., Beato, M., Berjon, R. (2015)  
Open *Linked Data* and mobile devices as e-tourism tools. A practical approach to collaborative e-learning.  
Universidad Pontificia de Salamanca. Salamanca, España.
11. **[Emaldi, Lázaro, Laiseca, Lopez 2012]**  
Emaldi, M., Lázara, J., Laiseca, X., Lopez, D. (2012)  
LinkedQR: Improving Tourism Experience through *Linked Data* and QR Codes.  
Deusto Institute of Technology. Bilbao, España.
12. **[Curé, 2013]**  
Curé, O. (2013)  
On the design of a self-medication web application built on linked open data.  
Université Paris-Est Marne la Vallée. Paris, Francia.
13. **[Galiotou, Fragkou, 2013]**  
Galiotou, E., Fragkou, P. (2013)  
Applying *Linked Data* Technologies to Greek Open Government Data: A Case Study.  
Technological Educational Institute of Athens. Atenas, Grecia.
14. **[Kitchenham, 2004]**  
Kitchenham, B. (2004)  
Systematic reviews. Proceedings.10th International Symposium on.
15. **[Bizer, Cyganiak, Heath, 2007]**  
Bizer, C., Cyganiak, R., Heath, T. (2007)  
How to publish *Linked Data* on the Web.
16. **[Pastor Sánchez, 2011]**  
Pastor Sanchez, J. (2011)  
Tecnologías de la web semántica.  
1ª edición. Barcelona, Editorial UOC.
17. **[Fensel, 2003]**  
Fensel, D. (2003)  
Spinning the semantic Web : bringing the World Wide Web to its full potential.  
MIT Press. Cambridge, Massachusetts.
18. **[Jizba, 2007]**  
Jizba, R. (2007)  
Measuring Search Effectiveness  
Creighton University, Nebraska, Estados Unidos.



**19. [Heath, Bizer, 2011]**

Heath, T., Bizer, C. (2011)

*Linked Data: Evolving the Web into a Global Data Space*

**Páginas Web**

**1. [Stanford Protégé]**

Universidad de Standford – Protégé.

Disponible en: <http://protegewiki.stanford.edu/wiki/Protege>

Consulta: 03 de Agosto de 2015.

**2. [W3C RDF]**

World Wide Web Consortium W3C – Resource Description Framework.

Disponible en: <http://www.w3.org/RDF/>

Consulta: 03 de Agosto de 2015.

**3. [W3C OWL]**

World Wide Web Consortium W3C – Web Ontology Language.

Disponible en: <http://www.w3.org/2001/sw/wiki/OWL>

Consulta: 03 de Agosto de 2015.

**4. [MySQL PROD]**

MySQL – Products Editions.

Disponible en: <http://www.mysql.com/products/>

Consulta: 03 de Agosto de 2015.

**5. [D2RQ Plat]**

*D2RQ Platform* - Accessing Relational Databases as Virtual RDF Graphs.

Disponible en: <http://d2rq.org/>

Consulta: 11 de Setiembre de 2015.

**6. [D2RQ Map]**

D2RQ – Accessing Relational Databases as Virtual RDF Graphs. Mapping Language.

Disponible en: <http://d2rq.org/d2rq-language#introduction>

Consulta: 03 de Agosto de 2015.

**7. [W3C SPARQL]**

World Wide Web Consortium W3C – SPARQL Query Language for RDF.

Disponible en: <http://www.w3.org/TR/rdf-sparql-query/>

Consulta: 03 de Agosto de 2015.

**8. [UML Model]**

UML - Unified Modeling Language.

Disponible en: <http://www.uml.org/>

Consulta: 03 de Agosto de 2015.

9. **[VISUAL]**  
Microsoft Visual Studio.  
Disponible en: <http://www.visualstudio.com>  
Consulta: 13 de Noviembre de 2015.
10. **[ASPNET]**  
ASP.NET.  
Disponible en: <http://www.asp.net/get-started>  
Consulta: 13 de Noviembre de 2015.
11. **[NETRDF]**  
Librería dotNetRDF – Introduction  
Disponible en: <http://www.dotnetrdf.org/>  
Consulta: 13 de Noviembre de 2015.
12. **[CommonKADS Basics]**  
CommonKADS – Basics.  
Disponible en: <http://commonkads.org/basics/>  
Consulta: 04 de Agosto de 2015.
13. **[CommonKADS Principles]**  
CommonKADS – Introduction – Principles.  
Disponible en: <http://commonkads.org/introduction/>  
Consulta: 04 de Agosto de 2015.