

WEB SEMÁNTICA (PARTE I): VISTA GENERAL DE RDF

Por M.A. Ing. Sergio José Rodríguez Méndez, srodriguez142857@gmail.com

RESUMEN

Este artículo tiene como objetivo presentar una introducción general al Marco de trabajo de descripción de recurso (*Resource Description Framework RDF*) para asistir a diseñadores de sistemas y desarrolladores de aplicaciones a entender los conceptos básicos de *RDF* y cómo usarlos. Se orienta específicamente a presentar una introducción conceptual a la *Semantic Web (SW)* y el marco de trabajo *RDF* de conceptos y modelo de datos (grafos).

DESCRIPTORES

Marco de trabajo de Descripción de recursos. Desarrollo de aplicaciones. Web Semántica. Grafos.

ABSTRACT

Resource Description Framework RDF is an informatics tool which assists system designers and applications developers. In this article, author presents basic concepts for uses of RDF and how to uses them, oriented specifically to conceptual introduction of Semantic Web and concepts and models related with graph.

KEYWORDS

Resources Description Framework. Informatics applications development. Semantic Web. Graph.

1. WEB SEMÁNTICA (PARTE I): VISTA GENERAL DE RDF

La *Web Semántica* (*Semantic Web*, *SW*) proporciona un marco común de trabajo que permite compartir y reutilizar datos a través de diferentes aplicaciones, organizaciones y comunidades. Es un esfuerzo colaborativo liderado por el Consorcio de la *World Wide Web* (*World Wide Web Consortium*, *W3C*) con la participación de un gran número de investigadores y socios industriales. Está basada en el estándar “Marco de trabajo de descripción de recurso” (*Resource Description Framework*, *RDF*).

La *SW* es una “red” (*Web*) de datos que relaciona de una forma significativa datos diversos que no necesariamente son parte de la *Web*. La red de datos que se persigue con la *SW* es independiente a los límites o dominios de las aplicaciones, las cuales controlan la información de una forma aislada: cada aplicación se “guarda” para sí misma los datos que maneja.

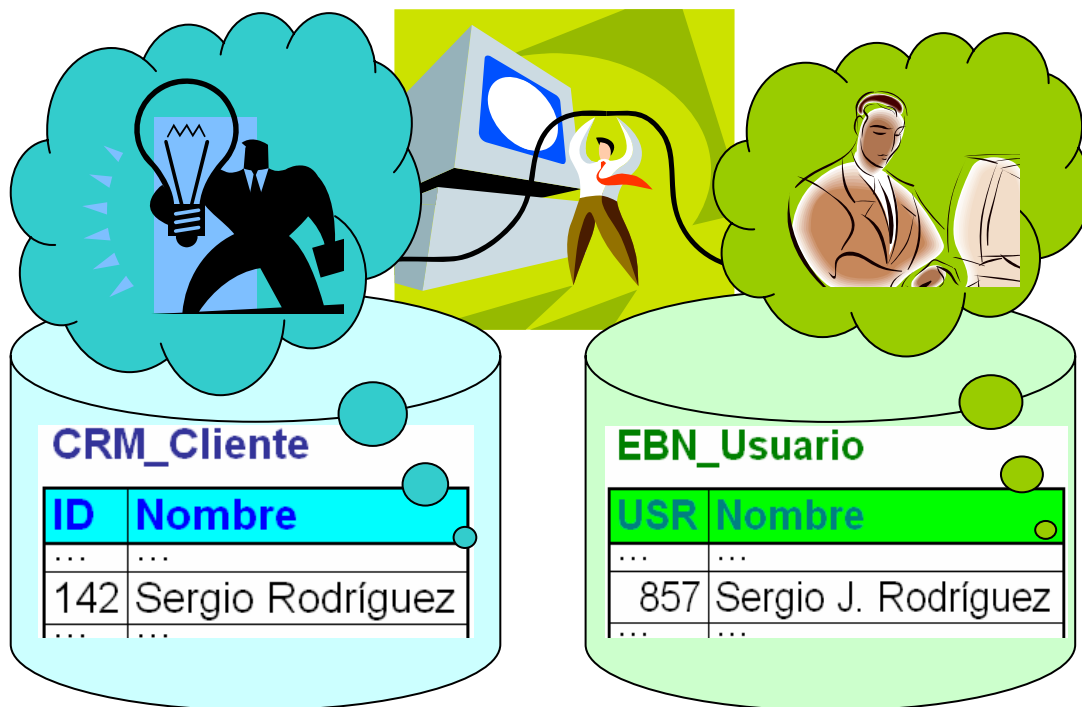


Fig. 1. Representación de la Web Semántica SW

La *SW* es acerca de dos aspectos:

1. Formatos comunes para integración y combinación de datos obtenidos de diversas fuentes, a diferencia de la *Web* original (*WWW*) la cual se enfoca principalmente en el intercambio de documentos.
2. Lenguaje para grabar cómo los datos se relacionan con objetos del mundo real. Esto permite a una persona o máquina iniciar desde una base de datos y luego moverse a través de un sin fin de bases de datos las cuales están conectadas, no por “alambres” sino, por ser o referirse acerca del mismo tópico/tema.

1.1. Especificaciones de la Web Semántica.

En Febrero de 2004, la W3C liberó *RDF* y *OWL* (Lenguaje de Ontologías de la Web *OWL*) como recomendaciones:

- *RDF* es utilizado para representar información y para intercambiar conocimiento en la *Web*.
- *OWL* es utilizado para publicar y compartir conjuntos de términos llamados ontologías, soportando búsqueda *Web* avanzada, agentes de *software* y administración del conocimiento (*Knowledge Management, KM*).

En este artículo se brinda una vista general de *RDF* basado en la parte introductoria de la Recomendación del W3C *RDF Primer*.

1.2. Introducción a RDF [RDF Primer]

El “Marco de trabajo de descripción de recurso” (*Resource Description Framework, RDF*) es un lenguaje para representar información sobre recursos en la “*World Wide Web*” (*Web*). Su enfoque específico es el representar metadatos acerca de recursos *Web*, tales como el título, autor, fecha de modificación, derechos reservados e información de licenciamiento acerca de un documento *Web* (*HTML*) o la disponibilidad calendarizada para un recurso compartido. Sin embargo, a través de la generalización del concepto de un “recurso *Web*”, *RDF* también puede ser utilizado para representar información acerca de artefactos que pueden ser IDENTIFICADOS en la *Web*, aún y cuando no puedan RECUPERARSE (u obtener). Ejemplos de éstos incluyen información acerca de ítems disponibles en sistemas de compra en línea (como información acerca de especificaciones, precios y disponibilidad), o la descripción de las preferencias de un usuario *Web* para entrega de información.

RDF está enfocado en utilizarse en situaciones en donde la información descriptiva (metadatos) de un recurso (concepto generalizado) necesita ser procesada por las aplicaciones, en vez de únicamente ser desplegada a las personas. *RDF* proporciona un marco de trabajo común para expresar esta información de tal forma que pueda ser intercambiada entre diferentes aplicaciones sin perder significado. Puesto que es un marco de trabajo común, los diseñadores de aplicaciones pueden tomar ventaja de la capacidad de analizadores gramaticales *RDF* comunes y herramientas de procesamiento. La habilidad para intercambiar información entre aplicaciones diferentes significa que la información puede ser dispuesta a otras aplicaciones totalmente ajenas para la cual fue creada originalmente.

RDF está basado en la idea de identificar recursos utilizando identificadores *Web* (llamados Identificadores Uniformes de Recursos, *Uniform Resource Identifiers - URIs*), y de describir a los mismos en términos de propiedades sencillas y valores de propiedades. Esto permite representar declaraciones sencillas acerca de recursos como un grafo de nodos y arcos que representan los recursos, sus propiedades y valores.

Así, por ejemplo, el grupo de declaraciones “existe una Persona identificada por <http://www.w3.org/People/EM/contact#me>, cuyo nombre es Eric Miller, cuya dirección de correo electrónico es em@w3.org, y cuyo título académico es Dr.”, puede representarse a través de un grafo *RDF* como se presenta en la Figura #2.

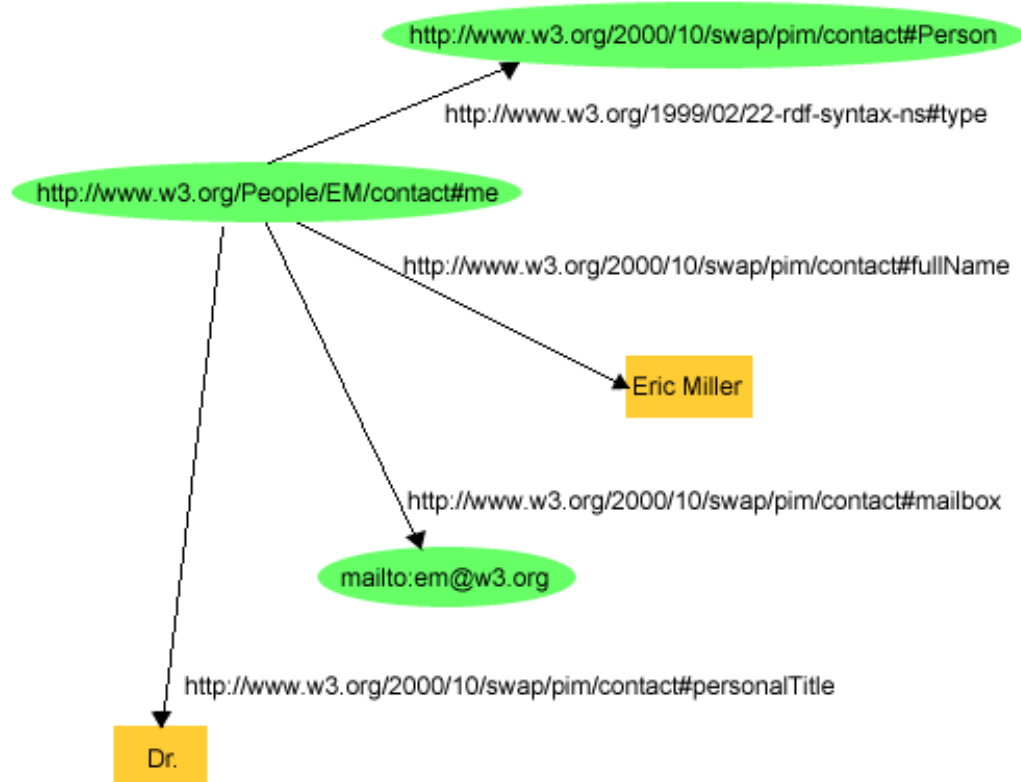


Fig. 2. Grafo RDF para representar un objeto

La figura #2 ilustra que *RDF* emplea *URIs* para identificar:

- Individuos, por ejemplo, Eric Miller, identificado por <http://www.w3.org/People/EM/contact#me>.
- Tipos de objetos (clases¹), por ejemplo, Persona, identificado por <http://www.w3.org/2000/10/swap/pim/contact#Person>.
- Propiedades de los objetos, por ejemplo, correo electrónico, identificado por <http://www.w3.org/2000/10/swap/pim/contact#mailbox>.
- Valores de dichas propiedades, por ejemplo, <mailto:em@w3.org> como el valor de la propiedad correo electrónico (*RDF* también utiliza cadenas de caracteres, tales

¹ Corresponde al mismo concepto de Programación Orientada a Objetos: una colección o grupo de objetos con características (atributos) y comportamiento (operaciones) similares → concepto, sustantivo en singular.

como “Eric Miller”, y valores de otros tipos de datos como enteros y fechas, como valores de propiedades).

RDF también proporciona una sintaxis basada en *XML* (llamada *RDF/XML*) para registrar/definir e intercambiar estos grafos. A continuación, se presenta un pequeño ejemplo de *RDF* escrito en *RDF/XML* correspondiente al grafo de la figura #2.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
]>
<rdf:RDF xmlns:rdf = "&rdf;"
          xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

Fig. 3. Sintaxis XML de un Marco de trabajo de descripción de recurso

Observar que este documento *RDF/XML* también contiene *URIs*, así como propiedades como `mailbox` y `fullName` (en forma abreviada) y sus respectivos valores `em@w3.org` y `Eric Miller`.

De la misma forma que *HTML*, este *RDF/XML* es procesable por computadoras y, empleando *URIs*, puede enlazar a piezas de información a través de la *Web*. Sin embargo, a diferencia del hipertexto convencional, los *URIs* de *RDF* pueden referenciar a cualquier objeto u artefacto identificable (recurso), incluyendo aquellos que no son directamente obtenidos en la *Web* (como la persona Eric Miller). El resultado es que en adición a describir recursos tangibles electrónicamente (documentos *HTML*, imágenes, archivos de multimedia, etc.), *RDF* también puede describir a entidades concretas y abstractas del mundo real como carros, negocios, personas, eventos de noticias, planetas, conceptos, etc. Además, las propiedades *RDF* tienen intrínsecamente *URIs*, para identificar precisamente las relaciones que existen entre los ítems o recursos enlazados.

Los siguientes documentos contribuyen a definir la especificación de *RDF*:

- Conceptos y sintaxis abstracta de *RDF* <<http://www.w3.org/TR/rdf-concepts/>>.
- Especificación de la sintaxis *RDF/XML* <<http://www.w3.org/TR/rdf-syntax-grammar/>>.
- Lenguaje de la descripción del vocabulario *RDF* 1.0: Esquema *RDF* <<http://www.w3.org/TR/rdf-schema/>>.
- Semánticas *RDF* <<http://www.w3.org/TR/rdf-mt/>>.
- Casos de prueba *RDF* <<http://www.w3.org/TR/rdf-testcases/>>.

1.3. Metas de diseño de RDF [RDF Concepts and Abstract Syntax]

Las metas de diseño de RDF son:

- Modelo simple de datos [2].
 - Fácil procesamiento y manipulación para las aplicaciones.
 - El modelo de datos es independiente de cualquier sintaxis específica de serialización.
- Semánticas formales e inferencia demostrable.
 - Proporcionan una base dependiente para el razonamiento acerca del significado de una expresión RDF.
 - Particularmente, apoya nociones rigurosamente definidas que proporcionan una base para definir reglas confiables de inferencia en datos RDF.
- Uso de un vocabulario extensible basado en URIs.
 - Totalmente extensible.
 - URIs con identificadores de fragmento opcionales.
 - Las referencias URIs son utilizadas para nombrar todo tipo de recursos y conceptos en RDF.
 - El otro tipo de valor que aparece en datos RDF es una literal.
- Uso de una sintaxis basada en XML.
 - La forma de serialización recomendada de RDF está basada en un vocabulario XML, el cual puede ser utilizado para codificar el modelo de datos para el intercambio de información entre aplicaciones.
- Soportar el uso de los tipos de datos definidos en la especificación de Esquemas XML.
 - RDF puede usar valores representados de acuerdo a los tipos de datos definidos en Esquemas XML, asistiendo así, el intercambio de información entre RDF/XML y otras aplicaciones XML.
- Permitir que cualquier persona pueda hacer declaraciones acerca de cualquier recurso.
 - Para facilitar las operaciones a escala de Internet, RDF es un marco de trabajo abierto y mundial que permite a cualquier persona el hacer declaraciones acerca de cualquier recurso.
 - En general, no se asume que información completa acerca de cualquier recurso está disponible.
 - RDF no previene el hacer aserciones “tontas” o inconsistentes con otras declaraciones.
 - Los diseñadores de aplicaciones que emplean RDF deben construir arquitecturas basadas bajo esta premisa → **se deben diseñar las aplicaciones para tolerar fuentes incompletas o inconsistentes de información.**

[2] La palabra “modelo” tiene un significado diferente a su uso dentro de la teoría de modelos (literatura matemática y lógica).

1.4. Conceptos de RDF [RDF Concepts and Abstract Syntax].

RDF usa los siguientes conceptos clave:

1.4.1 Modelo grafo de datos

La estructura subyacente de cualquier expresión RDF es una colección de tripletas, cada una conformada de un **sujeto**, un **predicado** y un **objeto**. Un conjunto de dichas tripletas es llamado un **grafo RDF**, el cual define la **sintaxis abstracta de RDF** [3]. Esto puede ilustrarse a través de un diagrama de nodos y arcos dirigidos, en la cual cada tripleta es representada como un enlace nodo-arco-nodo (de allí el término grafo [4]).

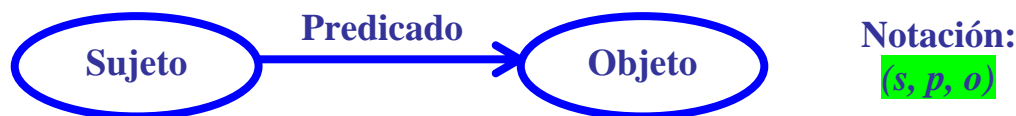


Fig. 4. Diagrama de nodos y arcos

Cada tripleta representa una declaración de una relación entre los artefactos denotados por los nodos que enlaza. Sus tres componentes son:

- Un sujeto que es una referencia URI de RDF o un nodo blanco,
- Un predicado que es una referencia URI de RDF (también llamado una propiedad) el cual denota una relación.
- Un objeto que es una referencia URI de RDF, una literal o un nodo blanco,

Una tripleta RDF se escribe convencionalmente en el orden: sujeto (s), predicado (p), objeto (o). La dirección del arco es significativo: siempre apunta hacia el objeto. Los nodos de un grafo RDF son sus sujetos y objetos.

La evaluación de una tripleta RDF menciona que alguna relación, indicada por el predicado, se mantiene/existe entre las entidades de información denotados por el sujeto y objeto de la tripleta. La evaluación de un grafo RDF acumula todas las evaluaciones de las tripletas que incluye, por lo que el significado de un grafo RDF es la conjunción (Y-lógico) de las declaraciones correspondientes a todas las tripletas que contiene.

[3] La sintaxis abstracta es aquella sobre la cual se definen las semánticas formales. Implementaciones del modelo tienen libertad de representar grafos *RDF* en cualquier otra forma equivalente.

[4] El conjunto de **nodos** de un grafo *RDF* es el conjunto de sujetos y objetos de tripletas en el grafo.

1.4.2. Vocabulario basado en URIs.

Un nodo puede ser un *URI* (con identificador opcional de fragmento), una literal o un valor en blanco. Las propiedades son referencias *URI*. *RDF* no utiliza *URIs* relativas, solamente maneja direcciones absolutas.

Una referencia *URI* o literal utilizada como un nodo, identifica lo que dicho nodo representa. Una referencia *URI* utilizada como un predicado, identifica una relación entre entidades representadas por los nodos que conecta.

Una referencia *URI* como predicado puede ser también un nodo en el grafo.

Un nodo con valor en blanco es un nodo que no es una referencia *URI* o literal. En la sintaxis abstracta *RDF*, un nodo en blanco es un nodo único que puede ser usado en una o más declaraciones *RDF*, pero no tiene ningún nombre intrínseco.

Una convención empleada por algunas representaciones lineales de un grafo *RDF* para permitir que varias declaraciones hagan referencia al mismo recurso no identificado es el usar un identificador de nodo en blanco, el cual es un identificador local que puede distinguirse entre todos los *URIs* y literales. Cuando los grafos son unidos o fusionados, los nodos en blanco deben mantenerse distintos si el significado es necesario preservarlo; esto puede implicar una realocución de los identificadores de los nodos en blanco. Nótese que los identificadores de los nodos en blanco no forman parte de la sintaxis abstracta *RDF*, y la representación de tripletas conteniendo nodos en blanco depende completamente de la sintaxis particular usada.

1.4.3. Tipos de datos (Esquemas XML).

Los tipos de datos son utilizados por *RDF* en la representación de valores tales como enteros, números de punto flotante y fechas. La abstracción de tipo de dato en *RDF* es compatible con la abstracción de tipo de dato en Esquema *XML*. Un tipo de dato consiste de los siguientes elementos:

- Un espacio de valores → representa el dominio semántico del tipo de dato (valores).
- Un espacio léxico → es un conjunto de cadenas de caracteres UNICODE.
- Un mapeo léxico a valor → es un conjunto de pares cuyo primer elemento pertenece al espacio léxico del tipo de dato y el segundo elemento pertenece al espacio de valores.
 - Cada miembro del espacio léxico está vinculado/mapeado con exactamente un miembro del espacio de valores.

- Cada miembro del espacio de valores puede estar vinculado con cualquier número (incluyendo cero) de miembros del espacio léxico (representaciones léxicas para dicho valor) [5].

Por ejemplo, el mapeo léxico a valor para el tipo de dato de Esquema *XML* `xsd:boolean`, donde cada miembro del espacio de valores (representado aquí como "T" y "F") tiene dos representaciones léxicas, se presenta a continuación:

Espacio de valores	{ T, F }
Espacio léxico	{ "0", "1", "true", "false" }
Mapeo léxico a valor	{ <"true", T>, <"1", T>, <"0", F>, <"false", F> }

RDF predefine solamente un tipo de dato `rdf:XMLLiteral`, usado para definir *XML* embebido en *RDF*.

En *RDF* no existe el concepto de números, fechas u otros valores comunes predefinidos. En vez de ello, *RDF* maneja una definición independiente de tipos de datos e identificados con referencias *URIs*. Los tipos de datos predefinidos en Esquema *XML* se esperan que sean ampliamente utilizados para este propósito⁶.

RDF no proporciona un mecanismo para definir nuevos tipos de datos. La especificación de "Esquema *XML* Parte 2: Tipos de Datos" proporciona un marco de trabajo extensible adecuado para definir nuevos tipos de datos para usar en *RDF*.

1.4.4. Literales

Las literales son utilizadas para identificar valores tales como números y fechas a través de una representación léxica. Cualquier entidad representada por una literal puede también ser representada por un *URI*, pero frecuentemente es más conveniente o intuitivo el uso de literales. Una literal puede ser el objeto de una declaración *RDF*, pero no el sujeto o el predicado.

Las literales pueden ser planas o tipificadas:

[5] Matemáticamente, se puede pensar en la siguiente definición: $f(x) = y = [función\ de\ mapeo]$, en donde $x = \{espacio\ léxico\}$, e , $y = \{espacio\ de\ valores\}$.

[6] Sin embargo, algunos tipos de datos predefinidos en Esquema *XML* no son adecuados para utilizarse dentro del contexto de *RDF*. Por ejemplo, el tipo de dato *QName* (nombre calificado) requiere la declaración de un espacio de nombres *XML* para su ámbito durante el mapeo, por lo que no es recomendable emplearlo en *RDF*.

- Una literal plana es una cadena de caracteres combinada con una marca opcional de lenguaje. Ésta puede utilizarse para texto plano en un lenguaje natural. Así como se recomienda en la especificación de semánticas formales de *RDF*, las literales planas son auto-denotadas.
- Una literal tipificada es una cadena de caracteres combinada con un tipo de dato *URI*. Denota al miembro del espacio de valores del tipo de dato identificado obtenido a través de aplicar el mapeo léxico a valor a la cadena de caracteres de la literal.

Continuando con el ejemplo anterior, las literales tipificadas que pueden definirse usando el tipo de dato `xsd:boolean` de Esquema *XML* son:

Literal tipificada	Mapeo léxico a valor	Valor
<code><xsd:boolean, "true"></code>	<code><"true", T></code>	T
<code><xsd:boolean, "1"></code>	<code><"1", T></code>	T
<code><xsd:boolean, "false"></code>	<code><"false", F></code>	F
<code><xsd:boolean, "0"></code>	<code><"0", F></code>	F

Para contenido textual que puede contener marcación, se deben emplear literales tipificadas con el tipo `rdf:XMLLiteral`. Si se requiere alguna anotación de lenguaje, debe incluirse explícitamente como marcación, usualmente a través de un atributo `xml:lang`, definido dentro de un elemento `` o `<div>` de *XHTML*, el cual se incluye dentro de *RDF* para este propósito.

1.4.5. Sintaxis de serialización XML.

La representación computacional del modelo de *RDF* se realiza a través de un vocabulario *XML*, para obtener los beneficios que ofrece *XML* como tecnología para estructurar información en la *Web*.

1.4.6. Expresión de hechos simples o básicos.

Algunos hechos simples (*facts*) indican una relación entre dos entidades (recursos – tangibles o intangibles). Un hecho puede ser representado como una tripleta *RDF* en el cual los predicados nombran la relación, y el sujeto y el objeto denota las dos entidades o recursos.

Una representación familiar de un hecho puede observarse en una tabla de una base de datos relacional: la tabla tiene dos columnas, correspondientes al sujeto y el objeto de la tripleta *RDF*; el nombre de la tabla corresponde al predicado de la tripleta *RDF*. Otra representación familiar puede ser como un predicado de dos lugares (dos columnas) en la lógica de primer orden.

Las bases de datos relacionales pueden permitir a una tabla el tener un número arbitrario de columnas, por tanto, existen filas que expresan información correspondiente a un predicado en lógica de primer orden con un número arbitrario de lugares. De esta forma, una fila, o predicado, tiene que ser descompuesto para poder ser representado como tripletas *RDF*. Una forma simple de descomposición introduce un nuevo nodo en blanco, correspondiente a la fila, y una nueva tripleta es introducida para cada celda en la fila. El sujeto de cada tripleta es el nuevo nodo en blanco, el predicado corresponde al nombre de la columna, y el objeto corresponde al valor en la celda. El nuevo nodo en blanco puede tener también una propiedad *rdf:type* cuyo valor corresponde al nombre de la tabla.

Como ejemplo, considere la siguiente figura:

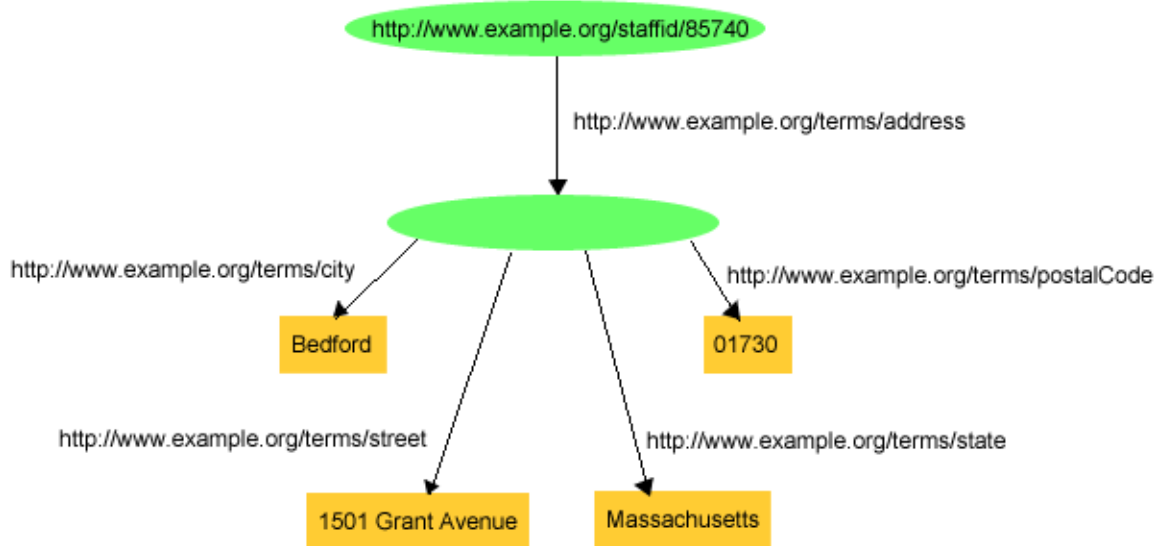


Fig. 5. Representación de tripletas *RDF*

Así, esta información puede corresponder a una fila en la tabla “*STAFF_ADDRESSES*”, con una llave primaria “*STAFFID*”, y columnas adicionales “*STREET*”, “*POSTALCODE*”, “*CITY*” y “*STATE*”.

De esta manera, un hecho de mayor complejidad es expresado en *RDF* usando una conjunción (Y-lógico) de relaciones binarias simples. *RDF* no proporciona mecanismos para expresar negación (NO-lógico) o disyunción (O-lógico).

Con su uso de vocabularios extensibles basados en *URIs*, *RDF* proporciona el medio para la expresión de hechos de temas arbitrarios; es decir, aserciones o evaluaciones de características/propiedades nombradas acerca de entidades/recursos nombrados específicos. Un *URI* puede ser construido para cualquier entidad o recurso (tangibles o intangibles) que pueda ser nombrado, por lo que los hechos *RDF* pueden referirse, consecuentemente, a cualquier cosa.

1.4.7. “Entailment” (exigencia)

Las ideas sobre el significado e inferencia en *RDF* se fundamentan a través del concepto formal de “*entailment*” (exigencia), tal y como se presenta en la especificación de semánticas formales de *RDF*. Brevemente, este concepto indica lo siguiente: una expresión “*A*” *RDF* se dice que “exige” a otra expresión “*B*” *RDF* si cada posible arreglo de entidades en el mundo que hacen a “*A*” verdadero también hace verdadero a “*B*”. Sobre esta base, si la verdad de “*A*” es presunta o demostrada entonces la verdad de “*B*” puede ser inferida.

2. REFERENCIAS BIBLIOGRÁFICAS

W3C. Semantic Web Activity. Consultar en:
<http://www.w3.org/2001/sw/>

W3C. RDF Primer. Consultar en:
<http://www.w3.org/TR/rdf-primer/>

W3C. Resource Description Framework (RDF):
Concepts and Abstract Syntax. Consultar en:
<http://www.w3.org/TR/rdf-concepts/>

RODRIGUEZ MÉNDEZ, SERGIO JOSÉ



Master en Finanzas e Ingeniero en Informática y Sistemas, graduado Magna Cum Laude de la Universidad Rafael Landívar. Experiencia laboral de 8 años en el ciclo de vida de desarrollo de software de arquitecturas distribuidas y orientadas a servicios basadas en tecnologías Web. Ha liderado la construcción de sistemas electrónicos, principalmente en el ámbito bancario. Catedrático de las Universidades Rafael Landívar y San Carlos de Guatemala. Ha impartido cursos y seminarios académicos de: Introducción a la Programación, Tecnologías orientadas a objetos, Programación en Internet, Tecnologías .NET y Familia de estándares XML, entre otros. Actualmente, es consultor informático del Proyecto SIAF-SAG