



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Intercambio de información entre instituciones
de la UNAM usando la Red Semántica

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
LICENCIADO EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:
HUGO ALBERTO RODRÍGUEZ PEÑA

DIRECTOR DE TESIS:
DR. SERGIO RAJSBAUM GORODEZKY



2010

Hoja de Datos del Jurado

1. Datos del alumno Rodríguez Peña Hugo Alberto 56 10 73 35 Universidad Nacional Autónoma de México Facultad de Ciencias Ciencias de la Computación 302128970
2. Datos del tutor Dr. Sergio Rajsbaum Gorodezky
3. Datos del Sinodal 1 Mat. Mónica Leñero Padierna
4. Datos del Sinodal 2 L. en C.C. Francisco Lorenzo Solsona Cruz
5. Datos del Sinodal 3 Dra. Elisa Viso Gurovich
6. Datos del Sinodal 4 Dr. Víctor Hugo Jacobo Armendáriz
7. Datos del trabajo escrito Intercambio de información entre instituciones de la UNAM usando la Red Semántica 32 páginas 2010

Índice general

1. Antecedentes	5
1.1. Introducción	5
1.1.1. ¿Qué se ha hecho?	5
1.1.2. El problema	6
1.1.3. La solución	6
1.2. Red Semántica	6
1.2.1. Definición	7
1.2.2. Detalles	8
1.2.3. Casos de uso	11
1.2.4. Importancia de la Red Semántica	12
2. Desarrollo	14
2.1. Sistemas a comunicar	14
2.2. Casos de uso	14
2.2.1. IMATE - DGAPA	14
2.2.2. IMATE - Facultad de Ciencias	15
2.3. Vocabularios	15
2.3.1. Persona UNAM	15
2.3.2. Curso	16
2.3.3. Currículum	16
2.3.4. Ejemplares	16
2.4. Implementación	17
2.4.1. Demostraciones en Java	17
2.4.1.1. Llenado de formas	17
2.4.1.2. Generador de RDF	17
2.4.1.3. Generador de generadores RDF	17
2.4.2. Demostraciones en Plone	18
2.4.2.1. Obtener datos usando RDF	18
2.4.2.2. Generador de RDF	18
2.4.2.3. <i>Data type</i> Persona	19
2.5. Manuales de uso	19
2.5.1. Demostraciones en Java	19
2.5.1.1. Instalación	19
2.5.1.2. Llenado de formas	19

2.5.1.3.	Generador de generadores	20
2.5.2.	Demostraciones en Plone	23
2.5.2.1.	Instalación	24
2.5.2.2.	Obtener datos usando RDF	24
2.5.2.3.	Generador de RDF	26
2.5.2.4.	<i>Data type</i> Persona	27
3.	Conclusiones	30

Capítulo 1

Antecedentes

1.1. Introducción

Este proyecto surgió por la necesidad de tener una forma más eficiente de intercambiar información entre las instituciones de la Universidad Nacional Autónoma de México. Se propone solucionar esto haciendo uso de la Red Semántica, actualmente se tienen implementadas soluciones que si bien no son malas son poco eficientes y poco escalables.

1.1.1. ¿Qué se ha hecho?

Se ha intentado en varias ocasiones normalizar la información que se maneja de las personas (estudiantes, docentes, etc.) de las distintas instituciones de la UNAM, pero ha resultado improductivo pues los datos que algunas instituciones consideran relevantes no lo son tanto para otras, además de que tienen sus sistemas de información organizados de formas muy diferentes, por lo cual no se ha podido llegar a un consenso sobre los datos que se deben manejar, si de igual forma se quisiera llegar a un consenso en los datos a compartir llegaríamos a un problema similar.

La DGAPA¹ realizó un proyecto para centralizar los datos del personal de la UNAM y compartirlos, llamado RUPA². El mecanismo para difundir la información consiste en un proceso que requiere de una computadora que cumpla con ciertas características, la computadora se mantiene en espera y al cabo de cierto tiempo devuelve la información de la RUPA. El escenario de actualización de los datos es un proceso similar. La RUPA es una buena solución pero es poco eficiente para compartir información.

¹Dirección General de Asuntos del Personal Académico

²Registro Único de Personal Académico

1.1.2. El problema

No se ha podido llegar a un consenso respecto a los datos que se requieren para cada institución, lo que crea dificultades para las personas que laboran en más de una institución. Tienen que llenar formas con los mismos datos más de una vez en distintas ocasiones, lo cual crea redundancia e inconsistencias en la información de la misma persona. Además el proceso de capturar los mismos datos en más de una ocasión puede llegar a ser molesto para los usuarios de los sistemas.

El medio por el cual se comparte la información es otra parte del problema. Las soluciones que se han presentado, son poco prácticas o muchas veces se requiere de hacer grandes modificaciones a los sistemas para lograr la interacción entre dos de ellos. Se incrementan las dificultades cuando la interacción es entre múltiples sistemas, más si cada uno de ellos cuenta con sus propios protocolos para compartir datos.

1.1.3. La solución

En este seminario se propone una solución a este problema haciendo uso de la Red Semántica como intermediario entre los sistemas.

Es una solución diseñada para compartir datos entre sistemas que es fácilmente expansible. Utilizando la Red Semántica se pueden crear soluciones para comunicar a dos instituciones. Si una o más quieren comunicarse no es necesario cambiar el método, simplemente se extiende lo que ya se tiene hecho de acuerdo a sus necesidades, dando a conocer estos cambios para que los demás estén enterados de las modificaciones y puedan aprovechar todos los datos que la institución está publicando. Con esto se elimina el problema de tener que llegar a un consenso entre todas las instituciones, logrando un intercambio de información efectivo.

1.2. Red Semántica

Hoy en día la WWW³ provee una gran cantidad de información disponible y utilizable para usuarios finales. Sea que deseemos buscar información de algún tema en específico en las páginas de internet, por ejemplo la Wikipedia, o precios de algunos artículos en las páginas de los proveedores.

Toda esta información tiene un formato amigable al ojo humano para que podamos encontrar rápidamente la información deseada. Cuando queremos encontrar información en la red usamos un motor de búsqueda, como Google. Después de unos cuantos *clicks* encontramos lo que nos interesa.

Cuando se trata de automatizar una búsqueda haciendo uso de una computadora las cosas se complican, pues la computadora no entiende el significado de las cadenas de texto que encuentra. Lo mejor que una computadora puede hacer para encontrar información es buscar en las cadenas de texto lo que se

³World Wide Web

desea, pero sin saber realmente si lo que devuelve se refiere al concepto que se buscaba originalmente. Este proceso es poco eficiente y no garantiza que el resultado encontrado sea el deseado.

1.2.1. Definición

La Red Semántica es una extensión de la Red actual dotándola de significado. Esto es un espacio donde la información tiene un significado bien definido, de forma que ésta pueda ser interpretada tanto por humanos como por computadoras. Como bien dice el precursor de la idea, Tim Berners Lee⁴:

“The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”.

La Red Semántica surge debido a la falta de significado entendible para una computadora de la inmensa información que está disponible en la Red. Con esta nueva Red podemos usar computadoras para automatizar búsquedas precisas de información, similar a lo que se puede hacer en una base de datos, convirtiendo de esta forma la WWW en una red de datos increíblemente grande capaz de ser navegada “inteligentemente” por una computadora.

Para añadir el significado de la Red Semántica se usa RDF⁵ que es un método para especificar metadatos sobre recursos, RDFS⁶ la cual es una especificación para describir vocabularios haciendo uso de RDF, OWL⁷ es un lenguaje hecho para crear ontologías. OWL fue desarrollado como una extensión al RDF. Usando estos lenguajes, se agrega el contenido semántico a la Red usando su sintaxis en XML.

RDF Schema y OWL (vocabularios y ontologías)

En el contexto de RDF un vocabulario es la descripción de conceptos y sus relaciones, entendible por computadoras y por humanos. Una ontología⁸ hace uso de vocabularios para definir objetos.

RDF Schema es la especificación de un vocabulario para describir vocabularios y ontologías de RDF, provee mecanismos para describir conjuntos de recursos (clases) y sus relaciones, también puede definir los recursos como ejemplares de una o más clases, además permite que se puedan organizar las clases en forma jerárquica.

Un vocabulario definido por un RDF Schema está escrito en RDF usando los términos dados en la especificación⁹.

⁴El creador de la World Wide Web, el protocolo utilizado para organizar el contenido que se despliega en la navegación por internet

⁵Resource Description Framework

⁶Resource Description Framework Schema

⁷Web Ontology Language

⁸Representación formal de conceptos dentro de un dominio y sus relaciones.

⁹Para más detalles acerca de RDF Schema refiérase a su página web <http://www.w3.org/TR/rdf-schema/>

OWL es un lenguaje para publicar y compartir ontologías, y al igual que RDF Schema puede usarse para describir el significado de recursos y las relaciones entre ellos. Fue desarrollado como una extensión al RDF aunque es un lenguaje con más poder expresivo que éste. Posee más funcionalidades para expresar el significado y semántica que XML, RDF y RDFS. Tiene tres sub-lenguajes (Lite,DL,Full) que incrementan su expresividad pero complican su computabilidad.

1.2.2. Detalles

Ya sabemos cuáles son los componentes para agregar significado a la Red semántica; ahora veremos como implementarlo.

El XML y XML Schema juegan un papel importante en la implementación de la Red Semántica; nos dan la sintaxis y la estructura necesaria para poder definir los documentos, haciendo uso de las representaciones de RDF Schema y OWL. En XML, se monta la Red Semántica.

RDF

El RDF está basado en la idea de identificar cosas usando URI¹⁰ que son cadenas de caracteres usados para nombrar recursos en la Red. Este concepto es importante pues nos permite identificar de forma única los elementos de la Red Semántica, debido a que un URI no puede estar asociado a más de un recurso.

En la Red Semántica la información es representada como enunciados compuestos de tres partes: sujeto, predicado y objeto; también se les conoce como tripletas.

El sujeto es lo que el enunciado describe.

El predicado describe la relación entre el sujeto y el objeto.

El objeto es a lo que se relaciona.

Por ejemplo:

La página <http://www.example.org/index.html> tiene un creador cuyo nombre es John Smith.

- El sujeto es la página (*<http://www.example.org/index.html>*).
- El predicado es tener un creador (creador).
- Y el objeto es el nombre (John Smith).

Los elementos del enunciado pueden ser URI o literales (cadenas, enteros, etc.).

En el ejemplo anterior sabemos que el creador es la persona que, valga la redundancia, crea algo, en este caso la página. Esto es información que la computadora no sabe al leerlo, lo mismo sucede con el nombre John Smith.

Para que las computadoras puedan entender la información como lo hacen las personas, usamos una tripleta de URI.

¹⁰Uniform Resource Identifier (Identificador Uniforme de Recurso)

<http://www.example.org/index.html>,
<http://purl.org/dc/elements/1.1/creator>,
<http://www.example.org/staffid/85740>

- <http://www.example.org/index.html> es la dirección de nuestra página.
- <http://purl.org/dc/elements/1.1/creator> es la dirección de una definición de la relación creador dentro de un vocabulario.
- <http://www.example.org/staffid/85740> es la dirección que describe a una persona cuyo nombre debemos esperar que sea John Smith.

Se pueden definir múltiples enunciados sobre un mismo sujeto:

1. *http://www.example.org/index.html* tiene un creador cuyo valor es *John Smith*
2. *http://www.example.org/index.html* tiene una fecha de creación *creation-date* cuyo valor es *August 16, 1999*
3. *http://www.example.org/index.html* tiene un lenguaje cuyo valor es *English*

Sus respectivas tripletas serían:

1. <http://www.example.org/index.html> ,
<http://purl.org/dc/elements/1.1/creator> ,
<http://www.example.org/staffid/85740>
2. <http://www.example.org/index.html> ,
<http://www.example.org/terms/creation-date> , "August 16, 1999"
3. <http://www.example.org/index.html> ,
<http://purl.org/dc/elements/1.1/language> , "en"

En estos casos se usaron valores literales para la fecha de creación y el lenguaje de la página, el tipo de valor que debemos de poner para definir el objeto es algo que está definido en la ontología de la cual estamos creando un ejemplar.

Escribir completa la dirección de los vocabularios hace difícil de leer el documento, por lo que podemos hacer uso de espacios de nombres.

prefijo ex:, espacio de nombre URI: http://www.example.org/
prefijo dc:, espacio de nombre URI:
http://purl.org/dc/elements/1.1/
prefijo exterms:, espacio de nombre
URI: http://www.example.org/terms/
prefijo exstaff:, espacio de nombre URI:
http://www.example.org/staffid/

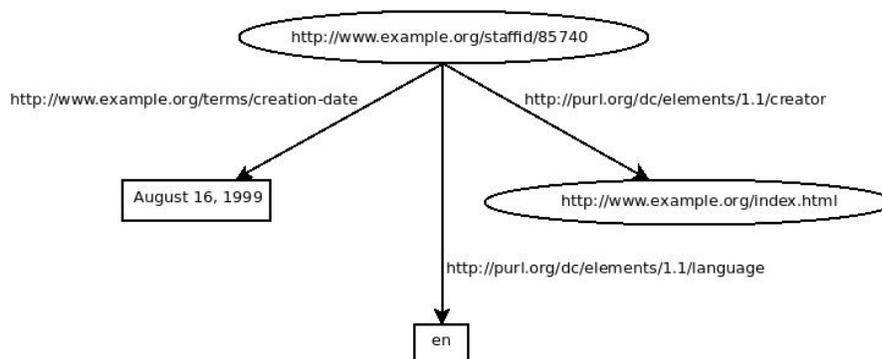
Así podemos reescribir nuestros enunciados de forma más corta y fácil de leer:

ex:index.html, dc:creator, exstaff:85740

ex:index.html, exterm:creation-date, "August 16, 1999"

ex:index.html, dc:language, "en"

Este es un ejemplo sencillo de las triplas que representan la información en RDF. Cabe mencionar que pueden ser vistas como gráficas, la gráfica de este ejemplo es:



Y su representación usando XML es:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:exterm="http://www.example.org/terms">
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <exterm:creation-date>August 16, 1999</exterm:creation-date>
    <dc:language>en</dc:language>
    <dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
  </rdf:Description>
</rdf:RDF>
```

Aprovechando la Red Semántica

Ya que se tienen vocabularios y estructuras para representar nuestra información, la pregunta sería: ¿Cómo aprovechamos estos esfuerzos? Para eso se tiene SPARQL¹¹ que es un protocolo y lenguaje de consultas para RDF. Tiene una sintaxis muy parecida a SQL. Existen razonadores lógicos para realizar inferencias aprovechando la estructura dada por las ontologías y encontrar relaciones no explícitas entre distintas clases. También hay visualizadores que, aprovechando la representación gráfica de los RDF permiten ver la Red Semántica de una forma más amigable al ojo humano.

¹¹Protocol And RDF Query Language

1.2.3. Casos de uso

La Red Semántica no es sólo una herramienta teórica, es algo que ya tiene cerca de una década en desarrollo. Hay una gran variedad de sitios que ya hacen uso de la Red Semántica:

Twine

Sitio que, haciendo uso de las relaciones definidas entre contenidos de la página (usuarios, noticias, favoritos, etc.), infiere posibles contenidos que le pueden interesar a una persona y se los muestra, tomando en cuenta sus últimos contenidos visitados, los de sus amigos, etc.

WEASEL

Es un sistema que implementó Vodafone¹² para compartir los datos entre sus sistemas, los cuales contemplan varios sitios web, bases de datos e información corporativa de fuentes heterogéneas no estructuradas. La información necesita ser accedida de forma uniforme, por lo que se desarrollaron ontologías para dotar de estructuras a toda su red corporativa. De esta forma se pueden realizar búsquedas más sofisticadas que simplemente consultar los contenidos de las bases de datos.

FOAF

Friend Of A Friend¹³ es una ontología para publicar datos personales de una persona así como su relación con otras personas.

SWOOGLE

Es como la tecnología de Google para la Red Semántica, tiene un motor de recuperación de datos que procesa, analiza y cataloga documentos RDF y OWL.

SCHEMAWEB

Esto es un sitio web que se dedica a almacenar ontologías creadas para su uso con RDF.

SIMILE

Es una iniciativa del MIT¹⁴ que contiene varios proyectos que hacen uso de la Red Semántica.

¹²Una de las empresas líderes en telefonía de Europa

¹³<http://www.foaf-project.org/>

¹⁴Massachusetts Institute of Technology

CALAIS

Es una empresa que se dedica al desarrollo de un conjunto de herramientas para incorporar funciones semánticas a blogs, CMS, páginas web, etc.

1.2.4. Importancia de la Red Semántica

La Red Semántica ha llamado la atención de empresas y gobiernos que en los últimos años han invertido millones de dólares en soluciones incorporándola a sus bases de datos, sitios web, etc. con el propósito de ofrecer un mejor servicio y generar mayores ganancias.

La CBS¹⁵ ha decidido contratar un servicio de CALAIS que agrega funciones de la Red Semántica a su sitio web CBS Interactive / CNET para realizar análisis semánticos de sus noticias sobre productos tecnológicos, premios otorgados, blogs, etc. La finalidad es poder brindar notas de mayor interés, publicidad mucho más rentable y aumentar el mercado que capta.

Por su parte Yahoo y Google conocidos por sus motores de búsqueda, han decidido incorporar las funcionalidades de la Red Semántica, tanto a sus sitios como a sus motores de búsqueda, para de esta manera obtener resultados de mejor calidad, mejor organización y clasificación, igualmente alentará a las personas que publican sitios web a que agreguen contenido semántico.

El gobierno de Estados Unidos tiene la iniciativa de agregar contenido semántico al sitio Data.gov (encargado de la transparencia en el uso de su presupuesto), para que toda la información que contiene este sitio pueda ser consultada más eficazmente. Ha invertido millones de dólares para el desarrollo e investigación de la Red Semántica.

El gobierno del Reino Unido siguió la línea de la idea del estadounidense, y pronto saldrá a la luz la versión beta del sitio Data.gov.uk, que al igual que el estadounidense planea aprovechar los beneficios de la Red Semántica. Por otra parte tienen su proyecto MIB¹⁶ en conjunto con las compañías Parcel Force Worldwide, British Gas Business, AXA, Clydesdale and Yorkshire Bank (NAGE), 3M y pH Group. Este proyecto, haciendo uso de la Red Semántica, planea crear inteligencia de mercado a un bajo costo a partir de las innumerables y crecientes colecciones de datos de las compañías así como colecciones externas para proveer a las compañías del Reino Unido con visión de su mercado de ventas, dinámica y datos relevantes que permitan tomar decisiones informadas sobre los clientes potenciales y las tendencias de compra que éstos tienen.

La revista italiana Quattroruote¹⁷ optó por agregar contenido semántico a sus páginas web y servidores, lo que les permitió clasificar la información de autos vendidos, características y sus precios desde 1980; de esta forma pueden proveer a sus usuarios un punto de vista único, sobre los clasificados de venta

¹⁵Columbia Broadcasting System es una de las cadenas de televisión y radio más grande de Estados Unidos.

¹⁶U.K. Government's Technology Program's Market Blended Insight

¹⁷Revista online italiana líder para aficionados a los carros y compradores de carros.

de autos: al agregarlos a sus bases de datos los clasifican acertadamente, y con esto pueden dar una opinión automática y objetiva del precio que tiene el carro.

Los anteriores son ejemplos de compañías y gobiernos que se han dado cuenta de los beneficios que trae la implementación de la Red Semántica para la manipulación de bancos de datos. En las palabras de David Siegel¹⁸ sobre la Red Semántica:

“Creo que se puede comparar a la web que utilizan hoy en día. Piense en donde todo el mundo estaba en 1992: nadie sabía lo que era un servidor, un sitio web, incluso un nombre de dominio. Las empresas tuvieron que adquirir todos estos nuevos activos, las habilidades y capacidades para tomar ventaja de la red, y hay enormes beneficios. No creo que ninguna empresa quiere volver a los días antes de la red, cuando se tenía que enviar correos masivos e ir a un montón de exposiciones comerciales para encontrar nuevos clientes. Así que habrá nuevos costos y nuevas oportunidades para el crecimiento y beneficio. Creo que sera costoso para las empresas que no participen, ya que van a perder las oportunidades por venir.”

Los usuarios finales de la Red Semántica también nos beneficiaremos ya que tendremos la información que necesitamos más a la mano de lo que la tenemos ahora.

¹⁸Se le atribuye ser uno de los primeros empresarios y diseñadores en el emergente negocio del diseño de sitio web

Capítulo 2

Desarrollo

Comenzamos por obtener y leer documentación acerca de la Red Semántica, entender las herramientas que tenemos disponibles para facilitar la creación de vocabularios y ejemplares, así como identificar los casos de uso a implementar en los sistemas de las instituciones a comunicar.

2.1. Sistemas a comunicar

La página web del IMATE¹ está desarrollada con Plone² que es un CMS³ desarrollado en su totalidad en Python, usando como base de datos para guardar la información Zope, que es una base de datos orientada a objetos. Plone es completamente orientado a objetos y tiene una curva de aprendizaje muy empinada por la forma en que manejan los contenidos. Como Plone está desarrollado con Python, tenemos herramientas confiables para manejar RDF y realizar consultas sobre los documentos. El problema es que la forma en que maneja los contenidos de archivos complica la comunicación entre sistemas que usan archivos XML.

La página web de la Facultad de Ciencias y la de la DGAPA están desarrolladas en Java. Java nos provee de herramientas muy confiables para realizar consultas a documentos RDF y permite un buen manejo de archivos XML para la publicación de los datos a compartir.

2.2. Casos de uso

2.2.1. IMATE - DGAPA

Se encontró un caso de uso muy práctico para comenzar a compartir información entre estas instituciones. Siempre que los profesores del IMATE desean registrarse para algún proyecto en la DGAPA deben llenar formas con la misma

¹Instituto de Matemáticas, UNAM

²<http://plone.org>

³*Content Management System*

información. Se planea agilizar este proceso dando la oportunidad de tomar estos datos de una dirección que contenga un RDF con toda la información acerca del profesor; de esta manera el llenado de formas se hará de forma automática, dando la posibilidad de revisar que los datos obtenidos del RDF sean correctos.

2.2.2. IMATE - Facultad de Ciencias

Cada semestre los profesores del IMATE que desean impartir una materia en la Facultad de Ciencias deben realizar el debido registro en la Facultad. Cuando el profesor desea agregar los datos de este curso a su curriculum en la página del IMATE debe de recapturar todos los datos a mano; este es un proceso innecesario pues la información está en alguna parte del sistema de la Facultad. Lo que se busca es publicar estos datos en un RDF y dar la opción, en la sección de cursos del IMATE, de usar un RDF para poder llenar automáticamente estos datos.

2.3. Vocabularios

Los vocabularios son muy importantes para el desarrollo del proyecto pues se debe llegar a un consenso entre las partes para definirlos de tal forma que cubra las necesidades de los sistemas. Si esto no se realiza correctamente puede que los datos a compartir sean insuficientes o erróneos.

Para definir los vocabularios tuvimos varias reuniones con personal de la DGAPA y de la Facultad de Ciencias, encargadas de los sistemas, para saber cuáles eran las necesidades que tenían referente a los datos que necesitaban que fueran cubiertos por nuestros vocabularios.

El espíritu de la Red Semántica es extender vocabularios ya existentes para normalizar la representación de los datos a través del mundo de la Red Semántica, logrando de esta manera que cualquier persona que tenga conocimiento de estos vocabularios, en cualquier parte del mundo, pueda acceder a la información que tenemos publicada sin necesidad de realizar un esfuerzo adicional, porque las dos partes conocen el significado de los vocabularios y cómo interpretarlos.

Siguiendo este espíritu para la comunicación con DGAPA, extendimos una descripción de persona que es muy usada: FOAF. Para la comunicación con la Facultad de Ciencias se creó un vocabulario para describir cursos, ya que no se encontró uno que se adecuara a nuestras necesidades.

Para crear los vocabularios usamos una herramienta gráfica llamada Protégé⁴ que nos permite, con una interfaz gráfica amigable, definir vocabularios para que posteriormente nos genere un RDF en su representación XML.

2.3.1. Persona UNAM

Para este vocabulario extendimos FOAF, primero realizando la extensión de éste para describir una persona de México a la que llamamos FOAFMexico.

⁴Herramienta para desarrollar vocabularios y ontologías, <http://protege.stanford.edu/>

Realizamos esto porque los datos que se toman en cuenta para describir una persona en FOAF tienen detalles que no se apegan a los formatos normales del registro de una persona en México, por ejemplo: la forma en que se manejan los nombres, apellidos y otros datos como el CURP.

Dejamos este vocabulario FOAFMexico con un formato más general, para que si otra persona en México quiera describir personas, pueda tomar en cuenta este documento y extenderlo de acuerdo a sus necesidades.

Después, tomando FOAFMexico extendimos éste a un caso más particular para persona dentro de la UNAM, agregando definiciones como nombramiento académico, número de cuenta, área de investigación, etc.

Tomó varias revisiones con el personal de DGAPA para que el vocabulario quedara de acuerdo a sus necesidades. Finalmente se optó por tener un enfoque minimalista para la descripción de una persona y poder empezar a trabajar con los vocabularios.

Por experiencia, hacer que algo abarque exactamente las necesidades de una institución, este nivel de detalle puede hacerlo inútil para otras. En un futuro, si se necesita hacer una modificación a este vocabulario la extensión del mismo es muy sencilla.

2.3.2. Curso

Para este vocabulario buscamos si ya existía otro que definiera cursos, pero no se encontró uno que se apegara a nuestras necesidades, por lo que no era posible extender un vocabulario. Creamos este vocabulario ajustándolo a las necesidades propuestas por el personal de la Facultad de Ciencias; así como las del IMATE, para que pudieran publicar los datos que necesitamos para tener la información completa.

2.3.3. Currículum

Realizamos este vocabulario porque en algunas reuniones que tuvimos se presentó el tema de un lenguaje que describiera el currículum de una persona. Buscando automatizar el llenado de algunas formas que serían prácticas para los profesores del IMATE, extendimos un vocabulario que encontramos, Resume-RDF, el cual ya contaba con la mayoría de los datos que necesitábamos, simplemente añadimos algunas definiciones que no tomaba en cuenta, así como la referencia a una PersonaUNAM como definición de la persona a la que pertenece el currículum.

2.3.4. Ejemplares

Para cada uno de los vocabularios realizamos ejemplares para fines de prueba, así como para mostrarlos en las reuniones que tuvimos; Protégé tiene una herramienta amigable para generar estos ejemplares.

2.4. Implementación

Se llevaron a cabo implementaciones de los casos de uso en Java usando JSP⁵ y Jena⁶. En particular usamos el motor que permite realizar consultas usando SPARQL. De igual forma se realizaron implementaciones en Plone, usando una biblioteca para Python llamada rdflib la cual nos da una funcionalidad muy parecida a Jena.

2.4.1. Demostraciones en Java

Se desarrollaron tres demostraciones usando Java y JSP con NetBeans⁷ esencialmente para mostrar lo sencillo que era compartir la información usando RDF en las reuniones.

2.4.1.1. Llenado de formas

Para esta demostración creamos una forma en HTML simulando los datos que se necesitarían para el registro de una persona en algún sitio web, teniendo la posibilidad de rellenarla a mano, o bien proveer una dirección con un documento RDF describiendo a la persona, del cual se sacan los datos para rellenar la forma. Se busca evitar el tener que escribir nuevamente datos que ya tenemos capturados en alguna parte de la web.

Usamos Jena para hacer las consultas en el documento RDF, generamos un XML sencillo con los datos que obtenemos con Jena y posteriormente lo utilizamos para rellenar los campos de la forma de los cuales se obtuvo información.

2.4.1.2. Generador de RDF

Se desarrolló esta demostración para ayudar a que las personas generen sus documentos RDF. Consiste en una forma en HTML, la cual pide los datos necesarios para generar el RDF de la persona. Cuando esta forma es enviada, toma los datos ingresados en los campos y a partir de ellos genera la descripción de la persona en RDF y la coloca en un campo de texto para que pueda ser copiado y guardado.

2.4.1.3. Generador de generadores RDF

Esto fue un ejercicio que se realizó para poder generar más RDF tomando como referencia un ejemplar de RDF ya fuese con los mismos datos o bien con un subconjunto de ellos. Consiste en dos partes, la primera que lee un RDF y aprovechando que este documento está en formato XML, extraemos, usando BISON, los campos que publica y crea una forma de HTML con estos campos, de tal forma que permite seleccionar qué campos se desean conservar. Una vez que seleccionados los campos que se desean generar y nos muestra el código en

⁵Java Server Pages

⁶Framework de java para crear aplicaciones para la Red Semántica

⁷IDE para desarrollar en java

XML con la especificación del RDF a generar. Podemos guardar este código para evitar volver a realizar la selección.

La segunda parte recibe una dirección con el XML que contiene la especificación del RDF a generar (creada por la primera parte del programa), lo interpreta y despliega una forma para que rellenemos con los datos que se quieren, genera el RDF con la descripción que se deseaba y lo despliega para poder copiar y guardarlo.

2.4.2. Demostraciones en Plone

Para el desarrollo de las demostraciones en Plone nos encontramos con el problema que nuestros conocimientos de formas en HTML no servían, pues los desarrolladores de Plone decidieron definir otras maneras de crear formas, las cuales son un tanto rebuscadas, ya que todo lo generan a partir de código Python. Otro problema fue cómo tener acceso a un archivo XML: el problema de estos archivos es que si subimos un archivo al sitio podemos acceder a él pero no lo podemos ver como contenido de la página. Si lo subimos como contenido de la página no podemos acceder a él con scripts, ya que Plone maneja los archivos de una forma especial.

2.4.2.1. Obtener datos usando RDF

Usando ArchGenXML⁸ se desarrolló un *data type*⁹ con los datos de un maestro, para dar la opción a crear contenidos de este tipo, usando un RDF como fuente de los datos y no la manera convencional de Plone. Ésta presenta una forma con campos de los datos que se requieren. Usamos el producto *z3c.form*¹⁰ para crear una forma sencilla, la cual sólo pide una dirección de RDF que usa para obtener los datos y a partir de éstos crear el *data type* deseado.

2.4.2.2. Generador de RDF

Como Plone maneja de forma especial los archivos en este punto no consideramos crear un tipo de contenido que arreglara este problema.

Usando ArchGenXML creamos un *data type* que sólo tiene título y un espacio para introducir texto; el espacio lo utilizamos para llenarlo con el RDF generado para que al igual que la demostración en Java se copiara y pegara el RDF a otro archivo; este código generado fue guardado en la página. Para esta demostración usamos nuevamente el producto *z3c.fom* para generar la forma con los datos que se necesitan para crear el RDF. Posteriormente con estos datos se crea el contenido descrito anteriormente.

⁸Generador de código que genera productos Plone usando modelos UML

⁹Tipo de dato que Plone maneja

¹⁰Producto Plone para generar formas

2.4.2.3. *Data type* Persona

Como último desarrollo dentro del seminario creamos un *data type* persona usando *paster-archetypes*¹¹ donde implementamos las dos opciones anteriores: Llenar los datos a partir de una dirección RDF y con los datos que tenemos generar un RDF el cual puede ser accedido con scripts. Para esto modificamos las vistas que tienen por omisión los *data types* y creamos vistas que permitieran el comportamiento deseado.

2.5. Manuales de uso

En esta sección se describe cómo hacer uso de las demostraciones creadas para el seminario.

2.5.1. Demostraciones en Java

De las demostraciones realizadas en Java (1.4.1.1), (1.4.1.2) y (1.4.1.3) sólo abarcamos el uso de la demostración para llenar formas y el generador de generadores de RDF, ya que el generador de RDF se puede ver como un subcaso del generador de generadores.

2.5.1.1. Instalación

Las demostraciones realizadas en Java fueron desarrolladas en NetBeans, por lo que la instalación de estos productos es bastante sencilla. Si se tiene NetBeans instalado simplemente se carga el proyecto y ejecuta.

2.5.1.2. Llenado de formas

Al cargar la página principal de esta demostración, aparece una interfaz con los campos para ser rellenos con la información de una persona y un campo para ser relleno con la dirección de un RDF (Figura 2.1), del cual se obtendrán los datos para relleno de los otros campos, acción que se lleva a cabo al presionar el botón de (Figura 2.2).

¹¹Herramienta que facilita la creación contenidos Plone

Rellenar información

A continuación puede rellenar los siguientes datos o puede ingresar el url de un archivo RDF.

Dirección RDF

- **Grado Máximo de Estudios**
- **Nombre**
- **Primer Apellido**
- **Segundo Apellido**
- **Teléfono**
- **Correo**
- **Homepage**
- **CURP**
- **Institución Actualmente Adscrita**
- **Resumen Curricular**

Figura 2.1: Proporcionamos la dirección de un RDF en el respectivo campo.

Rellenar información

A continuación puede rellenar los siguientes datos o puede ingresar el url de un archivo RDF.

Dirección RDF

- **Grado Máximo de Estudios**
- **Nombre**
- **Primer Apellido**
- **Segundo Apellido**
- **Teléfono**
- **Correo**
- **Homepage**
- **CURP**
- **Institución Actualmente Adscrita**
- **Resumen Curricular**

Figura 2.2: Presionamos el botón Insert y rellena el resto de los datos.

2.5.1.3. Generador de generadores

Al cargar la página principal de la demostración nos muestra una interfaz con un campo para proporcionar el RDF del cual queremos crear otros ejemplares, lo llenamos y presionamos el botón de `generar` (Figura 2.3). Después carga una lista de los campos que tiene el RDF para seleccionar los que queremos que

tengan los nuevos ejemplares, seleccionamos las que deseamos y presionamos (Figura 2.4).

La siguiente página nos muestra el código XML con los campos que elegimos para guardarlos, si así deseamos, y una liga al generador (Figura 2.5). El generador lee el XML anterior y genera una forma con los datos a llenar, llenamos con los datos y presionamos (Figura 2.6).

Por último nos muestra el RDF generado (Figura 2.7).

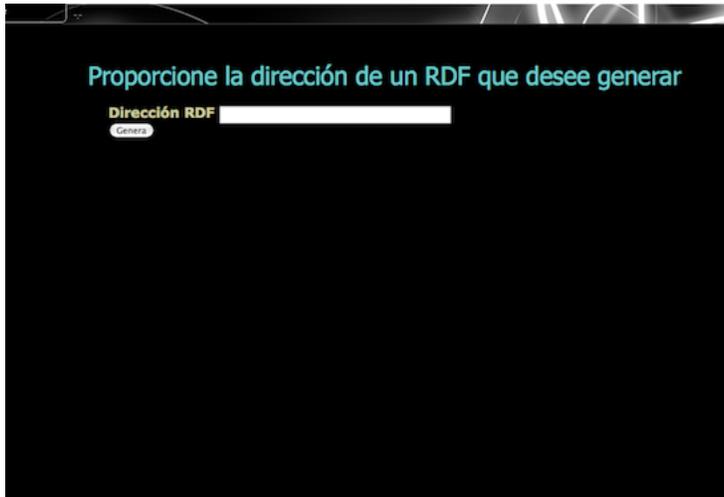


Figura 2.3: Proporcionamos la dirección RDF



Figura 2.4: Seleccionamos los campos deseados

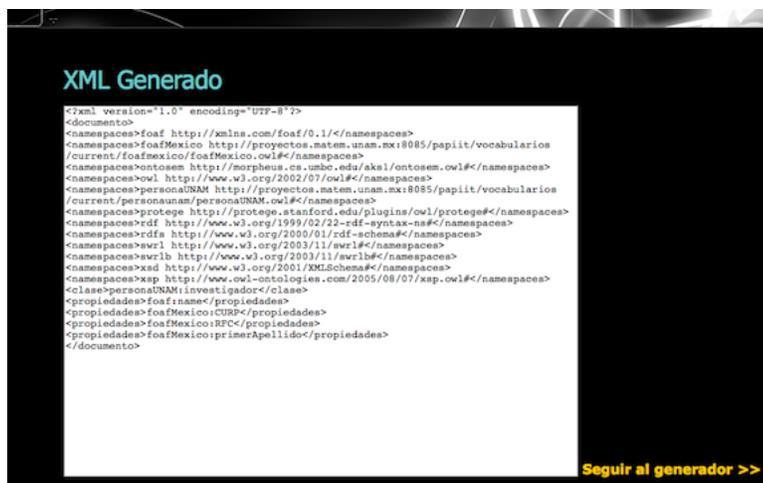


Figura 2.5: XML generado con la información de lo que seleccionamos

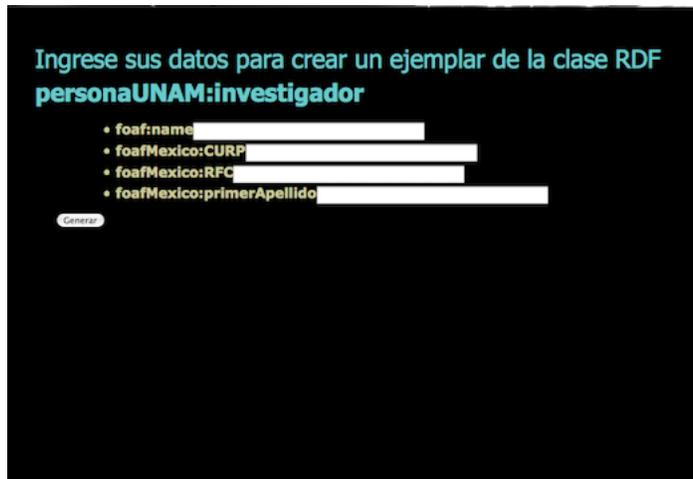


Figura 2.6: Llenamos la forma con los datos que pide



Figura 2.7: RDF Generado

2.5.2. Demostraciones en Plone

Describimos el uso de las tres demostraciones realizadas en Plone (1.4.2.1), (1.4.2.2) y (1.4.2.3). El uso de los dos primeros no es tan intuitivo pues a las vistas creadas no se puede acceder con un botón en la interfaz, tenemos que acceder a ellas a través de una URL.

2.5.2.1. Instalación

Para instalar las demostraciones en Plone para los *data types* desarrollados usando ArchGenXML, simplemente se pegan dentro de la carpeta productos/ de Plone las formas generadas para las dos primeras demostraciones y el *data type* (desarrollado usando *paster-archetypes*), se instalan poniendo el código de los productos dentro de la carpeta src/ de Plone y registrándolos en el buildout.cfg.

Suponiendo que el nombre del producto es mi.producto se registran en el buildout de la siguiente forma:

```
[buildout]
...
eggs = mi.producto
...
develop = src/mi.producto
[instance]
...
zcml = mi.producto
```

2.5.2.2. Obtener datos usando RDF

Para acceder a la vista debemos usar la dirección:

http://localhost:8080/Plone/@@persona_add_form.

Muestra la forma que pide una liga a un RDF FOAF válido; llenamos con una liga y presionamos el botón (Figura 2.8). Crea un *data type* con los datos obtenidos del RDF (Figura 2.9).

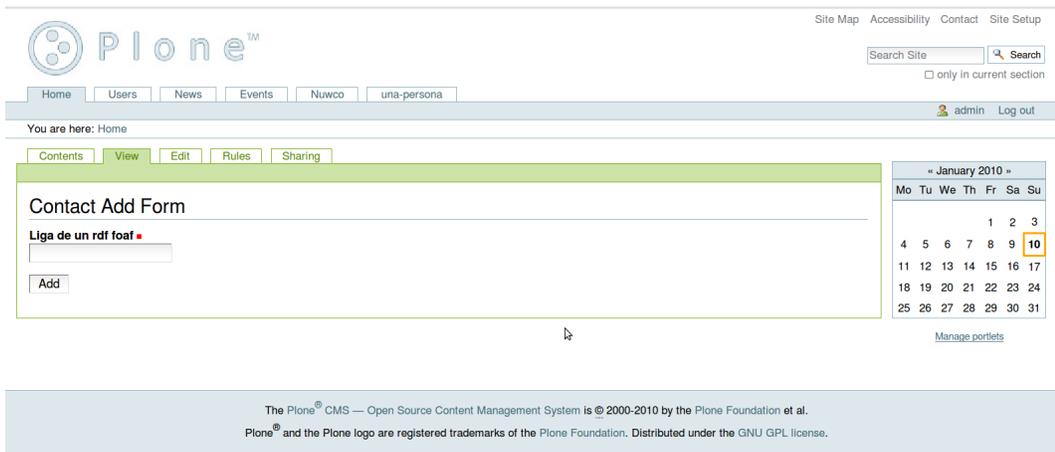


Figura 2.8: Damos la dirección de un RDF

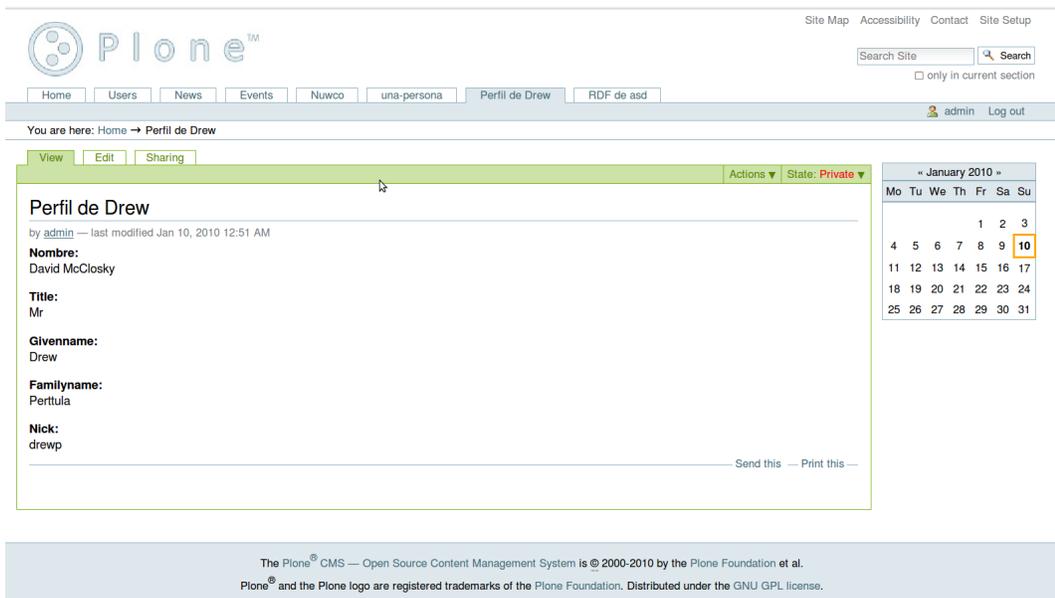


Figura 2.9: Data type Generado

2.5.2.3. Generador de RDF

Para acceder a la vista se debe usar la dirección:

`http://localhost:8080/Plone/@@contact_add_form.`

Muestra una forma que pide los datos de una persona. Rellenamos y presionamos el botón **Add** (Figura 2.10). Crea un *data type* que contiene en un campo de texto un XML con la descripción en RDF de la persona a la que pertenecen los datos de la forma anterior.

The screenshot shows the Plone Contact Add Form interface. The form is titled "Contact Add Form" and contains the following fields:

- Nombre**: Hugo
- Primer apellido**: Rodriguez
- Segundo apellido**: Pena
- CURP**: RCPH861109HDFDX000
- RFC**: RCPH861109
- Telefono**: 56 10 73 36
- Email**: jnd4ke@gmail.com
- Pagina personal**: www.sn4ke.com
- Adscripcion actual**: Fac. ciencias
- Grado maximo de estudios**: Licenciatura

An **Add** button is located at the bottom of the form. The interface also includes a Plone logo, navigation menu, search bar, and a calendar for January 2010.

Figura 2.10: Llenamos con los datos de la persona

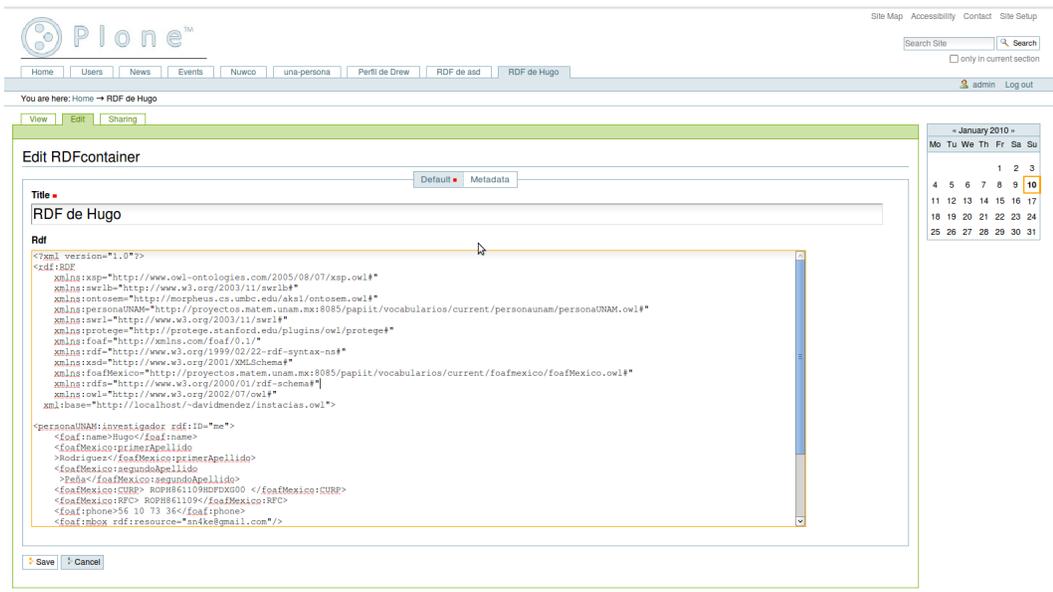


Figura 2.11: El data type con el RDF generado

2.5.2.4. *Data type* Persona

Usando la barra de Plone agregamos una nueva persona (Figura 2.12). Nos muestra la pantalla de edición del contenido donde podemos agregar a mano los datos de la persona, o bien dar una dirección RDF con estos datos, llenamos el campo de RDF y presionamos el botón para usarlo (Figura 2.13). Revisamos los datos obtenidos del RDF y guardamos (Figura 2.14). **Guardar** nos manda a la vista del contenido, en dicha pantalla vemos los datos de la persona (Figura 2.15).

Si se desea guardar la información de este documento en un RDF se puede presionar el botón **Generar RDF** y genera un archivo con el RDF.

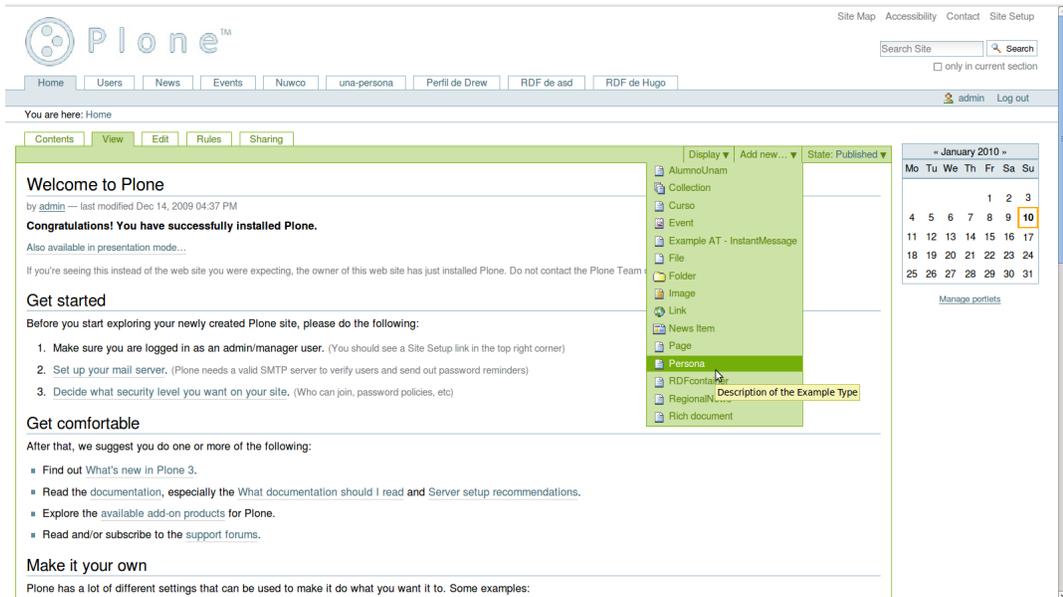


Figura 2.12: Agregamos una Persona

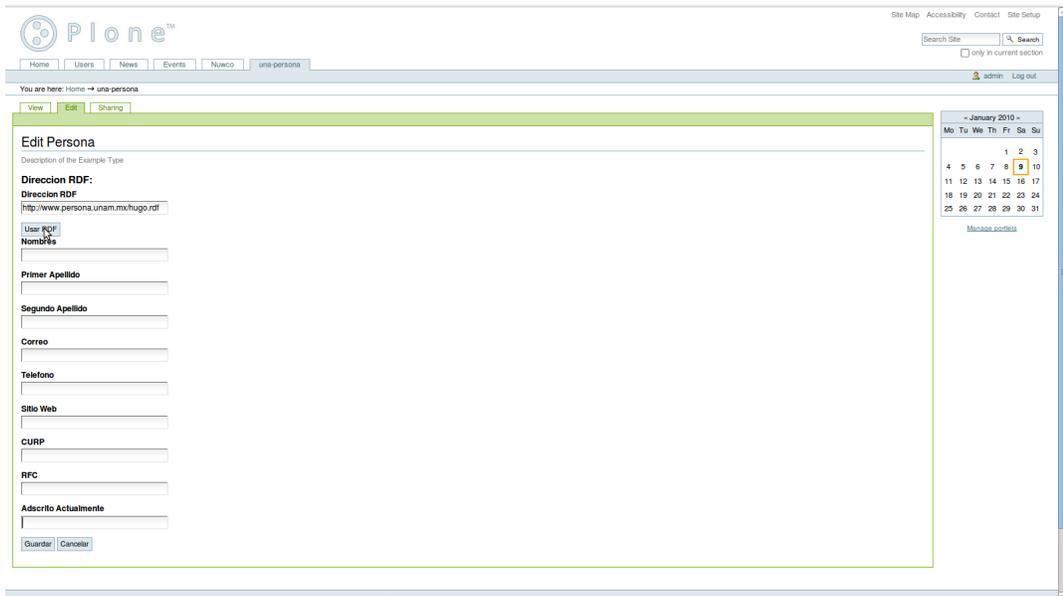


Figura 2.13: Usamos el botón Usar RDF

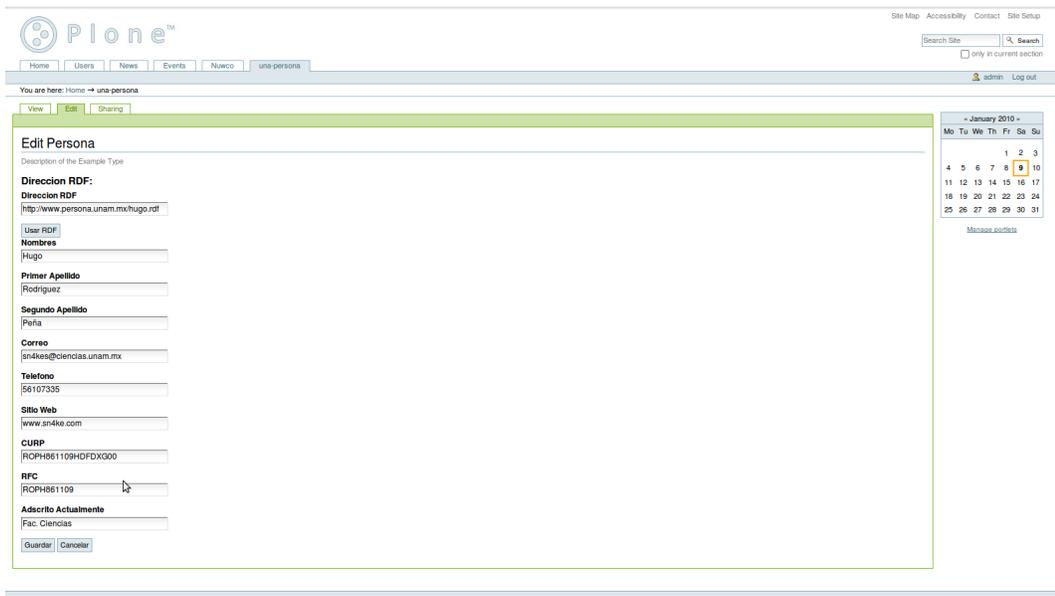


Figura 2.14: Revisamos los datos y guardamos

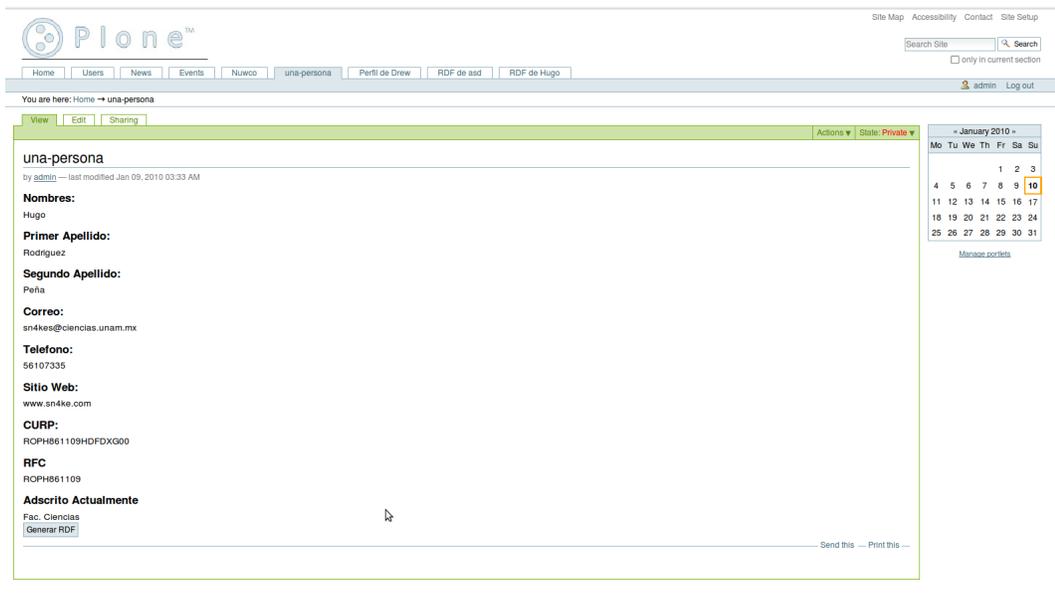


Figura 2.15: Vista de la persona

Capítulo 3

Conclusiones

La Red Semántica me parece un tema muy interesante, porque si bien hoy en día Internet es una herramienta muy poderosa para encontrar información, si se complementa con la Red Semántica, será una mejor herramienta ya que podremos realizar búsquedas más complejas recibiendo resultados eficaces además de los usos potenciales que tiene para hacer exámenes estadísticos sobre el cúmulo de información contenida en la Red.

El participar en el proyecto de intercambio de información entre instituciones fue interesante porque somos el primer esfuerzo que se hace para implementar la Red Semántica en la UNAM. Recibimos gran apoyo por parte del personal de la DGAPA y con esto logramos hacer vocabularios funcionales para la Universidad. Dichos vocabularios podrán ser usados y extendidos por otras instituciones en el futuro.

Trabajar con Plone me frustró un poco debido a la poca documentación que tiene, en ocasiones cuando encontrábamos documentación que nos era útil nos llevábamos la sorpresa que era documentación vieja y los métodos que explicaban ya no funcionaban.

Lamentablemente, cuando quisimos implementar las demostraciones realizadas en Plone sobre los productos que utilizan en el IMATE para el manejo de los cursos y personas, estos productos estaban en proceso de renovación, motivo por el cual no pudimos trabajar sobre ellos y creamos la última demostración de Persona. También al final del semestre, por coincidir con final de año, las instituciones con las que trabajábamos tuvieron que trabajar en otros proyectos más urgentes, dejando pendiente nuestro proyecto.

Me gustaría terminar de implementar los cambios pertinentes a los productos del IMATE y reanudar la comunicación con la DGAPA para terminar de implementar el intercambio de la información entre el IMATE y la DGAPA, que es el objetivo central del seminario.

Considero importante que se impulse más la implementación de la Red Semántica en la UNAM ya que realmente es un modo eficaz y eficiente de compartir datos. A nosotros nos costó trabajo realizar las demostraciones en Plone por estar poco familiarizados con esta plataforma, pero tomando en cuenta que cada

institución tiene su departamento dedicado a mantener sus servidores este personal podría fácilmente realizar la publicación de los datos que tiene en RDF, así como hacer las modificaciones pertinentes para hacer uso de ellos. De esta forma se agilizarían muchos trámites internos en la Universidad.

Bibliografía

- [1] John Hebler, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez. *Semantic Web Programming*. Wiley Publishing. Primera edición, 2009.
- [2] Miniwatts Marketing Group. *Internet World Usage Stats - Internet and Population Statistics*. Enero 2010. <<http://www.internetworldstats.com>>.
- [3] Tim Berners-Lee, James Hendler, Ora Lassila. *The Semantic Web*. Scientific American Magazine. Publicado mayo 2001. <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>.
- [4] Tim Berners-Lee. *RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax*. Network Working Group. Publicado enero 2005. <<http://tools.ietf.org/html/rfc3986>>.
- [5] Tim Bray, Jean Paoli. *Extensible Markup Language (XML) 1.0*. W3C recommendation. Publicado junio 2006. <<http://www.w3.org/TR/xml>>.
- [6] Andy Mckay, Alan Runyan. *The Definitive Guide to Plone*. Apress Publishing. Segunda Edición. Septiembre 2007.
- [7] Dave Beckett. *RDF/XML Syntax Specification (Revised)*. W3C recommendation. Publicado febrero 2004. <<http://www.w3.org/TR/REC-rdf-syntax/>>.
- [8] Thomas Baker, Tom Heath, Susie Stephens. *Semantic Web Case Studies and Use Cases*. W3C document. Publicado agosto 2009. <<http://www.w3.org/2001/sw/sweo/public/UseCases/>>.