



Maestría en Ingeniería en Sistemas y Cómputo Inteligente

**Inserción de información semántica a
datos del repositorio institucional de la UPPue**

Paulo Daniel Vázquez Mora

María Auxilio Medina Nieto

Jorge de la Calleja Mora

Antonio Benitez Ruiz



Inserción de información semántica a datos del repositorio institucional de la UPPue

Paulo Daniel Vázquez, María Auxilio Medina,
Jorge de la Calleja, Antonio Benitez

Maestría en Ingeniería en Sistemas y Cómputo Inteligente

{paulo.vazquez4303, maria.medina, jorge.delacalleja, antonio.benitez}@uppuebla.edu.mx

Tercer Carril del Ejido Serrano S/N, San Mateo Cuanalá, Juan C. Bonilla, Puebla, México

1. Introducción

La implementación de *repositorios institucionales* (RIs) satisface necesidades de conservación y publicación de la producción académica y científica. En los RIs, los mecanismos de consulta emplean datos descriptivos (o metadatos) y palabras clave.

Este cartel propone la inserción de información semántica a datos provenientes del RI de la Universidad Politécnica de Puebla (RI-UPPue) en una instancia de la ontología Onto4AIR[1], el propósito es mostrar que las tecnologías semánticas extienden los mecanismos de búsqueda en los RIs y sirven para validar automáticamente la consistencia lógica de los datos. El acceso a estos datos se realiza mediante la implementación de un servicio web denominado *recuperación semántica-01*, en adelante, RS-01.

2. Objetivo

Gestionar información semántica relacionada con los sinodales en un conjunto de datos de la colección de tesis utilizando la ontología Onto4AIR y el servicio web RS-01.

3. Metodología

La metodología consiste en utilizar la ontología Onto4AIR descrita en [1] e implementar el servicio RS-01; las características generales de este servicio se presentan en [2]. La Figura 1 muestra los pasos principales de esta metodología, actualmente se trabaja en el 5 y 6.

1. Análisis de requerimientos
2. Diseño de RS-01
3. Poblado de la ontología Onto4AIR
4. Instalación del módulo DSpace RDF en el RI-UPPue
5. Codificación del RS-01
6. Pruebas de funcionalidad y usabilidad

Figura 1. Pasos para gestionar información semántica

3.1 Análisis de requerimientos

A la fecha de elaboración de este documento, el Repositorio Nacional (RN) [3] integra datos de 88 RIs, su interoperabilidad se basa en la implementación de la versión 2.0 del Protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), la cual utiliza el formato estándar de metadatos Dublin Core (DC).

De acuerdo con los lineamientos técnicos del RN, el elemento DC que almacena el nombre del estudiante de una tesis es *dc:creator* y para el director, directora o sinodal, se emplea *dc:contributor*, lo cual ocasiona que no haya forma de saber cuál es el rol de los docentes en la elaboración de una tesis. Los requerimientos del servicio web RS-01 que atiende esta problemática se muestran en la Tabla 1.

Tabla 1. Requerimientos del servicio RS-01.

Descripción	Prioridad
Consulta de información semántica	Alta
Capacidad de almacenamiento de ternas en RDF	Alta
Exportación de datos en formatos no propietarios como JSON o RDF	Media
Integración de datos provenientes de otros RIs	Media

Previo al diseño del RS-01, se revisó el catálogo de servicios del RN, éstos son utilizados por usuarios con perfil de administrador, algunos de los datos a recuperar se incluyen en la Tabla 2.

Tabla 2. Ejemplo de servicios web soportados por el RN.

Datos a recuperar desde servicios web del RN	
Áreas de conocimiento	Licencia
Campos de conocimiento	Localidad
Disciplinas de conocimiento	Municipio
Sub-disciplinas de conocimiento	Nivel de acceso
Audiencia	País
Estado	Persona
Institución	Formato
Idioma	Programa

3.2 Diseño de RS-01

La Figura 2 muestra las entidades principales que intervienen en la implementación del servicio RS-01, se basa en el diseño disponible en [2].

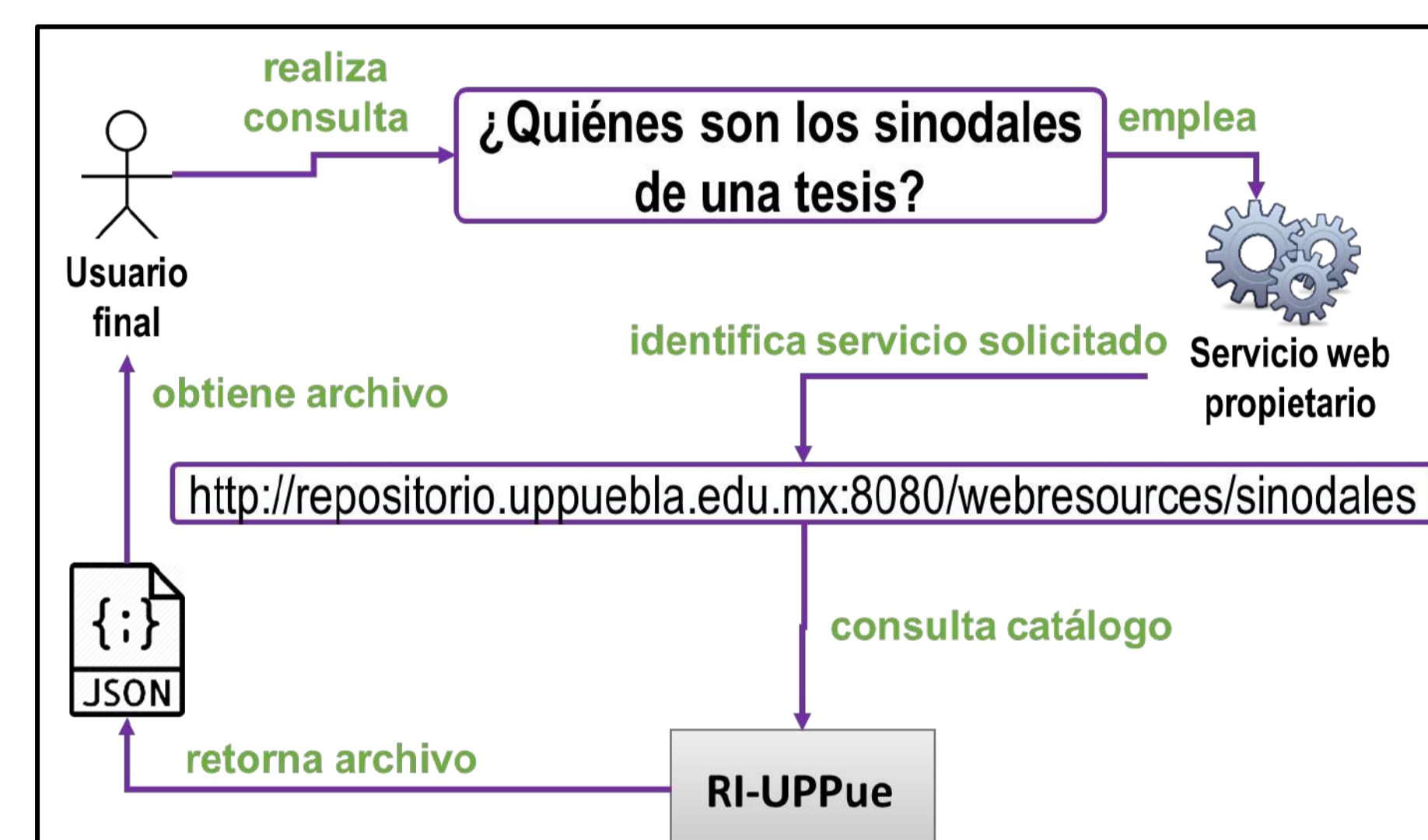


Figura 2. Arquitectura para el servicio RS-01

El servicio RS-01 se llama al introducir un URL como los que se muestran en la Tabla 3; un ejemplo del resultado se visualiza en la Figura 3, note que se utiliza como salida el formato JSON. Los resultados se pueden procesar al utilizar la librería *request* de *Python* y almacenarlos en cualquier tipo de estructura como una lista o tupla. Las peticiones POST, DELETE o PUT se utilizarían para agregar, borrar o actualizar la información.

Tabla 3. Entradas y salidas de RS-01.

Entrada	Salida en formato JSON
http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales	Lista de todos los sinodales
<a href="http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales/<cveSinodal>">http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales/<cveSinodal>	Datos del sinodal con identificador "cveSinodal"



Figura 3. Archivo tipo JSON resultante.

El diseño del servicio hace uso de patrones tales como MVC (modelo-vista-controlador), *Factory* (Constructor de objetos) y *Web service broker* (atención a solicitudes de servicio web) con el propósito de que sea seguro, flexible y portable.

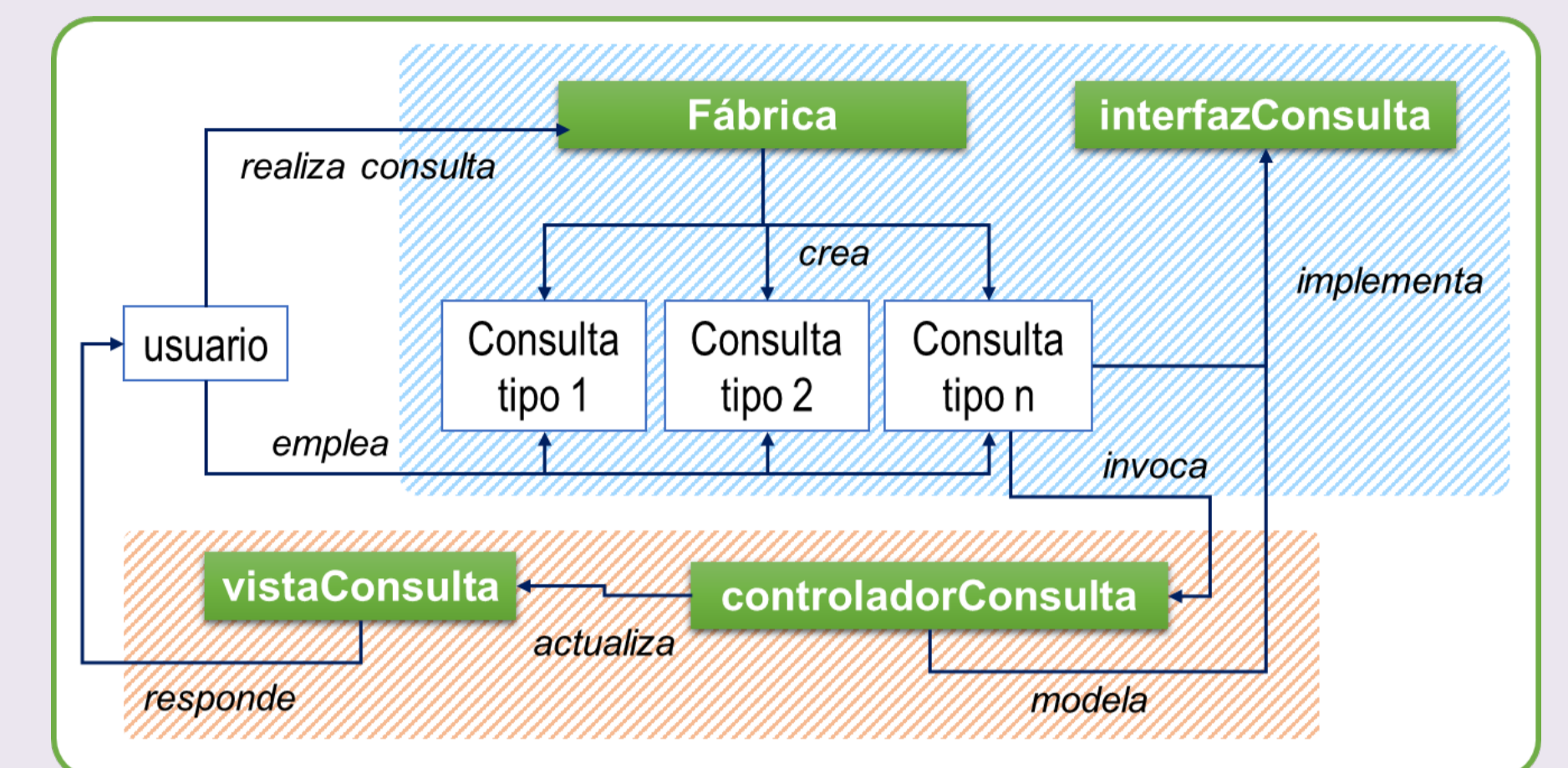


Figura 4. Diagrama de clases propuesto para el servicio RS-01.

En [2] se describe el diseño de clases y patrones de diseño integrados en el servicio web de la Figura 5.

3.4 Poblado de la ontología

Se agregaron instancias de tesis de maestría con el objetivo de realizar pruebas.

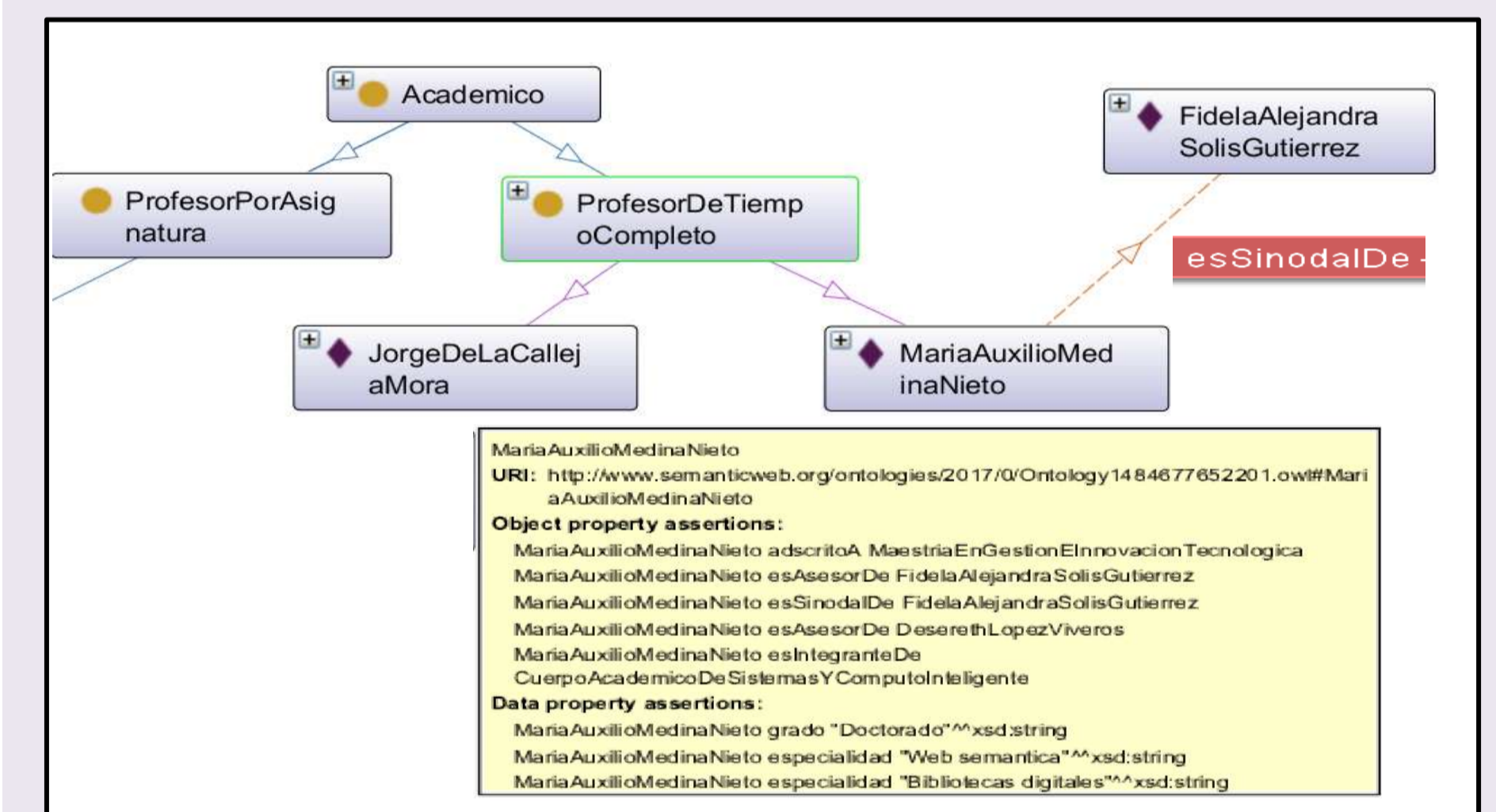


Figura 5. Gráfico de instancias relacionadas.

4. Resultados parciales

```

1 # Librerías empleadas
2 import requests
3 import json
4 # Petición HTTP al servicio RS-01 con GET:
5 r = requests.get("http://repositorio.uppuebla.edu.mx:8080/webresources/autor")
6 # Si el código de estado HTTP es 200 (OK):
7 if r.status_code == 200:
8     # Convierte JSON a lista
9     data = json.loads(r.text)
10    data.extend(lista)
11    for element in data:
12        pass
13        ""
14    Integración de datos de sinodales con los extraídos del RI-UPPue realizando comparaciones entre dc:Author y las instancias "autor" de Onto4RI-UPPue

```

Figura 6. Código reducido del servicio de inserción de datos RS-01

5. Conclusiones

La inserción de información semántica a dos RIs incrementa la posibilidad de reutilización, dado que extiende los servicios de búsqueda vigentes y permite validar la consistencia lógica de los datos de forma automática. A corto plazo, se espera integrar un catálogo de servicios de recuperación de información semántica en el RI-UPPue, de manera que distribuya los datos como datos abiertos enlazados.

Agradecimientos

A CONACYT por ser becario de este programa académico (No. de BECA: 863914) y a todas las personas que apoyan este proyecto incluyendo a mi familia.

Referencias

- [1] Medina N. M. A., S. J. (2017). Representación semántica de conocimiento operativo y de dominio para repositorios institucionales. Registro Público del derecho de autor 03-2017-042511235500-01.
- [2] Vázquez M. P. D., Medina N. M. A., De la Calleja M. J., Benítez R. A., Vidal T. M., Alanís U. J. S. (2018). Diseño de un servicio web para la extracción de información semántica del RI-UPPue. SOMI, Congreso de Instrumentación, ISSN 2395-8499
- [3] CONACyT, (2019). Repositorio Nacional. Disponible en: <http://catalogs.repositorionacionalcti.mx/> [Consulta: el 30 Mar. 2019].
- [4] CONACyT, (2019). Catálogo de Servicios REST. Disponible en: <http://catalogs.repositorionacionalcti.mx/> [Consulta: el 30 Mar. 2019].



Este material se distribuye bajo los términos de la
Licencia *Creative Commons* CC BY-NC-ND 2.5 MX

2019