

Modelado de la operatividad de un repositorio institucional con acceso abierto utilizando Protégé

María Auxilio Medina*, J. Alfredo Sánchez**,
Ofelia Cervantes V.***, Jorge de la Calleja****, Antonio Benitez*****

1. Introducción

Algunas Instituciones de Educación Superior (IES) almacenan parte de su producción intelectual en un repositorio institucional (RI), el cual es una “plataforma digital centralizada que sigue estándares internacionales de interoperabilidad, mantiene, preserva y disemina la información científica, tecnológica y de innovación derivada de investigaciones, productos educativos, académicos y desarrollos tecnológicos [1]”. Para las IES, contar con un RI ofrece beneficios como preservación de la producción académica y científica, acceso a texto completo a documentos en formato digital del propio RI u otros, incremento en la visibilidad de la institución y de los autores, establecimiento de políticas de distribución de acceso abierto para los documentos y metadatos.

Este documento propone el modelado de un RI y su operatividad utilizando el editor de ontologías Protégé, con el propósito de apoyar la adquisición de conocimiento de dominio a través del uso de un vocabulario controlado y de una representación formal, no ambigua de los términos y sus relaciones que puede compartirse y reutilizarse entre usuarios y computadoras. La versión completa del modelo es una ontología que se ha registrado formalmente ante el Instituto Nacional de Derecho de Autor (IndAutor) como obra de base de datos número 03-2017-042511235500-01 [6].

El documento está organizado como sigue: los beneficios del modelado semántico para RIs se listan en la sección 2. La sección 3 describe el modelado semántico u ontología. Finalmente, la sección 4 presenta las conclusiones y el trabajo a futuro.

2. Beneficios del modelado semántico para RIs

El uso de un modelo semántico para RIs apoyará la apropiación del conocimiento operativo y de dominio, así como la implantación de RIs en IES, amplía la posibilidad de revisar de manera automática la integridad y consistencia de los datos, lo cual representa un aumento en su valor

* maria.medina@uppuebla.edu.mx. Universidad Politécnica de Puebla.

** alfredo.sanchez@udlap.mx. Universidad de las Américas Puebla.

*** ofelia.cervantes@udlap.mx. Universidad de las Américas Puebla.

**** jorge.delacalleja@uppuebla.edu.mx. Universidad Politécnica de Puebla.

***** antonio.benitez@uppuebla.edu.mx. Universidad Politécnica de Puebla.

y veracidad. En comparación con un modelo de datos de tipo relacional en donde se definen entidades, relaciones y objetos, la expresividad de un modelo semántico es mayor por poseer características como las siguientes:

La representación de propiedades o relaciones entre clases incluye herencia, composición o asociación.

Las relaciones se definen a nivel de clases o instancias.

Las propiedades de datos asocian un atributo o característica de una instancia con un tipo de datos del esquema de nombres de XML.

Las propiedades de objeto definen relaciones binarias entre cualesquiera dos instancias

Junto con restricciones de integridad, se modelan restricciones de cardinalidad a las que se les asocia un valor máximo o mínimo o la propiedad exactamente igual, esto es, las relaciones no sólo son 1-1, 1-N o M-N, sino que la cardinalidad la adapta el desarrollador del modelo.

Las relaciones entre instancias y tipos de datos o las relaciones entre clases, pueden tener asociado un dominio y rango, así como valores fijos o predeterminados.

Se pueden definir reglas o restricciones que utilicen cuantificadores universales o existenciales, éstas se utilizan para validar consistencia e inferir conocimiento.

Es factible tener instancias que pertenezcan a una, ninguna o más clases si las últimas no son disjuntas.

Es viable modelar clases disjuntas.

Se obtiene conocimiento nuevo a partir de mecanismos de inferencia, es decir, se extrae conocimiento implícito a partir de conocimiento explícito.

La validación de la consistencia de un modelo utiliza razonadores lógicos, programas conocidos también con el nombre de razonadores.

3. Modelado de la operatividad de un RI

El modelado de la operatividad de un RI se divide en dos etapas: identificación de un RI y representación de componentes. El objetivo de la primer fase es identificar de manera unívoca a un RI, de forma que pueda distinguirse de cualquier otro en la Red Mexicana de Repositorios Institucionales (REMERI) o en redes de ámbito internacional. En la segunda fase, se representan los usuarios, documentos y términos relacionados con la temática del acceso abierto. A diferencia de los atributos de la fase uno, los componentes de la segunda fase pueden modificarse y adaptarse a distintos contextos de acuerdo a las necesidades de las comunidades de usuarios.

3.1 Fase 1 del modelado: identificación de un RI

Las Tablas 1 y 2 contienen los atributos que describen a un RI. De acuerdo con [2] y [3], un atributo es mandatorio si es un requisito incluirlo o recomendado si aporta información deseable.

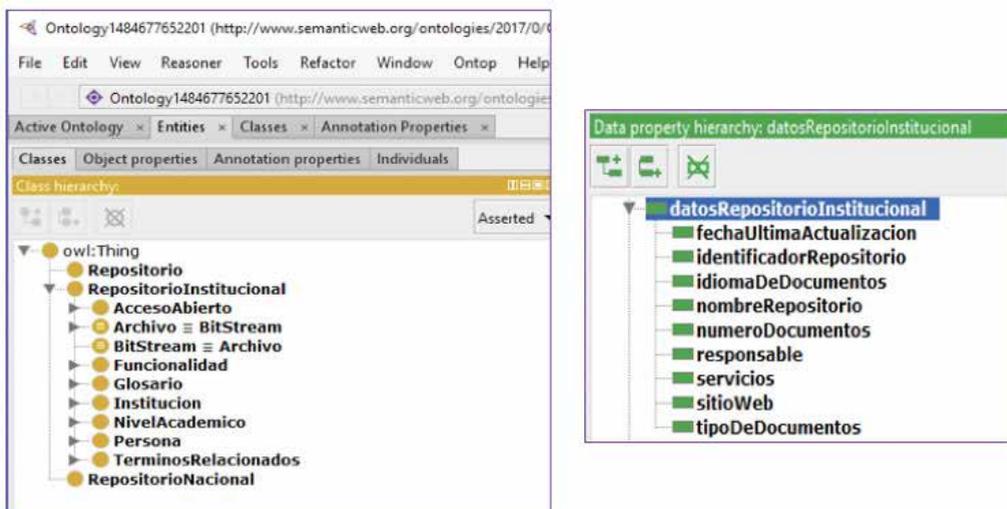
Tabla 1. Atributos de un RI.

Atributo	Descripción	Tipo de atributo
fechaUltimaActualizacion	Fecha de la última actualización de los datos	Mandatorio
identificadorRepositorio	Cadena que distingue a un RI de otro en la misma institución	Mandatorio
idiomaDeDocumentos	Abreviación que indica los idiomas en los que están escritos los documentos, por ejemplo, español: es, inglés: en.	Recomendado

Tabla 2. Atributos de un RI.

Atributo	Descripción	Tipo de atributo
nombreRepositorio	Cadena que contiene el nombre del repositorio, éste aparecerá en página de inicio.	Mandatorio
númeroDocumentos	Número de documentos almacenados.	Recomendado
responsable	Persona encargada del RI.	Mandatorio
sitioWeb	URL que contiene la dirección web de la página de inicio	Mandatorio
tipoDeDocumentos	Enumeración (lista) que indica el tipo de documentos que almacena el RI, por ejemplo, reportes técnicos, artículos, carteles o tesis.	Recomendado

La clase `Repositorio` es análoga a la clase `RepositorioInstitucional` o `RepositorioNacional`, esto con el propósito de resaltar que en las IES cualquier repositorio deberá seguir los lineamientos técnicos establecidos en [2] y [3] para interoperar. La Figura 1 muestra del lado izquierdo, los componentes de un RI, mismos que se modelan como clases en el editor Protégé; el lado derecho contiene los atributos de la Tabla 1 agrupados en la propiedad `datosRepositorioInstitucional`; esta propiedad está relacionada con la clase `RepositorioInstitucional`.

Figura 1. Sub-clases y atributos de `RepositorioInstitucional`

3.2 Fase 2 del modelado: representación de componentes

La fase 2 sigue los pasos para construir un modelo semántico u ontología de la metodología 101, propuesta en [4], los cuales son:

1. Determinar el dominio y alcance
2. Reutilizar ontologías existentes
3. Enumerar términos importantes
4. Definir clases y su jerarquía
5. Definir las propiedades de las clases (axiomas de propiedades de datos)
6. Definir las características (o facetas) de las propiedades
7. Crear instancias o individuos

La Figura 2 muestra la implementación del paso 4. El modelo está editado en la versión 5.1.0 de Protégé [5]. La implementación del paso 5 implica la definición de propiedades de datos y de objeto, en tanto, el paso 6, se refiere a seleccionar las facetas (características) de las propiedades, las cuales son: funcional, simétrica, inversa funcional, asimétrica, reflexiva e irreflexiva. El paso 7 se conoce también como poblar el modelo. Una vez que se cuente con instancias, el usuario final puede hacer uso de las interfaces de consulta del propio editor para recuperar información específica, así como de los razonadores para verificar automáticamente la consistencia.

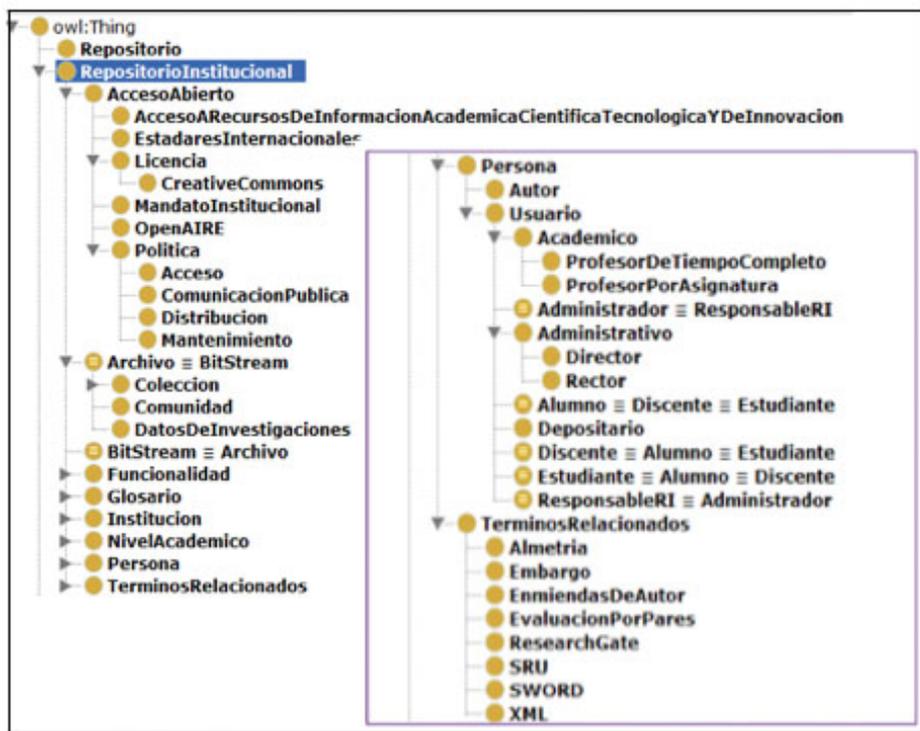


Figura 2. Componentes de un RI representados en Protégé en jerarquía de clases

Conclusiones

Este documento presentó algunos elementos de un modelo semántico para repositorios institucionales. El modelo integra conceptos relacionados con el acceso abierto y la operatividad de los RIs en un vocabulario común, no ambiguo, formal y consistente que puede ser compartido entre usuarios y computadoras. Los términos representados se agruparon en una jerarquía de clases y se relacionaron entre sí mediante propiedades de objeto. El modelo es flexible, de forma que diferentes IES podrían adoptarlo para atender necesidades de integración de datos. Como trabajo a futuro, se propone la migración de datos provenientes de diferentes RIs a instancias del modelo con el propósito de validar la consistencia de forma automática y así, ampliar la posibilidad de realizar análisis de datos en contextos semánticos.

Referencias

- [1] Estrategia de acceso abierto a la información científica, tecnológica y de innovación repositorios institucionales. Términos de referencia para desarrollar los repositorios institucionales de acceso abierto a la información científica, tecnológica y de innovación. Fecha de consulta: 30 de Abril del 2016. Disponible en: <http://www.conacyt.mx/index.php/el-conacyt/convocatorias-y-resultados-conacyt/convocatorias-direccion-adjunta-de-planeacion-y-evaluacion/convocatoria-2015-para-desarrollar-los-repositorios-institucionales-de-acceso-abierto-a-la-informacion-cientifica-tecnologica-y-de-innovacion/10724-terminos-de-referencia-repositorios-institucionales-2015/file>.
- [2] Lineamientos generales para el repositorio nacional y los repositorios institucionales. Documento disponible desde el sitio web del Sistema Integrado de Información sobre Investigación Científica, Desarrollo Tecnológico e Innovación (SIICYT). Fecha de consulta: 15 de Marzo del 2017. Dirección web: https://www.repositorionacionalcti.mx/docs/Lineamientos_generales.pdf.
- [3] Lineamientos técnicos para el repositorio nacional y los repositorios institucionales. Documento disponible desde el sitio web del Sistema Integrado de Información sobre Investigación Científica, Desarrollo Tecnológico e Innovación (SIICYT). Fecha de consulta: 15 de Marzo del 2017. Dirección web: <http://www.siicyt.gob.mx/index.php/normatividad/2-conacyt/4-conacyt/1499-lineamientos-tecnicos-para-el-repositorio-nacional-y-los-repositorios-institucionales/file>
- [4] Noy N. F., McGuinness D. L. 2001. Ontology development 101: A guide to create your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, [en línea]. [Fecha de consulta: 16 de Enero del 2017]. Disponible en: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- [5] Musen, M.A. The Protégé project: A look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015. DOI: 10.1145/2557001.25757003.
- [6] Medina N. M. A., Sánchez J.A, Cervantes, O., Medina R. R. C., De la Calleja, M. J., Benitez A. 2017. Representación semántica de conocimiento operativo y de dominio para

repositorios institucionales. Registro público del derecho de autor. Número: 03-2017-042511235500-01. Fecha de registro: 26 de Abril del 2017.