

Use of an Ontological Model to Assess the Relevance of Scientific Production

J. Guerrero-Sosa, V. Menéndez-Domínguez, M. Castellanos-Bolaños, and J. Gómez-Montalvo

Abstract—Several institutions do scientific publications, which are used to know the academic and technological level of a teacher, a research group and the institution. In this article, the use of an ontological model is proposed for the representation of the relevance of the scientific production of the most important public university in the southeastern of Mexico. The architecture of the system is presented considering the information gathering from repositories of digital resources, the storage of the data in a document-oriented database and in an ontological one to later represent through SPARQL queries some relevant information about scientific production. The presented results allow to observe the behavior of the scientific production in the analyzed institution and through statistics and graphic representations it is easier to make the decisions corresponding to the production. The shown SPARQL queries are the active researchers, the multidisciplinary publications, the quartiles of citations and the researchers who work in collaboration with other ones that are attached to the same institution. It is intended to establish rules and restricted classes in the ontology to facilitate the representation of the problem and infer instances through an ontological reasoner.

Index Terms—Scientific production ontology, SPARQL, Scientific relevance, Digital repositories.

I. INTRODUCCIÓN

LOS investigadores realizan diversas actividades cuyos resultados se exponen a la comunidad científica por medio de publicaciones, y éstas pueden ser artículos de revistas, libros, capítulos de libro, ponencias, entre otras.

Además, pueden recibir estímulos económicos, apoyos para la investigación y otros beneficios al ser evaluados por su propia institución o alguna externa. Dichas evaluaciones incluyen indicadores cuantitativos para la producción científica utilizando como base las publicaciones correspondientes de la persona correspondiente.

En México, algunas instancias evaluadoras de la producción científica son el Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP). El SNI fue creado para reconocer los diversos trabajos hechos por personas dedicadas al ámbito científico y tecnológico desempeñado en México, sin importar la nacionalidad [1]. El PRODEP es el encargado de profesionalizar a los profesores de tiempo completo en

actividades de investigación-docencia, desarrollo tecnológico e innovación para la consolidación de grupos de investigación, denominados cuerpos académicos, con la finalidad de generar una comunidad científica activa [2].

Entre los indicadores se encuentran el factor de impacto de la revista, el índice de citas, el impacto de la investigación, el índice h y el factor de prestigio de la revista [3], [4]. Algunas de estas publicaciones se encuentran en repositorios de acceso abierto, los cuales permiten ingresar al documento, además, presentan la información descriptiva (metadatos) relacionada de cada recurso. Otras publicaciones se encuentran en repositorios restringidos y el acceso únicamente se permite realizando el pago de una suscripción o perteneciendo a una institución educativa o científica. Sin embargo, repositorios como Scopus [5] permiten consultar y recuperar los metadatos del recurso. Como los indicadores requieren únicamente la información respecto a la producción y no el documento en sí, es posible calcular sus valores recuperando la información necesaria a través de los repositorios de acceso abierto, siempre y cuando dispongan de una herramienta que permita la solicitud de peticiones para la recuperación de la información, como el protocolo para la interoperabilidad entre repositorios OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), el cual consiste en seis verbos para la recuperación de los metadatos de los recursos y del repositorio [6]:

- 1) *GetRecord*. Recupera el registro individual de un recurso del repositorio.
- 2) *Identify*. Recupera la información acerca del repositorio.
- 3) *ListIdentifiers*. Recupera únicamente los identificadores de los recursos.
- 4) *ListMetadataFormats*. Recupera los formatos de metadatos disponibles en el repositorio.
- 5) *ListRecords*. Recupera los metadatos de los recursos.
- 6) *ListSets*. Recupera los grupos en los cuales se encuentran organizados los recursos del repositorio.

Una ontología es una especificación formal y explícita de una conceptualización compartida [7] utilizando conceptos en un dominio del discurso, propiedades de cada concepto describiendo sus características y atributos, además de las restricciones en las propiedades [8]. Para que una ontología pueda ser compartida y reutilizada es elemental que se ponga a disposición a través de la Web, dando lugar al término Web semántica, cuya finalidad es la integración y combinación de datos extraídos de diversas fuentes y el uso del lenguaje para el registro de la relación entre datos en el mundo real para su publicación por medio de diversas aplicaciones a través de los metadatos semánticos [9], [10]. Las ontologías pueden utilizar

J. Guerrero-Sosa, Universidad Autónoma de Yucatán, Jaredgs93@gmail.com.

V. Menéndez-Domínguez, Universidad Autónoma de Yucatán, mdoming@correo.uady.mx.

M. Castellanos-Bolaños, Universidad Autónoma de Yucatán, enriqueta.c@correo.uady.mx.

J. Gómez-Montalvo, Universidad Autónoma de Yucatán, jgomez@correo.uady.mx.

elementos o la estructura de otras que cumplen con la representación que se desea, manteniendo la consistencia del mundo modelado. Para que una ontología pueda ser compartida y reutilizada es elemental que se ponga a disposición a través de la Web, debido a que pasó de ser un repositorio de texto e imágenes a ser un proveedor de servicios y de información con diversos fines [11], dando lugar al término Web semántica, cuya finalidad es la publicación de datos para ser utilizados por diversas aplicaciones por medio de los metadatos semánticos.

Las ontologías se han implementado en diversos ámbitos de estudio, como en el uso de un algoritmo de similitud que calcula el porcentaje de adherencia de los profesores de una universidad utilizando una ontología de conceptos [12]; para crear recomendaciones de mejores prácticas basadas en elementos empresariales para la mejora de procesos de software [13]; en el proceso de generación y validación para modelos de la técnica de Diseño de Software, Arquitectura dirigida por modelos [14] y para el cálculo de similitud entre documentos de tesis [15].

Para que exista coherencia y consistencia entre los datos que se comparten en la Web semántica se utilizan estándares como XML (lenguaje para la descripción de recursos por medio de etiquetas) [16], RDF (formato de grafos dirigido y etiquetado) [17] y OWL (Ontology Web Language), siendo el tercero de interés para las ontologías, ya que es el lenguaje estandar aceptado por la World Wide Web para publicar y compartir ontologías por medio de la Web [18]. Lo que distingue a OWL sobre XML y RDF es que estos últimos se enfocan en ser únicamente un formato de mensaje, mientras que OWL representa conocimiento. Aún así, la sintaxis normativa de OWL es RDF/XML.

Existe la semántica entre OWL y RDF, ya que OWL es una extensión de RDF, lo que significa que cualquier grafo RDF forma una ontología completa RDF [18]. Esto hace posible la realización de las consultas SPARQL (Simple Protocol And RDF Query Language) desde diversos orígenes de datos RDF con su propia sintaxis y semántica [17]. Este tipo de consultas incluyen conjuntos de patrones de tripletas llamadas patrón de grafo básico, los cuales extraen subgrafos de nodos relacionados de grafos en RDF [19]. Los subgrafos que devuelven la consulta SPARQL siguen la estructura *sujeto-predicado-objeto*.

El objetivo de este trabajo es presentar un modelo ontológico para la representación de la producción científica de un investigador, y a partir de ello generar consultas SPARQL que generen indicadores que resultan relevantes para el reconocimiento de la calidad investigadora de un individuo o una institución.

El contenido del artículo está organizado de la siguiente manera: en la sección I se presentó la introducción a la problemática identificada y las temáticas afines. En la sección II se muestra la arquitectura del sistema que permite recuperar los datos para la representación de la producción científica. En la sección III se describe la metodología utilizada, misma que se utiliza en la sección IV. En la sección V se describen las consultas realizadas a partir de la ontología y los resultados obtenidos. Finalmente, en la sección VI se indican las conclusiones y sugerencias para trabajo a futuro.

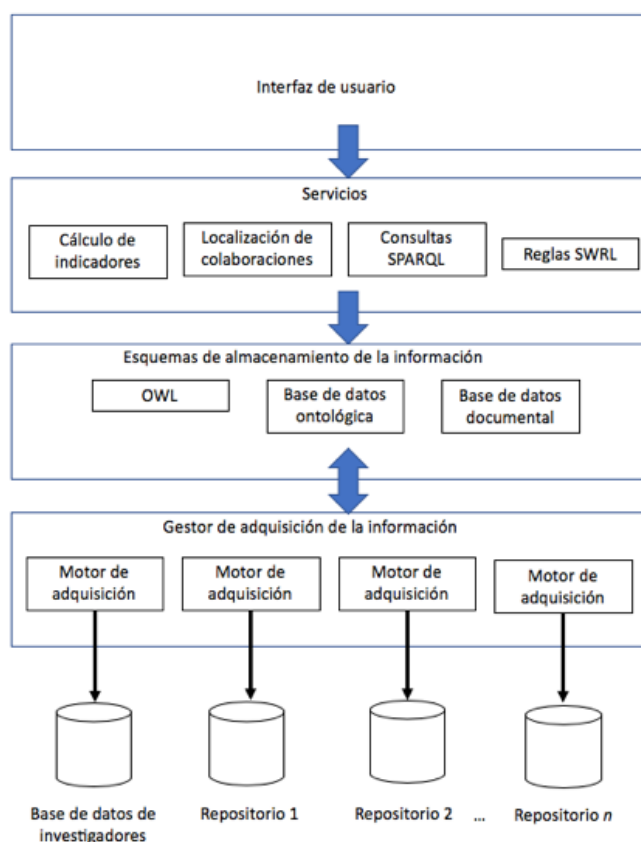


Fig. 1. Arquitectura del sistema.

II. ARQUITECTURA DEL SISTEMA

En la Figura 1 se presenta la arquitectura del sistema que propone una solución para la recuperación de la información de los investigadores con su respectiva producción científica a partir de los repositorios digitales. La arquitectura se ha diseñado en capas, facilitando el flujo de la información entre ellas y de manera ordenada. Cada capa tiene una función específica.

La capa de la *interfaz del usuario* presenta la información obtenida por los servicios a partir de los datos obtenidos. El sistema presenta gráficos interactivos como grafos y mapas de árbol que resumen la información relacionada con la producción académica. En la Figura 2 se muestra la distribución de la relevancia de los grupos de investigación de la Universidad Autónoma de Yucatán.

La capa de *servicios* se compone de módulos que trabajan en conjunto para la representación de los resultados por medio de la interfaz del usuario y entre sus funciones se encuentran el cálculo de indicadores, la localización de colaboraciones entre investigadores, las consultas SPARQL y las reglas SWRL (Semantic Web Rule Language), que son reglas que utilizan la información de la ontología en formato OWL definidas por una serie de antecedentes (condiciones a cumplir), y el consecuente, que es el resultado [20]. Las consultas SPARQL se encargan de obtener información relevante de los datos almacenados en la ontología. Las reglas SWRL filtran las instancias que cumplan con las condiciones establecidas por

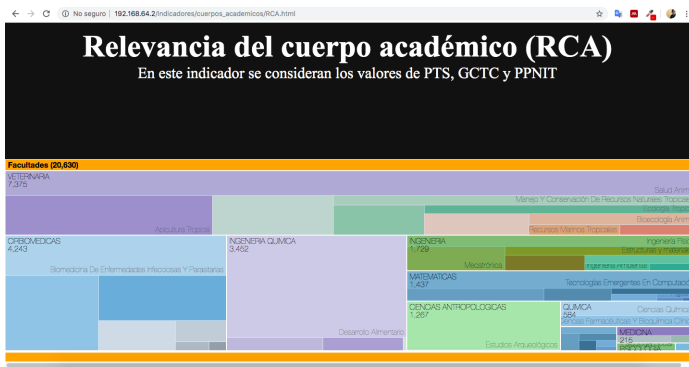


Fig. 2. Visualización de la relevancia de los grupos de investigación en el interfaz del usuario.

la ontología.

La capa de *esquemas de almacenamiento de la información* contiene las herramientas que se encargan de almacenar la información de los investigadores y de su producción como OWL, una base de datos ontológica y una base de datos no relacional.

La capa *gestor de adquisición de la información* contiene motores de solicitud que se conectan directamente a diversos repositorios por el medio correspondiente, ya sea a través de la cosecha de metadatos con OAI-PMH o alguna API, si el repositorio lo requiere.

Finalmente, el origen de los datos son los repositorios que contienen información acerca de la producción académica y científica de los investigadores.

III. METODOLOGÍA

Cristiani y Cuel [21] realizaron un estudio acerca de las diversas metodologías existentes para el desarrollo de una ontología. Entre las que presentan, se encuentran DOLCE [22], Ontology Development 101 [8], OTK Methodology [23], Methontology [24], DILIGENT Methodology [25], entre otras.

En el presente documento se propone utilizar Ontology Development 101. Si bien Noy y McGuinness [8] reconocen que no existe una metodología específica para el diseño de una ontología, creen necesario que seguir los pasos que se presentan a continuación, facilitan la tarea y se evitan interpretaciones inadecuadas de la información representada. Ontology Development 101 se divide en dos partes: la primera se enfoca en realizar la ontología a través del diseño y especificaciones de las clases, propiedades y restricciones. La segunda parte se encarga de asegurar que el diseño realizado en la primera parte sea consistente y no presente anomalías para crear instancias y realizar inferencias.

1) Primera fase.

- Determinar el dominio y alcance de una ontología. Se basa en cuatro preguntas, cuyas respuestas cambiarán a lo largo del desarrollo de la ontología: *¿cuál es el dominio a cubrir con la ontología?*, *¿qué uso se dará a la ontología?*, *¿qué tipo de preguntas debe responder la ontología?* y *¿quién usará y dará mantenimiento a la ontología?*

- Considerar reutilizar las ontologías existentes. Consiste en realizar una revisión de las ontologías que resuelvan o se aproximan a la solución del modelado que se plantea y utilizar los elementos que resulten útiles.
- Enumerar términos importantes en la ontología. Se puede considerar los términos de los que se desea hablar, las propiedades que dichos términos tienen, lo que se puede decir de los mismos, entre otras cosas.
- Definir clases y jerarquías. Se puede realizar iniciando con la definición de los conceptos más generales en el dominio y realizar las especificaciones correspondientes.
- Definir las propiedades de las clases. Como lo que se ha especificado no brindará información suficiente para las preguntas del paso 1, es necesario describir la estructura de los conceptos. Para ello, existen algunos tipos de propiedades de objeto para la ontología y son intrínsecas, extrínsecas, y las relaciones con los otros individuos.
- Definir las facetas de las propiedades. Es elemental definir el tipo de valor que contendrán las propiedades (tipo valor como *string*, *double*, *int* o tipo *objeto*), los valores permitidos, la cardinalidad y el dominio y rango.
- Crear instancias. A partir de lo establecido en los pasos anteriores, se crean instancias individuales.

2) Segunda fase

- Asegurar que la jerarquía de clases es la correcta.
- Analizar clases hermanas en la jerarquía.
- Analizar la herencia múltiple. Considerar aspectos relacionados con las propiedades, restricciones y relaciones que una subclase puede tener y una superclase no.
- Definir si es necesario definir una nueva clase o propiedad de acuerdo con las instancias y la relevancia de la representación a partir de la clase o propiedad.
- Decidir si definir una instancia o una clase. Consiste en especificar cuáles son los elementos más específicos que se desean representar en la base del conocimiento.
- Limitar el alcance de la ontología. Se revisa un par de reglas: no especializar más de lo que se requiere para una aplicación y establecer relaciones entre conceptos que ya se han incluido en la ontología.

IV. IMPLEMENTACIÓN

De acuerdo con la arquitectura del sistema propuesto, los servicios de consultas SPARQL y reglas SWRL se relacionan directamente con la ontología. En este caso, el enfoque va a las consultas SPARQL. La ontología se diseñó con el programa Protégé, que es un editor y marco de ontologías de código abierto y gratuito para construir sistemas inteligentes [26]. Se crearon las clases, las propiedades y sus respectivas características. En la Figura 3 se presentan las clases y subclases

(donde *isa* significa que una clase es subclase de otra) que representan las publicaciones y sus tipos (libros, capítulos de libro y artículos de revista), investigadores e identificadores, tanto para las publicaciones (ISSN para artículos de revistas, ISBN para libros y capítulos de libro, y DOI, para cualquier tipo de publicación), como para los investigadores, utilizando los que considera el Consejo Nacional de Ciencia y Tecnología (CONACYT) en México [1] para el Sistema Nacional de Investigadores (SNI). Esto incluye el Curriculum Vitae Único (CVU), la Clave Única de Registro de Población (CURP) y el número de registro en el Instituto Nacional Electoral (INE).

En las Figuras 4, 5 y 6 se presentan las propiedades que tienen las instancias de las clases y el tipo de valor de cada una.

En la Figura 4, se observa que las clases *Libro* y *Articulo* son clases heredadas de *Publicacion*, debido a que todo tipo de publicación cuenta con título, fecha de publicación, DOI (al tratarse de publicaciones disponibles en formato digital), citas recibidas y otros indicadores. En el caso de los capítulos de libro, estos heredan la subclase *Libro*, considerando que es diferente el título del capítulo al del libro, además de la necesidad de la distinción del autor y el editor del libro.

En la Figura 5 se presentan los identificadores considerados para su representación en la ontología. A pesar de contener instancias de objetos de diferente tipo (investigadores y publicaciones), se optó por crear una clase padre *Identificador* que tiene una sola propiedad denominada *tieneValor*. Dicha clase cuenta con dos subclases: *InvestigadorIdentificador* y *PublicaciónIdentificador*. La primera cuenta con tres subclases: *CVU*, *CURP* e *INE*, pero heredan la propiedad de su clase padre *InvestigadorIdentificador*, *esIdentificadorDeInvestigador*. Se definen las tres subclases con la finalidad de distinguir el tipo de identificador, además de que permite la creación de propiedades para estos identificadores en el futuro.

La subclase *PublicacionIdentificador* cuenta con una propiedad, *esIdentificadorDePublicacion*, la cual tiene como rango el identificador de la publicación a la que se encuentra asociada. Las tres subclases son *ISSN*, *DOI* e *ISBN* y cuentan con una propiedad para asociar al identificador con la publicación correspondiente.

En la Figura 6 se expone la estructura de la clase *Investigador* y sus respectivas propiedades. Entre los datos que representan a los investigadores se encuentran: información básica (nombre, institución, área de conocimiento y género), algunos indicadores (índice de citación y relevancia), instancias mexicanas que evalúan su relevancia científica (SNI [1] y PRODEP [2]) y otros asociados a su producción científica (publicaciones y colaboraciones).

Como caso de estudio se crearon instancias a partir de la producción de profesores investigadores de la Universidad Autónoma de Yucatán (UADY) utilizando el Repositorio Digital de la misma mediante el protocolo OAI-PMH para recolectar sus metadatos, y la API de Scopus para recuperar la producción científica indizada. La Universidad Autónoma de Yucatán es la más importante universidad pública del sureste de México. Cuenta con más de 25,000 estudiantes distribuidos en 15 facultades que imparten 45 programas educativos de licenciatura y 63 programas educativos de posgrado [27].

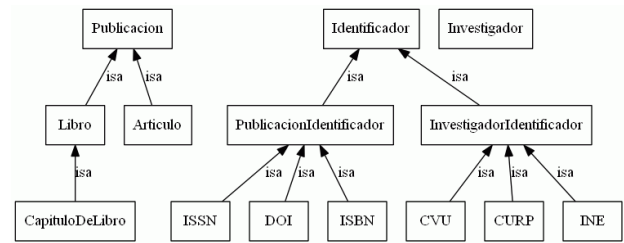


Fig. 3. Jerarquía de clases del modelo ontológico.

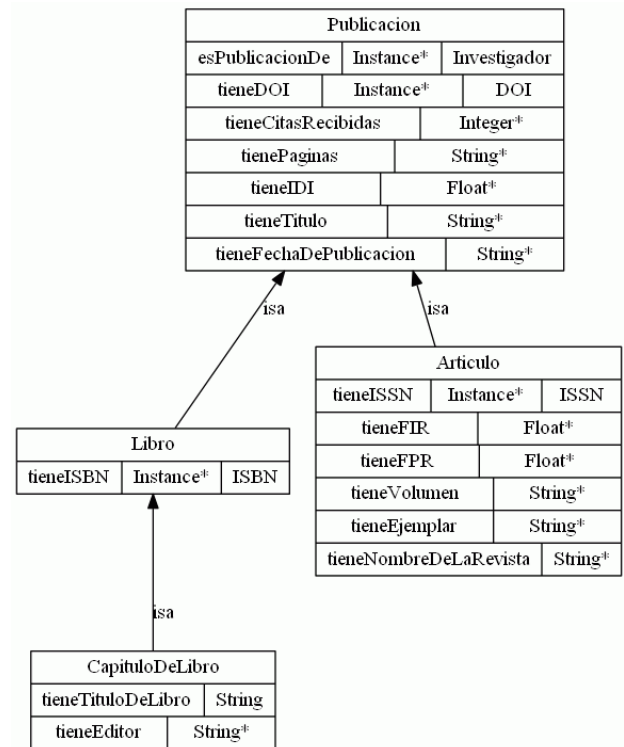


Fig. 4. Propiedades de las clases de las publicaciones.

Los datos recuperados de ambos repositorios se almacenaron en MongoDB [28], que es una base de datos documental que mantiene sus registros en formatos JSON, donde cada documento puede tener diferentes campos y las estructuras de datos son modificables.

Cabe resaltar que las propiedades de la ontología se diseñaron de acuerdo con la estructura de los documentos de la base de datos en MongoDB con la finalidad de mantener consistencia en los datos. Para crear las instancias de las clases de la ontología se diseñó un script en Python que tiene la función de utilizar los datos almacenados en los documentos de MongoDB y enviarlos a la ontología, misma que se almacenó en formato .OWL. Incluso al abrir el archivo con dicha extensión en Protégé es posible visualizar las instancias creadas, como se presenta en la Figura 7, así como los valores de sus propiedades, como se muestra en la Figura 8.

Por otra parte, la ontología se almacenó en una base de datos de tripletas llamada Virtuoso [29] que contiene un motor de consultas SPARQL, que permitieron presentar algunos datos relevantes acerca de la producción científica en la UADY.

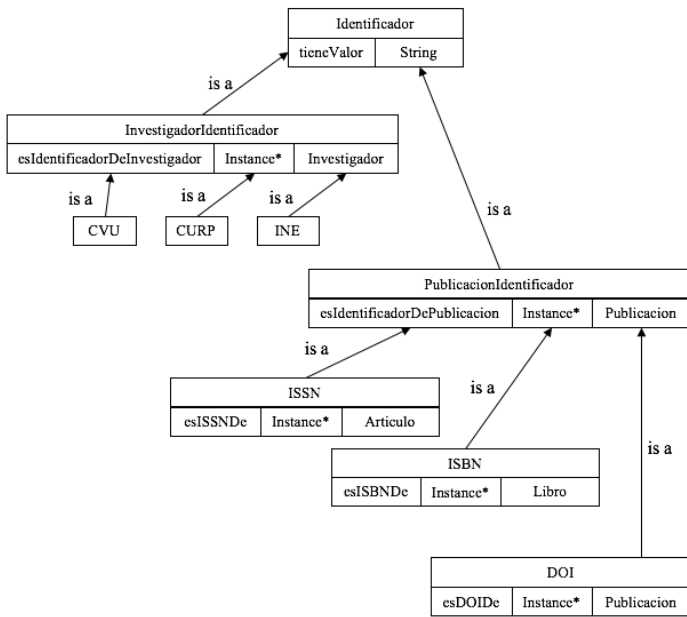


Fig. 5. Propiedades de las clases de los identificadores.

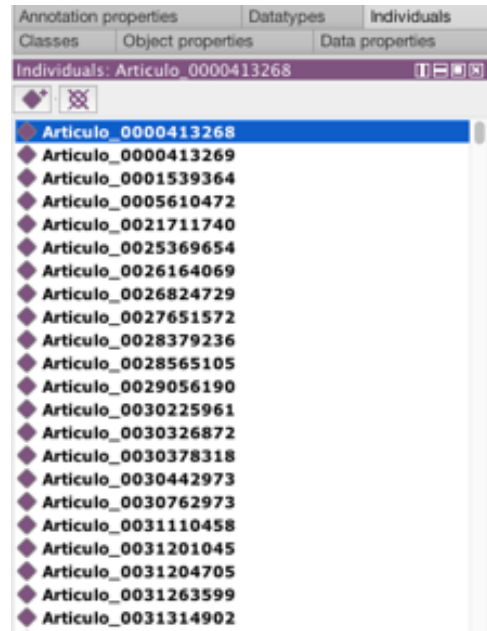


Fig. 7. Ejemplo de visualización de instancias en Protégé.

Investigador		
tieneInstitucion	Instance*	Institucion
tienePublicacion	Instance*	Publicacion
tieneInvestigadorIdentificador	Instance*	InvestigadorIdentificador
tieneNombre		String*
tieneIndiceDeCitacion		Integer*
investigadorEsRelevante		Boolean*
tieneAreaDeConocimiento		String*
tienePRODEP		Boolean*
tieneSNI		String*
tieneGenero		String*
tieneTituloAcademico		String*
tieneApellidos		String*
colaboraCon	Instance*	Investigador

Fig. 6. Propiedades de la clase Investigador.

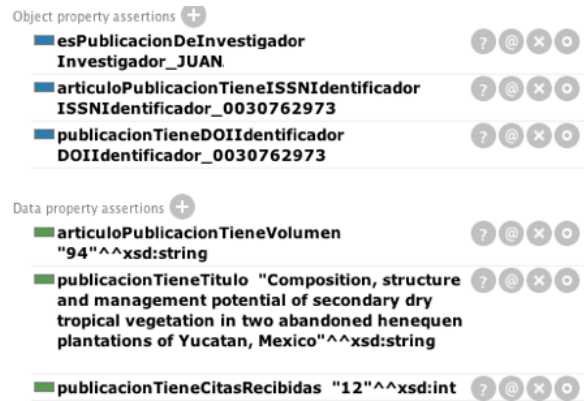


Fig. 8. Ejemplo de visualización de las propiedades de una instancia en Protégé.

Los resultados de la consulta se presentan en una interfaz del usuario, tal como se observa en la Figura 9.

V. RESULTADOS

Se diseñaron 7 consultas, las cuales se realizaron a través del motor de consultas SPARQL:

A. Investigadores Activos

La primer consulta consiste en presentar a los investigadores que tienen al menos tres publicaciones en los últimos tres años. Los resultados que devuelve la consulta incluyen el nombre del investigador, el nivel del SNI al que pertenece (puede ser candidato, nivel 1, 2 ó 3, siendo el 3 el reconocimiento máximo por parte del CONACYT), si cuenta con el perfil de profesor de tiempo completo (PRODEP) y el número de publicaciones realizadas en el lapso indicado. En la Tabla I se presenta parte de los resultados.

Consulta	Descripción
Investigadores activos	Todos los investigadores que tienen al menos tres publicaciones en los últimos tres años
Citas por área de conocimiento	Son las áreas de conocimiento con su respectivo índice de citación
Publicaciones multidisciplinarias	Son todas aquellas publicaciones cuyos autores pertenecen a por lo menos 3 áreas de conocimiento diferentes
Primer cuartil de citas	Son los investigadores que el total de sus citas es un valor entre el 75% y 100% respecto al índice de citación más alto
Segundo cuartil de citas	Son los investigadores que el total de sus citas es un valor entre el 50% y 74% respecto al índice de citación más alto
Tercer cuartil de citas	Son los investigadores que el total de sus citas es un valor entre el 25% y 49% respecto al índice de citación más alto
Cuarto cuartil de citas	Son los investigadores que el total de sus citas es un valor entre el 0% y 24% respecto al índice de citación más alto
Colaboradores internos	Son los investigadores de una institución que tienen publicaciones con otros investigadores pertenecientes a la misma institución

Fig. 9. Interfaz del usuario para las consultas SPARQL.

TABLA I
PARTE DE LOS RESULTADOS DE LOS INVESTIGADORES ACTIVOS

Clave	SNI	PRODEP	Publicaciones
Investigador 1	3	True	23
Investigador 2	2	True	20
Investigador 3	3	True	17
Investigador 4	3	True	16
Investigador 5	2	True	13

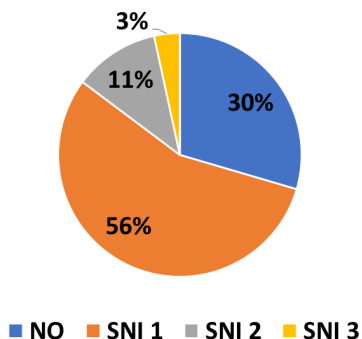


Fig. 10. Porcentaje de los investigadores activos considerando el nivel del SNI.

El total de resultados devueltos por la consulta es de 94 investigadores, y a partir de los mismos es posible obtener algunas estadísticas que resulten útiles, como el porcentaje de los investigadores activos considerando el nivel del SNI al que pertenecen. En la Figura 10 se puede observar que de los investigadores activos, el 30% no pertenecen al SNI, el 56% son nivel 1, el 11% son nivel 2 y el 3% son nivel 3.

B. Publicaciones Multidisciplinarias

Los resultados de esta consulta representan las publicaciones que involucran por lo menos tres áreas de conocimiento diferentes. Se consideran como áreas las siete establecidas por el CONACYT: Físico-Matemáticas y Ciencias de la Tierra; Biología, Química y Ciencias de la Vida; Medicina y Ciencias de la Salud; Humanidades y Ciencias de la Conducta; Ciencias Sociales; Biotecnología y Ciencias Agropecuarias; e Ingenierías [1]. La información que devuelve la consulta son el título de la publicación y el número de áreas de conocimiento involucradas. Parte de los resultados se presentan en la Tabla II.

TABLA II
PARTE DE LOS RESULTADOS DE LAS PUBLICACIONES MULTIDISCIPLINARIAS.

Publicación	Áreas involucradas
Publicación 1	6
Publicación 2	6
Publicación 3	6
Publicación 4	6
Publicación 5	6

El total de resultados devueltos por la consulta es de 556 publicaciones. A partir de esto es posible obtener el porcentaje por número de áreas de conocimiento involucradas, como se presenta en la Figura 11, donde se observa que de las

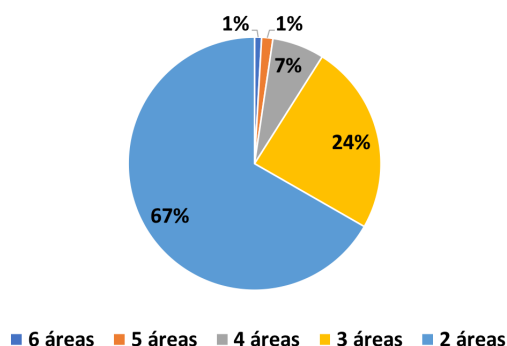


Fig. 11. Porcentaje de publicaciones multidisciplinarias por número de áreas.

publicaciones multidisciplinarias, el 1% involucra a 6 áreas de conocimiento, el 1% a 5 áreas, el 7% a 4 áreas, el 24% a 3 áreas y el 67% a 2 áreas.

C. Cuartiles de Citas

Los resultados representan a los investigadores cuyo número de citas recibidas es un valor entre el 75% y el 100% (primer cuartil, Tabla III), entre el 50% y el 74% (segundo cuartil, Tabla IV), entre el 25% y el 49% (tercer cuartil, Tabla V) y entre el 0% y el 24% (cuarto cuartil, Tabla VI) respecto al índice de citación más alto en la Universidad, que es 1266. Los resultados devuelven el nombre del investigador, su nivel del SNI, si pertenece al PRODEP y el número de citas recibidas.

TABLA III
RESULTADOS DEL PRIMER CUARTIL DE CITAS

Investigador	SNI	PRODEP	Citas recibidas
Investigador 1	3	False	1266
Investigador 2	3	True	1106

TABLA IV
RESULTADOS DEL SEGUNDO CUARTIL DE CITAS

Investigador	SNI	PRODEP	Citas recibidas
Investigador 1	3	True	889
Investigador 2	1	True	779
Investigador 3	3	True	768

TABLA V
RESULTADOS DEL TERCER CUARTIL

Investigador	SNI	PRODEP	Citas recibidas
Investigador 1	2	False	574
Investigador 2	2	True	532
Investigador 3	3	True	490
Investigador 4	2	True	411
Investigador 5	2	True	404
Investigador 6	2	True	363
Investigador 7	1	True	320

Respecto a los cuartiles, es posible generar una estadística considerando el porcentaje de investigadores por cuartil, como se observa en la Figura 12 donde el 1% pertenece al primer cuartil, el 1% al segundo cuartil, el 3% al tercer cuartil y el 95% al cuarto cuartil.

TABLA VI
PARTE DE LOS RESULTADOS DEL CUARTO CUARTIL DE CITAS

Investigador	SNI	PRODEP	Citas recibidas
Investigador 1	1	True	286
Investigador 2	1	True	282
Investigador 3	1	True	258
Investigador 4	2	True	252
Investigador 5	1	True	251

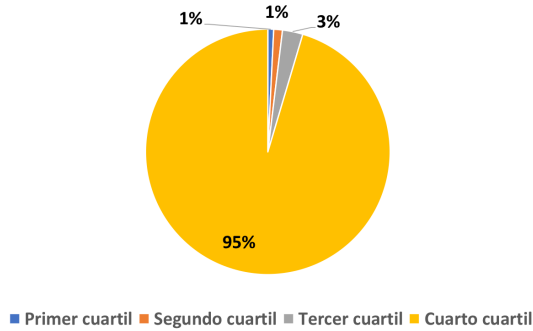


Fig. 12. Índice de citación por cuartiles.

D. Colaboradores Internos

La consulta representa a los investigadores que tienen publicaciones en conjunto con otros que pertenecen a la misma institución. Esto devuelve 173 resultados, representando el 28.78% de las instancias de los investigadores de la UADY considerando el nombre del investigador, su nivel del SNI, si pertenece al PRODEP y el número de colaboradores internos. En la Tabla VII se visualizan 5 de los 173 resultados.

TABLA VII
PARTE DE LOS RESULTADOS DE LOS COLABORADORES INTERNOS

Investigador	SNI	PRODEP	Colaboradores
Investigador 1	3	True	103
Investigador 2	2	True	73
Investigador 3	1	True	70
Investigador 4	2	False	65
Investigador 5	1	True	60

VI. CONCLUSIONES

Se observa que la mayoría de los investigadores activos pertenecen al nivel 1 del SNI, mientras que el 30% no cuenta con dicho reconocimiento. Lo mismo ocurre con el perfil PRODEP, donde la mayoría cuenta con el perfil. Ambas afirmaciones se han validado con las publicaciones de Scopus y la UADY, que previamente ha obtenido esta información con procedimientos realizados de forma manual. Respecto a las publicaciones multidisciplinarias, la mayoría involucra a dos áreas de conocimiento diferentes, lo cual representa que en la UADY existe colaboración entre expertos de diversas áreas para realizar trabajos que en conjunto han brindado resultados. Con relación a los cuartiles, la mayoría pertenece al cuarto cuartil, pero se debe a que el índice de citación mayor (1266) es muy alto con relación al promedio de citas (67) lo cual es normal debido a que los investigadores con nivel 3 del SNI (que es el más alto) tienen muy alta producción y únicamente

son el 2% de los investigadores de la UADY y se encuentran en el primer y segundo cuartil.

De acuerdo con los resultados obtenidos a través de las consultas SPARQL, se concluye que a partir de una ontología, que además de representar e inferir información de acuerdo con las restricciones establecidas, se puede obtener un panorama más amplio acerca de la producción científica de la investigación y generar estadísticas que permiten observar el estado actual de la investigación y localizar algunas fortalezas y debilidades en el área.

En algunos resultados se espera que los investigadores más reconocidos por la institución o por otra instancia (por ejemplo, SNI) sean activos, que pertenezcan a los cuartiles de citación más altos, o tengan un alto índice de colaboración. Sin embargo, en ocasiones se observa que incluso algunos investigadores que no tienen un nivel alto de reconocimiento se encuentran produciendo publicaciones que resultan relevantes, lo cual es un indicador útil para que reciba algún estímulo o beneficio económico para la realización de proyectos.

Como se mencionó con anterioridad, los resultados son óptimos ya que han sido validados por la institución estudiada y son similares a los obtenidos por la misma previamente pero a través de un procedimiento manual. El beneficio del modelo ontológico se encuentra en el diseño que proporciona utilidad a las consultas, ya que considera aspectos de evaluación por el Sistema Nacional de Investigadores (SNI) y el Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP), ambas en México, recuperando los datos para las instancias de la ontología desde diversos repositorios que estas mismas instituciones utilizan para sus procesos de evaluación y selección de investigadores reconocidos. Además, la arquitectura está diseñada de forma independiente a cualquier tecnología, lo cual facilita la flexibilidad del funcionamiento y modificación.

Se presentó la problemática y los temas relacionados, así como la propuesta de una arquitectura de un sistema para valorar la relevancia de la producción científica, misma que obtiene los datos a partir de repositorios de acceso abierto y de bases de datos de publicaciones indizadas.

Se expuso la metodología para el diseño de una ontología considerando la construcción y el mejoramiento de la misma. Posteriormente se mencionó la implementación, desde el uso de un software para la construcción de modelos ontológicos, los repositorios utilizados, las bases de datos, tanto documental como ontológica, mismas que dan lugar a la realización de las pruebas de las consultas SPARQL que tienen como finalidad presentar datos relevantes acerca de la producción científica de una institución y a partir de los resultados, generar estadísticas para facilitar la toma de decisiones.

A partir del modelo ontológico propuesto se está trabajando en la correlación con otras técnicas de medición considerando los niveles de producción, las colaboraciones, el impacto y prestigio de las publicaciones, y con base en ellos, establecer nuevas reglas y clases restringidas en la ontología que permitan inferir resultados útiles para una solución más amplia de la problemática.

AGRADECIMIENTOS

Este trabajo ha sido desarrollado gracias al apoyo del Consejo Nacional de Ciencia y Tecnología (CONACYT, México) a través de la beca con número (CVU/Becario): 853088/630948.

REFERENCIAS

- [1] CONACYT, “Archivo histórico,” 2018. [Online]. Available: <https://www.conacyt.gob.mx/index.php/el-conacyt/sistema-nacional-de-investigadores/archivo-historico>
- [2] DGESU, “Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP),” 2019. [Online]. Available: <http://www.dgesu.ses.sep.gob.mx/PRODEP.htm>
- [3] G. Buela-Casal, “Evaluación de la calidad de los artículos y de las revistas científicas: Propuesta del factor de impacto ponderado y de un índice de calidad,” *Psicothema*, vol. 15, no. 1, pp. 23–35, 2003.
- [4] O. P. Sharma, “Quality Indicators of Scientific Research,” *Indian Journal of Microbiology*, vol. 52, no. 2, pp. 305–306, 2012.
- [5] Elsevier, “The largest database of peer-reviewed literature-Scopus,” 2019. [Online]. Available: <https://www.elsevier.com/es-mx/solutions/scopus>
- [6] M. Nelson and S. Warner, “The Open Archives Initiative Protocol for Metadata Harvesting,” aug 2015. [Online]. Available: <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [7] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowledge Acquisition*, vol. 5, pp. 199–220, 1993.
- [8] N. F. Noy and Deborah L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology,” Stanford, California, Tech. Rep., 2001.
- [9] T. Berners-Lee and J. Hendler, “Publishing on the semantic web,” pp. 1023–1024, 2001.
- [10] W3C, “W3C Semantic Web Activity,” 2013. [Online]. Available: <https://www.w3.org/2001/sw/>
- [11] S. McIlraith, T. Son, and Honglei Zeng, “Semantic Web services,” *IEEE Intelligent Systems*, vol. 16, no. 2, pp. 46–53, mar 2001. [Online]. Available: <http://ieeexplore.ieee.org/document/920599/>
- [12] S. Freitas, E. Canedo, and D. Jesus, “Calculating similarity of curriculum lattes,” *IEEE Latin America Transactions*, vol. 16, no. 6, pp. 1758–1764, 2018.
- [13] E. Muñoz, M. Muñoz, E. García, and J. Mejía, “Knowledge management in process improvement and best practices sharing,” *IEEE Latin America Transactions*, vol. 12, no. 3, pp. 469–474, 2014.
- [14] N. Silega, M. Noguera, and D. Macías, “Ontology-based Transformation from CIM to PIM,” *IEEE Latin America Transactions*, vol. 14, no. 9, pp. 4156–4165, 2016.
- [15] R. Vences Nava, V. H. Menéndez-Domínguez, and J. R. Gómez-Montalvo, “A Document Recommendation System Using a Document-Similarity Ontology,” *IEEE Latin America Transactions*, vol. 14, no. 7, pp. 3329–3334, 2016.
- [16] L. Quin, “Extensible Markup Language (XML),” 2011. [Online]. Available: <https://www.w3.org/XML/>
- [17] E. Prud’hommeaux and A. Seaborne, “SPARQL Query Language for RDF,” 2008. [Online]. Available: <https://www.w3.org/TR/rdf-sparql-query/>
- [18] M. K. Smith, C. Welty, and D. L. McGuinness, “OWL Web Ontology Language Guide,” 2004. [Online]. Available: <https://www.w3.org/TR/2004/REC-owl-guide-20040210/>
- [19] G. Fletcher, J. Van Den Bussche, D. Van Gucht, and S. Vansummeren, “Towards a theory of search queries,” *ACM Transactions on Database Systems*, vol. 35, no. 4, 2010.
- [20] M. Horrocks, Ian Patel-Schneider, Peter F. Boley, Harold Tabet, Said Grosf, Benjamin Dean, “SWRL: A Semantic Web Rule Language Combining OWL and RuleML,” 2004.
- [21] M. Cristani and R. Cuel, “A Survey on Ontology Creation Methodologies,” *International Journal on Semantic Web and Information Systems*, vol. 1, no. 2, pp. 48–68, 2005. [Online]. Available: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jswis.2005040103>
- [22] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider, “WonderWeb Deliverable D17,” 2002.
- [23] Y. Sure, “Methodology, tools & case studies for ontology based knowledge management,” Ph.D. dissertation, Universität Friderician zu Karlsruhe, 2003.
- [24] M. Fernández, A. Gómez-Pérez, and N. Juristo, “Methontology: from ontological art towards ontological engineering,” in *Proceedings of the AAAI Spring Symposium on Ontological Engineering*, Stanford, California, 1997, pp. 33–40.
- [25] Y. Sure, C. Tempich, S. Pinto, and S. Staab, “A case study in supporting distributed, loosely-controlled and evolving engineering of ontologies,” in *Intelligent Learning Infrastructure for Knowledge Intensive Organizations: A Semantic Web Perspective*, M. Lytras and A. Naeve, Eds. IGI Global, 2005, ch. 14, pp. 357–368.
- [26] Stanford University, “protégé,” 2016. [Online]. Available: <https://protege.stanford.edu/>
- [27] UADY, “Universidad Autónoma de Yucatán,” 2019. [Online]. Available: <http://www.uady.mx/>
- [28] MongoDB, “¿Qué es MongoDB?” 2018. [Online]. Available: <https://www.mongodb.com/es/what-is-mongodb>
- [29] OpenLink, “OpenLink Virtuoso,” 2015. [Online]. Available: <https://virtuoso.openlinksw.com/>



Jared Guerrero-Sosa es Ingeniero en Tecnologías de la Información y Comunicaciones por el Instituto Tecnológico de Chetumal, México. Es estudiante de la Maestría en Ciencias de la Computación en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Sus intereses de investigación se enfocan en temas relacionados con los repositorios digitales, la representación del conocimiento y la Informática Educativa.



Víctor Menéndez-Domínguez es Doctor en Tecnologías Informáticas Avanzadas por la Universidad de Castilla-La Mancha, España. Es Profesor Titular en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con repositorios de documentos digitales, la representación del conocimiento y la gestión de Objetos de Aprendizaje.



María Castellanos-Bolaños es Maestra en Gestión de Tecnología de Información por la Universidad Anáhuac Mayab, México. Es Profesora de tiempo completo en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con la gestión del conocimiento y la Ingeniería Web.



Jorge Gómez-Montalvo es profesor e investigador en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Es Doctor en Redes, Telecomunicaciones, Sistemas y Arquitecturas por el Instituto Nacional de Ciencias Aplicadas de Toulouse, Francia. Su trabajo de investigación se centra en marcos de trabajo autónomos y semánticos para servicios distribuidos y sus aplicaciones para el desarrollo de nuevas tecnologías para el ambiente de vida asistida.