



**Extracción y enriquecimiento de perfiles de investigación
usando ontologías**

*Para obtener el grado de
Maestro en Ciencias de la Computación*

Presentado por

Ing. Isabel Cruz Ruiz

Asesores: Dra. Maricela Claudia Bravo Contreras
Dr. José Alejandro Reyes Ortiz

Enero 2019

Dedicado a:

Mis padres, María de la Luz Ruiz Trejo y Mario Cruz Gutiérrez, que me han apoyado en este difícil camino y han estado a mi lado apoyándome. Especialmente para ti mamita...TE AMO!

Carlos Javier Domínguez Cano, gracias por tu apoyo incondicional en estos dos largos años, tus palabras de ánimo, tu paciencia, por creer en mí y no dejarme rendir a pesar de las dificultades.

Mi hermana Chio, mi sobrina Payos, tía Ali, tías, tíos, primas y al resto de la familia por su apoyo y buenos deseos.

Mis amigas: Ximena Cerna Landa, Esther Delgado Castro, Miriam Roa porque también me apoyaron, me brindaron sus consejos y experiencias para seguir adelante.

A mis asesores, la Dra. Maricela Bravo Contreras y el Dr. Alejandro Reyes Ortiz, gracias por su experiencia, su trabajo, sus palabras, todo el apoyo que me brindaron en estos dos años de desvelos, esfuerzo y sobre todo, años llenos de satisfacciones.

Al coordinador de la Maestría, el Dr. Luis Fernando Hoyos Reyes, por todo el apoyo brindado.

Agradecimientos

Gracias a Dios por darme la oportunidad de vivir y experimentar esta etapa de mi vida.

Gracias a mis padres por el regalo de tener una carrera que amo.

Gracias a la Empresa THSD, especialmente a Roberto Guerrero, Israel Flores, Bruno Ríos, Esaú Maganda; a mis compañeros: José Granillo, Omar Guzmán, Citlali Martínez, Edwin Aguilar, Dany Correa, Irot Rivas, Fabiola Sánchez y Francisco García, gracias por su apoyo y sobre todo su paciencia en el tiempo que me tomó terminar la maestría, en verdad... muchísimas gracias!

Gracias a todos los profesores que me acompañaron en este tiempo de formación, compartiendo sus conocimientos y por todo el apoyo que me brindaron.

Lista de figuras

| | | |
|------|---|----|
| 2.1 | Representación de una ontología | 7 |
| 2.2 | Ejemplo de archivo XML | 10 |
| 2.3 | Ejemplo de etiquetas de un artículo DBLP | 13 |
| 2.4 | Segundo ejemplo de etiquetas de un artículo DBLP | 13 |
| 4.1 | Arquitectura de solución | 27 |
| 4.2 | Ontología de Persona (<i>Person</i>) | 31 |
| 4.3 | Tipos de personas que pueden ingresar | 31 |
| 4.4 | División de Empleado (<i>Employee</i>) y Estudiante (<i>Student</i>) | 32 |
| 4.5 | Representación de todas las clases de la ontología de Persona (<i>Person</i>) | 32 |
| 4.6 | <i>Data properties</i> de la Ontología de Persona (<i>Person</i>) | 35 |
| 4.7 | <i>Object properties</i> de la ontología de Persona | 36 |
| 4.8 | Ontología de Publicación (<i>Publication</i>) | 37 |
| 4.9 | Propiedades de datos de la ontología de Publicación (<i>Publication</i>) | 39 |
| 4.10 | Ontología de Perfil de investigación (<i>Researcher Profile</i>) | 40 |
| 4.11 | Clases de la ontología de Perfil de investigación (<i>Researcher Profile</i>) | 41 |
| 4.12 | <i>Data properties</i> de la ontología de Perfil de investigación (<i>Researcher Profile</i>) | 42 |
| 4.13 | <i>Object properties</i> de la ontología de Perfil de investigación (<i>Researcher Profile</i>) | 43 |
| 4.14 | Diagrama de flujo del proceso de poblado de la ontología de Persona | 47 |

| | | |
|------|---|----|
| 4.15 | Ejemplo de un individuo de la ontología de Persona (<i>Person</i>) | 48 |
| 4.16 | Ejemplo de un individuo de la ontología de Persona (<i>Person</i>) | 49 |
| 4.17 | Diagrama de flujo del método que regresó la coincidencia entre un autor de una Publicación y un Profesor | 51 |
| 4.18 | Diagrama de flujo del proceso de poblado de Publicación | 54 |
| 4.19 | Diagrama de flujo del método que instancia una Publicación | 55 |
| 4.20 | Ejemplo de un individuo de la ontología de Publicación (<i>Publication</i>) | 56 |
| 4.21 | Segundo ejemplo de un individuo de la ontología de Publicación (<i>Publi- cation</i>) | 56 |
| 4.22 | Ontología de Perfil de investigación (<i>Researcher profile</i>) | 57 |
| 4.23 | Representación de la relación entre ontologías (Persona y Publicación) | 58 |
| 4.24 | Ejemplo de relación de colaboración entre profesores | 60 |
| 4.25 | Segundo ejemplo de relación de colaboración entre profesores | 60 |
| 4.26 | Tercer ejemplo de relación de colaboración entre profesores | 60 |
| 4.27 | Diagrama de clases | 61 |
| 5.1 | Ejemplo de etiquetas de un archivo DBLP | 65 |
| 5.2 | Segundo ejemplo de etiquetas de un archivo DBLP | 66 |
| 5.3 | Respuesta a primer pregunta de competencia | 72 |
| 5.4 | Respuesta a segunda pregunta de competencia | 72 |
| 5.5 | Respuesta a tercer pregunta de competencia | 73 |
| 5.6 | Respuesta a cuarta pregunta de competencia | 73 |
| 5.7 | Respuesta a quinta pregunta de competencia | 74 |
| 5.8 | Respuesta a sexta pregunta de competencia | 74 |
| 5.9 | Respuesta a séptima pregunta de competencia | 75 |

Lista de tablas

| | | |
|-----|--|----|
| 3.1 | Tabla comparativa del Estado del arte. | 24 |
| 3.2 | Continuación de tabla comparativa del Estado del arte. | 25 |
| 4.1 | Clases principales de las ontologías. | 30 |
| 5.1 | Tabla de profesores por departamento. | 66 |
| 5.2 | Resultados del primer escenario. | 68 |
| 5.3 | Resultados del segundo escenario. | 69 |
| 5.4 | Resultados del tercer escenario. | 70 |
| 5.5 | Resultados del cuarto escenario. | 71 |
| 5.6 | Tabla de resultados primer escenario. | 75 |
| 5.7 | Tabla de resultados segundo escenario. | 75 |
| 5.8 | Tabla de resultados tercer escenario. | 76 |
| 5.9 | Tabla de resultados cuarto escenario. | 76 |

Tabla de contenido

| | |
|---|-----------|
| Lista de figuras | IV |
| Lista de tablas | VI |
| 1: Introducción | 1 |
| 1.1 Descripción del problema | 4 |
| 1.2 Objetivo general | 6 |
| 1.3 Objetivos específicos | 6 |
| 1.4 Organización del documento | 6 |
| 2: Marco teórico | 7 |
| 2.1 Representación de perfiles de investigación | 7 |
| 2.1.1 Elementos de cualquier ontología | 8 |
| 2.1.2 Componentes de cualquier ontología | 8 |
| 2.1.3 Perfil de investigación | 9 |
| 2.2 Lenguajes y modelos de representación de perfiles | 9 |
| 2.2.1 XML - <i>Extensible Markup Language</i> | 9 |
| 2.2.2 OWL - <i>Web Ontology Language</i> | 10 |
| 2.3 DBLP - Fuente de datos pública para la extracción de investi- gación | 11 |

| | |
|--|-----------|
| 3: Estado del arte | 14 |
| 3.1 Búsqueda personalizada basada en ontología. | 15 |
| 3.2 Sistema de búsqueda de experiencia de investigadores utilizando minería de datos en una ontología. | 15 |
| 3.3 La ontología y la arquitectura para una red social académica | 15 |
| 3.4 Búsqueda de investigadores similares en entornos académicos | 16 |
| 3.5 Mejora del perfil de usuario basado en ontología | 17 |
| 3.6 Un modelo basado en ontología para búsquedas y clasificaciones expertas | 17 |
| 3.7 Contexto y conocimiento del contexto de los humanos y los sistemas AmI | 18 |
| 3.8 Arquitectura basada en ontologías para la adaptación de servicios de contexto | 18 |
| 3.9 Modelo ontológico genérico de alto nivel para aplicaciones orientadas al contexto en ambientes inteligentes | 19 |
| 3.10 Modelo de perfil de usuario ontológico para la personalización de aplicaciones contextualizadas | 20 |
| 3.11 Una ontología para el cómputo en ambientes pervasivos conscientes de contexto | 20 |
| 3.12 Ontologías para administrar perfiles de usuario en la entrega personalizada de servicios móviles | 21 |
| 3.13 Una red social semántica de investigadores académicos | 21 |
| 3.14 ScholarLens: extracción de competencias de publicaciones de investigación para la generación automática de perfiles de usuario semánticos . . . | 22 |
| 3.15 Conclusión del estado del arte | 26 |
| 4: Metodología de solución | 27 |
| 4.1 Modelado de ontologías | 28 |
| 4.1.1 Ontología de Persona (<i>Person</i>) | 30 |
| 4.1.2 Ontología de Publicación (<i>Publication</i>) | 36 |
| 4.1.3 Ontología de Perfil de Investigador (<i>Researcher profile</i>) | 39 |
| 4.2 Poblado de ontologías | 43 |

| | | |
|-----------|--|-----------|
| 4.2.1 | Poblado de la ontología de Persona (<i>Person</i>) | 45 |
| 4.2.2 | Poblado de ontología de Publicación (<i>Publication</i>) | 49 |
| 4.2.3 | Poblado de ontología Perfil de investigación (<i>Researcher profile</i>) | 56 |
| 4.3 | Enriquecimiento de ontología Perfil de investigador (<i>Researcher profile</i>) | 57 |
| 4.4 | Diagrama de clases | 61 |
| 5: | Experimentación | 63 |
| 5.1 | Conjunto de datos | 63 |
| 5.1.1 | Estructura DBLP | 64 |
| 5.1.2 | Ejemplo de artículo | 65 |
| 5.2 | Plan experimental | 66 |
| 5.2.1 | Descripción de escenarios | 66 |
| 5.2.1.1 | Escenario 1.- Inserción de profesores del departamento de Sistemas en la ontología de Persona (<i>Person</i>) y una parte del archivo DBLP | 67 |
| 5.2.1.2 | Escenario 2.- Inserción de profesores del departamento de Sistemas en la ontología de Persona (<i>Person</i>) y archivo completo de DBLP | 68 |
| 5.2.1.3 | Escenario 3.- Inserción de profesores del departamento de Electrónica en la ontología de Persona (<i>Person</i>) y archivo completo de DBLP | 69 |
| 5.2.1.4 | Escenario 4.- Inserción de profesores del departamento de Sistemas y Electrónica en la ontología de Persona (<i>Person</i>) y archivo completo de DBLP | 70 |
| 5.3 | Evaluación del modelo ontológico | 71 |
| 5.4 | Análisis de resultados | 75 |
| 5.4.1 | Escenario 1 | 75 |
| 5.4.2 | Escenario 2 | 75 |
| 5.4.3 | Escenario 3 | 76 |
| 5.4.4 | Escenario 4 | 76 |

| | |
|---|------------|
| 6: Conclusión y trabajo a futuro | 77 |
| Anexos: Artículos publicados | 79 |
| Referencias | 119 |

1 Introducción

En los últimos años, las ontologías se han vuelto extremadamente populares como un medio para representar el conocimiento de forma semántica y legible (o interpretable) por la máquina. El rápido crecimiento de la Web y el problema de sobrecarga de información que ha causado, ha desencadenado una investigación significativa en el desarrollo de soluciones prácticas de extracción de información, las cuales procesan el contenido de la Web. Sin embargo, la dificultad de extraer información de la Web, que se produjo principalmente para visualizar información, ha impulsado el nacimiento de la Web Semántica, la cual tiene como objetivo adjuntar información semántica a los recursos e información en la Web para que éstos sean interpretables por la máquina [1].

Actualmente, para tener acceso a información, buscar o localizar datos o documentos sobre un tema específico, se recurre a los buscadores tradicionales como *Google*. De igual manera, cuando un investigador busca información sobre algún tema de investigación en particular, realiza búsquedas en bases de datos de bibliografía científica, lo que resulta en una tarea complicada, tediosa y tardada ya que toma mucho tiempo en organizar la cantidad de información que le devuelven los buscadores. Además del tiempo requerido para buscar, seleccionar y organizar la información, el hecho de que el investigador haya encontrado ciertos resultados no le garantiza que la información seleccionada (aun después de analizarla) sea relevante para su investigación.

El problema de que la información recuperada a través de los buscadores no sea relevante con respecto a los intereses del investigador, tiene diferentes causas:

1. La información académica no se encuentra representada de manera semántica lo que imposibilita su inferencia y enriquecimiento.
2. Los buscadores no consideran el perfil de investigación del usuario que está realizando la búsqueda.

3. Los buscadores utilizan algoritmos basados en enfoques sintácticos y no semánticos.
4. La forma como se recupera y organiza la información está basada principalmente en técnicas de minería de datos o bases de datos.

En este documento se presenta una solución basada en el uso de ontologías para la representación de perfiles de investigación de académicos y profesores de la Universidad Autónoma Metropolitana - unidad Azcapotzalco, así como la incorporación de métodos de extracción y enriquecimiento de información. El modelo ontológico desarrollado permite que la información de los perfiles de investigación se almacene y se enriquezca a través de la ejecución de reglas de inferencia, generando información que es de interés para los investigadores.

En la actualidad existen diversas definiciones de lo que una ontología significa, de acuerdo con Tom Gruber (Gruber, 1993) [2], una ontología es una especificación explícita de una conceptualización. Es una descripción formal de los conceptos y las relaciones entre conceptos. Puede decirse que define los términos básicos y sus relaciones a partir del vocabulario, y las reglas para su combinación.

Una ontología es un “catálogo” de los tipos de entidades que existen en un dominio bajo la perspectiva de una persona que utiliza un lenguaje para describir el dominio en cuestión (Sowa, 2001) [3]. Los tipos o categorías de una ontología representan los predicados, es decir, los conceptos que esas categorías incluyen o abarcan y las relaciones establecidas para discutir los temas del dominio.

También puede decirse que una ontología [4] es un instrumento para representar el conocimiento dentro de un área determinada y que permite recuperar ese conocimiento por medios informáticos.

Una ontología describe formalmente conceptos generales o sobre un dominio; consiste en definir una lista de términos y las relaciones básicas entre ellos. Los términos denotan conceptos importantes y las relaciones típicamente incluyen jerarquías de clases (Antoniou & Van Harmelen, 2008 [5]).

Es decir, las ontologías son herramientas para especificar claramente los conceptos de un dominio concreto, sus propiedades y sus relaciones; permitiendo que la información publicada se encuentre en formatos inteligibles para que agentes máquinas puedan localizar y gestionar de forma precisa la información [6].

En el presente trabajo se utilizan las ontologías para la representación de perfiles de investigación.

Un perfil de investigación se refiere a la información relevante sobre la formación académica, como cantidad de publicaciones en distintos años, título académico, actividades en redes sociales académicas, habilidades, especialidades e intereses de investigación de cada individuo. En este trabajo se aborda el perfil de investigación de personal académico, el cual se refiere a la experiencia y conocimientos reflejados a través de las publicaciones científicas.

Se utilizó el lenguaje XML (por sus siglas en inglés *Extensible Markup Language*), el cual es un metalenguaje que permite a los usuarios definir lenguajes de marcado personalizados y contiene símbolos de marcado para describir el contenido de una página o archivo [7].

Durante el diseño de esta ontología se tomaron en cuenta cinco cuestiones clave:

- Claridad: una ontología debe comunicar de manera efectiva el significado de sus términos. Las definiciones deben ser objetivas y explicarse en lenguaje natural [8].
- Coherencia: una ontología debe permitir hacer inferencias que sean consistentes con las definiciones [8].
- Extensibilidad: se anticipa un nuevo uso de la ontología para permitir extensiones y especializaciones [8].
- Especificidad: se debe especificar a nivel de conocimiento, sin que dependa de una codificación particular a nivel de símbolo [8].
- Precisión: debe hacerse la menor cantidad de “suposiciones” acerca del mundo modelado [8].

1.1. Descripción del problema

Actualmente, en todo dominio académico se realizan búsquedas de información de muchas formas; ya sea usando computadoras de escritorio o portátiles, desde un teléfono inteligente o una tablet, por citar algunos ejemplos. Los buscadores comunes devuelven la información solicitada considerando principalmente las palabras claves introducidas en el buscador, la cual se puede ir refinando o mejorando de acuerdo a la interacción del usuario con el buscador. Sin embargo, ningún buscador actualmente incorpora en la búsqueda el perfil del usuario que está realizando la búsqueda. En particular, en un dominio académico, los perfiles de los usuarios pueden ser de estudiantes, trabajadores o perfiles de investigadores. Esta propuesta se centra en los perfiles de los investigadores, quienes al realizar búsquedas sobre algún tema de investigación en particular, utilizan bases de datos de bibliografía científica.

Este tipo de información, sobre investigaciones y datos académicos, se encuentra en la Web, pero carece de una estructura semántica que sea procesable por computadoras, para hacer posible la creación de buscadores semánticos. Esto representa un reto para el área de extracción de información y representación de conocimiento.

Algunos problemas que surgen cuando la información académica no está estructurada son los siguientes:

1. Los perfiles de los investigadores no se encuentran disponibles debido a que la información académica no está organizada.
2. Los buscadores no consideran el perfil de investigación del usuario que está realizando la búsqueda.
3. Los buscadores utilizan algoritmos basados en enfoques sintácticos y no semánticos.
4. La forma como se recupera y organiza la información está basada principalmente en técnicas de minería de datos o bases de datos.

Para resolver el problema es necesario construir la infraestructura computacional que, a futuro, permita realizar búsquedas centradas en el perfil de los investigadores y que la información devuelta sea realmente lo que el usuario está buscando. En este proyecto se busca crear dicha infraestructura semántica. La necesidad de contar con este tipo de infraestructura trae consigo diversos aspectos a considerar:

1. Representar perfiles de investigación mediante modelos que otorguen una estructura semántica la información.
2. Automatizar la extracción y almacenamiento de los perfiles de investigación.
3. Implementar mecanismos que permitan aumentar y mejorar la información almacenada.

1.2. Objetivo general

Diseñar e implementar una metodología para la extracción de información académica y enriquecimiento de perfiles de investigadores usando ontologías.

1.3. Objetivos específicos

Diseñar e implementar:

- Modelo basado en ontologías para la representación de perfiles de investigación.
- Método para la extracción de información académica a partir de datos (semi) estructurados.
- Método para realizar el enriquecimiento de la información en los perfiles de investigación.

1.4. Organización del documento

El resto del presente trabajo está organizado de la siguiente manera:

- Capítulo 2 - Se presenta el marco teórico, es decir, se detalla la terminología de las características principales que se utilizaron en el presente trabajo.
- Capítulo 3 - Resumen de trabajos que utilizan una o más de las técnicas y/o métodos que son similares a los utilizados para realizar este proyecto.
- Capítulo 4 - Contiene la metodología de solución, es decir, se describe el modelado de las ontologías utilizadas para solucionar el problema principal, el poblado de las mismas y el enriquecimiento de la ontología de Perfiles de investigación.
- Capítulo 5 - En este capítulo se describe la experimentación, es decir, la explicación sobre el conjunto de datos utilizados y la evaluación realizada a las ontologías.

Finalmente se abordan las conclusiones, trabajo a futuro y la bibliografía que se utilizó como apoyo para la realización del presente trabajo.

2 Marco teórico

Esta sección incluye las definiciones principales que se utilizarán en la representación de perfiles de investigación.

2.1. Representación de perfiles de investigación

De acuerdo con Tom Gruber (Gruber, 1993) [2], en el contexto de las ciencias de la computación, una ontología define un conjunto de representaciones, las cuales son clases (o conjuntos), atributos (o propiedades) y relaciones (entre miembros de una clase), con las que se modela un dominio de conocimiento o discurso. En la figura 2.1 se muestra un ejemplo de representación de una ontología.

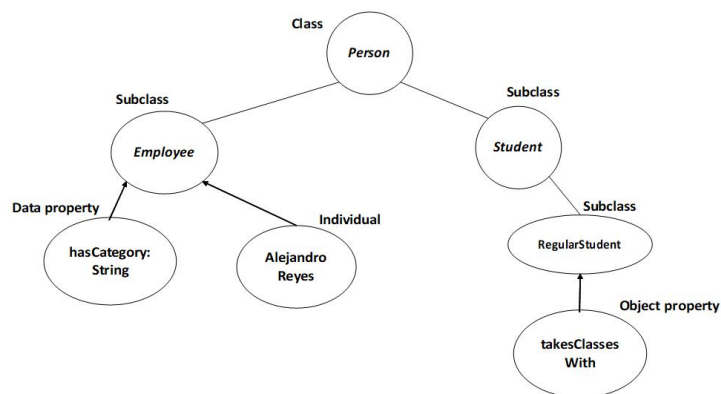


Figura 2.1: Representación de una ontología

Las definiciones de las representaciones incluyen información acerca de su significado y restricciones sobre su aplicación lógicamente consistente. En el contexto de los sistemas de bases de datos, la ontología puede ser vista como un nivel de abstracción de modelos

de datos, análogo a los modelos jerárquicos y relacionales, pero destinado a modelar el conocimiento sobre los individuos, sus atributos y sus relaciones con otros individuos.

Las ontologías se especifican utilizando lenguajes formales que permiten la abstracción de estructuras de datos y estrategias de implementación. En la práctica, los lenguajes de las ontologías son más cercanos en poder expresivo a la lógica de primer orden que los lenguajes para modelar bases de datos. Por esta razón, se dice que las ontologías están en el nivel “semántico”, mientras que el esquema de base de datos son modelos de datos a nivel “lógico” o “físico”. Debido a su independencia de los modelos de datos de nivel inferior, las ontologías se utilizan para integrar bases de datos heterogéneas, permitiendo la interoperabilidad entre sistemas y especificando interfaces a servicios independientes basados en el conocimiento. Existen ahora lenguajes estándar y una variedad de herramientas comerciales y de código abierto para crear y trabajar con ontologías. Los elementos de una ontología son: un dominio del discurso (también llamado clase o concepto), características de la clase, atributos del concepto y restricciones del concepto. Para efectos de este trabajo, se utilizará una ontología para representación y enriquecimiento de perfiles de investigación.

2.1.1. Elementos de cualquier ontología

- Conceptualización.- Ideas básicas que se formalizan. Pueden ser clases de objetos, métodos, estrategias, procesos, etc.
- Relaciones taxonómicas.- Interacción y enlace entre conceptos.
- Reglas (axiomas y restricciones).-Teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología.
- Individuos.- Representan objetos de un concepto.

2.1.2. Componentes de cualquier ontología

- Clase. Son grupos abstractos, conjuntos o colecciones de objetos.
- Propiedades de tipo de dato (*Data properties*). Relacionan individuos con tipos de datos (por ejemplo, cadenas, números, tipos de fecha, etc.)

- Propiedades de tipo de objeto (*Object properties*). Son aquellas que relacionan individuos con otros individuos.
- Individuos. Son los componentes básicos a “nivel de suelo” de una ontología, es decir, son instancias u objetos de una clase.

2.1.3. Perfil de investigación

Un perfil de investigación se refiere a la información relevante sobre la experiencia laboral, formación académica, actividades en redes sociales académicas, habilidades, especialidades e intereses de investigación de cada individuo. En este trabajo se abordará el perfil de investigación de personal académico de la Universidad Autónoma Metropolitana - unidad Azcapotzalco, el cual se refiere a la experiencia y conocimientos reflejados a través de las publicaciones científicas.

2.2. Lenguajes y modelos de representación de perfiles

En esta sección se presentan dos lenguajes que se tomarán en cuenta para el manejo de información y para la representación de perfiles de investigación.

2.2.1. XML - Extensible Markup Language

Este lenguaje se utiliza para el manejo de información, es decir, es el formato en que se encuentran las etiquetas de los artículos de investigación.

XML es un metalenguaje que permite a los usuarios definir lenguajes de marcado personalizados, especialmente para mostrar documentos en Internet. Contiene símbolos de marcado para describir el contenido de una página o archivo. Los datos XML se conocen como autodescriptivos o autodefinidos, lo que significa que la estructura de los datos está integrada con los datos, por lo que cuando llegan los datos no es necesario precompilar la estructura para almacenar los datos; se entiende dinámicamente dentro del XML. El formato XML puede ser utilizado por cualquier individuo que quiera compartir información de manera consistente. XML es en realidad un subconjunto más simple y fácil de usar del lenguaje de marcado generalizado estándar para la creación de documentos (SGML ¹, por

¹ SGML permite que la estructura de un documento pueda ser definida en base a la relación lógica de sus partes [9].

```

1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <!DOCTYPE dblp SYSTEM "dblp.dtd">
3 <dblp>
4 <author>Friedemann Andreas R&ouml;l&szlig;ler</author>
5 <phdthesis mdate="2012-04-18" key="phd/de/Rossl&euml;r2009">
6 <title>Bridging the gap between volume visualization and medical applications.</title>
7 <year>2009</year>
8 <school>University of Stuttgart</school>
9 <ee>http://elib.uni-stuttgart.de/opus/volltexte/2010/5428/</ee>
10 <url>http://elib.uni-stuttgart.de/opus/volltexte/2010/5428/</url>
11 <note type="urn">urn:nbn:de:bsz:93-opus-54287</note>
12 <note type="dnb">http://d-nb.info/1004286740</note>
13 </phdthesis>
14 <phdthesis mdate="2012-04-18" key="phd/de/Scheidig2003">

```

Figura 2.2: Ejemplo de archivo XML

sus siglas en inglés *Standard Generalized Markup Language*).

En la figura 2.2 se muestra un ejemplo de la estructura general que tiene un archivo XML. La estructura consta de una serie de etiquetas, entre ellas está incluido el significado de cada una de ellas. Por ejemplo, entre las etiquetas de autor (*author*) se encuentra el nombre de autor del artículo.

2.2.2. OWL - Web Ontology Language

El W3C *Web Ontology Language* (OWL) es un lenguaje de Web Semántica diseñado para representar un conocimiento rico y complejo sobre cosas, grupos de cosas y relaciones entre cosas. OWL es un lenguaje basado en lógica computacional tal que el conocimiento expresado en OWL puede ser explotado por programas de computadora, por ejemplo, para verificar la consistencia de ese conocimiento o para hacer explícito el conocimiento implícito. Los documentos OWL, conocidos como ontologías, pueden publicarse en la World Wide Web y pueden referirse o ser referidos desde otras ontologías OWL. Los entregables que conforman la especificación OWL 2 incluyen una descripción del documento, que sirve como una introducción a OWL 2, describe la relación entre OWL 1 y OWL 2 y proporciona un punto de entrada a los resultados restantes a través de lo que se conoce como *Documentation Roadmap* [10].

Actualmente, OWL tiene tres variantes:

- OWL Lite
- OWL DL

- OWL Full

Estas variantes incorporan diferentes funcionalidades, y en general, OWL Lite es más sencillo que OWL DL, y OWL DL es más sencillo que OWL Full. OWL Lite y OWL DL esta construido de tal forma que toda sentencia pueda ser resuelta en tiempo finito, la versión más completa de OWL Full puede contener “bucles” infinitos [11].

Una ventaja que tiene OWL sobre XML es que el primero es un lenguaje con más poder expresivo porque permite establecer relaciones entre términos, como por ejemplo la cardinalidad y definir su significado. Se trata de un lenguaje diseñado para usarse cuando la información contenida en los documentos necesita ser procesada por programas o aplicaciones, en oposición a situaciones donde el contenido solamente necesita ser presentado a los seres humanos. OWL surge como una revisión al lenguaje DAML-OIL y es mucho más potente que éste [8].

El documento XML posee una estructura que consta de una serie de declaraciones, elementos, comentarios, entre otros, y se indican mediante marcas explícitas. Los documentos XML tienen una estructura jerárquica respecto a las etiquetas que delimitan sus elementos; esto significa que estos elementos deben estar correctamente anidados [12].

2.3. DBLP - Fuente de datos pública para la extracción de perfiles de investigación

Significa *Digital Bibliography & Library Project*. La bibliografía informática de DBLP es la referencia en línea para obtener información bibliográfica sobre las principales publicaciones de informática. Ha evolucionado desde un pequeño servidor web experimental inicial hasta un popular servicio de datos abiertos para la comunidad informática. La misión del DBLP es apoyar a los investigadores en ciencias de la computación en sus esfuerzos diarios brindando acceso gratuito a metadatos bibliográficos de alta calidad y enlaces a las ediciones electrónicas de publicaciones [13].

A partir de mayo de 2016, DBLP indexa más de 3,3 millones de publicaciones, publicadas por más de 1,7 millones de autores. Con este fin, DBLP indexa alrededor de 32,000 volúmenes de revistas, más de 31,000 procedimientos de conferencias o talleres, y más de 23,000 monografías.

El servicio proporcionado por DBLP inició en el grupo de investigación de sistemas de bases de datos y programación lógica en la Universidad de Trier, Alemania, y se centró

en publicaciones de este campo de investigación. A través de los años DBLP se expandió gradualmente hacia todos los campos de la informática, mientras que el acrónimo sobrevivió. En ocasiones, la etiqueta “Bibliografía digital y proyecto de biblioteca” se ha adoptado como un contexto para DBLP.

Ahora puede aceptar “*DBLP computer science bibliography*”, o simplemente “DBLP”, como el nombre propio de este servicio web [13].

Algunas estadísticas de DBLP son:

- Número de publicaciones: 4,178,231
- Número de autores: 2,094,645
- Número de conferencias: 5,410
- Número de diarios(journals): 1,584

En la figura 2.3 y la figura 2.4 se muestra un ejemplo de la estructura del DBLP. En cuanto a artículos científicos, esta estructura consta de los siguientes elementos:

- Article / Inproceeding
- Author
- Title
- Pages
- Year
- Volume
- Journal
- EE
- Url

```

1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <!DOCTYPE dblp SYSTEM "dblp.dtd">
3 <dblp>
4 <author>Friedemann Andreas R&ouml;l&szlig;ler</author>
5 <phdthesis mdate="2012-04-18" key="phd/de/Rossler2009">
6 <title>Bridging the gap between volume visualization and medical applications.</title>
7 <year>2009</year>
8 <school>University of Stuttgart</school>
9 <ee>http://elib.uni-stuttgart.de/opus/volltexte/2010/5428/</ee>
10 <url>http://elib.uni-stuttgart.de/opus/volltexte/2010/5428/</url>
11 <note type="urn">urn:nbn:de:bsz:93-opus-54287</note>
12 <note type="dnb">http://d-nb.info/1004286740</note>
13 </phdthesis>
14 <phdthesis mdate="2012-04-18" key="phd/de/Scheidig2003">

```

Figura 2.3: Ejemplo de etiquetas de un artículo DBLP

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE dblp SYSTEM "dblp.dtd">
<dblp>
<article mdate="2017-05-28" key="journals/cce/TrokanasC16">
<author>Nikolaos Trokanas</author>
<author>Franjo Cecelja</author>
<title>Ontology evaluation for reuse in the domain of Process Systems Engineering.</title>
<pages>177-187</pages>
<year>2016</year>
<volume>85</volume>
<journal>Computers & Chemical Engineering</journal>
<ee>https://doi.org/10.1016/j.compchemeng.2015.12.003</ee>
<url>db/journals/cce/cce85.html#TrokanasC16</url>
</article>
</dblp>

```

Figura 2.4: Segundo ejemplo de etiquetas de un artículo DBLP

3 Estado del arte

En esta sección se presenta el estado del arte, el cual es una revisión de los trabajos relacionados con el tema de esta investigación. En particular, se revisan los trabajos que abordan la representación de perfiles de investigación, las tareas de extracción de información, poblado y enriquecimiento de ontologías para los perfiles de investigación. A continuación, se describen los trabajos analizados y, posteriormente, se presenta un análisis comparativo y una discusión sobre éstos.

Para seleccionar los trabajos relacionados se buscaron aquellos que involucraran operaciones con ontologías:

- Extracción de información
- Poblado
- Enriquecimiento
- Evaluación

3.1. *Búsqueda personalizada basada en ontología.*

En este trabajo se estudia la manera de modelar los intereses de los usuarios y muestra cómo estos modelos, también llamados perfiles, pueden ser desplegados para una recuperación y filtrado de información de manera eficaz. Los autores explican que un perfil se crea analizando las páginas que el usuario visita en un periodo de tiempo para identificar su contenido y asociarlo a la longitud del documento y al tiempo que invierte en él. Ellos muestran cómo se pueden utilizar los perfiles para lograr mejoras en el rendimiento de la búsqueda. Un perfil de usuario consta de dos partes: la historia de navegación y la relación entre los intereses calculados y los intereses reales del usuario. Además, se presenta una evaluación de los perfiles basada en una noción de convergencia. En términos de privacidad, el sistema existente almacena el perfil en la máquina del usuario. Dado que los resultados de búsqueda se procesan después, no hay necesidad de un servidor de perfiles central [14].

3.2. *Sistema de búsqueda de experiencia de investigadores utilizando minería de datos en una ontología.*

Los autores en este artículo exponen que los perfiles indican la experiencia de un investigador. Dicha información se puede obtener de diversas fuentes de información, por ejemplo, el currículum vitae, publicaciones, blogs, sitios Web y descripciones de proyectos de investigación. Por ello, los autores presentan un sistema de búsqueda de experiencia, que utiliza una ontología de clasificación de habilidades y perfiles de investigadores en el proceso de recuperación de información. Estos perfiles, almacenados dentro de la ontología, son expresados en el lenguaje OWL. Además, se proponen métodos de similitud basada en contenido y clasificación de textos para encontrar correspondencias semánticas entre los perfiles. Ellos, también determinan que un perfil de investigación contendrá la experiencia obtenida y la calificación relevante de los investigadores [15].

3.3. *La ontología y la arquitectura para una red social académica*

En este trabajo el solicitante tiene que buscar manualmente y comparar todos los programas de doctorado de las universidades, cursos y áreas de investigación de los académicos individualmente para encontrar un programa con los criterios requeridos usando la web actual. La persona debe encontrar las relaciones entre los datos. Para que este proceso sea más eficiente, los datos deben estar vinculados. El uso de máquinas de metadatos puede

leer el contenido mirando fotos y leyendo descripciones como una persona. Los autores utilizan: FOAF (por sus siglas en inglés, *Friend of a Friend*), el cual provee un mecanismo para enlazar a la gente con sus propios archivos y los ayuda a definir su red social. FOAF se ha convertido en un vocabulario estándar ampliamente aceptado para representar la red social y adopta un enfoque evolutivo de las extensiones de información. Además, utilizan SIOC (por sus siglas en inglés, *Semantically-Interlinked Online Communities*), tecnología de la web semántica que provee métodos para interconectar foros de discusión como blogs y listas de mail. Un formato abierto, estándar y legible por máquina para expresar la información contenida explícitamente en los métodos de discusión en Internet [16].

3.4. Búsqueda de investigadores similares en entornos académicos

En este artículo los autores proponen modelos para calcular la similitud entre las investigaciones basadas en perfiles de experiencia extraídos de sus publicaciones y páginas web académicas. Los perfiles de experiencia de los investigadores son modelados usando varias representaciones: vectores de términos basados en los documentos creados por el investigador, distribución de probabilidad de los temas en los que trabajó o atributos estructurales que describen al investigador. Los autores definen formalmente la recomendación del investigador, una instancia de búsqueda de entidad similar para el dominio académico. Además, proponen modelos para la representación de perfiles de investigadores y la similitud entre ellos:

1. Un perfil de investigador representado por un vector de términos derivados del contenido asociado con el investigador.
2. Un perfil de investigador representado en términos de una distribución de probabilidad.
3. Un modelo probabilístico de temas.
4. Similitud basada en rastreo que modela conceptos como vectores que usando matrices de densidad de los términos. Se menciona que en esta investigación se utiliza ArnetMiner y CiteSeerX como fuentes de datos para la extracción de perfiles de investigación, publicaciones y páginas de cursos. Ambos conjuntos de datos contienen información relevante al perfil de experiencia del investigador [17].

3.5. Mejora del perfil de usuario basado en ontología

El objetivo principal de este artículo es investigar técnicas que construyan una ontología basada en perfiles de usuario. Los perfiles se construyen automáticamente sin la interacción con el usuario, monitoreando los hábitos de navegación del mismo. Los perfiles se representan como una jerarquía en la que los nodos de hoja se consideran los intereses del usuario a corto plazo, mientras que los nodos de los padres son más generales y se consideran los intereses del usuario a largo plazo. Los autores construyen el perfil de usuario recolectando datos como URLs, fecha, tiempo que pasa en cada página. El proceso para construir un perfil de usuario consta de 3 fases:

1. Entrenamiento del clasificador;
2. Recolección de datos de usuario:
3. Clasificación de las páginas web recogidas.

El número de conceptos en los perfiles de usuario aumenta continuamente, y el número total de conceptos en el tiempo no muestra ninguna convergencia. Los perfiles de usuario tampoco mostraron convergencia fuerte cuando el número total de conceptos en los perfiles se representó en función del número de páginas web recopiladas [18].

3.6. Un modelo basado en ontología para búsquedas y clasificaciones expertas

Los autores proponen un modelo ontológico para encontrar y clasificar a los expertos de un dominio en particular. Una red social académica es construida basada en la información proporcionada por la base de conocimiento de un tema dado. La parte más importante de la red social académica es la extracción de relaciones para expandir y modelar la red para una aplicación específica. Por ello, en este artículo se propone un nuevo enfoque para modelar la red social académica teniendo en cuenta los diversos criterios y explorar las relaciones de clasificación de los candidatos expertos. Estos enfoques incluyen ArnetMiner, Fine-Grained, modelos híbridos y evidencia combinada. Todos ellos llevan a cabo la búsqueda de expertos mediante la consideración de métodos probabilísticos y el enfoque de minería de texto. Los autores presentan un modelo basado en ontologías para encontrar y clasificar a los investigadores en un dominio en particular, básicamente áreas de investigación en ciencias de computación y campos de ingeniería. El objetivo de [6] es ampliar un esquema de red social de FOAF para extraer perfiles de investigadores de

la web y se proponen tres métodos para modelar tópicos de documentos, autores y lugar de publicación. Ellos utilizan una interfaz de consulta y una base de conocimiento para la representación ontológica de la investigación científica. El poblado de las ontologías se realiza a partir de recursos web como DBLP y Citeseer. Las palabras clave son identificadas de cada publicación y representadas como un conjunto de características para cada documento utilizando la ponderación TF-IDF ¹ [20].

3.7. Contexto y conocimiento del contexto de los humanos y los sistemas AmI

En este trabajo los autores presentan el perfil de usuario representado por: identidad, hábitos, intereses, gustos, categoría social, cultura, y estado social. En la medida en que la información es proporcionada por el usuario, el perfil se va actualizando de vez en cuando y en general no necesita interpretación adicional. Al agregar el perfil de usuario, hay datos de contexto que se pueden derivar de información existente, que se puede obtener de bases de datos o bibliotecas digitales, como los mapas. En [7] se modelan los aspectos del mundo que se convierten en contexto a través de la forma en que los desarrolladores de sistemas o modeladores de ontología los utilizan en la interpretación y no debido a sus propiedades inherentes, los modelos de contexto, en particular los representados a través del formalismo ontológico deben evaluarse basándose en su exhaustividad, dinamismo, fidelidad con los fenómenos del mundo real, exactitud, consistencia interna, robustez, coherencia, por nombrar algunos criterios. Sin embargo, lo que sí es cierto es que no hay certeza de que la investigación sobre sistemas conscientes del contexto logre superar la cuestión de que los modelos contextuales implementados son siempre limitados [21].

3.8. Arquitectura basada en ontologías para la adaptación de servicios de contexto

La aplicación consciente del contexto puede capturar dinámicamente un rango de información desde su entorno y esta información representa un contexto, la aplicación adapta su ejecución de acuerdo con este contexto. Las ontologías presentan el instrumento más prometedor para el modelado y la gestión del contexto debido a su alta y formal expresividad y a las posibilidades de aplicar técnicas ontológicas de razonamiento. En este

¹ TF-IDF, por sus siglas en inglés *Term frequency Inverse document frequency* es la capacidad de los términos para representar el contenido de los documentos en la colección, que permitan identificar cuáles son relevantes o no ante la consulta del usuario. Al valor e índice que es capaz de determinar este extremo se le denomina “peso del término” o “ponderación del término” y su cálculo implica determinar la “Frecuencia de aparición del término TF” y la “Frecuencia inversa del documento para un término IDF” [19]

artículo se propone SOCAM (Por sus siglas en inglés *Service Oriented Context-Aware Middleware*) es un middleware para el conocimiento del contexto que proporciona una infraestructura para crear aplicaciones sensibles al contexto utilizando ontologías. SOCAM quita a los desarrolladores de servicios de contexto la tarea de gestión de contexto proporcionando un modelo para describir el contexto. Sin embargo, no hay forma de que los desarrolladores elijan los sensores utilizados para recopilar el contexto. La adaptación no se tiene en cuenta y se deja a los desarrolladores la tarea de implementar las adaptaciones de cada servicio. Además, los autores realizan la recuperación del contexto y asimilación, para la cual eligen OWL como un lenguaje de descripción de ontología diseñado para publicar y compartir ontologías y permitir la representación del conocimiento y el razonamiento basado en un axioma formal [22].

3.9. Modelo ontológico genérico de alto nivel para aplicaciones orientadas al contexto en ambientes inteligentes

Aunque se han creado numerosas ontologías de alto nivel para modelar el contexto dentro de entornos inteligentes, enfrentando muchos obstáculos, todavía existen una serie de desafíos. La mayoría de las ontologías de contexto de alto nivel que existen son de dominio específico o han sido desarrolladas para un único dominio, como las casas inteligentes. Por lo tanto, en este artículo se caracteriza el contexto en un entorno inteligente mediante cuatro componentes clave necesarios para facilitar un entorno “inteligente”, los cuales son: la capacidad de supervisar, interpretar, inferir y retroalimentar. Estos componentes están estrechamente relacionados entre sí y dependen del apoyo a los servicios de un entorno inteligente. Primero es necesario, en tiempo real, monitorear el ambiente actual y el comportamiento de los sistemas u ocupantes dentro de él. También hay una serie de elementos básicos que forman entornos inteligentes. Estos elementos centrales pueden definirse como sensores, objetos, elementos naturales, usuarios, temporales y espaciales. El contexto también tiene una serie de elementos básicos que se pueden dividir en persona o usuario, lugar, objeto, aplicación y situación. Los autores crean una ontología, que hacen disponible a otros investigadores para usar y desplegarla en una serie de escenarios en tiempo real. Estos escenarios incorporan actividades a lo largo de una casa inteligente, una sala de reuniones inteligente y oficinas [23].

3.10. Modelo de perfil de usuario ontológico para la personalización de aplicaciones contextualizadas

En este artículo se describe la creación de una ontología de perfiles de usuario para la personalización de aplicaciones orientadas al contexto en entornos móviles. En este trabajo se caracteriza al usuario en aplicaciones contextuales mediante el modelado de perfil de usuario con la finalidad de la personalización; ya sea recuperación de información personalizada, interfaces de usuario adaptables o entrega de servicios personalizados de aplicaciones que detectan el contexto. El modelado de perfil de usuario implica la creación de una estructura de datos que puede contener los atributos característicos de un tipo de usuarios. Esto se hace definiendo manualmente los atributos como hábitos, preferencias e intereses o inferir información no observable a partir de datos observables relacionados con sus acciones, pensamientos o comportamientos. La estructura de datos suele denominarse un modelo de usuario que sirve como plantilla para generar perfiles de usuario específicos para diferentes individuos. Los autores utilizan las ontologías para representar un vocabulario controlado que se estructura en una taxonomía jerárquica, donde se encuentran los conceptos clave del dominio. Cada clase definida puede tener clase padre y/o hijos (operando a través de un enlace “es-un”) formando una jerarquía de conceptos relacionados. La ontología se define utilizando el editor de ontologías Protégé, y se representa en el estándar *Web Ontology Language* (OWL). Una característica convincente de la ontología es su enfoque en los aspectos dinámicos de un usuario. Para construir la ontología del perfil de usuario, en primer lugar, los términos clave para describir un usuario se identifican y modelan como clases ontológicas. Además, se presenta como un caso de estudio, la combinación de un servicio de creación de perfiles de usuario, un servicio de personalización, una interfaz de usuario y un gestor de diálogo, el proyecto Mobile-Sage, que tiene como objetivo facilitar la personalización de servicios contextuales para los usuarios a medida que se mueven entre entornos. Una característica distintiva del modelo propuesto es su enfoque en los conceptos dinámicos y temporales para permitir la personalización de aplicaciones a medida que el usuario se mueve entre entornos móviles [24].

3.11. Una ontología para el cómputo en ambientes pervasivos conscientes de contexto

En este artículo se refieren al contexto como la información sobre una ubicación, sus atributos ambientales (por ejemplo, nivel de ruido, intensidad de luz, temperatura y movimiento) y las personas, dispositivos, objetos y agentes de software que contiene. El

contexto también puede incluir capacidades del sistema, servicios ofrecidos y buscados, las actividades y tareas en las cuales las personas y las entidades de computación están comprometidas, y sus roles, creencias e intenciones. [25] Los autores confían en que las ontologías son requisitos clave para la construcción de sistemas conscientes del contexto por las siguientes razones:

1. Una ontología común permite el intercambio de conocimientos en un sistema distribuido abierto y dinámico.
2. Ontologías con semántica declarativa bien definida proporcionan un medio para los agentes inteligentes y razonar sobre la información contextual.
3. Las ontologías explícitamente representadas permiten a los dispositivos y agentes no expresamente diseñados para trabajar juntos e interoperar.

3.12. Ontologías para administrar perfiles de usuario en la entrega personalizada de servicios móviles

En este artículo refieren al término AmI-Ambiente Inteligente, al espacio donde la gente está rodeada de un entorno capaz de responder a las necesidades de las personas que la habitan y sin su conocimiento. AmI se puede describir como un paradigma centrado en el usuario que funciona de forma invisible para ayudar a un usuario en particular. Los autores presentan un modelado de usuarios que implica la construcción de una serie de conceptos relacionados para el dominio de la salud. Las ontologías son utilizadas como un "vocabulario de representación", donde los conceptos de usuario están estructurados en una taxonomía basada en varios aspectos. Además, presentan una caracterización de usuarios en entornos generalizados considerando sus preferencias, necesidades, deseos, atributos y hábitos. Los contextos temporales se refieren a las ubicaciones de los usuarios y a las actividades específicas del tiempo utilizando ontologías. Finalmente, presentan una administración automática de la información del perfil de usuario después de que se capturan y cómo razonar rutinariamente sobre esto para proporcionar servicios personalizados a un usuario bajo demanda [26].

3.13. Una red social semántica de investigadores académicos

En este artículo los autores presentan un marco de trabajo en diversos pasos, primero se construye un perfil para cada investigador individual utilizando la información recopilada

de diferentes fuentes, tales como DBLP, ACM, IEEE, y Springer. Después, los perfiles se enriquecen integrando el conocimiento semántico externo extraído de Wikipedia y, por lo tanto, las relaciones semánticas entre investigadores se pueden descubrir basada en los perfiles enriquecidos. Una red social se construye de acuerdo con las similitudes entre los perfiles de los expertos, y comunidades de investigación los cuales son detectados por una técnica de análisis de agrupamiento. La información del perfil de investigación se obtiene rastreando la web y extrayendo información para crear un perfil de páginas web relevantes a un individuo en particular. Los perfiles construidos de esta manera contienen información relevante como experiencia de trabajo, historia educativa, actividades sociales, habilidades, especialidades e intereses para cada individuo. En términos de evaluación, los autores realizan experimentos para demostrar la efectividad de la red social construida en conferencias académicas, específicamente, en el proceso de asignación de documentos a investigadores relevantes para el proceso de revisión [27].

3.14. ScholarLens: extracción de competencias de publicaciones de investigación para la generación automática de perfiles de usuario semánticos

Los científicos confían cada vez más en sistemas de información inteligentes para ayudarlos en sus tareas diarias, en particular, para administrar objetos de investigación, como publicaciones o conjuntos de datos. El campo de investigación relativamente joven de la publicación semántica aborda la cuestión de cómo pueden mejorarse las aplicaciones científicas a través de representaciones que son semánticamente ricas de objetos de investigación, para facilitar su descubrimiento y reutilización. Para complementar los esfuerzos en esta área, se propone un flujo de trabajo automático para construir perfiles de usuarios semánticos de académicos, de modo que las aplicaciones académicas, como las bibliotecas digitales o los repositorios de datos, puedan comprender mejor los intereses, tareas y competencias de sus usuarios al incorporar sus perfiles en su diseño. Para que éstos perfiles se compartan entre las aplicaciones, se propuso construirlos basados en tecnologías semánticas estándar, en particular el Marco de Descripción de Recursos (RDF) para representar perfiles de usuario y fuentes de Datos Abiertos Vinculados (LOD) para representar temas de competencia. Para evitar el problema del arranque en frío, se sugiere rellenar automáticamente estos perfiles analizando las publicaciones (co) creadas por los usuarios, que según nuestra hipótesis reflejan sus competencias de investigación.

ScholarLens, desarrolla un enfoque novedoso que puede generar automáticamente perfiles de usuario semánticos para los autores de literatura académica. Para modelar las compe-

tencias de los usuarios y grupos académicos, se analizaron varios vocabularios de datos abiertos vinculados existentes. De acuerdo con las mejores prácticas de LOD, se propone un modelo basado en el esquema RDF (RDFS) para los registros de competencia que reutilizan los vocabularios existentes cuando sea apropiado. Para automatizar la creación de perfiles de usuario semánticos, se desarrolló un flujo de trabajo automatizado y completo que puede generar perfiles de usuario semánticos mediante el análisis de artículos de investigación de texto completo a través de varias técnicas de procesamiento de lenguaje natural (PNL). Se utilizó un conjunto de artículos de investigación para un usuario determinado. Las competencias se derivan de la minería de texto de los artículos, incluidos los pasos de enlace sintácticos, semánticos y de LOD. Después se llena una base de conocimientos en formato RDF con perfiles de usuario que contienen las competencias extraídas. Se implementa el enfoque como una biblioteca de código abierto y se evalúa el sistema a través de dos estudios de usuarios, lo que resultó en una precisión promedio (MAP) de hasta el 95%. Como parte de la evaluación, también se analiza el impacto de la zonificación semántica de los artículos de investigación en la precisión de los perfiles resultantes. Finalmente, se demostró cómo estos perfiles de usuario semánticos pueden aplicarse en varios casos de uso, incluida la clasificación de artículos para la búsqueda personalizada y la búsqueda de científicos competentes en un tema, por ejemplo, para encontrar revisores para un artículo [28].

| Artículo | Maneja P.I. | Usa ontología | Extracción Inf. | Poblado ontológico | Enriquecimiento |
|-----------------------|-------------|---------------|-----------------|--------------------|-----------------|
| [15]Punnarut,et al. | Sí | Sí | Sí | Sí | No |
| [14]Pretschner,et al. | No | Sí | No | No | No |
| [16]Challenger | Sí | Sí | Sí | No | Sí |
| [17]Gollapalli,et al. | Sí | Sí | Sí | No | No |
| [18]Trajkova & Gauch | No | Sí | Sí | No | No |
| [20]Uddin,et al. | Sí | Sí | Sí | Sí | No |
| [21]Bibri, S.E. | No | Sí | No | No | No |
| [22]Guermah,et al. | No | Sí | No | No | No |
| [23]McAvoy,et al. | No | Sí | No | No | Sí |
| [24]Skillen,et al. | No | Sí | Sí | No | No |
| [25]Chen,et al. | No | Sí | No | No | No |
| [26]Skillen,et al. | No | Sí | Sí | No | No |
| [27]Davoodi,et al. | Sí | No | Sí | No | No |
| [28]Sateli,et al. | Sí | Sí | Sí | Sí | No |
| Este proyecto | Sí | Sí | Sí | Sí | Sí |

Tabla 3.1: Tabla comparativa del Estado del arte.

| Artículo | Fuente de datos | Método de evaluación |
|-----------------------|---|--|
| [15]Punnarut,et al. | Artículos y proyectos de investigación | Términos de experiencia aplicados a ontología |
| [14]Pretschner,et al. | Páginas web que visita el usuario | Jerarquía de conceptos |
| [16]Challenger | Datos DBLP, SIOC y FOAF | SPARQL Query. Consultas |
| [17]Gollapalli,et al. | Arnet Miner y la colección UvT | Datos y reglas para comparar búsqueda |
| [18]Trajkova & Gauch | Urls del usuario | Det. perfil, construir 3 perfiles y presentarlos |
| [20]Uddin,et al. | DBLP, CiteSeer, ODP, Google Scholar | Agrupación de juicios |
| [21]Bibri, S.E. | No aplica | Evaluación en base a diversos criterios |
| [22]Guermah,et al. | Ontologías | No se evalúa |
| [23]McAvoy,et al. | Datos de sensores | No se evalúa |
| [24]Skillen,et al. | Ontologías | No se evalúa |
| [25]Chen,et al. | Dispositivos, agentes y sensores | Mecanismo llamado Tabling |
| [26]Skillen,et al. | Ontologías | Reglas SWRL |
| [27]Davoodi,et al. | Wikipedia, Ontología WordNet, DBLP, ACM, IEEE, Springer | k-means clusters |
| [28]Sateli,et al. | Artículos de investigación | Estudios de usuarios y analizar impacto |
| Este proyecto | DBLP y Excel | Diseño de ontologías |

Tabla 3.2: Continuación de tabla comparativa del Estado del arte.

3.15. Conclusión del estado del arte

En el presente trabajo se hace un manejo de los perfiles de investigación que se integran con los datos académicos de los profesores, de las publicaciones que tengan y la relación de co-autoría, para efectos del presente trabajo, son otros profesores. Al igual que los otros trabajos, se utilizaron ontologías, en las cuales se almacenó la información que se recabó, con la finalidad de hacer un enriquecimiento de la ontología descubriendo relaciones entre autores.

La extracción de la información en lo que respecta al archivo XML se hizo a través de un método que actuó como un *parser*, el cual obtuvo la información contenida en cada una de las etiquetas y posteriormente se acomodó en cada uno de los atributos en su respectivo tipo de dato y/o tipo de objeto ontológico, alimentando la ontología de Publicación (*Publication*).

El poblado ontológico se hizo a través de un método que tomó automáticamente cada uno de los atributos adquiridos tanto en la ontología de Persona (*Person*) como en la ontología de Publicación (*Publication*) y los convirtió en instancias de una ontología de manera formal. Dentro del poblado de las ontologías se crearon relaciones entre ellas, de manera que se relacionaron las instancias ontológicas de Persona a las instancias de Publicación y viceversa. La relación que va de Persona a Publicación se nombró *hasPublish* y su contraparte, es decir, la relación que va de Publicación a Persona se nombró *isPublishedBy*.

El enriquecimiento se realizó creando una nueva relación entre instancias, esta relación tiene por nombre *CollaborateWith* o “ColaboraCon”. Esto tuvo el propósito de relacionar autores entre una o más publicaciones. La regla se escribió en lenguaje SQWRL, hay que tomar en cuenta que esta regla puede generar redundancia y hacer que un autor sea co-autor de sí mismo. Para evitar que esta relación consigo mismo se registrara en la ontología se llamó a la función *differentFrom*, la cual omitió al autor que sea el mismo, es decir, la misma persona no fue su co-autor. De esta manera sólo se crearon relaciones entre autores diferentes.

4 Metodología de solución

En esta sección se presenta la metodología de solución que se siguió para dar respuesta al problema de extraer y enriquecer el perfil de investigación del personal académico de la Universidad Autónoma Metropolitana - unidad Azcapotzalco, utilizando ontologías. En la figura 4.1 se muestra esta metodología.

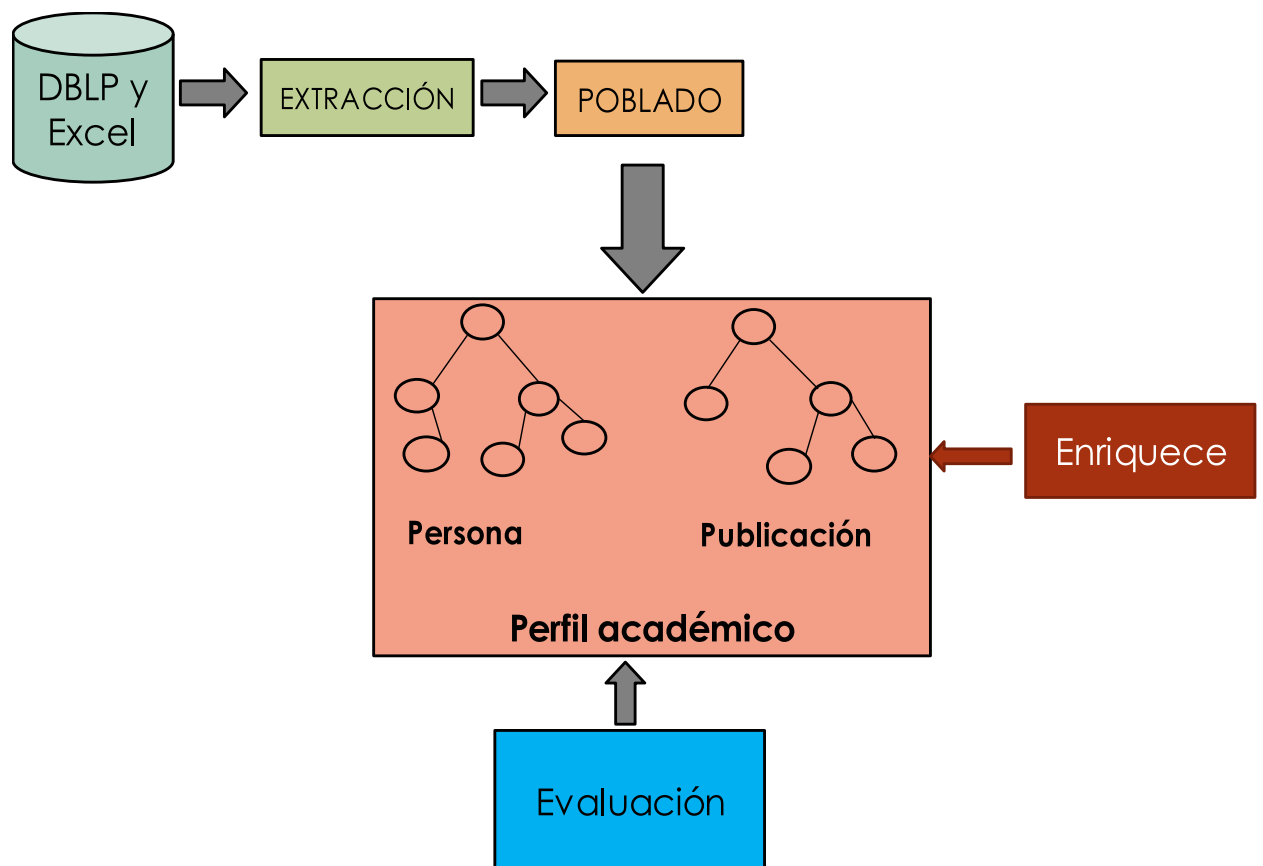


Figura 4.1: Arquitectura de solución

La metodología incluyó los pasos que se enlistan a continuación:

1. Se tuvo como fuente de datos un archivo Excel y el DBLP en lenguaje XML, los cuales sirvieron para proporcionar la información que se almacenó dentro de la ontología de Persona (*Person*), en el caso del archivo Excel, y en la ontología de Pulicación (*Publication*) para el XML.
2. Extracción de información. Una vez que se tuvieron las fuentes de datos, se creó un método para hacer la extracción de la información de dichas fuentes. Esta información se almacenó dentro de un arreglo de objetos; cada objeto almacenó una propiedad que se encuentra en cada uno de los archivos (fuente de datos).
3. Poblado de ontologías. Una vez que los objetos se llenaron con las propiedades de cada uno de los individuos, se creó un método que instanció cada objeto junto con sus propiedades y fueron insertados en la ontología. El poblado incluye a las ontologías de:
 - Persona (*Person*)
 - Publicación (*Publication*)
 - Perfil de investigación (*Researcher profile*)
4. Enriquecimiento. Para realizar esta tarea se emplearon las reglas de competencia ¹ para la creación de la relación de co-autoría.
5. Evaluación. Para realizar la evaluación de los modelos ontológicos, se utilizó el principio de competencia ontológica, el cual establece que la competencia de una ontología se puede determinar mediante un conjunto de preguntas que la ontología debe ser capaz de responder.

4.1. Modelado de ontologías

El modelado de ontologías se refiere a proveer descripciones y estructurar de manera formal la descripción de los conceptos explícitos e implícitos, relaciones utilizadas en el

¹Las reglas de competencia consisten en un conjunto de preguntas formuladas y contestadas (las respuestas a menudo se pasan por alto durante la especificación de la ontología) en lenguaje natural, de modo que la ontología debe poder responderlas correctamente [1]. Desempeñan un papel importante en el ciclo de vida del desarrollo de la ontología, ya que representan los requisitos de la ontología.

dominio de la documentación, lo que en la práctica permite representar, con los ajustes necesarios, de manera adecuada buena parte de los procesos centrales del Perfil de investigadores [29].

En este proyecto se planteó el desarrollo de un modelo ontológico que sirvió como apoyo a la recolección, organización y análisis de información acerca de los perfiles de investigadores, los tipos de relaciones entre estos, y cómo se manifiestan estas relaciones en el uso de los recursos u objetos y el uso de herramientas para la creación de las ontologías involucradas en esta tarea [30].

Para resolver el problema mencionado anteriormente, se diseñaron tres ontologías para almacenar la información recabada de las fuentes de datos. Estas ontologías son las de Persona (*Person*), Publicación (*Publication*) y Perfil de Investigación (*Researcher Profile*). Se llegó a esta conclusión debido a ciertas preguntas que se realizaron en lenguaje natural como por ejemplo:

- ¿Qué profesor tiene más publicaciones en el departamento de Electrónica?
- ¿Quién colabora con el Profesor Alejandro Reyes?
- ¿Qué departamento publicó más artículos en el año 2015?
- ¿Cuántas relaciones de colaboración hay en el departamento de Sistemas?
- ¿Cuántos profesores con Doctorado del departamento de electrónica han publicado?
- ¿Cuántas publicaciones hubo en el año 2017?
- ¿Cuántas mujeres del departamento de sistemas tienen al menos una publicación?
- ¿Cuántas publicaciones tiene el departamento de sistemas en el año 2014?

Una vez elaboradas las preguntas de competencia, se hizo un análisis de los conceptos relevantes entre estas preguntas y se agruparon de acuerdo a ciertas características de similitud para formar las clases principales, separándolas de las respectivas subclases. De la misma manera se analizaron sus atributos y el tipo de dato de cada uno de ellos para declararlo dentro de cada una de las ontologías. De no hacerlo de esta manera, en el momento de hacer la evaluación en la ontología se debe realizar un paso, que se explicará más adelante, para hacer una “pre - evaluación” de la estructura de la ontología y todos

los axiomas que se declaren para verificar que todas las ontologías sean consistentes.

Para la definición de cada ontología se llevaron a cabo los siguientes pasos:

1. Identificar las clases de cada ontología
2. Establecer la jerarquía de clases
3. Definir las relaciones semánticas entre las clases
4. Establecer los axiomas de cada clase (*Data properties*)
5. Establecer los axiomas de las relaciones semánticas (*Object properties*)
6. Establecer los axiomas de las clases (*Disjoint*)
7. Definir de manera completa cada clase

Una vez concluido este proceso de diseño, la tabla 4.1 muestra las clases principales resultantes:

| |
|-------------------------|
| Clase |
| Persona |
| Profesor |
| Perfil de investigación |

Tabla 4.1: Clases principales de las ontologías.

Cada una de estas clases se convirtió en una ontología individual y la clase de **Perfil de investigación** (*Researcher Profile*) es aquella que contiene a las otras dos; es decir, es una ontología que importa a dos ontologías. A continuación se describe cada una de ellas.

4.1.1. *Ontología de Persona (Person)*

Dentro de la Universidad Autónoma Metropolitana, existen ciertas jerarquías que deben respetarse, en el diseño de esta ontología se tomó en cuenta para crear el modelo en la ontología de Persona *Person*.



Figura 4.2: Ontología de Persona (*Person*)

La siguiente división que se hizo fue separar los tipos de personas que pueden ingresar a la Universidad Autónoma Metropolitana - unidad Azcapotzalco, los cuales son:

- Empleado
- Visitante
- Estudiante

La figura 4.3 muestra su representación dentro de la ontología.

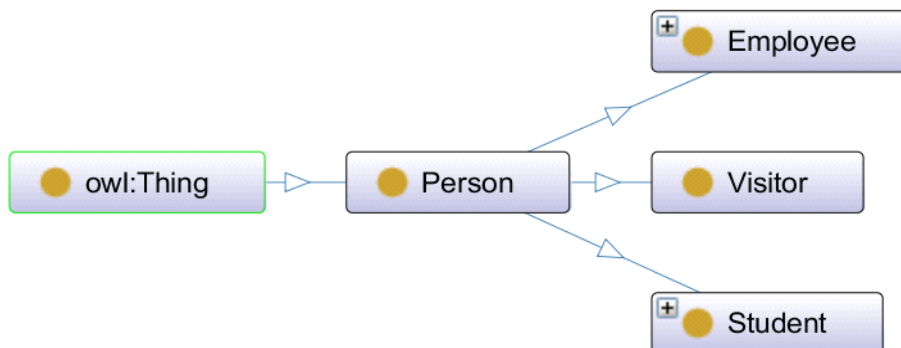


Figura 4.3: Tipos de personas que pueden ingresar

La subclase Empleado (*Employee*) tiene dos divisiones, son Académico (*Academic*) y Administrativo (*Administrative*). La subclase Estudiante (*Student*) tiene sólo una división que es Estudiante regular (*Regular Student*). En la figura 4.4 se muestra esta representación ontológica.

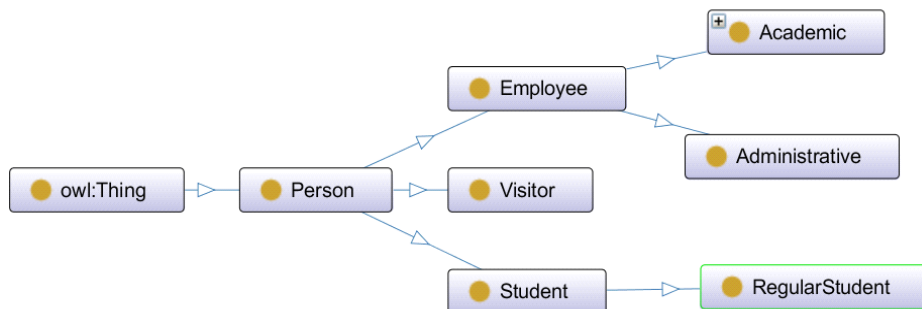


Figura 4.4: División de Empleado (*Employee*) y Estudiante (*Student*)

Como último paso se hizo una división más en la clase de Académico (*Academic*) la cual consistió en dividirla en Ayudante (*Assistant Student*) y Profesor (*Professor*). En la figura 4.5 se muestra la representación ontológica final de la ontología de Persona (*Person*).

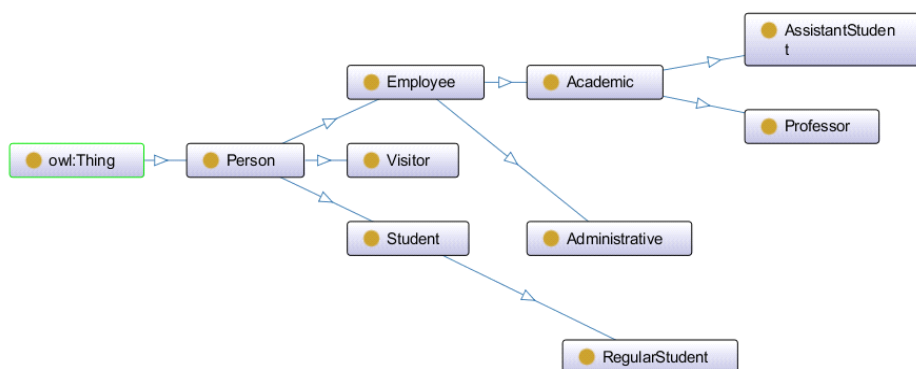


Figura 4.5: Representación de todas las clases de la ontología de Persona (*Person*)

Las *data properties* (atributos) correspondientes a la clase de Persona (*Person*) se enlistan a continuación. Para cada una de estas *data properties* se indicó un dominio y un rango, en algunos casos se declararon propiedades extra para evitar la inconsistencia de la ontología.

- tieneCategoria (*hasCategory*)
 - Dominio: Empleado (*Employee*)
 - Rango: *String*

- Descripción: Esta *data property* se refiere a la categoría que tiene un empleado dentro de la Universidad Autónoma Metropolitana - unidad Azcapotzalco.
- tieneNumeroEconomico (*hasEconomicNumber*)
 - Dominio: Empleado (*Employee*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al identificador único de cada uno de los empleados.
 - tieneCorreoElectronico (*hasEmail*)
 - Dominio: Profesor (*Professor*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere a la dirección de correo electrónico que se otorga a cada empleado. Generalmente se compone de un *nick name* (sobrenombre) y al final las siglas de la Universidad.
 - tieneGenero (*hasGender*)
 - Dominio: Persona (*Person*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al género de cada persona, es decir, si es femenino o masculino.
 - tieneNombre (*hasName*)
 - Dominio: Persona (*Person*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere a la designación o denominación verbal que se le da a una persona, para distinguirlo de otros. Se compone del nombre, apellido paterno y apellido materno.
 - tienePosicion (*hasPosition*)
 - Dominio: Administrativo (*Administrative*)
 - Rango: *String*

- Descripción: Esta *data property* se requiere para diferenciar a los “Administrativos” y “Académicos”, es decir, este dato indica su puesto administrativo.
- tieneProyecto (*hasProject*)
 - Dominio: Académico (*Academic*)
 - Rango: *String*
 - Descripción: Esta *data property* es un proyecto de tipo académico (“Proyecto Docente”): clases, asesorías, investigación. El año sabático es un caso especial, sin embargo, los profesores en esta situación continúan con los proyectos académicos.
 - tieneMatricula (*hasStudentID*)
 - Dominio: Estudiante (*Student*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al identificador único que es propio de cada estudiante.
 - tienePropositoDeVisita (*hasVisitPrupose*)
 - Dominio: Visitante (*Visitor*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al motivo por el cual una persona externa se registra en las instalaciones de la Universidad.
 - daAsesoria (*providesCounseling*)
 - Dominio: Ayudante (*Assistant Student*)
 - Rango: *String*
 - Descripción: Esta *data property* se refiere a los ayudantes (*AssistantStudent*) que dan apoyo a los estudiantes sobre tareas, temas que no se comprenden en clase o sustituir a algún profesor que no acuda a su respectiva clase.

En la figura 4.6 se muestra su representación dentro de *Protegé*.

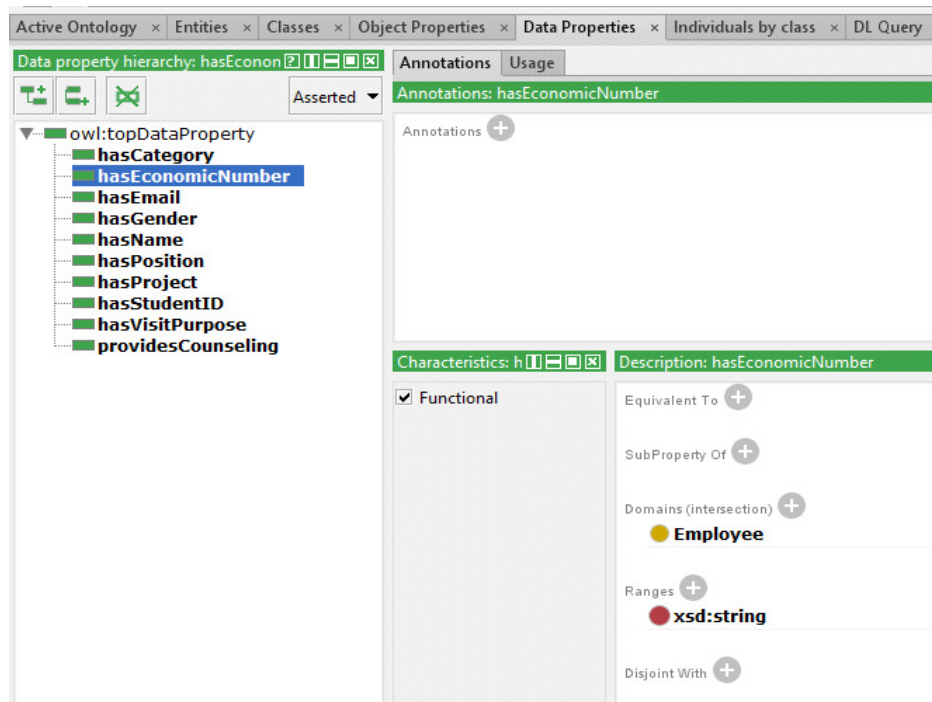


Figura 4.6: *Data properties* de la Ontología de Persona (*Person*)

En el caso de las *object properties*, en la ontología de Persona (*Person*) tenemos:

- *tomaClasesCon* (*takeClassWith*)
 - Dominio: Estudiante regular (*Regular Student*)
 - Rango: Profesor
 - Descripción: Esta *object property* se refiere a un estudiante que toma una o más clases con un profesor.

- *tieneDepartamento* (*hasDepartment*)
 - Dominio: Profesor(*Professor*)
 - Rango: Departamento (*Department*)
 - Descripción: Esta *object property* se refiere al área o división al que corresponde cada uno de los profesores, es decir, si pertenecen a Ciencias Básicas, Artes y Humanidades o Diseño.

- tieneTituloAcademico (*hasAcademicTitle*)
 - Dominio: Profesor(*Professor*)
 - Rango: TituloAcademico (*AcademicTitle*)
 - Descripción: Esta *object property* se refiere al grado académico de cada uno de los profesores: licenciatura, maestría, doctorado, etc.

En la figura 4.7 se muestran las *object property* en el software *Protegé*.

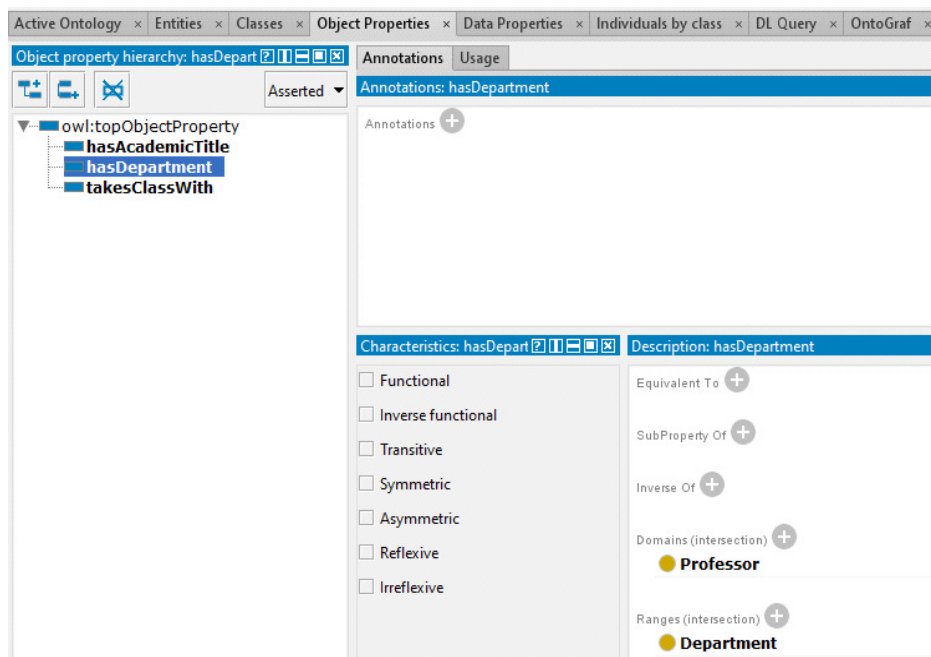


Figura 4.7: *Object properties* de la ontología de Persona

4.1.2. Ontología de Publicación (*Publication*)

En la actualidad, algunas publicaciones de artículos sobre Ciencias de la Computación hechas en revistas, congresos o eventos de índole científico quedan registradas en páginas de bibliografía científica como el DBLP.

De esta página se obtuvo un archivo XML con la información de estas publicaciones. Para el diseño de esta ontología se determinó que solamente incluye una clase principal, a la

cual se le denominó *Publicación (Publication)*. En la figura 4.20 se muestra su representación en *Protegé*.

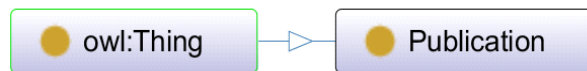


Figura 4.8: Ontología de *Publicación (Publication)*

La siguiente lista enumera las *data properties* que se establecieron para la ontología de *Publicación (Publication)*.

- tieneAutor (*hasAuthor*)
 - Dominio: *Publicación*
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al redactor o redactores de la publicación.

- tieneEE (*hasEE*)
 - Dominio: *Publicación*
 - Rango: *String*
 - Descripción: Esta *data property* es un tipo de URL, es decir, un identificador único para cada una de las publicaciones.

- tieneLlave (*hasKey*)
 - Dominio: *Publicación*
 - Rango: *String*
 - Descripción: Esta *data property* es una cadena que se forma con el tipo de publicación, si es *journal* o *paper*, etc. Contiene los apellidos de los autores.

- tienePaginas (*hasPages*)
 - Dominio: *Publicación*
 - Rango: *String*

- Descripción: Esta *data property* indica las páginas que abarca la publicación.
- tieneTitulo (*hasTitle*)
 - Dominio: Publicación
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al nombre que se le da a la publicación.
- tieneUrl (*hasUrl*)
 - Dominio: Publicación
 - Rango: *String*
 - Descripción: Esta *data property* es una cadena con cierto texto que incluye los apellidos de los autores y una ruta donde se almacena la publicación.
- tieneVolumen (*hasVolume*)
 - Dominio: Publicación
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al volumen correspondiente a la publicación como si fuera el volumen de un libro.
- tieneAño (*hasYear*)
 - Dominio: Publicación
 - Rango: *String*
 - Descripción: Esta *data property* se refiere al año en que se crea la publicación.

En la figura 4.20 se muestra la representación de las *data properties* en el software *Protegé*.

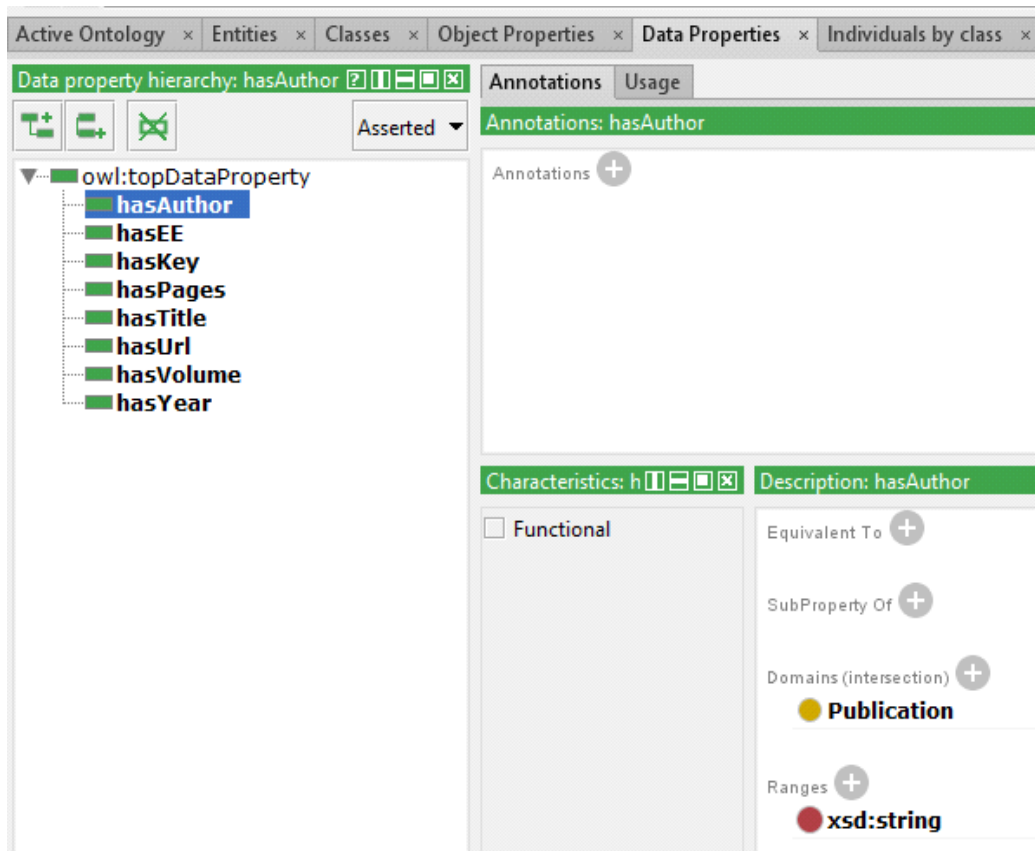


Figura 4.9: Propiedades de datos de la ontología de Publicación (*Publication*)

4.1.3. Ontología de Perfil de Investigador (*Researcher profile*)

Tiene como característica principal ser una ontología que importa a las ontologías de Persona (*Person*) y Publicación (*Publication*). La razón principal de la existencia de esta ontología es debido a que el perfil de un investigador se conforma de los datos académicos que componen la base de datos en Excel y que conforman las propiedades de la ontología de Persona (*Person*), concretamente de la clase de Profesor (*Professor*), además de las publicaciones que le corresponden del archivo DBLP, las cuales son las propiedades de la ontología de Publicación (*Publication*) y ya que estas ontologías contienen propiedades completamente diferentes entre sí, es decir, las propiedades de la ontología de Persona (*Person*) y la ontología de Publicación (*Publication*) son diferentes unas de otras, se decidió que el perfil de investigación las englobara para representarlo de una manera más completa. Debido a esto, la ontología de Perfil de Investigador (*Researcher profile*) importa tanto a la ontología de Persona (*Person*) como a la ontología de Publicación (*Publi-*

tion). Debe existir una manera de relacionar ambas ontologías (Persona y Publicación) y esto se logró a través de la creación de una relación, en este trabajo dicha relación es de **colaboración** relacionando a uno o más autores con uno u otros autores diferentes. A esta relación de una manera formal la llamaremos “ColaboraCon” (“*CollaborateWith*”), esto para indicar que un Profesor (que es de tipo Persona) ha escrito una o más publicaciones (ontología de Publicación) y en algunos casos se ha unido con otros autores y si en algún momento un Profesor necesita hacer una consulta, o desea crear una nueva publicación acerca de un tema en específico con otro profesor que ha investigado sobre el mismo tema, puede encontrarlo con ayuda de la relación de colaboración. En la figura 4.10 se muestra la representación de esta ontología en el software *Protegé*.

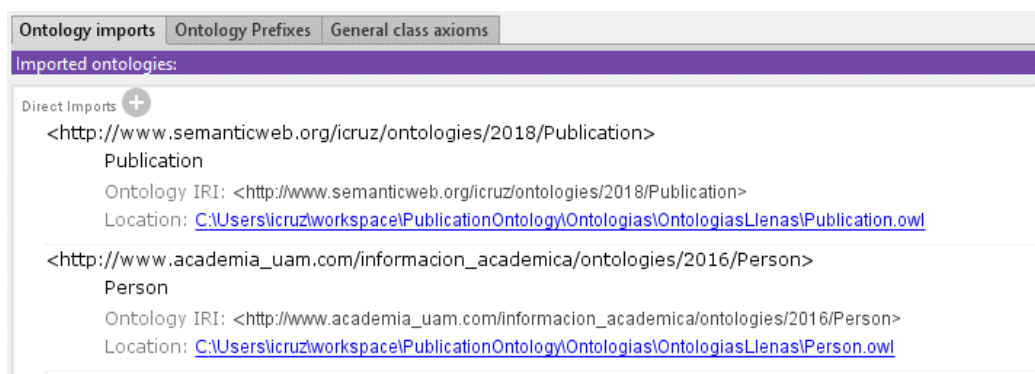


Figura 4.10: Ontología de Perfil de investigación (*Researcher Profile*)

En la figura 4.11 se muestra un diagrama de la ontología de Perfil de Investigador (*Researcher profile*). En este diagrama podemos observar las clases que importa la ontología de Perfil de investigador (*Researcher Profile*), es decir, todas las clases que pertenecen tanto a la ontología de Persona (*Person*) y Publicación (*Publication*) las contiene, así como sus *data properties* (figura 4.12) y *object properties* (figura 4.13).

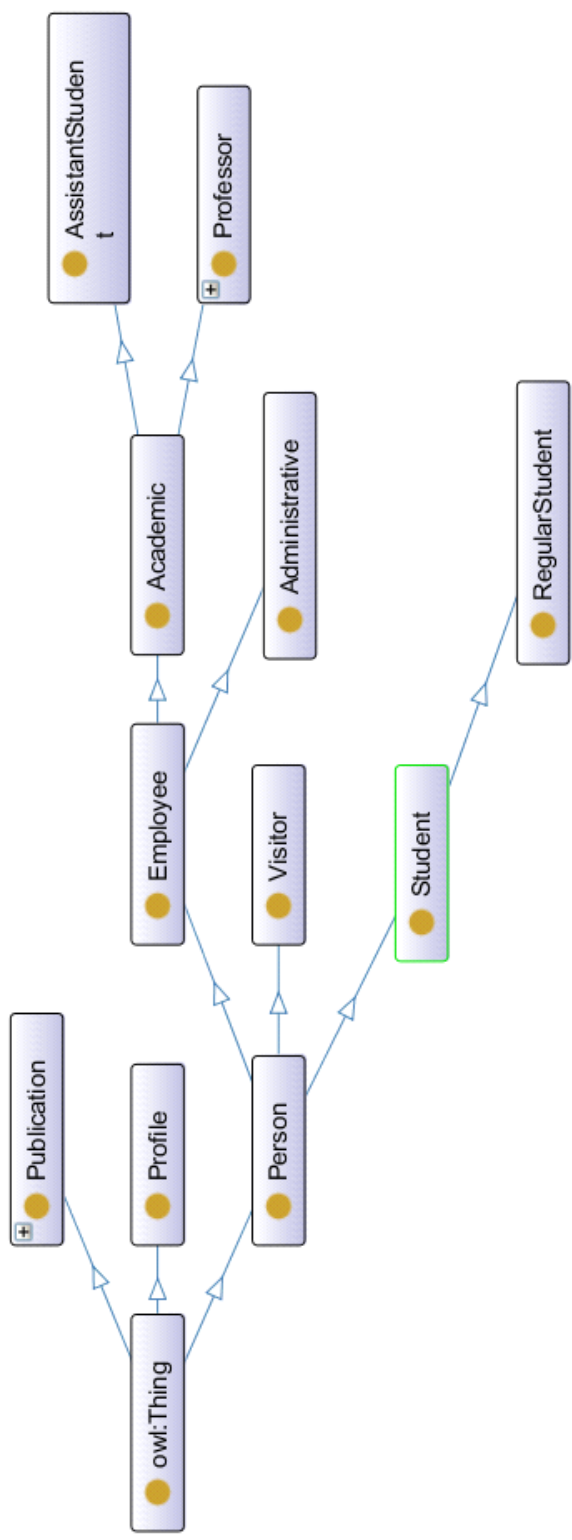


Figura 4.11: Clases de la ontología de Perfil de investigación (*Researcher Profile*)

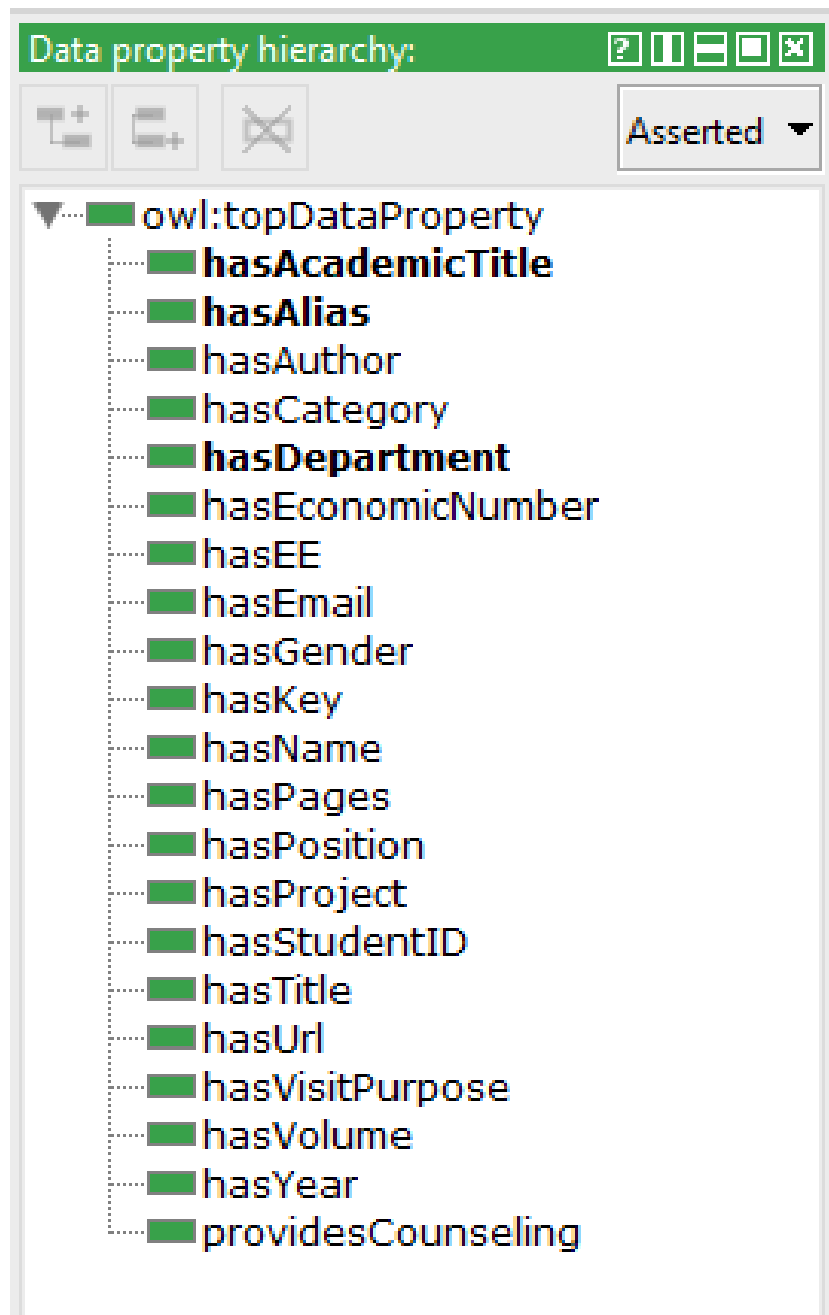


Figura 4.12: *Data properties* de la ontología de Perfil de investigación (*Researcher Profile*)

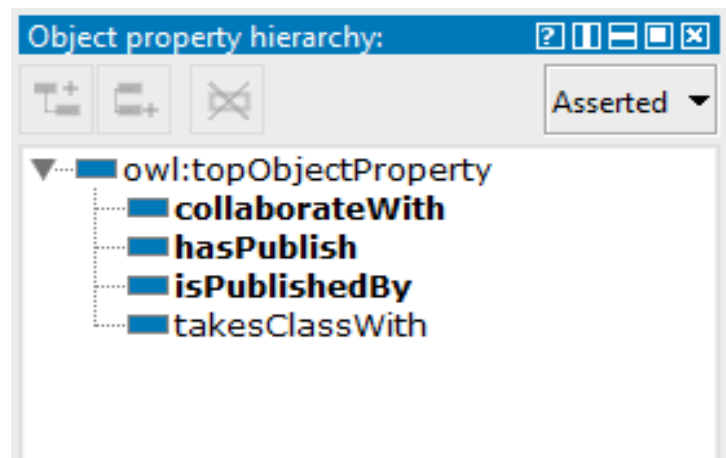


Figura 4.13: *Object properties* de la ontología de Perfil de investigación (*Researcher Profile*)

4.2. Poblado de ontologías

En esta sección, se aborda el poblado de ontologías sobre la base de documentos de texto. El poblado de ontologías implica encontrar instancias tanto de relaciones entre clases como relaciones entre los conceptos y/o sus atributos. El problema de encontrar relaciones entre las instancias es ciertamente muy difícil y prácticamente requiere de una comprensión completa del lenguaje natural. Esto está fuera del alcance de la presente tesis. Un objetivo más modesto es la extracción de un conjunto de relaciones predefinidas. Este es el objetivo de la tarea de extracción de información definida originalmente en el contexto de la comprensión de mensajes (MUC) [31].

La población ontológica es la tarea de agregar nuevas instancias de conceptos a la ontología. La asociación de un conjunto de instancias con un concepto se conoce como población ontológica o etiquetado ontológico. La población ontológica es el proceso de agregar nuevas instancias de conceptos y/o relaciones en una ontología, por lo general al ubicar los objetos y/o términos y sinónimos correspondientes en el corpus.

En una visión simplificada, una ontología puede considerarse como un conjunto de conceptos, relaciones entre los conceptos y sus instancias. Una instancia conceptual es la realización del concepto en el dominio, por ejemplo, la ejemplificación del concepto como una frase en un corpus textual. El proceso de población de ontología no cambia la estructura de una ontología, es decir, la jerarquía de conceptos y las relaciones no taxonómicas no se modifican. Lo que cambia es el conjunto de realización (instancias) de conceptos y

relaciones en el dominio. La población de ontología requiere una ontología inicial que se completará y un motor de extracción de instancias.

El motor de extracción es responsable de localizar instancias de conceptos y relaciones en un corpus. El motor de extracción procesa dicho corpus para localizar el concepto y/o la relación. La lista de instancias de concepto- relación extraídas se utiliza posteriormente para poblar la ontología. Recordando la idea de “pastel de capas”, el proceso de población involucra algunas capas. En particular, se trata de la adquisición de relaciones (es decir, objetos y relaciones alternativas - sinónimos) de conceptos y relaciones. Un enfoque típico es usar relaciones conocidas y asociadas con conceptos - relaciones que pueden haber sido identificadas durante la formación del concepto - relación, para localizar los objetos - sinónimos correspondientes en un corpus. El resultado es un corpus anotado, que se puede usar para construir extractores de instancias más generales, utilizando el aprendizaje automático. Un aspecto interesante de la población ontológica, que no se trata adecuadamente en la literatura, es el manejo de la redundancia. La eliminación de redundancia en el conjunto de instancias requiere eliminar la ambigüedad de entidades, que es el proceso de identificación de instancias que se refieren al mismo objeto o evento real. Si una ontología se rellena con una instancia sin verificar, si el objeto real o evento representado por la instancia ya existe en la ontología, se insertarán instancias redundantes. En el peor de los casos, las instancias redundantes contienen información contradictoria, lo que puede conducir a una ontología inconsistente.

Hasta donde sabemos, solo tres enfoques abordan este problema. El sistema Artequakt [1] aplica heurística escrita manualmente, para fusionar instancias que se refieren al mismo objeto o evento real. Estas heurísticas se evalúan después de que un lote de instancias haya poblado la ontología. El sistema SOBA [1], por otro lado, realiza comprobaciones simples utilizando reglas especiales de mapeo, durante la creación de la instancia (es decir, antes de que las instancias se inserten dentro de la ontología), para reutilizar instancias que se refieren al mismo objeto real o evento. El enfoque seguido por BOEMIE [1] mejora el de Artequakt, a través del uso del aprendizaje automático en lugar de la heurística desarrollada manualmente. Los sistemas de población de ontología están estrechamente relacionados con los sistemas de extracción de información basados en ontología, ya que estos últimos proporcionan mecanismos para asociar fragmentos de los datos con conceptos de una ontología. Por lo tanto, cada sistema de extracción de información basado en ontologías se puede ver como un sistema de población ontológica, ya que se puede extender para asimilar instancias extraídas en la ontología [1].

Dentro de esta sección se muestra el proceso para insertar cada uno de los individuos instanciados en cada una de las ontologías. Para lograr este paso, se crearon objetos y

sus respectivos atributos y/o propiedades para que el objeto pase completo y se convierta en una instancia ontológica y de esta manera se inserte dentro de ella. Esta operación se realizó tantas veces como registros haya en cada una de las fuentes de datos mencionadas anteriormente (Excel y DBLP).

En las siguientes subsecciones se explica el pseudocódigo que se utilizó para el poblado de cada una de las ontologías.

4.2.1. Poblado de la ontología de Persona (*Person*)

El poblado de la ontología de Persona se describe a continuación:

El archivo Excel es la fuente de datos única y principal para esta ontología. Este archivo contiene la información de Profesores de la división de CBI (Ciencias Básicas e Ingeniería), el cual contiene datos académicos de Profesores como su nombre, departamento al que pertenece, correo electrónico, título académico, entre otros.

Los pasos que se siguieron para el poblado de la ontología de Persona (*Person*) es el siguiente:

1. Archivo Excel. Ir a la carpeta donde se encuentra el archivo con los datos de los Profesores.
2. Objeto tipo Profesor. Se creó el objeto que almacenó los atributos de un Profesor, es decir, lo que se convierte en *data properties*.
3. Creación de arreglo de cadena. Este arreglo se utilizó para almacenar todos los alias de cada uno de los Profesores (para el presente trabajo solamente se buscaron los alias de los Profesores del departamento de Sistemas).
4. Se creó un contador para llevar el número de registros que se insertaron con éxito
5. Comienza la iteración fila por fila.
6. Al principio de cada iteración se “limpia” el objeto de tipo Profesor y el arreglo de alias.
7. Se programó un switch para indicarle al programa el tipo de dato que se encuentra en cada columna del archivo fuente. Es decir:

- Columna 0: Nombre del Profesor
 - Columna 1: Género
 - Columna 2: Departamento
 - Columna 3: Correo electrónico
 - Columna 4: Identificador único de profesor.
 - Columna 5: Título académico
 - Columna 6: Proyecto al que está asignado
 - Default: En las columnas subsecuentes se encuentran los alias.
8. Se llamó a la instanciación y se pasó como parámetro el objeto de tipo Profesor y la clase para indicar a qué clase pertenece este objeto.
 9. Se incrementó en uno el contador
 10. Cerrar el libro Excel.

En la figura 4.14 se muestra el diagrama de flujo de este proceso de poblado.

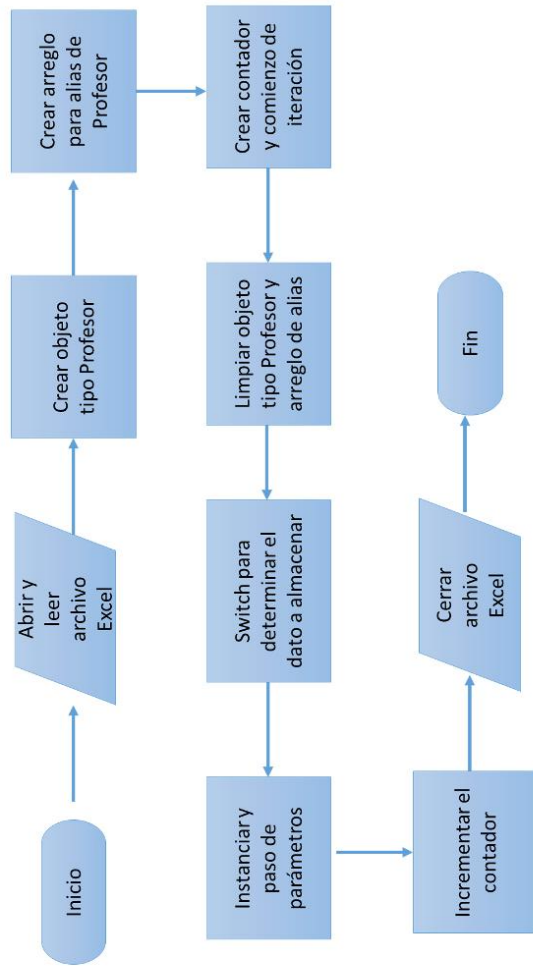


Figura 4.14: Diagrama de flujo del proceso de poblado de la ontología de Persona

Método para instanciar un objeto tipo Profesor y convertirlo en instancia ontológica

- Instanciar el objeto de acuerdo al tipo de clase que se indica en el parámetro.
 - Se creó la IRI (Identificador de Recurso Internacional, *Internationalized Resource Identifiers*, por sus siglas en inglés), la cual funcionó como un identificador de clase.
 - Se creó la siguiente IRI pero ahora a nivel de individuo.
 - Se instanció la clase y al individuo
 - El método insertó al individuo en la ontología
 - Se guardó la ontología
- Una vez que el individuo se encuentra dentro de la ontología se procedió a insertar cada una de las *object properties* y *data properties*.
- Para el caso de los alias se recorrió cada uno de los elementos del arreglo y se agregó su *data property* “tieneAlias” (*hasAlias*).

En la figura 4.15 y la figura 4.16 se muestra un ejemplo de instancias de la ontología de Persona (*Person*).

The screenshot displays an ontology editor interface. The top menu bar includes 'Active Ontology', 'Entities', 'Classes', 'Object Properties', 'Data Properties', 'Individuals by class', 'DL Query', 'OntoGraf', 'SWRLTab', and 'SQWRLTab'. The main window is divided into several panes:

- Class hierarchy:** Shows a tree structure starting from 'owl:Thing', with sub-classes like 'AcademicTitle', 'Department', 'Person', 'Employee', 'Academic', 'AssistantStudent', 'Professor', and 'Administrative'. 'Professor' is highlighted.
- Instances:** A list of individuals for the 'Professor' class, including 3359, 341, 34214, 34327, 34592, 34786, 35009, 35691 (highlighted), and 35693.
- Annotations:** Shows 'Annotations: 35691'.
- Property assertions:** Lists object property assertions for the selected instance: 'hasAcademicTitle DOCTORADO' and 'hasDepartment SISTEMAS'.
- Data property assertions:** Lists data property assertions for the selected instance, such as 'hasAlias "BRAVO CONTRERAS MARICELA CLAUDIA"', 'hasEmail "mcbc@correo.azc.uam.mx"', 'hasAlias "MARICELA CLAUDIA BRAVO CONTRERAS"', 'hasEconomicNumber "35691"', 'hasAlias "MARICELA BRAVO"', 'hasProject "CLASES"', 'hasAlias "MARICELA C. BRAVO"', 'hasName "BRAVO CONTRERAS MARICELA CLAUDIA"', and 'hasGender "FEMENINO"'. Each assertion is preceded by a green checkmark.
- Negative object property assertions:** A section for negative assertions, currently empty.

Figura 4.15: Ejemplo de un individuo de la ontología de Persona (*Person*)

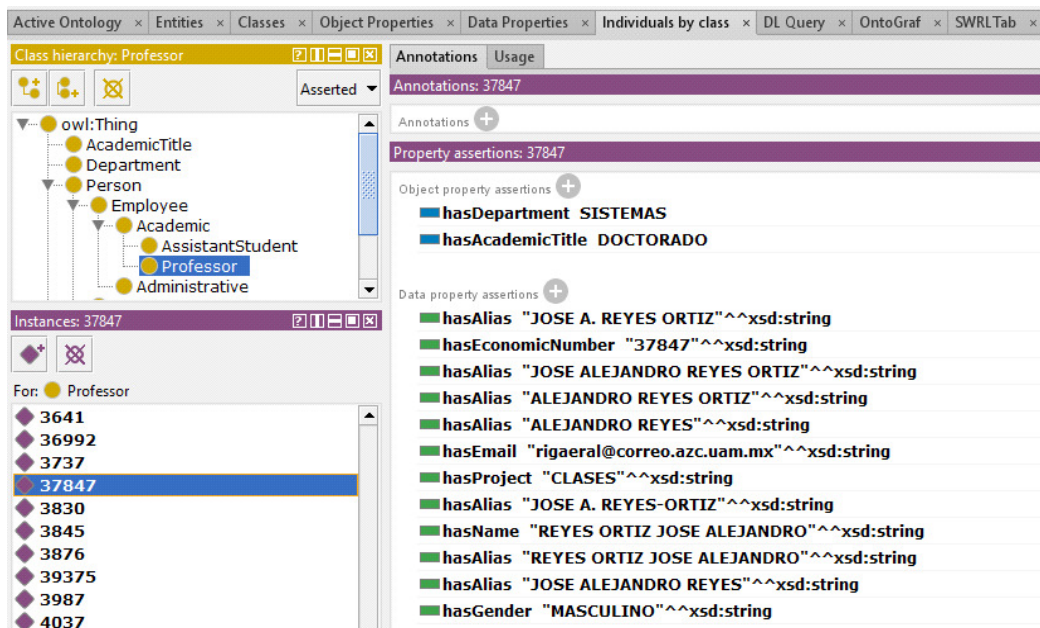


Figura 4.16: Ejemplo de un individuo de la ontología de Persona (*Person*)

4.2.2. Poblado de ontología de Publicación (*Publication*)

El *parser*, es decir, el método que separó cada una de las etiquetas en el archivo XML para que sean procesadas en el poblado de esta ontología se enlista a continuación:

1. Archivo XML. Se posicionó en la carpeta donde se encuentra el archivo con los datos de las Publicaciones.
2. Objeto tipo Profesor. Este objeto se utilizó como apoyo en el proceso de crear la relación de co-autoría (se explica más adelante).
3. Se creó un arreglo de cadenas para almacenar a los autores correspondientes en cada artículo.
4. Se creó un arreglo de objetos de tipo Profesor (*Professor*). Este arreglo se “llenó” con un método, el cual obtuvo todas las instancias de la ontología de Persona (*Person*), específicamente todas las instancias de tipo Profesor (*Professor*) y los insertó en una lista para tener acceso a ellos y encontrar las relaciones de co-autoría (se hizo en pasos posteriores). Este método se describe a continuación:

- Se creó un arreglo de objetos de tipo Profesor.
- Se creó una lista de instancias ontológicas para almacenar a cada uno de los Profesores.
- Comienza un ciclo que va desde la primer instancia a la última.
- Se creó un objeto individual de tipo Profesor (*Professor*).
- Se insertó cada *data property* en los atributos del objeto.
- Cada objeto de tipo Profesor se agregó en la lista de objetos.
- Termina el ciclo.
- El método regresó la lista de objetos de tipo Profesor con los atributos de nombre, identificador único del profesor (se utilizó como llave), género, correo electrónico, proyecto y alias.

En la figura 4.17 se muestra el diagrama de flujo de este proceso.

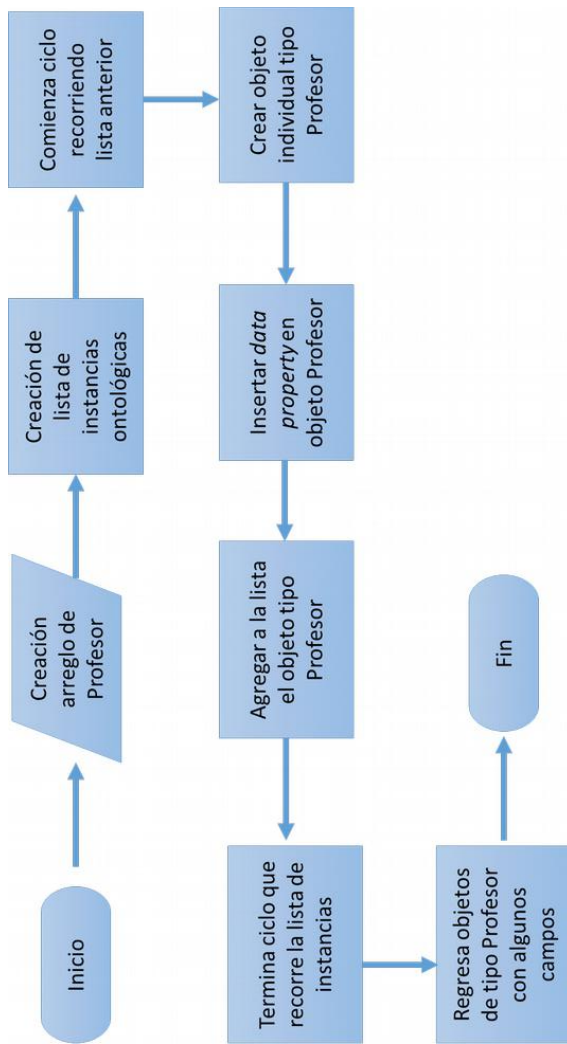


Figura 4.17: Diagrama de flujo del método que regresó la coincidencia entre un autor de una Publicación y un Profesor

5. Comienza el método a buscar en el archivo XML indicado la palabra “*inproceedings*” o “*article*”, que es donde se encontraron los datos de cada uno de los artículos.
6. Se creó un contador para llevar el número de registros que se insertaron con éxito y comienza la iteración.
7. Al principio de cada iteración se creó un nuevo objeto de tipo Publicación (*Publication*) y el arreglo de autores.
8. Se comparó cada etiqueta que se encuentra en el XML y dependiendo de ésta, se “*seteó*” dentro de un atributo del objeto Publicación (*Publication*). Los atributos que se tomaron en cuenta son:
 - Llave (*Key*)
 - Título (*Title*)
 - Páginas (*Pages*)
 - Año (*Year*)
 - Volumen (*Volume*)
 - Revista (*Journal*)
 - Número (*Number*)
 - URL
 - EE
9. Comienza el ciclo que itera sobre la lista de autores.
10. Para cada autor del DBLP se buscó si coincidió con algún alias de la lista del paso anterior.
11. Si se encontró coincidencia entonces continua en el siguiente paso, de lo contrario sale del ciclo y el artículo se omite.
12. Cuando hay coincidencia se procedió a instanciar e insertar el objeto de tipo Publicación (*Publication*) en la ontología.
13. Se incrementó en uno el contador del número de artículos insertados en la ontología.

14. Una vez instanciado el objeto, se agregó la *object property* de **tienePublicacion** (*hasPublish*) entre la instancia de tipo Publicación (*Publication*) y la instancia de Profesor (*Professor*).
15. De igual manera se creó la *object property* de la instancia de tipo Profesor (*Professor*) hacia la instancia de tipo Publicación (*Publication*) **esPublicadoPor** (*isPublishedBy*).
16. Fin del ciclo de leer etiquetas del archivo fuente (XML).
17. Fin del método.

En la figura 4.18 se muestra el diagrama de flujo de este proceso.

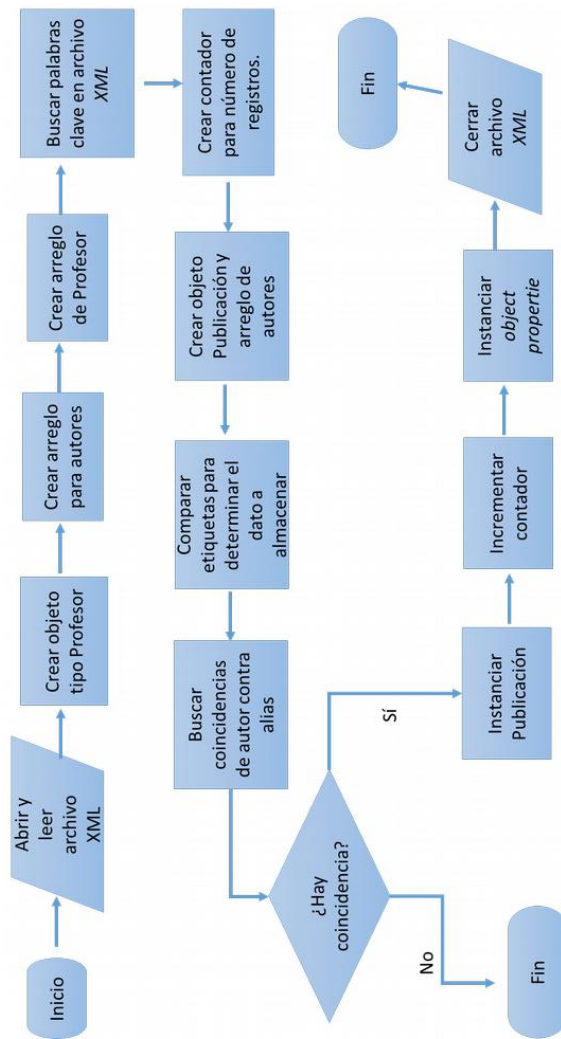


Figura 4.18: Diagrama de flujo del proceso de poblado de Publicación

A continuación se explica el método al que se hace referencia en el paso 13, es decir, el método que recibió como parámetros un objeto de tipo *Publicación* (*Publication*) y una cadena indicando la clase (que en este caso es la clase *Publicación*).

1. Como primer paso, este método manda a llamar a otro método al que se le pasó como parámetros un individuo y una clase.
 - Se creó la IRI de la clase.
 - Se creó la IRI del individuo.
 - Creación de la instancia del objeto en individuo de ontología.
 - Se insertó la instancia en la ontología (axioma).
 - Se guardó la ontología.
2. Se agregó cada una de las *data properties* después de haber instanciado el objeto.
3. Termina el método

En la figura 4.19 se muestra el diagrama de flujo de este proceso.

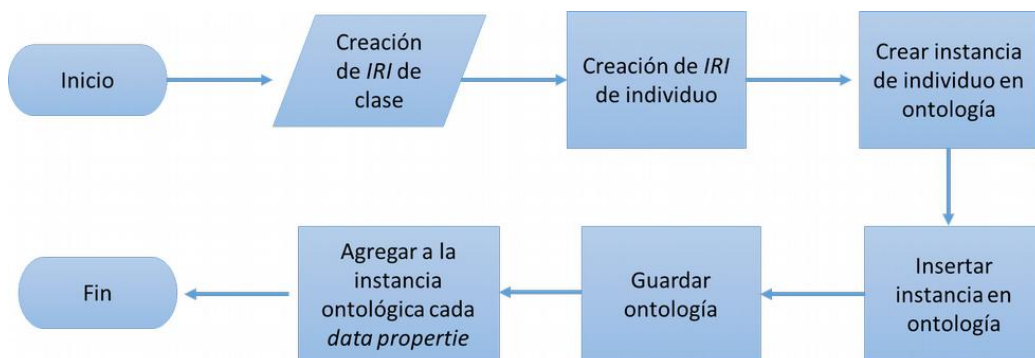


Figura 4.19: Diagrama de flujo del método que instancia una *Publicación*

En la figura 4.20 y la figura 4.21 se muestra un ejemplo de una instancia de la ontología de *Publicación* (*Publication*).

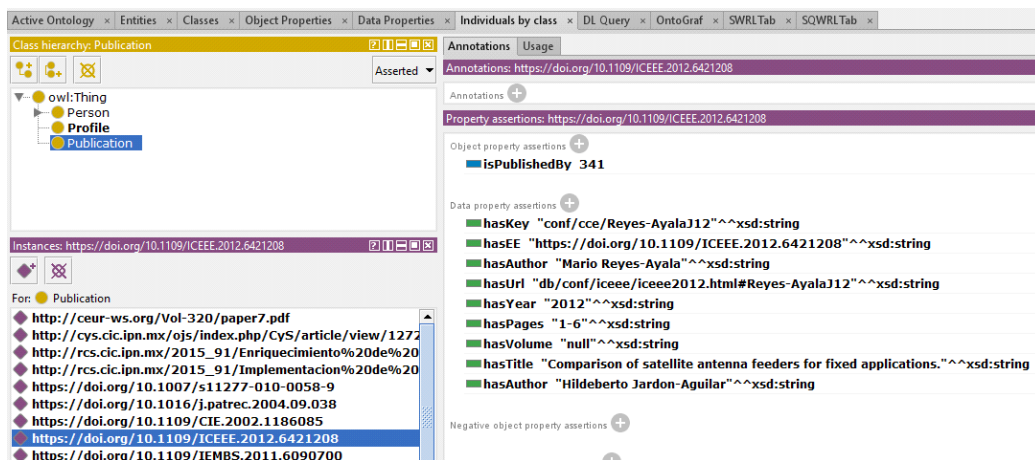


Figura 4.20: Ejemplo de un individuo de la ontología de *Publicación (Publication)*

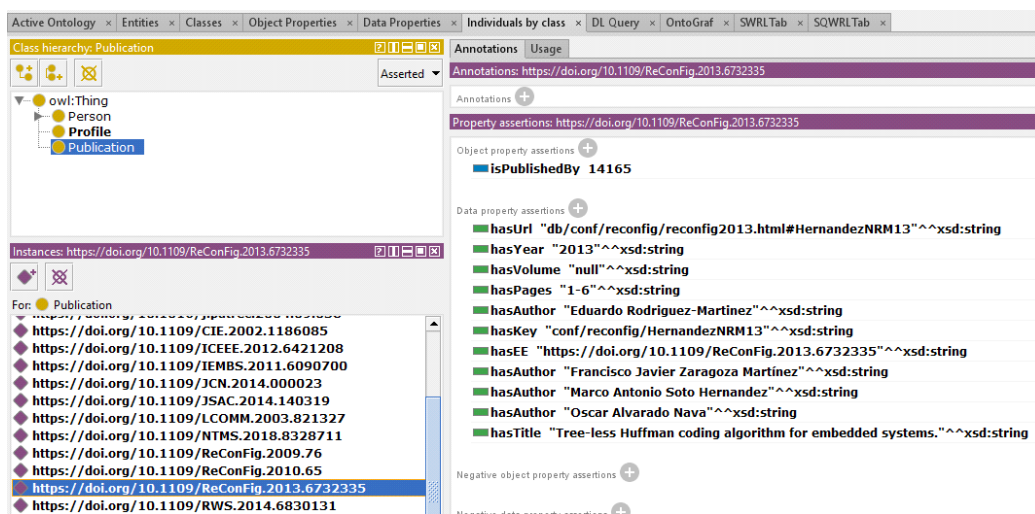


Figura 4.21: Segundo ejemplo de un individuo de la ontología de *Publicación (Publication)*

4.2.3. Poblado de ontología Perfil de investigación (Researcher profile)

Dentro de la ontología de Perfil de investigación (*Researcher profile*) no se hizo como tal un poblado, debido a que esta es una ontología que importa tanto a *Persona (Person)*

como a Publicación (*Publication*), por lo tanto los individuos que pertenecen a cada una de ellas se encuentra también dentro de la ontología de Perfil de investigación (*Researcher profile*).

En la figura 4.22 se muestra la ontología que importa a las otras dos (Persona - *Person* y Publicación - *Pulication*).

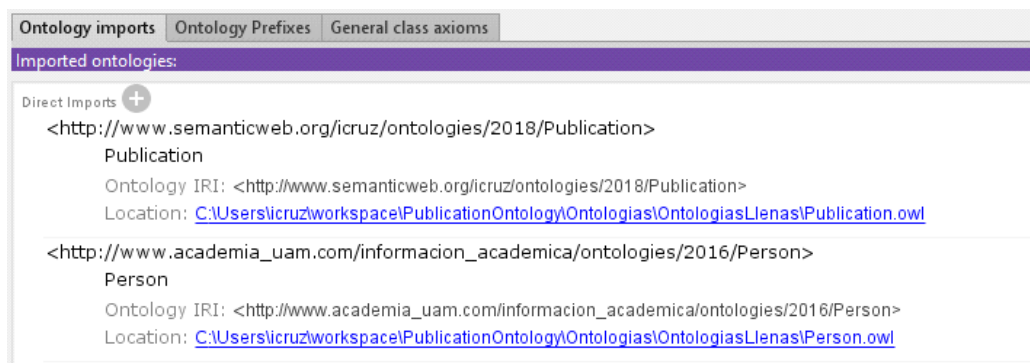


Figura 4.22: Ontología de Perfil de investigación (*Researcher profile*)

4.3. Enriquecimiento de ontología Perfil de investigador (*Researcher profile*)

El enriquecimiento ontológico es el proceso de extender una ontología con nuevos conceptos, relaciones y reglas. Se realiza cada vez que el conocimiento del dominio existente no es suficiente para explicar la información extraída del corpus. Por lo tanto, se espera que la actividad de enriquecimiento ontológico amplíe los conocimientos previos, a fin de explicar mejor la información extraída en el futuro.

Dado que se pueden agregar nuevos conceptos y relaciones durante el enriquecimiento, la estructura de la ontología cambia. Recordando nuestra discusión sobre el “pastel de capas”, el proceso de enriquecimiento involucró a todas las capas, a diferencia de la población de ontología, que se refiere solo a las capas inferiores. El enfoque principal adoptado por los métodos del estado del arte comienza con la identificación de los objetos y sus realizaciones - sinónimos alternativos. Cada objeto, junto con un posible conjunto de realizaciones alternativas, es un concepto candidato para ser agregado a una ontología. Avanzando a la tercera capa del “pastel”, cada conjunto propuesto de objetos y realizaciones alternativas que posiblemente representen un concepto se evaluó para decidir si constituye un concepto o no.

Si el objeto no representó un concepto, dicho concepto debe formularse creando una definición intencional y posiblemente aumentada con evidencia o instancias que justificaron

la adición del nuevo concepto. En la siguiente capa, las relaciones (ya sean taxonómicas o no taxonómicas) se identificaron entre conceptos, generalmente basados en información espacio-temporal para modalidades como imagen, video o información lingüística (ya sea sintáctica o semántica) para texto. Finalmente, para apoyar el razonamiento y derivar hechos no codificados explícitamente, pero derivables de la ontología, se adquirieron reglas y restricciones [1].

Para realizar el enriquecimiento de la ontología de Perfil de investigación (*Researcher profile*) se consideró la relación de co-autoría, para la creación de dicha relación se utilizaron los datos del archivo DBLP como fuente de datos y, por lo tanto, se obtuvo la información de las publicaciones de un usuario.

En la figura 4.23 se muestra la manera en que se relacionaron las ontologías de Publicación (*Publication*), Persona (*Person*) y Perfil de investigación (*Researcher profile*).

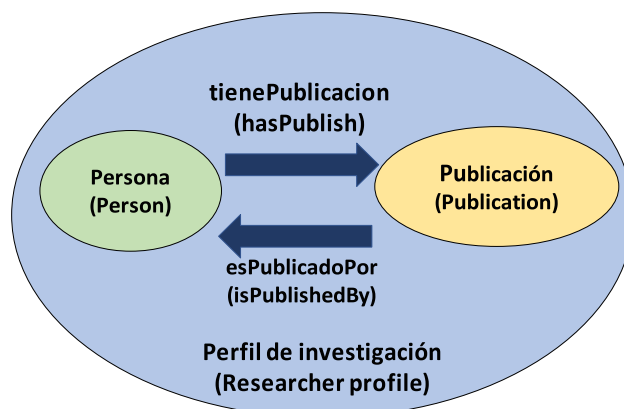


Figura 4.23: Representación de la relación entre ontologías (Persona y Publicación)

Después de identificar el alias del profesor con el nombre del autor del artículo, se tomó el identificador único en la ontología de Persona (*Person*) y sirvió para vincularse con la ontología de Publicación (*Publication*).

La relación que va de Persona (*Person*) a Publicación (*Publication*) se llama tienePublicacion (*hasPublish*), lo cual significa que un Profesor tiene una o más publicaciones; es decir, los artículos y sus autores se evalúan uno por uno y se encontraron tantas relaciones como coincidencias de autores contra alias. La relación que va de Publicación

(*Publication*) a Persona (*Person*) se llama *esPublicadoPor* (*isPublishedBy*), significa que un artículo tiene uno o más autores. Para el presente trabajo únicamente se tomaron en cuenta a los profesores de la Universidad Autónoma Metropolitana - unidad Azcapotzalco. Los artículos escritos por personas que no pertenecen a esta comunidad se descartaron.

Las dos relaciones anteriores se declararon como *objects properties* dentro del software *Protegé* con el cual se crearon las ontologías.

La regla que se creó para obtener los registros de la relación de coautoría es la siguiente:

$$Publication(?pub) \wedge isPublishedBy(?pub, ?prof1) \wedge isPublishedBy(?pub, ?prof2) \wedge differentFrom(?prof1, ?prof2) \rightarrow collaborateWith(?prof1, ?prof2, ?pub)$$

Esta regla significa que se tomó una publicación junto con la *object property* *esPublicadoPor* (*isPublishedBy*) la cual tiene dos parámetros que son la publicación y el autor, es decir, el profesor; comparando con una nueva relación de *esPublicadoPor* (*isPublishedBy*) con el mismo artículo pero diferente Profesor, hay que diferenciar al primer profesor con el segundo, esto para evitar que se almacene una relación de un profesor consigo mismo y con ayuda de la instrucción *differentFrom* se omitió la duplicidad y esta parte de crear esta última relación. Como encabezado de esta regla hacemos que la regla muestre al profesor 1, profesor 2 y la publicación.

En la pestaña de SWRL se ejecutó la regla anterior en tres pasos:

1. Se transfirieron las reglas de SWRL y los objetos o instancias que fueron requeridos por las reglas.
2. Se ejecutó el motor de reglas.
3. Se transfirieron los objetos o instancias inferidas del motor de reglas al conocimiento de OWL.

Una vez hechos estos tres pasos, se revisaron los individuos y después de que se creó la relación de *colaboraCon* (*collaborateWith*), la figura 4.24, 4.25 y 4.26 muestra un ejemplo de esta nueva *object property* que enriqueció la ontología de Perfil de investigación (*Researcher profile*).

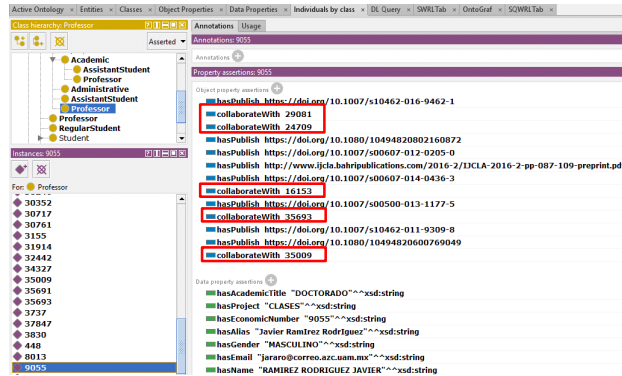


Figura 4.24: Ejemplo de relación de colaboración entre profesores

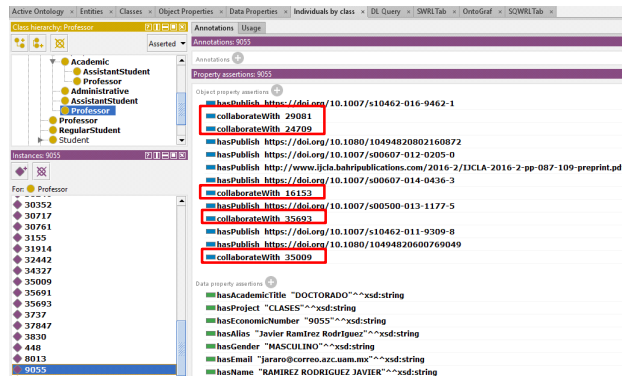


Figura 4.25: Segundo ejemplo de relación de colaboración entre profesores

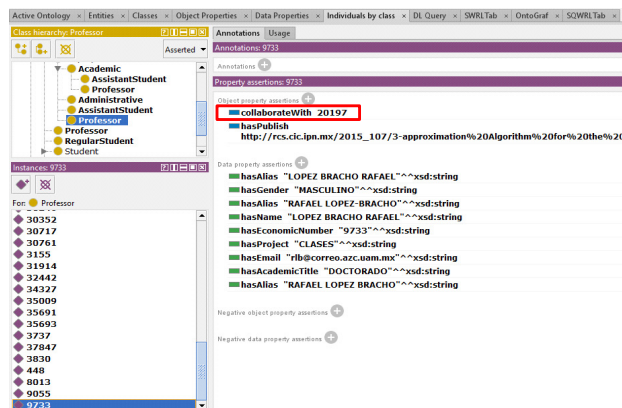


Figura 4.26: Tercer ejemplo de relación de colaboración entre profesores

4.4. Diagrama de clases

En esta sección se muestra el diagrama de clases con los atributos y métodos de cada una de las clases que se utilizó para dar solución al problema planteado. En la figura 4.27 se muestra dicho diagrama.

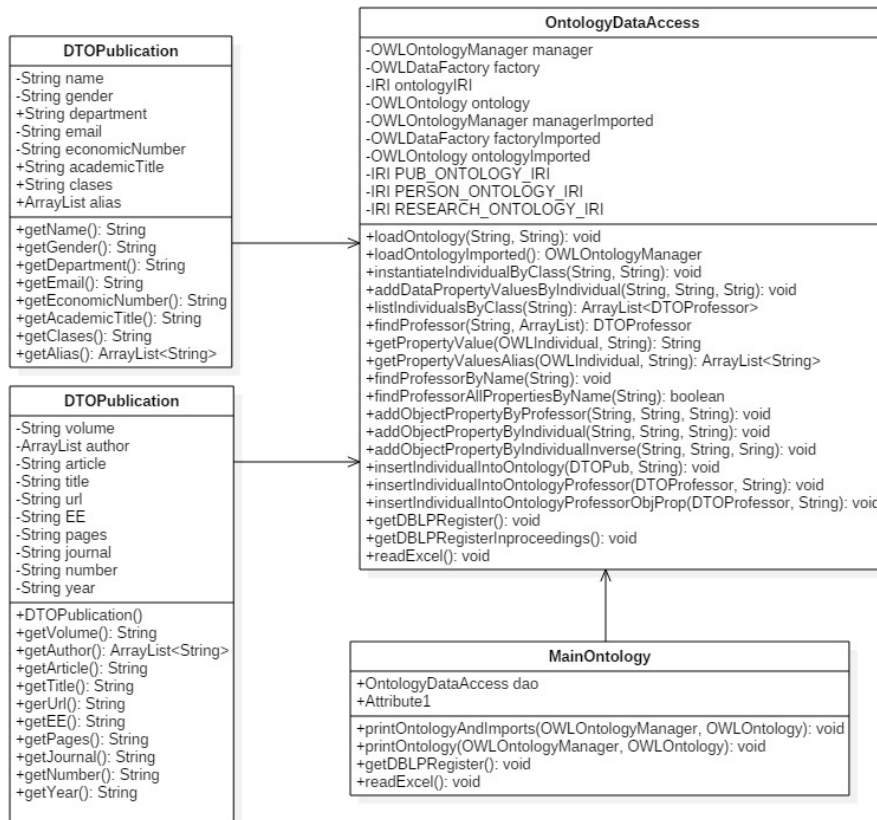


Figura 4.27: Diagrama de clases

El diagrama de clases está compuesto por lo siguiente:

- *DTOPublication*. Tiene la lista de atributos y accesoros (métodos *get* y *set*) de una Publicación (*Publication*).
- *DTOProfessor*. Tiene la lista de atributos y accesoros (métodos *get* y *set*) de un Profesor (*Professor*).
- *OntologyDataAccess*. Es la clase que contiene todos los métodos de extracción de información (Publicaciones y Profesores), poblado (ontología de Publicación, Profesor y Perfil de investigador), enriquecimiento, relaciones entre individuos de las diferentes ontologías. Esta clase es la que realiza las operaciones de extracción de información, poblado, relaciones entre individuos ontológicos y enriquecimiento.
- *MainOntology*. Clase principal que llama a cada uno de los métodos incluidos en la clase de *OntologyDataAccess*. En esta clase se invocan los métodos de leerExcel (*readExcel()*) y leerDBLP (*getDBLPRegister*), los cuales hacen el poblado automático y estos a su vez invocan los métodos de relaciones entre individuos, enriquecimiento ontológico, etc.

5 Experimentación

En esta sección se muestran las fuentes de datos utilizadas para dar solución al problema planteado sobre la representación de perfiles de investigación utilizando ontologías, así como la explicación de la estructura de cada una de las fuentes de datos. Se describe cada uno de los escenarios de pruebas con las cuales se evaluaron las ontologías y se presentan los resultados de estas pruebas. Dichos perfiles de investigación corresponden al personal académico de la Universidad Autónoma Metropolitana - unidad Azcapotzalco.

5.1. Conjunto de datos

En el presente trabajo se utilizaron dos fuentes de datos:

1. Una base de datos en Excel con datos de los profesores (nombre, correo, título).
2. Archivo XML con etiquetas, las cuales indican cada uno de los atributos de una Publicación, esta fuente de datos se describe y explica más adelante.

En la ontología de Persona (*Person*) hay una ramificación hacia empleado, académico y después se encuentra la clase Profesor. El archivo tiene la siguiente información:

- Nombre
- Género
- Departamento
- Correo electrónico
- Identificador único del profesor

- Proyecto en el que se encuentra
- Alias

De acuerdo al lenguaje de ontologías, cada uno de estos atributos se convirtió en *data property* o en *object property*, además de indicar el tipo de dato, es decir, si el campo contiene un campo de tipo cadena, entero, booleano, double, fecha, etc.

5.1.1. Estructura DBLP

En esta sección se mostrará la estructura del archivo DBLP en formato XML. Este archivo tiene diferentes etiquetas que representan un atributo de cierto tipo de objetos. Dentro del DBLP encontramos artículos científicos, tesis, *journals*, entre otros.

En el presente trabajo se trabajará con los bloques que están entre las etiquetas de “*article*” y entre las etiquetas “*inproceedings*”. La estructura que corresponde al bloque de etiquetas “*article*” es la siguiente:

- *Article*, esta etiqueta contiene una fecha de registro y una llave que se conforma del tipo de publicación, el apellido paterno del autor y un identificador formado de letras y números.
- *Author*, esta etiqueta contiene el nombre(s) del(los) autor(es) que están involucrados en la publicación.
- *Title*, en esta etiqueta encontramos el título de la publicación.
- *Pages*, en esta etiqueta se encuentra el número de páginas de la publicación.
- *Year*, etiqueta que indica el año de la publicación.
- *Volume*, etiqueta que contiene el volumen de la publicación.
- *Journal*, en esta etiqueta se encuentra la revista a la que pertenece la publicación.
- *Number*
- *EE*, en esta etiqueta se encuentra la URL que identifica a la publicación.

- *URL*, en esta etiqueta se indica el lugar donde se almacena la publicación.

La estructura de los “*inproceedings*” es la siguiente:

- *Inproceedings*
- *Author*, esta etiqueta contiene el nombre(s) del(los) autor(es) que están involucrados en la publicación.
- *Title*, en esta etiqueta encontramos el título de la Publicación.
- *Pages*, en esta etiqueta se encuentra el número de páginas de la Publicación.
- *Year*, etiqueta que indica el año de la publicación.
- *Booktitle*
- *EE*, en esta etiqueta se encuentra la URL que identifica a la publicación.
- *Crossref*
- *URL*, en esta etiqueta se indica el lugar donde se almacena la publicación.

5.1.2. Ejemplo de artículo

En la figura 5.1 y la figura 5.2 se muestran dos ejemplos del conjunto de etiquetas que se utilizan para describir un artículo.

```
<article mdate="2018-05-29" key="journals/jifs/Reyes-OrtizB18">
<author>Jos&eacute; A. Reyes-Ortiz</author>
<author>Maricela Bravo</author>
<title>Enhancing patterns with linguistic information for criminal event recognition.</title>
<pages>3027-3036</pages>
<year>2018</year>
<volume>34</volume>
<journal>Journal of Intelligent and Fuzzy Systems</journal>
<number>5</number>
<ee>https://doi.org/10.3233/JIFS-169487</ee>
<url>db/journals/jifs/jifs34.html#Reyes-OrtizB18</url>
</article>
```

Figura 5.1: Ejemplo de etiquetas de un archivo DBLP

```

<article mdate="2017-05-26" key="journals/ijwgs/BravoOHSRL06">
<author>Maricela Bravo</author>
<author>Joaquin Perez Ortega</author>
<author>Jose Conrado Velazquez Hernandez</author>
<author>Victor Jesus Sosa Sosa</author>
<author>Azucena Montes Rendon</author>
<author>Maximo Lopez</author>
<title>Design of a shared ontology used for translating negotiation primitives.</title>
<pages>237-259</pages>
<year>2006</year>
<volume>2</volume>
<journal>IJWGS</journal>
<number>3</number>
<ee>https://doi.org/10.1504/IJWGS.2006.011356</ee>
<url>db/journals/ijwgs/ijwgs2.html#BravoOHSRL06</url>
</article>

```

Figura 5.2: Segundo ejemplo de etiquetas de un archivo DBLP

5.2. Plan experimental

Para las pruebas correspondientes, se crearon cuatro escenarios diferentes. La distribución de profesores se muestra en la tabla 5.1:

| Departamento | Cantidad |
|------------------|----------|
| Ciencias básicas | 144 |
| Electrónica | 60 |
| Energía | 67 |
| Materiales | 52 |
| Sistemas | 50 |

Tabla 5.1: Tabla de profesores por departamento.

5.2.1. Descripción de escenarios

En esta sección se describe cada uno de los escenarios que se probaron. La fuente de datos de los profesores se dividió por departamentos, para una primera prueba, se tomaron 15 profesores del departamento de sistemas, se instanciaron en la ontología y se corrió el método para encontrar coincidencias entre todas las publicaciones que contiene el archivo XML. En este escenario de prueba se encontraron 65 publicaciones en un tiempo de 90 minutos.

Para los siguientes escenarios, se tomaron todos los profesores del departamento de sistemas, en otro escenario se tomaron los profesores del departamento de electrónica y los de ciencias básicas.

5.2.1.1. Escenario 1.- Inserción de profesores del departamento de Sistemas en la ontología de Persona (Person) y una parte del archivo DBLP

En la ontología de Persona (*Person*) de un total de 50 profesores que pertenecen al departamento de Sistemas, se insertaron los 50 profesores, es decir el 100 % se convirtió en una instancia de la ontología de manera exitosa.

Una vez instanciados los profesores se procede a correr el método de poblado de la ontología de Publicación (*Publication*). El archivo XML correspondiente contiene aproximadamente un millón de artículos, es decir, 35 millones de líneas.

De todo el archivo DBLP se obtuvieron aquellas secciones que iniciaran y terminaran con la palabra *article* o la palabra *inproceeding*. El archivo resultante se dividió en 13 y 15 archivos y se evaluó el tiempo que le tomó al programa procesar cada uno de ellos y el número de registros que se insertaron de manera exitosa.

Para este primer escenario, únicamente se tomaron en cuenta las secciones con la etiqueta *article* y con 13 archivos XML en total, el resultado se indica a continuación:

| Archivo | Registros insertados | Tiempo (min) |
|------------|----------------------|--------------|
| Archivo 1 | 1 | 1 |
| Archivo 2 | 0 | 1 |
| Archivo 3 | 6 | 2 |
| Archivo 4 | 4 | 1 |
| Archivo 5 | 2 | 66 |
| Archivo 6 | 4 | 2 |
| Archivo 7 | 3 | 2 |
| Archivo 8 | 1 | 1 |
| Archivo 9 | 1 | 1 |
| Archivo 10 | 0 | 45 |
| Archivo 11 | 0 | 1 |
| Archivo 12 | 0 | 1 |
| Archivo 13 | 0 | 1 |
| Total | 22 | 125 |

Tabla 5.2: Resultados del primer escenario.

Para los siguientes escenarios se utiliza todo el DBLP, es decir, buscar en las más de 35 millones de líneas de este archivo para comparar si existen más artículos que se pueden insertar en la ontología de Publicación (*Publication*).

5.2.1.2. Escenario 2.- Inserción de profesores del departamento de Sistemas en la ontología de Persona (Person) y archivo completo de DBLP

En el segundo escenario, se toma un total de 53 profesores, es decir, el 100% de los profesores del departamento de Sistemas y todo el archivo de DBLP, el cual se dividió en 12 archivos y uno por uno se evaluó en el método que registra las instancias en la ontología de Publicación (*Publication*). El resultado es el siguiente:

| Archivo | Registros insertados | Tiempo (min) |
|------------|----------------------|--------------|
| Archivo 1 | 5 | 1 |
| Archivo 2 | 9 | 1 |
| Archivo 3 | 34 | 3 |
| Archivo 4 | 17 | 1 |
| Archivo 5 | 25 | 65 |
| Archivo 6 | 8 | 1 |
| Archivo 7 | 3 | 1 |
| Archivo 8 | 11 | 1 |
| Archivo 9 | 24 | 1 |
| Archivo 10 | 9 | 68 |
| Archivo 11 | 0 | 1 |
| Archivo 12 | 0 | 1 |
| Total | 145 | 145 |

Tabla 5.3: Resultados del segundo escenario.

5.2.1.3. Escenario 3.- Inserción de profesores del departamento de Electrónica en la ontología de Persona (Person) y archivo completo de DBLP

En el siguiente escenario, se probó con los profesores del departamento de electrónica, con un total de 144 profesores, se examinó todo el DBLP para encontrar coincidencias, al igual que en el escenario anterior, el DBLP se dividió en 12 archivos diferentes. El resultado es el siguiente:

| Archivo | Registros insertados | Tiempo (min) |
|------------|----------------------|--------------|
| Archivo 1 | 0 | 1 |
| Archivo 2 | 0 | 1 |
| Archivo 3 | 4 | 1 |
| Archivo 4 | 1 | 1 |
| Archivo 5 | 7 | 46 |
| Archivo 6 | 0 | 1 |
| Archivo 7 | 0 | 1 |
| Archivo 8 | 0 | 1 |
| Archivo 9 | 0 | 1 |
| Archivo 10 | 1 | 122 |
| Archivo 11 | 0 | 1 |
| Archivo 12 | 0 | 1 |
| Total | 13 | 178 |

Tabla 5.4: Resultados del tercer escenario.

5.2.1.4. Escenario 4.- Inserción de profesores del departamento de Sistemas y Electrónica en la ontología de Persona (Person) y archivo completo de DBLP

Para este escenario se agregaron alias a cada uno de los nombres de la muestra de 53 profesores de los departamento de sistemas. Estos alias son nombres alternativos que los profesores utilizan en sus publicaciones. Los archivos en que se dividió el DBLP continúan siendo 12 y el resultado se describe a continuación:

| Archivo | Registros insertados | Tiempo (min) |
|------------|----------------------|--------------|
| Archivo 1 | 5 | 1 |
| Archivo 2 | 10 | 1 |
| Archivo 3 | 34 | 3 |
| Archivo 4 | 17 | 1 |
| Archivo 5 | 26 | 120 |
| Archivo 6 | 8 | 1 |
| Archivo 7 | 3 | 1 |
| Archivo 8 | 12 | 1 |
| Archivo 9 | 25 | 1 |
| Archivo 10 | 10 | 70 |
| Archivo 11 | 0 | 1 |
| Archivo 12 | 0 | 1 |
| Total | 150 | 202 |

Tabla 5.5: Resultados del cuarto escenario.

5.3. Evaluación del modelo ontológico

Las siguientes preguntas de competencia se utilizaron para evaluar el modelo ontológico:

1. Pregunta de competencia:

¿Cuáles son las publicaciones del profesor José Alejandro Reyes Ortiz?

Pregunta en lenguaje SQWRL:

Professor(?prof1) \wedge hasName (?prof1, ?n) \wedge hasPublish(?prof1, ?pub) \wedge swrlb:equal(?e, "REYES ORTIZ JOSE ALEJANDRO") \rightarrow sqwrl:select(?n, ?pub)

En la figura 5.3 se muestra el resultado de la ejecución de esta regla

| n | pub |
|----------------------------|---|
| REYES ORTIZ JOSE ALEJANDRO | http://rcs.cic.ipn.mx/2015_97/Clasificacion%20de%20servicios%20Web%20mediante%20una%20red%20neuronal%20artificial%20usando%20n-gramas%20de%20palabras.pdf |
| REYES ORTIZ JOSE ALEJANDRO | https://doi.org/10.1007/978-3-662-45563-0_49 |
| REYES ORTIZ JOSE ALEJANDRO | http://rcs.cic.ipn.mx/2016_130/A%20Lexical%20Supervised%20Approach%20for%20Opinion%20Mining%20in%20the%20Domain%20of%20Laptops%20and%20Restaurants.pdf |
| REYES ORTIZ JOSE ALEJANDRO | https://doi.org/10.1109/ECOCN.2014.704823 |
| REYES ORTIZ JOSE ALEJANDRO | http://rcs.cic.ipn.mx/2017_134/Mineria%20de%20opiniones%20de%20entrada%20en%20temas%20de%20interior%20de%20textos%20en%20espa%20ol.pdf |
| REYES ORTIZ JOSE ALEJANDRO | http://rcs.cic.ipn.mx/2015_95/Plataforma%20de%20ontologias%20de%20perfiles%20de%20academicos%20a%20partir%20de%20textos%20en%20espa%20ol.pdf |
| REYES ORTIZ JOSE ALEJANDRO | http://www.micai.org/rcs/2013_68/Ontology-based%20Knowledge%20Representation%20for%20Supporting%20Medical%20Decisions.pdf |
| REYES ORTIZ JOSE ALEJANDRO | https://doi.org/10.3233/978-1-60947 |
| REYES ORTIZ JOSE ALEJANDRO | https://doi.org/10.1007/978-3-642-53842-1_11 |

Figura 5.3: Respuesta a primer pregunta de competencia

2. Pregunta de competencia:

¿Con quién colabora el Dr. José Alejandro Reyes Ortiz?

Pregunta en lenguaje SQWRL:

Professor(?p1) ∧ Professor(?p2) ∧ collaborateWith(?p1, ?p2) ∧ hasName(?p1, ?profesor1) ∧ hasName(?p2, ?profesor2) ∧ differentFrom(?p1, ?p2) ∧ swrlb:equal(?profesor1, "REYES ORTIZ JOSE ALEJANDRO") -¿sqwrl:select(?profesor1, ?profesor2)

En la figura 5.4 se muestra el resultado de la ejecución de esta regla

| profesor1 | profesor2 |
|----------------------------|----------------------------------|
| REYES ORTIZ JOSE ALEJANDRO | BRAVO CONTRERAS MARICELA CLAUDIA |
| REYES ORTIZ JOSE ALEJANDRO | HERRERA ALCANTARA OSCAR |

Figura 5.4: Respuesta a segunda pregunta de competencia

3. Pregunta de competencia:

¿Cuáles son las publicaciones de la Dra. Maricela Bravo?

Pregunta en lenguaje SQWRL:

Professor(?prof1) ∧ hasName(?prof1, ?nombre) ∧ hasPublish(?prof1, ?publication) ∧ swrlb:equal(?e, "BRAVO CONTRERAS MARICELA CLAUDIA") → sqwrl:select(?nombre, ?publication)

En la figura 5.5 se muestra el resultado de la ejecución de esta regla

En la figura 5.7 se muestra el resultado de la ejecución de esta regla

| | | |
|----------------|---|----|
| S6 | Professor(?prof) ^ hasDepartment(?prof, ?dep) ^ swrlb:contains(?dep, "SISTEMAS") ^ hasPublish(?prof, ?pub) -> sqwrl:count(?pub) | |
| SQWRL Queries | OWL 2 RL | S6 |
| count(?pub) | | |
| "123"^^xsd:int | | |

Figura 5.7: Respuesta a quinta pregunta de competencia

6. Pregunta de competencia:

¿Cuántas mujeres del departamento de sistemas tienen al menos una publicación?

Pregunta en lenguaje SQWRL:

*Professor(?prof) ^ hasGender(?prof, ?gen) ^ swrlb:equal(?gen, "FEMENINO")
 ^ hasDepartment(?prof, ?dep) ^ swrlb:equal(?dep, "SISTEMAS") ^ hasPublish(?prof, ?pub) → sqwrl:count(?prof)*

En la figura 5.8 se muestra el resultado de la ejecución de esta regla

| | | | |
|---------------|--|----|-----|
| S10 | Professor(?prof) ^ hasGender(?prof, ?gen) ^ swrlb:equal(?gen, "FEMENINO") ^ hasDepartment(?prof, ?dep) ^ swrlb:equal(?dep, "SISTEMAS") ^ hasPublish(?prof, ?pub) -> sqwrl:count(?prof) | | |
| SQWRL Queries | OWL 2 RL | S6 | S10 |
| count(?prof) | | | |
| "49"^^xsd:int | | | |

Figura 5.8: Respuesta a sexta pregunta de competencia

7. Pregunta de competencia:

¿Cuántas publicaciones tiene el departamento de sistemas en el año 2014?

Pregunta en lenguaje SQWRL:

Publication(?pub) ^ hasYear(?pub, ?year) ^ swrlb:equal(?year, "2014") → sqwrl:count(?pub)

En la figura 5.9 se muestra el resultado de la ejecución de esta regla

| S11 | | | | |
|--|----------|----|-----|-------|
| Publication(?pub) ^ hasYear(?pub, ?year) ^ swrlb:equal(?year, "2014") -> sqwrl:count(?pub) | | | | |
| SQWRL Queries | OWL 2 RL | S6 | S10 | S11 |
| | | | | count |
| "16"^^xsd:int | | | | |

Figura 5.9: Respuesta a séptima pregunta de competencia

5.4. Análisis de resultados

Para el análisis de los resultados se comprobó de una manera manual, a través de la página del DBLP, las publicaciones de algunos profesores. De esto, podemos concluir lo siguiente:

5.4.1. Escenario 1

En el primer escenario con los profesores del departamento de Sistemas se insertaron un total 22 registros de Publicaciones. En la tabla 5.6 se muestran los resultados de este primer escenario:

| Profesores | Total de artículos | Artículos insertados | Porcentaje | Tiempo total (min) |
|------------|--------------------|----------------------|------------|--------------------|
| 53 | 150 | 22 | 14.66 % | 125 |

Tabla 5.6: Tabla de resultados primer escenario.

5.4.2. Escenario 2

En el segundo escenario con todos los profesores del departamento de Sistemas se insertaron un total 145 registros de Publicaciones. En la tabla 5.7 se muestran los resultados de este escenario:

| Profesores | Total de artículos | Artículos insertados | Porcentaje | Tiempo total (min) |
|------------|--------------------|----------------------|------------|--------------------|
| 53 | 150 | 145 | 96.66 % | 145 |

Tabla 5.7: Tabla de resultados segundo escenario.

5.4.3. Escenario 3

El tercer escenario se formó con todos los profesores del departamento de Electrónica, se insertaron un total de 13 registros de publicaciones y 144 profesores. En la tabla 5.8 se muestran los resultados de este escenario:

| Profesores | Total de artículos | Artículos insertados | Porcentaje | Tiempo total (min) |
|------------|--------------------|----------------------|------------|--------------------|
| 144 | 150 | 13 | 8.66% | 178 |

Tabla 5.8: Tabla de resultados tercer escenario.

5.4.4. Escenario 4

El cuarto escenario se formó con todos los profesores del departamento de Sistemas con los alias extra que se obtuvieron, se insertaron un total de 150 registros de Publicaciones y 53 profesores (50 profesores del departamento de Sistemas y 3 profesores del departamento de Electrónica). En la tabla 5.9 se muestran los resultados de este escenario:

| Profesores | Total de artículos | Artículos insertados | Porcentaje | Tiempo total (min) |
|------------|--------------------|----------------------|------------|--------------------|
| 53 | 150 | 150 | 100% | 200 |

Tabla 5.9: Tabla de resultados cuarto escenario.

6 Conclusión y trabajo a futuro

En el presente trabajo se utilizó el poblado de ontologías y el enriquecimiento de las mismas para construir el perfil de investigador. Para el poblado de la ontología de Persona (*Person*) se utilizó una muestra de 373 profesores, se tomaron dos departamentos, 50 son del departamento de Sistemas y 60 de Electrónica. Se insertaron el 100% de ellos en la ontología de Persona (*Person*), específicamente en la clase de Profesor (*Professor*).

Se dividieron las ontologías y se utilizó una para profesores del departamento de electrónica y otra para profesores del departamento de sistemas. La ontología de sistemas tuvo un total de 50 profesores. Una vez dividido el universo de profesores, se procedió a poblar la ontología de Publicación (*Publication*) con ayuda del archivo DBLP, el cual contiene aproximadamente un millón de artículos y más de 35 millones de líneas. El resultado del poblado de la ontología de publicación de profesores de sistemas dio como resultado un total de 135 publicaciones que coincidieron entre los alias de los profesores y los autores indicados dentro de la etiqueta *author* del DBLP. Al mismo tiempo se encontraron 116 relaciones de colaboración entre los profesores del departamento de sistemas.

La ontología de profesores del departamento de electrónica, con un total de 58 profesores se sometió a la misma prueba que la ontología de profesores del departamento de sistemas y se encontraron 22 publicaciones de la misma muestra del DBLP, es decir, se compararon todos los alias de los profesores de electrónica contra los autores del millón de artículos del DBLP. En esta ontología se encontraron 4 relaciones de colaboración, es decir, en cuatro publicaciones participaron dos o más profesores del mismo departamento.

Ventajas de utilizar ontologías sobre bases de datos:

- A partir de las ontologías se pueden obtener relaciones nuevas entre clases, primero

se deben establecer ciertas reglas, es decir, enriquecer la ontología haciendo un cruce entre los individuos (los que se indiquen en la regla).

- Las ontologías permiten trabajar con conceptos, en lugar de palabras clave.
- Simplificar la unificación (o traducción) de distintas representaciones.
- Las ontologías utilizan axiomas, los cuales son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “Si A y B son de la clase C, entonces A no es subclase de B”, “Para todo A que cumpla la condición C1, A es B”, etc.

Una desventaja que se observó al realizar las pruebas con los escenarios, es que entre más axiomas e individuos tenga la ontología, el enriquecimiento o el hecho de verificar la consistencia de la ontología puede tomar mucho tiempo y consumir casi la totalidad de la memoria RAM de la computadora que se esté utilizando y la memoria virtual de Java, en ocasiones estas tareas podrían no terminar nunca.

Como trabajo a futuro se pueden considerar otras fuentes para continuar la alimentación de las ontologías y el enriquecimiento de las mismas, como Arnet Miner que contiene los abstract y keywords de publicaciones y de esta manera relacionar los artículos con ciertos temas de interés para cada uno de los profesores de la ontología de Persona. De esta manera la ontología cumpliría con la característica de ser escalable y hacer que el perfil de cada profesor sea un poco más completo, específico, determinante y en caso de que otro profesor necesite investigar sobre un tema puede recurrir a la ontología para descubrir quién ha publicado sobre el tema de interés y colaborar con él o simplemente recibir asesoría. También se puede considerar tomar en cuenta a otros profesores o personal que no sea precisamente docente o empleado, basta con que tenga una o más publicaciones para que sea considerado y cuente con un perfil de investigador.

Anexos



The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks
(EUSPN 2018)

Ontology for Academic Context Reasoning

Maricela Bravo*, José A. Reyes-Ortiz, Isabel Cruz-Ruiz, Ariadna Gutiérrez-Rosales, and
Josué Padilla-Cuevas

Autonomous Metropolitan University, San Pablo 180, Azcapotzalco, CDMX, México

Abstract

Ontologies have gained popularity in the scientific community as representational mechanisms to support intelligent reasoning and execute inferences. In this paper we describe an ontology designed specifically to represent academic contexts at a public university. This model consists of a collection of ontologies designed to represent persons, physical space, sensor networks, events, etc. among other entities that exist in the academic environment. In particular, we describe the design requirements that guided the construction of the ontologies. The resulting ontology model is evaluated considering the competency of the ontology, and the concept domain coverage. Results are promising and the set of competency questions are translated to queries showing that the ontology model adheres to the requirements.

© 2018 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Selection and peer-review under responsibility of the scientific committee of EUSPN 2018.

Keywords: Context ontology; Ontology design; Ontology competency

1. Introduction

The myriad of daily activities occurring at universities generate a high demand for information services, which causes the requirement to process large volumes of data and offer information to users in a fast, concise, and pertinent manner. These information requirements vary according to the context of the user, the reasons why he / she is in a given building of the institution, and the time in which an event happens. Commonly, users who demand information services at the university are teachers, researchers, students, employees and visitors. The type of information requested ranges from course schedules, location of facilities, location of professors, etc. Therefore, the

* Corresponding author. Tel.: +52 55 53 18 95 32 ext. 149.

E-mail address: mcbc@correo.azc.uam.mx

processing of users contexts must consider various situations and characteristics: a) public universities count with wireless network technologies through which they offer Internet access and diverse applications for users at the university; b) most people at the university have mobile devices capable of wireless network connection; and c) universities count with sufficient computing resources to support the user's duties. Given these conditions, it is possible to develop the computational infrastructure to facilitate smart interconnection of devices and data processing, that is, to implement Internet of Things (IoT) technology. However, the implementation of the IoT presents a series of challenges that must be solved so that the interconnection of objects and people in a university can be efficiently exploited. Among the most relevant problems to solve are:

- a) Context management and service personalization. Context management consists of context acquisition, context processing, and context reasoning. Service personalization is the intelligent outcome of context management.
- b) Event detection, management, and activation. Event management requires the representation of events at the academic environment in such a way that it is possible to know *What* happened? *When* happened? *Where*? and *Who* was involved? Handling of events requires the detection, registration and activation of actions in response to events.
- c) The efficient interaction between intelligent information systems and hardware. This interaction requires interoperability between communication protocols, and the dynamic interconnection of mobile devices with wireless networks and sensors/actuators networks.

There are other relevant and derived problems. However, in this article we present an ontology solution approach for the representation and management of contexts and events in the university environment addressing particularly academic issues. Ontologies were selected as the formal representational mechanism as they are very popular in IoT based applications. Ontologies facilitate reusability, knowledge sharing, and execution of formal reasoning tasks such as satisfiability of concepts, consistency checking, classification and inference.

2. Related Work

The revision of the related work was done emphasizing the use of ontologies for context representation, management and reasoning. In 2003 Chen, Finin, and Joshi [1] described CoBrA, a context broker agent architecture that is capable of managing a shared model of the context and reasoning support for context-aware applications. Later in 2004 authors detailed the SOUPA ontology [2] which consists of vocabularies for describing person contact information, beliefs, desires, and intentions of an agent, actions, policies, time, space, and events. Razmerita, Angehrn, and Maedche [3] presented in 2003 OntobUM, a generic ontology-based user modeling architecture. This architecture integrates three ontologies: the user ontology, the domain ontology, and the log ontology. Later in 2007 [4] authors augmented their OntobUM model by representing the behavior of the user, such as: level of activity, type of activity, level of knowledge sharing, etc. Wang et al. [5] described in 2004 CONON, an ontology for modeling context in pervasive computing environments divided into upper ontology and specific ontology. The upper ontology model defines computational entity, location, person and activity as the most important entities of a context model. Later in 2004 [6] authors presented SOCAM, a Service-Oriented Context Aware Middleware architecture to support the construction of context-aware services in intelligent environments. SOCAM architecture incorporates CONON ontology. Preuveneers et al. [7] presented in 2004 CoDAMoS, an extensible context ontology for ambient intelligence, which describes four main concepts: user, environment, platform, and service. Authors described the requirements for ambient intelligence: application adaptability, resource awareness, mobile services, semantic service discovery, code generation, and context-aware user interfaces. In 2007 Ejigu et al. [8] presented an ontology-based context model to facilitate context reasoning by providing structure for contexts, rules and their semantics. This work takes the basic elements of a pervasive computing environment characterized by their dynamicity, heterogeneity, and ubiquity of users, devices and resources, ad hoc connection between devices; and the existence of logical and physical sensors. The sources of context information are implemented as classes: User, Device, Application, Physical Environment, Resource, Network, Location, and Activity; all these classes are subclasses of a general class Context. In 2010 Zainol and Nakata [9] presented a Generic Context Ontology Model to represent context information in general. The aim is to facilitate the common context representation, context matching and context reasoning. This approach is based on the idea that a well-defined context model will minimize the

complexity of context-aware systems enhancing their maintainability. This model consists of a formal specification of the semantics of context identifiers; allowing sharing knowledge among different resources. In 2010 Poveda-Villalón et al. [10] presented mIO! an ontology network for a mobile environment. mIO! Ontology consists of eleven modular ontologies: user, role, environment, location, time, service, provider, device, interface, source, and network. This ontology covers a wide range of concepts related with context representation. Skillen et al. [11] presented in 2012 a user profile model for context-aware application personalization; authors concentrated on concepts to model a dynamic context: user time, user location, user activity, and user context. In 2013 Guermah et al. [12] described an architecture for the development of context-aware services based on ontologies. This architecture is composed of three main elements: a meta-model of context, an ontology for the meta-model, and a reasoning engine. The context is modeled based on a meta-model that defines the context and sub-contexts. Context properties are gathered from sensors, each property has a context validity and context specification. In 2014 Nadozeva and Kiritsis [13] presented an ontology-based context model to capture general concepts about users and business. The aim of this work is to propose an ontology-based model and rules to classify the context of users and business. This ontology describes general concepts: space, matter, object, event and action. Domain specific extensions specify the vocabulary and properties related to a generic domain by specializing terms in the upper ontology. This model is composed of three sub-models: user context, business context, and information feature model. In 2015 Kayes, A. Han, J, Colman A. [14] presented Ontology-based Context-Aware Access Control (OntCAAC) a generic framework that models dynamic contexts and access control policies. The aim is to use a policy model for specifying and enforcing context-aware access control. The OntCAAC provides the capability to control access for software services and resources by taking into account the context information. This ontology defines the general context entity classes: User, Role, Source, Owner, Relationships, Place, EnvPerson and Device. It also defines the different types of dynamic context information: ContextInfo, LocationInfo, TemporalInfo. The place ontology identifies different buildings, departments and rooms of the hospital. The Relationship ontology represents the relations between users: Person-Centric or Location-Centric. The former is used for represent relationship between users, the latter represents that the concerned people are collocated. The Status Information Ontology stores the health status, the current personal status, the current location status. Miraoui et al. [15] proposed in 2015 an ontology based on modeling a smart living room environment and their contextual information for enabling a common understanding of context and enhancing its sharing. Authors propose a definition of context for service oriented systems as any information that triggers a service or changes the quality of a service if its values change. Based on this concept modeling a smart room starts by specifying the services that each equipment can provide and the set of information that triggers the service.

None of revised works fully achieve the requirements specification, majority of ontologies include information about person or users, but do not consider IoT-based identification such as RFID tags or MAC address, for instance [2], [3], [5], [6], and [7]. Majority of revised related ontologies consider the geographical localization, but do not correlate localization with physical spaces and persons [3] and [7] not allowing the automatic identification and localization of persons and objects. Another important requirement was the incorporation of networks (computer-based networks and sensor networks), which was not considered. And finally, events or activities management is important, few reported works considered. All these requirements are necessary for context reasoning.

3. Ontology Design Methodology

In this section, the methodology that was defined and executed for design, construction and evaluation of the ontology is described. An initial set of competency questions were used for term elicitation and for final competency evaluation. Ontology design encompasses three stages: specification of ontology requirements, ontology construction, and ontology evaluation.

3.1. Specification of Ontology Requirements

In order to specify a set of initial ontology requirements, we reviewed the concepts of context and events management. Abowd et al. [16] compared and analyzed different definitions of context, and presented their

definition as follows: “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves”. From this definition we consider that the entities (or objects) present at an academic environment are: **Person** (for example professors, students, staff, etc.), **Physical space** (classroom, building, laboratory, bathroom, etc.), and computing **Devices** and **Networks**, among others. Schilit et al. [17] stated that the three important aspects of context are: where you are, who you are with, and what resources are nearby. From this approach, a context should include information about the geographical localization of objects and persons; together with the accurate identification of persons and objects. Hanzal et al. [18] distinguish between objects and events (both important concepts related with the concept of context [16] [17]). They clarify the distinction between objects and events from a philosophical perspective stating that objects are *continuants* (they exist and persist through time) and events are *occurrents* (they happen or take place at some point). This clarification of the concept **event** emphasizes the need to represent not only objects or entities, but events occurring at a point, this means that the ontology model should include the concept of **Time**. From this initial analysis we have defined the concept coverage requirements of the ontology and defined the main objective of the ontology, which is to facilitate intelligent context processing in the academic environment.

3.1.1. Concept coverage

Based on the definitions of context and event given in [16] and [17] the ontology model should include the following concepts:

1. Person profile information to represent the user data that is possible to gather from public networks, public Web pages, or public data bases available such as DBLP.
2. Data for the identification of persons such as id credentials, RFID tags, MAC address, and passwords provided.
3. Data for the identification of objects (RFID Tags, MAC address, etc.)
4. Data for geographical localization of persons, objects or physical locations. In order to enable localization services, this data should consist of longitude, latitude and height.
5. Physical space to represent the buildings organization with their precise geographical coordinates.
6. Sensor networks to represent the intranet and private nets organization as they are currently arranged into the institution.
7. Environmental data to represent the physical sensor measures located at the different physical locations.
8. Device to represent any kind of hardware device available including personal computers, microcontrollers, cellular and any device capable of data processing.
9. Events that may occur at an academic environment which result of interest for the users at the university. Events should be correlated with time, physical space, the location, and the person involved.

4. Ontology Construction

The ontology was incrementally implemented using the Protégé ontology editor, and represented using the standard Web Ontology language (OWL). The resulting ontology model consists of a set of modular ontologies among which a set of semantic relations are defined to support intelligent context reasoning.

4.1. Person Ontology

Person ontology represents all possible human users that may be present at the university in a moment or period of time, such as: visitor, professor, student, employee, etc. Additionally, this ontology included the concepts of **Department** and **AcademicTitle** to improve the context descriptions with useful attributes for **Academics**. Figure 1 shows the class hierarchy of the **Person** ontology. An important characteristic of this ontology was to define a unique identifier for every type of person that would be present inside the sensor-enabled context. The concept **Person** is defined as an equivalence through the *hasName* and *hasGender* data properties, indicating that every person should provide his name and gender. The concept **Employee** is defined as a sub class of a **Person** that

hasEconomicNumber data property. Whereas the concept **Student** is defined as a sub class of **Person** that hasStudentId. An important concept is a **Professor** which is an **Academic**, is an **Employee** and is a **Person** that hasCategory, hasDepartment, and hasEmail; and inherits the data property of an **Academic** hasProject. The class hierarchy of the **Person** ontology shows the sub-classification of the class **Student** into **RegularStudent** and **AssistantStudent**. This classification addresses a particular need to represent the two types of students that exist in the university where the **AssistantStudent** is an **Academic**, is an **Employee** and a **Student**.

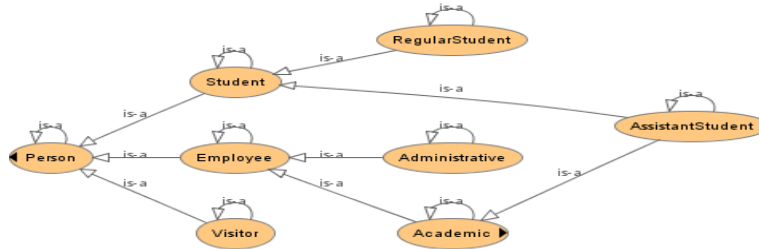


Fig. 1. Class hierarchy of the **Person** Ontology.

4.2. Physical Space Ontology

The **PhysicalSpace** ontology represents any kind of physical location such as buildings, cubicles, classrooms, offices, parking lots, plazas, green areas, etc. The class hierarchy of the ontology shown in Figure 2 contains as a main concept the **PhysicalSpace** class to conceptualize any kind of physical space which is divided into two main sub-class definitions: **Internal** and **External**. An internal physical space is used to represent closed rooms; whereas external physical space is used to represent open spaces such as: hallway, green areas, parking, etc. The set of data properties that were defined are: hasName, hasDoorState, hasAirConditioner, hasLampsNumber, hasLevel, hasPeopleCapacity, hasCarCapacity, hasProjector, hasService, hasWindow, isOpen, has Area, among other properties. Two important relationships between concepts (object properties) defined for this class are: isLocatedInto, this property is used to specify that a physical space is part of another physical space, enabling physical objects composed of physical objects. For instance a classroom is located into a building; and the property isBesideOf, which is used to specify neighboring localities, for instance a cubicle A isBesideOf cubicle B and cubicle B isBesideOf bathroom 2.

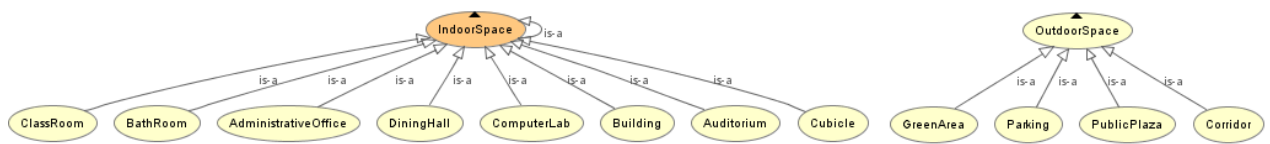


Fig. 2. Class hierarchy of the **PhysicalSpace** class.

4.3. Sensor Network Ontology

The **SensorNetwork** ontology represents a collection of physical sensors, mobile sensors and actuators. The objective of the network sensor is to obtain data from the physical context, user context and eventually activate some actuators. This ontology aims at representing environmental data such as temperature, lighting, humidity, and presence of humans into the environment. Another important objective of this model is the possible identification of the users and the data generated by user interaction with the environment. The following types of sensors are considered: *environmental sensors*, which are used to obtain data room temperature, humidity, luminosity, and presence of persons; *mobile and wearable sensors*, such as: accelerometer, gyroscope, magnetometer, proximity sensor, light sensor, barometer, thermometer, pedometer, heart rate monitor, fingerprint sensors, etc.; *identification*

sensors carried by the user, which will serve both the user ID to the acquisition of additional information; *actuators* represent the hardware devices through which actions are activated in order to achieve an ideal state.

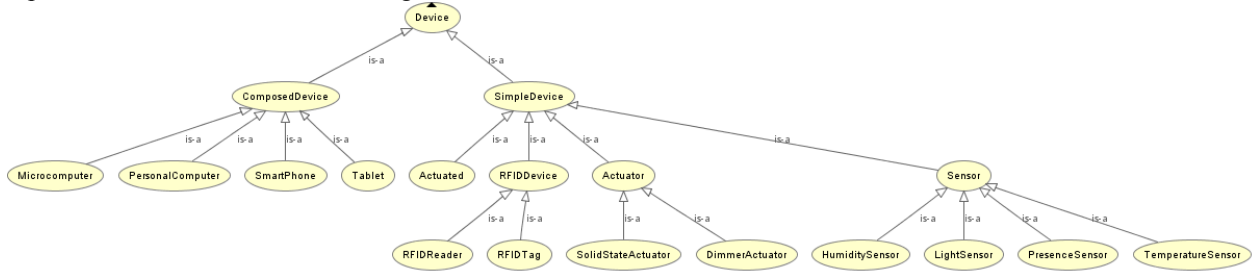


Fig. 3. Taxonomy of the *Device* class from the *SensorNetwork* Ontology.

4.4. Time Ontology

Time representation is an important requirement to manage events in the academic context. Figure 4 shows the structure of this ontology which consists of a class *TemporalEntity* sub divides into the concepts of *Instant* and *Interval* as follows: an *Instant* is defined as an individual that has the data attributes of *hasYear*, *hasMonth*, *hasDay*, *hasHour* and *hasMinute*; whereas *Interval* is defined with the object properties *hasBeginning* with range *Instant* and *hasEnd* with range *Instant*.



Fig. 4. Time Ontology.

5. Context Reasoning

Ontology reasoning consists of executing a program to infer logical consequences from a set of asserted facts or axioms. In order to realize the intelligent context reasoning the set of ontologies described above were imported and integrated into the *IntelligentEnvironment* ontology (see Fig. 5). This ontology was completed with additional class hierarchies, data properties and object properties. An important class hierarchy that was included into the ontology is the *Event* class to define a variety of events that may occur at the academic environment. A *Presence* event is sub divided into entering and leaving a physical space. These concepts are of great importance for the identification of persons into the academic intelligent environment. A *Presence Event* is defined through the data and object properties: *hasDescription*, *hasPersonInvolved* (*Person* class), *hasTime* (*TemporalEntity* class), and *happensIn* (*PhysicalSpace* class). All these semantic relationships are useful to answer any question regarding events occurring in the academic environment. What happened? When? Who was involved? At what time occurred the event?

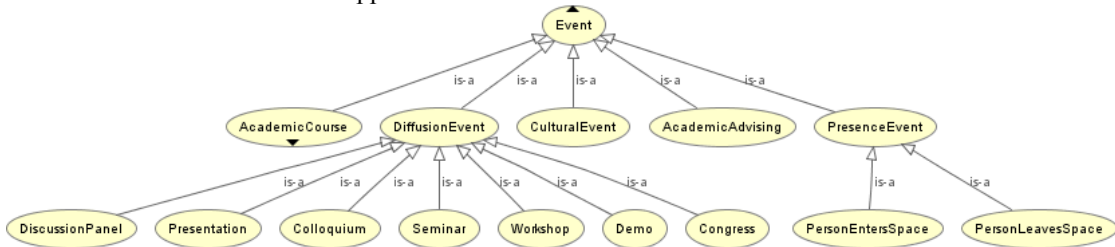


Fig. 5. The Event class hierarchy.

6. Ontology Evaluation

Ontology Evaluation [19] concerns the correct building of the ontology, ensuring that its definitions correctly implement the ontology requirements and competency questions. For evaluation two important aspects are considered: the *competence of the ontology*, that is, if it is able to respond to a set of competency questions; and the verification of *requirements compliance*. The following competency questions were correctly answered:

- Who is present in the classroom E313?
Person and (personLocatedIn value classRoomE313)
- Where is the cubicle of professor Alejandro Reyes?
Cubicle and (hasEmployeeAssigned some (hasName value "REYES ORTIZ JOSE ALEJANDRO"))
- Where is professor Maricela?
PhysicalSpace and (hasPersonDetected some (Professor and hasName value "BRAVO CONTRERAS MARICELA CLAUDIA"))
- At what time did Professor Alejandro Reyes Ortiz leave the Babbage computing laboratory?
Instant and (instantPersonLeaves some (hasPersonInvolved some (hasName value "REYES ORTIZ JOSE ALEJANDRO")))
- Who is the professor responsible of Distributed Systems course?
Professor and (isProfessorResponsibleOfCourse some (hasEventName value "Curso Sistemas Distribuidos"))

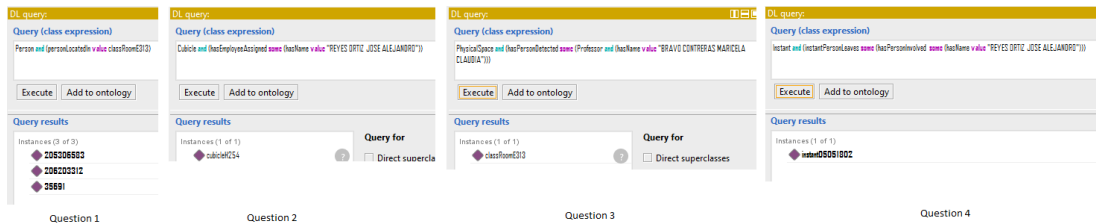


Fig. 6. Answers to competency questions using DL Query tool.

- How many publications does the professor with economical number “14233” has?
 $\text{Professor}(\text{?prof1}) \wedge \text{hasPublish}(\text{?prof1}, \text{?pub}) \wedge \text{hasEconomicNumber}(\text{?prof1}, \text{?e}) \wedge \text{swrlb:equal}(\text{?e}, "14233") \rightarrow \text{sqwrl:count}(\text{?pub})$
- With whom the professor identified with economical number “14233” collaborates with?
 $\text{Professor}(\text{?p1}) \wedge \text{collaborateWith}(\text{?p1}, \text{?p2}) \wedge \text{hasEconomicNumber}(\text{?p1}, \text{?e}) \wedge \text{swrlb:equal}(\text{?e}, "14233") \rightarrow \text{sqwrl:select}(\text{?p1}, \text{?p2})$
- How many women professors from Systems Department have published at least one article?
 $\text{Professor}(\text{?prof}) \wedge \text{hasGender}(\text{?prof}, \text{?gen}) \wedge \text{swrlb:equal}(\text{?gen}, "FEMENINO") \wedge \text{hasDepartment}(\text{?prof}, \text{?dep}) \wedge \text{swrlb:equal}(\text{?dep}, "SISTEMAS") \wedge \text{hasPublish}(\text{?prof}, \text{?pub}) \rightarrow \text{sqwrl:count}(\text{?prof})$

6.1. Evaluation of the Requirements

Ontology requirements were fully attended as follows: Person profile information and data for the automatic identification of persons were included in the *Person* ontology. The particular properties *hasRFIDTag* with domain *Employee* and *Student* and range *RFIDTag*, *hasMACAddress*, and *hasIPAddress* allow the automatic identification

of persons and objects at the intelligent environment. Geographical localization of persons, objects or physical locations is included in the **PhysicalSpace** ontology by means of the following properties: *isLocatedInto*, *isBesideOf*, *hasLatitude*, *hasLongitude*, and *hasAltitude*. Network and Device is represented at the **SensorNetwork** ontology. Environmental data is represented at the **PhysicalMeasure** ontology. Events are represented at the **IntelligentEnvironment** ontology, in this ontology events are correlated with time, physical space, and persons.

7. Conclusions

The ontology model reported in this paper is envisioned for a wireless networked environment, where users may be identified by their mobile device mac address or by an RFID card. Such an environment may be an office or laboratory into an academic institution or university, where users enter and leave the environment freely. The ontology was constructed from scratch, because the full list of requirements was not included in any of the reported works. For evaluation, a set of competency questions were translated to queries and executed to show that the ontology is complete and is capable of context reasoning. Results and the evaluation of the ontology model show promising advances towards the construction of an integral IoT platform.

References

- [1] Chen, H., Finin, T., & Joshi, A. An ontology for context-aware pervasive computing environments. *The knowledge engineering review*, 18 (03), 197-207, (2003).
- [2] Chen, H., Finin, T., & Joshi, A. (2005). The SOUPA ontology for pervasive computing. In *Ontologies for agents: Theory and experiences* (pp. 233-258). Birkhäuser Basel.
- [3] Razmerita, L., Angehrn, A., & Maedche, A. Ontology-based user modeling for knowledge management systems. In *International Conference on User Modeling* (pp. 213-217). Springer Berlin Heidelberg, (2003).
- [4] Razmerita, L. Ontology-based user modeling. In *Ontologies* (pp. 635-664). Springer US, (2007).
- [5] Wang, X. H., Zhang, D. Q., Gu, T., & Pung, H. K. Ontology based context modelling and reasoning using OWL. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on* (pp. 18-22). IEEE, (2004).
- [6] Gu, T., Wang, X. H., Pung, H. K., & Zhang, D. Q. An ontology-based context model in intelligent environments. In *Proceedings of communication networks and distributed systems modeling and simulation conference* (Vol. 2004, pp. 270-275), (2004).
- [7] Preuveneers, D., Van den Bergh, J., Wagelaar, D., Georges, A., Rigole, P., Clerckx, T., ... & De Bosschere, K. Towards an extensible context ontology for ambient intelligence. In *European Symposium on Ambient Intelligence* (pp. 148-159). Springer Berlin Heidelberg, 2004.
- [8] D. Ejigu, M. Scuturici and L. Brunie, An Ontology-Based Approach to Context Modeling and Reasoning in Pervasive Computing, *Pervasive Computing and Communications Workshops, 2007. PerCom Workshops '07. Fifth Annual IEEE International Conference on*, White Plains, NY, pp. 14-19, (2007).
- [9] Zainol, A., Nakata, K., Generic Context Ontology Modelling: A review and Framework, on *2nd International Conference on Computer Technology and Development (ICCTD)*, pp 126-130, (2010).
- [10] Poveda Villalon, M., Suárez-Figueroa, M. C., García-Castro, R., & Gómez-Pérez. A context ontology for mobile environments. (2010).
- [11] Skillen, K. L., Chen, L., Nugent, C. D., Donnelly, M. P., Burns, W., & Solheim, I. (2012, December). Ontological user profile modeling for context-aware application personalization. In *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 261-268). Springer Berlin Heidelberg.
- [12] Guermah, H., Fissaa, T., Hafiddi, H., Nassar, M., & Kriouile, A, Context modeling and reasoning for building context aware services, In *Computer Systems and Applications (AICCSA), 2013 ACS International Conference on*, pp. 1-7, (2013).
- [13] Nadoveza A., Kiritsis D., *Ontology-Based Approach for Context Modeling in Enterprises, Computer in Industry*, pp. 1218-1231, (2014).
- [14] Kayes, A., Han J., Colman, A. OntCAAS: An Ontology-Based Approach to Context-Aware Access Control for Software Services, *The Computer Journal*, v. 58, No 11, (2015).
- [15] Miraoui, M., El-etriby, S., Tadj, Ch., Zaid Abid, A., *Ontology-Based Context Modeling for a Smart Living Room, Proceedings of the World Congress on Engineering and Computer Science 2015 Vol I WCECS 2015, October 21-23, 2015, San Francisco, USA*
- [16] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggle, P. (1999, September). Towards a better understanding of context and context-awareness. In *International Symposium on Handheld and Ubiquitous Computing* (pp. 304-307). Springer, Berlin, Heidelberg.
- [17] Schilit, B., Adams, N., & Want, R. (1994, December). Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on* (pp. 85-90). IEEE.
- [18] Hanzal, T., Svátek, V., & Vacura, M. (2016, July). Event Categories on the Semantic Web and Their Relationship/Object Distinction. In *FOIS* (pp. 183-196).
- [19] Gómez-Pérez, A. Towards a framework to verify knowledge sharing technology. *Expert Systems with Applications*, 11(4), 519-529, (1996).

Ontology-based Population and Enrichment of Researcher Profiles

Isabel Cruz-Ruiz, Maricela Bravo and José A. Reyes-Ortiz

Systems Department, Autonomous Metropolitan University,
Av San Pablo 180, Reynosa Tamaulipas, 02200 Ciudad de México, México
icruizr2014@hotmail.com, {mcbc, jaro}@azc.uam.mx

Abstract. The representation, management, and exploitation of researcher profiles is an important task that every research institution must achieve. In this paper we investigate on the use of ontologies as the main solution approach to support the representation of researcher profiles in a given academic environment. We describe the ontology model design, the automatic ontology population processes, and the discovery and enrichment processes of interesting semantic relations between researcher profiles. The functional competency of the enriched ontology is evaluated utilizing a set of inference rules and queries.

Keywords. ontology population, ontology enrichment, researcher profile

1. Introduction

Currently, higher education institutions and research institutes have highly specialized human resources who count with high degrees of postgraduate studies. The capacity, expertise and talent accumulated by academic and research staff is one of the most important assets available to institutions. Representing, quantifying and knowing how to better manage these highly specialized human resources is a very important issue; however, it is not an easy task to perform, since it requires the acquisition, representation and intelligent treatment of large volumes of data. A good management of highly specialized human resources can be carried out through the administration of researcher profiles to enable: finding similar profiles to establish new collaborations, looking for specific profiles that allow to integrate a work team with specialists, discovering groups or classes of researchers that address similar topics, discovering groups of researchers that address different problems but that use similar approaches, among other possible applications.

A researcher profile consists of the relevant information regarding previous academic work experience in different research institutions, education and level of studies considering undergraduate, graduate, and specialization studies; an important aspect of a researcher profile is the scientific published articles, chapters, and books, as they represent the researcher topics of interest, and the researcher most active collaborations.

In this paper, we present an ontology solution approach for the acquisition, representation and management of researcher profiles. This ontology solution is evaluated using a set of competency questions through which the functional competency of the proposed solution is evaluated satisfactorily.

This paper reports a contribution in the area of ontology learning. Maedche and Staab [1] define *ontology learning* as the process of automatic or semi-automatic construction, enrichment and adaptation of ontologies. Accordingly, the main tasks involved in ontology learning are ontology enrichment, inconsistency resolution and ontology population.

- a) *Ontology enrichment* is the task of extending an existing ontology with additional concepts and semantic relations in the ontology.
- b) *Inconsistency resolution* is the task of resolving inconsistencies that appear in an ontology aiming at producing and maintaining a consistent ontology.
- c) *Ontology population*, is the task of adding new instances of concepts to the ontology.

The methodological process followed for the construction and evaluation of the proposed solution consist of four phases: ontology design, ontology population, ontology enrichment, and ontology evaluation.

The rest of this paper is organized as follows. Section 2 presents the related work, which is briefly described to compare them with the approach presented in this paper. Section 3 describes the specification of ontology requirements. Section 4 describes the ontology design methods. Section 5 presents the automatic ontology population of each ontology. Section 6 describes the ontology research profile enrichment. Section 7 presents an evaluation based on competency questions. Finally, Section 8 shows the conclusions and future work.

2. Related Work

In this section we first present the definition of ontology, and describe the related works that address the representation and management of researcher or academic profiles. We analyze their applications and concept coverage.

Over the last decades, different ontology definitions have been presented and discussed. According to Gruber [2] an ontology is an “explicit specification of a conceptualization”, an ontology is used to formally define the important concepts of a terminology and the semantic relationships that may exist between concepts. It is frequent that the set of formally defined concepts belong to a specific area of knowledge, and the set of rules and axioms defined are congruent with the particular area of knowledge. In [3] Sowa stated that an ontology represents a catalog of categories to classify entity types that exist in a given domain. In [4] Cámara explained that an ontology can be conceived as an instrument for knowledge representation in a particular topic area, through which knowledge recovery and information retrieval can be executed. Ontologies were selected as the formal representational mechanism as they facilitate reusability, knowledge sharing, and execution of formal reasoning tasks such as satisfiability of concepts, consistency checking, classification and inference.

Concerning researcher profile, Yao, Tang and Li [5] address the problem of researcher profiling by annotating a collection of researcher web pages, and defined a series of difficulties found using this approach. Authors identify tokens in the Web page heuristically, assign tags to each token (Position, Affiliation, Email, Address, Phone, Fax), using the tags, they perform the profiling extraction. In Liu et al [6] authors address the problem of finding experts with required expertise. They describe two ontologies: an *expert ontology*, which defines concepts such as: Person, Publication, Project, and Research Interests; and a *domain ontology* which stores the key concepts (research areas), the attributes of the concepts and the relations between concepts (for example, broader, narrower and part-of). In [7] authors address the problem of automatic extraction of topics of expertise of a person based on the

documents accessed by the person through information extraction techniques. They define a user profile using a set of topics with weights determining his level of interest. In [8] authors present a multi-agent paradigm supported by a semantic web architecture to address the challenges of researcher profiling and association. Authors describe an ontological model to represent information such as researcher profiles, conference papers, research centers, etc.

In [9] authors describe ArnetMiner, to address the following questions:

- a) How to automatically extract researcher profiles from the Web?
- b) How to integrate the extracted information (e.g., researchers' profiles and publications) from different sources?
- c) How to model different types of information in a unified approach?
- d) How to provide powerful search services based on the constructed network?

In ArnetMiner the schema of a researcher profile was proposed consisting in two main entities: Researcher and Publication. Based on the work reported in [9] in this project we address the same questions and present a solution approach based on the use of ontologies and reasoning tasks.

In [10] authors describe a skill classification ontology model containing skills of research in the area of computer and information science. Their main contributions are:

- a) A process to build the skill classification ontology.
- b) A methodology to determine expertise of the researcher using the skill classification ontology.
- c) A method to retrieve the relevant researchers who may have competency matched to the desired expertise.

Motivated by these related works, we propose an ontology-based solution approach for the acquisition, representation and management of researcher profiles.

3. Specification of Ontology Requirements

The main objective of the ontology model presented in this paper is to facilitate researcher profile processing and reasoning. Considering that every research institution requires the efficient management and dissemination of information relative to the professors and research activities, etc. The ultimate goal of this project is the smart and provisioning of services to researcher communities in which researchers search for specialized publications (such as publications, coauthors, conferences) and are also interested in establishing collaborations with other researchers. Considering this motivation, the following requirements were defined in order to guide the ontology design, construction and evaluation.

3.1. Scope of the Ontology

In order to specify the scope of the ontology, we reviewed the concepts of research profile. Yao, Tang and Li [5] described profiling as the process of obtaining the values associated with the different properties that constitute the person model. Authors define the schema of a researcher profile containing: name, affiliation, position, phone, address, email, research interests, and postgraduate studies. From this definition we consider that the entities (or objects) that constitute a researcher profile are: **Person** (for example professors, students, staff, etc.), and **Publication** (to extract research interests).

From this initial analysis we have defined the concept coverage requirements of the ontology and defined the main objective of the ontology, which is to facilitate researcher profile representation and processing in the academic environment.

3.2. Concept coverage

The ontology model should include the following concepts:

- a) Data for the identification of persons and researchers such as name, economical number, staff card, etc.
- b) Person profile information to represent the user data that is possible to gather from public networks, public Web pages, or public data bases available such as DBLP.
- c) Data to represent publications such as: thesis, chapters, journals, etc.

3.3. Competency of the Ontology

Gruninger and Fox [11] proposed six characteristics to evaluate a Business Model. These characteristics were proposed to answer the question of “How can one determine which model is correct for a given task?” To give a guideline on the operation of these characteristics, the authors define the concept of competence of the model as follows: given an appropriately instantiated model and a demonstrator of theorems, the competence of a model is the set of questions that the model can answer. Based on this definition, we may state that

The competence of an ontology model is the set of questions that the ontology can answer.

Evaluation of the competency of an ontology system is crucial to verify that a representational model is complete with respect to a given set of competency questions. During the phase of requirements specification, a team of experts and programmers defined the following set of competency questions for this model:

1. To know how many scientific works does a given researcher has published
2. To find groups of authorship collaborations
3. To know the researcher’s publications from a specialized topic with high degree studies
4. To answer about statistical data of publications
5. To know the scientific productivity of a given department
6. To know the number of female researchers from a given department with published scientific works.

4. Ontology Design

Ontology design is the process of selecting and applying methods, techniques and principles with the objective of producing an ontology model. In this section, the design considerations that were taken into account are described.

A good quality ontology design depends mainly on the selection and incorporation of design principles. Uschold and Gruninger [12] presented their initial ideas and detailed a set of ontology design criteria. For the design and construction of the researcher profile ontology the following design principles were taken in consideration:

- i. *Clarity* principle states that an ontology should effectively communicate the intended distinctions. Ambiguity should be minimized, distinctions should be

motivated, and examples should be given to understand definitions that lack necessary and sufficient conditions.

- ii. *Coherence* design principle specifies that an ontology should be internally consistent. Coherence should also apply to the parts of the definitions that are not axiomatic.
- iii. *Extensibility* principle states that an ontology should be designed anticipating possible uses of the shared vocabulary.

Additionally, an initial set of competency questions were used for term elicitation and for final competency evaluation.

4.1. Person Ontology

Person ontology was designed to represent all possible academics which hold a permanent or temporal position as professor or researcher at the university, such as: academic visitor, full time professor, external sabbaticals, etc. This ontology also represents postgraduate students, and research oriented undergraduate students, among others. Figure 1 shows the main class hierarchy of the *Person* ontology. An important characteristic of this ontology is that it uses a unique identifier for every type of person. The concept *Person* is defined as an equivalence through the *hasName* and *hasGender* data properties, indicating that every person individual is obligated to have name and gender to be classified as type of *Person* class. The concept *Employee* is defined as a sub class of a *Person* that *hasEconomicNumber* data property. Whereas the concept *Student* is defined as a sub class of *Person* that *hasStudentId*. An important concept is a *Professor* which is an *Academic*, is an *Employee* and is a *Person* that *hasCategory*, *hasDepartment*, and *hasEmail*; and inherits the data property of an *Academic* *hasProject*. The class hierarchy of the *Person* ontology shows the sub-classification of the class *Student* into *RegularStudent* and *AssistantStudent*. This classification addresses the particular need to represent the two types of students that exist in the university where an individual of the *AssistantStudent* class is considered to be an *Academic*, an *Employee* and a *Student*.

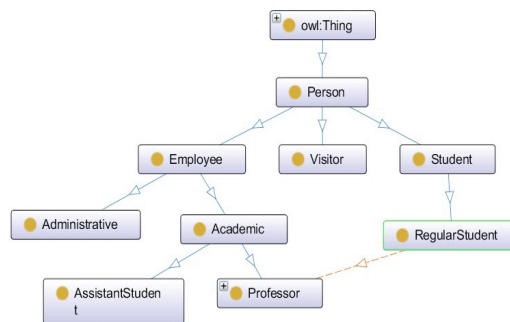


Fig. 1. Person ontology class hierarchy

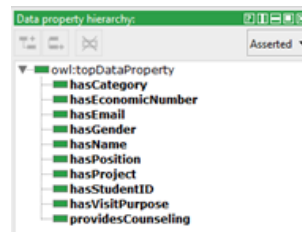


Fig. 2. Data type properties defined for the Person ontology.

The full list of data type properties defined for the *Person* ontology are shown in Figure 2.

4.2. Publication Ontology

Scientific published articles, chapters, and books are the most important sources of information in order to integrate a researcher profile. Scientific publications contain the author's topics of interest, conferences and journals of preference, the years of publications and periodicity; also the researcher most active collaborations. In order to build a researcher profile, the design and construction of a publication ontology considered as input the information extracted from the DBLP computer science bibliography on-line reference, extracting the most relevant bibliographic information on major computer science publications.

The *Publication* ontology defines the same attributes utilized in DBLP. Figure 3, shows those data type attributes.

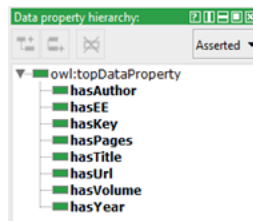


Fig. 3. Data properties of the Publication ontology.

4.3. Researcher Profile Ontology

Researcher Profile ontology was designed to incorporate conceptualizations from *Person* and *Publication* ontologies. From *Person* ontology imports *Professor* personal data, such as full name, and economical number; from the *Publication* ontology imports publications organized by year, type of publication, among others. Additionally, incorporates *Department* and *Academic Title* concepts. All these conceptualizations are used to complete the definition of a *Researcher Profile*, considering the associated publications, the affiliated department, the academic title obtained, and the rest of personal data. Figure 4 shows the main concepts that integrate the *Researcher Profile* Ontology.

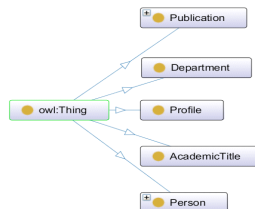


Fig. 4. Researcher Profile ontology class hierarchy.

5. Ontology Population

Ontology population is the process of adding (instantiating) new individuals in the ontology concepts (classes). Automated ontology population is desirable due to the large amount of data that must be extracted and instantiated in the *Person* and *Publication* ontologies.

5.1. Person Ontology Population

For *Person* ontology population, the data source comes from a set of excel files that the management staff of the university uses for different purposes. These excel files contain the information of all academic staff who are affiliated with the university, such as: professor's full name, gender, department, email, economical number, academic projects, and alias. For the automated population of the *Person* ontology, two Java modules were developed: a module to *parse and extract* the information from the source files; and another module to *interact* with the ontologies using the Java OWL Application Programming Interface (OWL API) to load and manipulate ontologies, creating new individuals, instantiating object properties and data properties with individuals, and register them in the ontology. Figure 5 shows the values of data type attributes registered for professor "REYES ORTIZ JOSE ALEJANDRO", and the recognized alias names. Alias names are important in order to facilitate the semantic association of the researcher individual with all his publications.

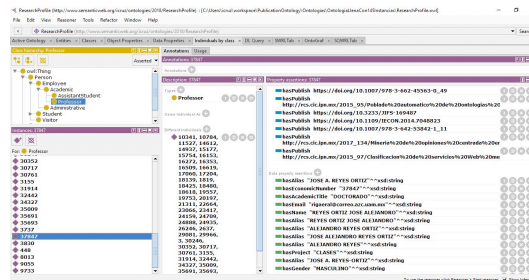


Fig. 5. Person ontology population.

5.2. Publication Ontology Population

For *Publication* ontology population, the data was extracted from the DBLP (Digital Bibliography & Library Project) [13], a compressed XML file, which contains more than a million of Computer Science publications. The XML file from the DBLP contains publication title, author names, publication year, volume, EE (a unique publication identifier), URL, and pages (see Figure 6). However, it does not provide the abstract and keywords of publications.

```
<article iddate="2018-05-29" key="journals/jifs/Reyes-OrtizB18">
  <author>JoseAlejandre A. Reyes-Ortiz</author>
  <author>Maricela Bravo</author>
  <title>Enhancing patterns with linguistic information for criminal event recognition.</title>
  <pages>3027-3036</pages>
  <year>2018</year>
  <volume>34</volume>
  <journal>Journal of Intelligent and Fuzzy Systems</journal>
  <number>5</number>
  <ee>https://doi.org/10.3233/JIFS-169487</ee>
  <url>db/journals/jifs/jifs34.html#Reyes-OrtizB18</url>
</article>
```

Fig. 6. DBLP XML file extract

A Java module was built to *interact* with the ontologies using the OWL API to load and manipulate ontologies, creating new individuals, instantiating object properties and data properties. Figure 7 shows the instantiation of new **Publication** individuals correlated with the identification of the author that published.

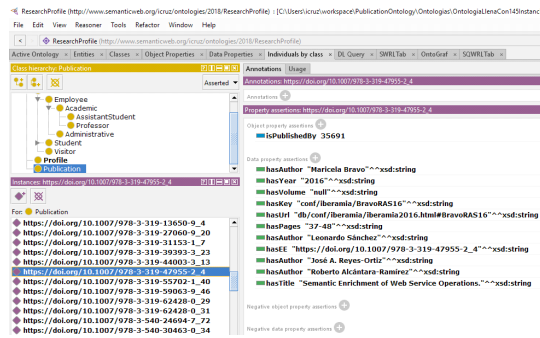


Fig. 7. Publication ontology population

6. Ontology Enrichment

Ontology enrichment is the automatic process of analyzing the population data values and discovering new interesting semantic relations between individuals. Of particular interest in this enrichment process is the automatic discovery of collaboration relations between authors of publications. For this, the following object properties and inference rules were defined.

collaborateWith is an object property with domain **Person** and range **Person**. This object property was defined to establish semantic relationships between authors of publications.

Figure 8 shows the SWRL rule that was executed to find collaborations between authors of publications is:

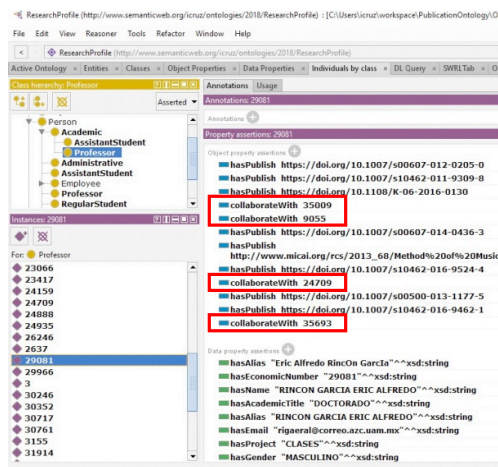
$$\text{Publication(?pub)} \wedge \text{isPublishedBy(?pub, ?prof)} \wedge \text{isPublishedBy(?pub, ?prof2)} \wedge \text{differentFrom(?prof, ?prof2)} \wedge \neg \text{collaborateWith(?prof, ?prof2)}$$


Fig. 8. Discovery of authorship collaborations

7. Ontology Evaluation

Ontology Evaluation [14] concerns the correct building of the ontology, ensuring that its definitions correctly implement the ontology requirements and competency questions. For evaluation the *competence of the ontology* was considered, that is, if it is able to respond to a set of competency questions; and the verification of *requirements compliance*. The following competency questions were coded in SWRL language and their results were correct.

7.1. Researcher Publications

To know how many scientific works does a given researcher has published, the following rule was defined and executed. Figure 9 shows the result of this competency question.

```
Professor(?prof1) ^ hasPublish(?prof1, ?pub) ^
hasEconomicNumber(?prof1, ?e) ^ swrlb:equal(?e, "14233") -> sqwrl:count(?pub)
```

| SWRL Queries | OWL 2 RL | SP |
|--|----------|-------------|
| Professor(?prof1) ^ hasPublish(?prof1, ?pub) ^ hasEconomicNumber(?prof1, ?e) ^ swrlb:equal(?e, '14233') -> sqwrl:count(?pub) | | |
| | | count(?pub) |
| | | 3 |

Fig. 9. Discovery of authorship collaborations

7.2. Collaboration between Researchers

To find groups of authorship collaborations, the following semantic Web rule is defined. The result is shown in Figure 10.

```
Professor(?p1) ^ collaborateWith(?p1, ?p2) ^
hasEconomicNumber(?p1, ?e) ^ swrlb:equal(?e, "14233") ->
sqwrl:select(?p1, ?p2)
```

| SWRL Queries | OWL 2 RL | SP |
|--|----------|-----|
| Professor(?p1) ^ collaborateWith(?p1, ?p2) ^ hasEconomicNumber(?p1, ?e) ^ swrlb:equal(?e, '14233') -> sqwrl:select(?p1, ?p2) | | |
| | p1 | p2 |
| | 14233 | 341 |

Fig. 10. Discovery of authorship collaborations

7.3. Qualified and Specialized Researchers

In order to know the researcher's publications from a specialized topic with high degree studies, the following semantic Web rule was utilized. Figure 11 shows the result of the execution.

```
Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^ swrlb:contains(?dep,
"ELECTRONICA") ^ hasAcademicTitle(?prof1, ?at) ^ swrlb:equal(?at, "DOCTORADO")
^ hasPublish(?prof1, ?pub) -> sqwrl:count(?pub)
```

| S13 | Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^ swrlb:contains(?dep, "ELECTRONICA") ^ hasAcademicTitle(?prof1, ?at) ^ swrlb:equal(?at, "DOCTORADO") ^ hasPublish(?prof1, ?pub) -> sqwr1:count(?pub) |
|---------------|---|
| SQWRL Queries | OWL 2 RL S8 S9 S18 |
| | count(?pub) |
| | *1**xsdint |

Fig. 11. Highly specialized researchers

7.4. Publications by Year

The ontology is capable of answering statistical data of publications, for instance: How many publications were there in the year 2017? Figure 12 shows the result of this rule.

```

Publication(?pub) ^
  hasYear(?pub, ?y) ^
  swrlb:equal(?y, "2017") -> sqwr1:count(?pub)

```

| S19 | Publication(?pub) ^ hasYear(?pub, ?y) ^ swrlb:equal(?y, "2017") -> sqwr1:count(?pub) |
|---------------|--|
| SQWRL Queries | OWL 2 RL S8 S9 S18 S19 |
| | |
| | *1**xsdint |

Fig. 12. Publications by year

7.5. Publications by Department

In order to know the scientific productivity of a given department, the following semantic Web rule is used. Figure 13 shows the result.

```

Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^ swrlb:contains(?dep,
  "SISTEMAS") ^ hasPublish(?prof1, ?pub) -> sqwr1:count(?pub)

```

| S5 | Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^ swrlb:contains(?dep, "SISTEMAS") ^ hasPublish(?prof1, ?pub) -> sqwr1:count(?pub) |
|---------------|--|
| SQWRL Queries | OWL 2 RL S6 |
| | count(?pub) |
| | *123**xsdint |

Fig. 13. Publications by department

7.6. Publications by Gender

To know the number of female researchers from a given department that have published scientific works, the following semantic Web rule was used. Figure 14 shows the results.

```

Professor(?prof) ^
  hasGender(?prof, ?gen) ^
  swrlb:equal(?gen, "FEMENINO") ^ hasDepartment(?prof, ?dep) ^ swrlb:equal(?dep,
  "SISTEMAS") ^ hasPublish(?prof, ?pub)
  -> sqwr1:count(?prof)

```

| S10 | Professor(?prof) ^ hasGender(?prof, ?gen) ^ swrlb:equal(?gen, "FEMENINO") ^ hasDepartment(?prof, ?dep) ^ swrlb:equal(?dep, "SISTEMAS") ^ hasPublish(?prof, ?pub) -> sqwr1:count(?prof) |
|---------------|--|
| SQWRL Queries | OWL 2 RL S6 S10 |
| | count(?prof) |
| | *49**xsdint |

Fig. 14. Publications by Gender

8. Conclusions

In the work reported in this paper, an automated ontology population was used to build researcher profiles. For the population of the *Person* ontology, a collection of 373 professors was used, two departments were instantiated in the *Department* class, 50 professor individuals are from the Systems *Department* and 60 from Electronics. 100% of them were correctly inserted in the *Person* ontology, specifically in the *Professor* class.

For experimentation purposes two ontologies were generated: one was used for professors of the electronics department and another for professors of the systems department. The systems ontology had a total of 50 researchers in total. Once the universe of professors was divided, the publishing ontology was populated using as a data source the DBLP file, which contains approximately one million articles and more than 56 million lines. The result of the population of the *Publication* ontology with professors of the systems department resulted in a total of 135 publications that coincided between the aliases of the professors and the authors indicated within the *<author>* label of the DBLP. At the same time 116 collaboration relationships were found among the professors of the systems department.

The ontology of professors of the electronics department, with a total of 58 professors, was subjected to the same test as the ontology of professors of the systems department and 22 publications were found from the same sample of the DBLP, that is, all the aliases were compared of the researchers of the electronics department against the authors of the million articles of the DBLP. In this ontology 4 collaborative relationships were found, that is, in two publications two or more professors from the same department participated.

As future work, other sources can be considered to continue enriching the ontologies with more semantic relationships, such as ArnetMiner [9] which contains abstracts and keywords of publications in order to enable the semantic relationship between publications, researchers and topics of interest. In this way the ontology would comply with the characteristic of being scalable and make the profile of each researcher a more complete.

References

1. Maedche, A., Staab, S.: Ontology Learning. In: Handbook on Ontologies (2004)
2. Gruber, Tom. What is an Ontology?, 1993 ksl.stanford.edu/kst/what-is-an-ontology.html
3. Sowa, J. F.: Guided Tour of Ontology, 2001 <http://www.jfsowa.com/ontology/guided.htm>.
4. Cámara, J. C.: Learning Metadata Standards: Ontologias Barcelona: UPF 2002 <http://www.iaa.upf.es/jblat/material/doctorat/students/jccbis/Ontologias.htm>
5. Yao, L., Tang, J., Li, J.: A unified approach to researcher profiling. In: Proceedings of the IEEE International Conference on Web Intelligence, pp. 359-366 (2007)
6. Liu, P., Liu, K., Liu, J.: Ontology-based expertise matching system within academia. In: Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing, WiCom, pp. 5431-5434 (2007)
7. Thiagarajan, R., Manjunath, G., Stumptner, M.: Finding experts by semantic matching of user profiles (Doctoral dissertation, CEUR-WS) (2008)
8. Adnan, S., Tahir, A., Basharat, A., de Cesare, S.: Semantic agent oriented architecture for researcher profiling and association (semora). In: Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, vol. 3, pp. 559-562 (2009)

9. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 990-998 (2008)
10. Punnarut, R., Sriharee, G.: A researcher expertise search system using ontology-based data mining. In: Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling, vol. 110, pp. 71-78 (2010)
11. Grüninger, M., Fox, M. S.: Methodology for the design and evaluation of ontologies (1995)
12. Uschold, M., King, M.: Towards a methodology for building ontologies Edinburgh. In: Artificial Intelligence Applications Institute, University of Edinburgh, pp. 15-30 (1995)
13. Ley, M.: The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives. In Proceedings of the 9th International Symposium on String Processing and Information Retrieval, pp. 1-10 (2002)
14. Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. Expert Systems with Applications, vol. 11, no. 4, pp. 519-529 (1996)

Researcher Profile Ontology for Academic Environment

Maricela Bravo¹, Alejandro Reyes-Ortiz¹, Isabel Cruz¹

¹ Autonomous Metropolitan University, San Pablo 180, CDMX, Mexico
{mcbc, jaro, icruz}@azc.uam.mx

Abstract. Universities must know the research profiles and professional competences of professors and researchers to be more competitive and know how to offer creative solutions (through courses, diplomas, research projects, etc.) to complex needs of society. Therefore the representation, management, and exploitation of researcher profiles are important tasks that must be addressed. In this paper we investigate on the use of ontologies as the main solution approach to support the representation of researcher profiles in a given academic environment. We describe the ontology model design, the automatic ontology population processes, and the discovery and enrichment processes of interesting semantic relations between researcher profiles. The functional competency of the enriched ontology is evaluated utilizing a set of inference rules and queries.

Keywords: Researcher profile, Ontology population, Ontology enrichment.

1 Introduction

Currently, higher education institutions and research centers hire highly specialized human resources who have postgraduate studies. The capacity, expertise and talent accumulated by academic and research staff is one of the most important assets available to institutions. Representing, quantifying and knowing how to better manage these highly specialized human resources is a very important issue; however, it is not an easy task to perform, since it requires the acquisition, representation and intelligent treatment of large volumes of data. A good management of highly specialized human resources can be carried out through the administration of researcher profiles to enable: finding similar profiles to establish new collaborations, looking for specific profiles that allow to integrate a work team with specialists, discovering groups or classes of researchers that address similar topics, discovering groups of researchers that address different problems but that use similar solution techniques and methods, among other possible applications.

A researcher profile is constituted with the relevant information regarding previous academic work experience in different research institutions, education and level of studies considering undergraduate, graduate, and specialization studies; an important aspect of a researcher profile is the set of scientific published articles, chapters, and books, as they represent the researcher topics of interest, and the researcher most active collaborations.

In this paper, we present an ontology solution approach for the acquisition, representation and management of researcher profiles. This ontology solution is evaluated using a set of competency questions through which the functional competency of the proposed solution is evaluated satisfactorily.

This paper reports a contribution in the area of ontology learning, which is defined as the process of automatic or semi-automatic construction, enrichment and adaptation of ontologies [1]. Accordingly, the main tasks involved in ontology learning are: ontology enrichment, extending an existing ontology with additional concepts and semantic relations; inconsistency resolution, resolving inconsistencies that appear in an ontology aiming at producing and maintaining a consistent ontology; and ontology population, adding new instances of concepts to the ontology. The methodological process followed for the construction and evaluation of the proposed solution consist of four phases: ontology design, ontology population, ontology enrichment, and ontology evaluation.

2 Analysis of Related Work

In this section we describe the related works that address the representation and management of researcher or academic profiles. We analyze their applications and concept coverage.

Yao, Tang and Li [2] address the problem of researcher profiling by annotating a collection of researcher web pages, and defined a series of difficulties found using this approach. Authors identify tokens in the Web page heuristically, assign tags to each token (Position, Affiliation, Email, Address, Phone, Fax), using the tags, they perform the profiling extraction.

In Liu et al [3] authors address the problem of finding experts with required expertise. They describe two ontologies: an expert ontology, which defines concepts such as: Person, Publication, Project, and Research Interests; and a domain ontology which stores the key concepts (research areas), the attributes of the concepts and the relations between concepts (for example, broader, narrower and part-of).

ArnetMiner is one of the most outstanding researcher profile searcher and mining tool presented in 2008 by Tang et al. [4]. ArnetMiner extracts automatically topics of expertise of a person based on the documents accessed by the person, it constructs a user profile using a set of topics with weights determining his level of interest. ArnetMiner represents all extracted data in a relational database, and the schema of a researcher profile consisting of two main entities: **Researcher** and **Publication**.

Thiagarajan et al. in 2008 [5] addressed the problem of expert finding by computing the similarity between two user profiles with an *ontology-based* Spreading Activation Network (SAN). In this work authors represent the user profile as a set of topics with weights determining the level of interest. As a data source for profile extraction they obtained a set of documents and information provided by participants and used WordNet for term expansion.

Sadaf Adnan et al. in 2009 [6] presented a multi-agent paradigm supported by a semantic web architecture to address the challenges of researcher profiling and association. Authors describe an ontological model to represent information such as researcher profiles, conference papers, research centers, etc.

In 2010 Punnarut and Sriharee [7] described a skill classification ontology model containing skills of research in the area of computer and information science. Their main contributions are: a process to build the skill classification ontology, a methodology to determine expertise of the researcher using the skill classification ontology, and a method to retrieve the relevant researchers who may have competency matched to the desired expertise.

In 2012 Kamsiang and Senivongse [8] described a solution approach to address the problem of identifying common research interests between researchers. Their approach is based on the generation of researcher profile ontologies from keywords in a particular subject area during a certain period; keywords were taken from ISI Web of Knowledge. To compare the similarity between profile ontologies, authors presented an ontology matching algorithm.

In 2013 Whaling et al. [9] described Lens, an open source faceted browser for research profiling. The main objective of this work is to make available a number of co-existing dimensions that can be simultaneously browsed by the user. The information was extracted from public available data about researchers at the University of Chicago. They developed a prototype for the university profiling system to help researchers to find collaborators.

Table 1 shows a comparison of related work considering the most important questions regarding the automatic extraction and representation of Researcher Profiling: What sources of information are utilized? What representation mechanism is used? Are the basic information data and contact information data possible to extract? What other important aspects of the researcher profile are being considered?

Table 1. Comparison of related work.

| Ref | Profile Information Source | Information Extraction Method | Representation Mechanism | Basic Information | | | Contact Information | | | | Educational History | Expertise or Research Interests | Research Projects | Publications |
|-----|--|--|-----------------------------|-------------------|-------------|----------|---------------------|-------|-----------|-----|---------------------|---------------------------------|-------------------|--------------|
| | | | | Photo | Affiliation | Position | Address | Email | Telephone | URL | | | | |
| [2] | Web pages and DBLP | Automatic extraction and semantic annotation | Ontologies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | No | Yes |
| [3] | Heterogeneous data sources | Wrappers | Ontologies | Yes | Yes | Yes | No | Yes | No | No | No | Yes | Yes | Yes |
| [4] | Homepages from the Web and online digital libraries | Probabilistic framework based on Hidden Markov Random Fields | Relational Data Base (RNKB) | Yes | Yes | Yes | No | No | No | Yes | Yes | Yes | No | Yes |
| [5] | Documents and information provided by participants and WordNet | Spreading Activation Network | Ontologies | No | No | No | No | No | No | No | No | Yes | No | Yes |

| | | | | | | | | | | | | | | |
|-----|---------------------------------------|--|------------|----|----|----|----|----|----|-----|----|-----|-----|-----|
| [6] | Google scholar CiteSeer DBLP | Agent-based | Ontologies | No | No | No | No | No | No | No | No | Yes | No | Yes |
| [7] | Research papers and research projects | Extracting text from titles and using probabilistic approach | Ontologies | No | No | No | No | No | No | Yes | No | Yes | Yes | Yes |

As Table 1 shows, authors of related work have utilized different data sources; it is possible to state that using all possible sources of information could produce a more complete profile; however, the consistency and the actualization of the information is a problem to address. It is also visible that majority of authors are selecting ontologies as the representation model. Motivated by these related works, we propose an ontology-based solution approach for the acquisition, representation and management of researcher profiles.

3 Methodology

The scientific methodology that was followed to build a solution is described in this section.

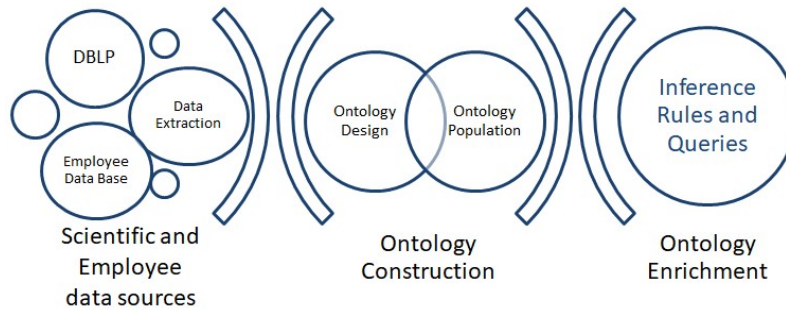


Fig. 1. Methodology for the extraction of researcher profile, ontology construction and enrichment.

3.1 Scientific Data Sources

In order to obtain data of researcher’s different sources can be utilized by crawling and extracting data from researcher’s home pages, URL’s and key words of Web pages, E-mail communications, organizational and institutional documents or databases. There are also different repositories publicly available to extract scientific publication data, for instance: DBLP, CiteSeer, Google Scholar, Scopus, and ArnetMiner. For reference: in [2] and [4] authors extracted information from

researcher's homepages and the digital libraries DBLP and CiteSeer. In [5] authors extracted information from documents provided by participants and WordNet.

In this work, we count with an excel database with researchers data records from two academic departments of the university. This database consists of researchers' full names, email, department, academic project, and academic title. Additionally, we selected DBLP (Digital Bibliography & Library Project) [10] to extract publication data and enrich the ontology model. DBLP is a computer science bibliography dataset available as one big XML file containing: *dblp.xml* an XML file which contains all bibliographic records, and *dblp.dtd* the corresponding document type definition to validate the XML file. The following code is a fragment of the *dblp.xml* file

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE dblp SYSTEM "dblp.dtd">
<dblp>
[... ]
<article key="journals/cacm/Gentry10" mdate="2010-04-26">
<author>Craig Gentry</author>
<title>Computing arbitrary functions of encrypted data.
</title>
<pages>97-105</pages>
<year>2010</year>
<volume>53</volume>
<journal>Commun. ACM</journal>
<number>3</number>
<ee>http://doi.acm.org/10.1145/1666420.1666444</ee>
<url>db/journals/cacm/cacm53.html#Gentry10</url>
</article>
```

3.2 Ontology Construction

The objective of this phase is to design an ontology specifying a set of initial requirements consisting of the concepts that the ontology must include and the competency. Once the ontology model has been axiomatically defined and onstructed, then a set of Java modules should be implemented to fully automate the population of the ontology considering the data sources available.

3.3 Ontology Enrichment

The objective of this phase is to automatically process and analyze the population of the ontology to discover new interesting semantic relations between individuals. In order to execute ontology enrichment in this project a set of semantic inference rules and queries were defined.

4 Specification of Ontology Requirements

The main objective of the ontology model reported in this paper is to represent, process, and reason about the information of researcher profiles, considering that every research institution requires the efficient management and dissemination of information relative to the professors and research activities. The ultimate goal of this project is the smart and provisioning of services to academic communities in which users search for specialized information (such as publications, coauthoring, conferences, etc.) and are also interested in establishing collaborations with other researchers. Considering this motivation, the following requirements were defined in order to guide the ontology design, construction and evaluation.

4.1 Concept Coverage

In order to specify the scope of the ontology, we reviewed the concepts of research profile. Yao, Tang and Li [2] described profiling as the process of obtaining the values associated with the different properties that constitute the person model. Authors define the schema of a researcher profile containing: name, affiliation, position, phone, address, email, research interests, and postgraduate studies. From this definition we consider that the entities (or objects) that constitute a researcher profile are: **Person** (for example professors, students, staff, etc.), and **Publication** (to extract research interests). From this initial analysis we have defined the concept coverage requirements of the ontology and defined the main objective of the ontology, which is to facilitate researcher profile representation and processing in the academic environment.

The ontology model should include the following concepts:

- a) A complete definition of **Person (Researcher)** consisting of full name, unique identifier, department, gender, affiliation department, academic title, etc.
- b) A complete definition of **Publication** consisting of: title, authors, key words, volume, year of publication, among others.
- c) A semantic relationship to correlate a **Researcher** as an **author** of a **Publication**.
- d) A semantic relationship to correlate **collaborations** between **Researcher** individuals.

4.2 Competency of the Ontology

Based on the definition of Gruninger and Fox [8] the competence of an ontology model is the set of questions that the ontology can answer. Evaluation of the competency of an ontology system is crucial to verify that a representational model is complete with respect to a given set of competency questions. The following constitutes the types of information that results of interest for the evaluation of the ontology model:

1. To know how many scientific works does a given researcher has published
2. To find groups of authorship collaborations

3. To know the researcher's publications from a topic with high degree studies
4. To answer about statistical data of publications
5. To know the scientific productivity of a given department
6. To know the number of female researchers from a given department with published scientific works.

5 Ontology Design

Ontology design is the process of selecting and applying methods, techniques and principles with the objective of producing an ontology model. In this section, the design considerations that were taken into account are described.

5.1 Person Ontology

Person ontology (shown in Figure 2) was designed to represent all possible academics that hold a permanent or temporal position as professor or researcher at the university, such as: academic visitor, full time professor, external sabbaticals, etc. This ontology also represents postgraduate students, and research oriented undergraduate students, among others. An important characteristic of this ontology is that it uses a unique identifier for every type of person. The concept *Person* is defined as an equivalence through the *hasName* and *hasGender* data properties, indicating that every person individual is obligated to have name and gender to be classified as type of *Person* class. The concept *Employee* is defined as a sub class of a *Person* that *hasEconomicNumber* data property. Whereas the concept *Student* is defined as a sub class of *Person* that *hasStudentId*. An important concept is a *Professor* which is an *Academic*, is an *Employee* and is a *Person* that *hasCategory*, *hasDepartment*, and *hasEmail*; and inherits the data property of an *Academic* *hasProject*. The class hierarchy of the *Person* ontology shows the sub-classification of the class *Student* into *RegularStudent* and *AssistantStudent*. This classification addresses the particular need to represent the two types of students that exist in the university where an individual of the *AssistantStudent* class is considered to be an *Academic*, an *Employee* and a *Student*.

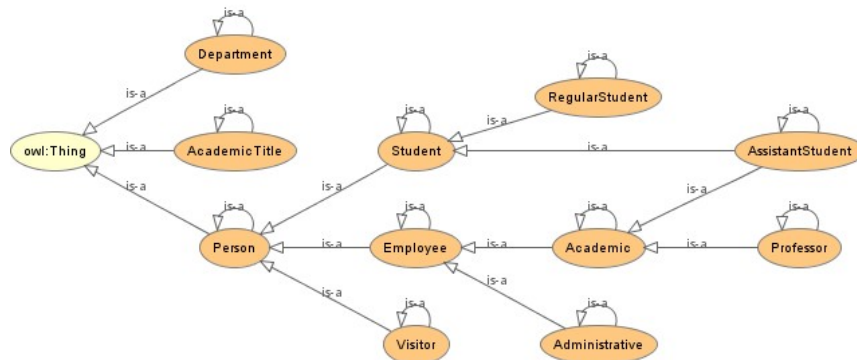


Fig. 2. Hierarchy class of the Person ontology.

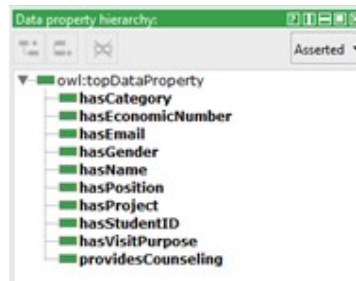


Fig. 3. Data properties defined for the Person ontology.

5.2 Publication Class

Scientific published articles, chapters, and books are the most important sources of information in order to fully integrate a researcher profile. Scientific publications contain the author's topics of interest, conferences and journals of preference, the years of publications and periodicity; also the researcher most active collaborations. In order to build a researcher profile, the design and construction of a **Publication** class considered as input the information extracted from the DBLP digital library, extracting the most relevant bibliographic information on major computer science publications. The **Publication Class** defines the same attributes utilized in DBLP: **hasAuthor** to post process semantic relations between **Professor** and **Publication** classes, **hasEE** is a DBLP unique identifier of the publication, **hasKey** to aggregate the set of keywords, **hasPages**, **hasTitle**, **hasURL**, **hasVolume**, and **hasYear**; all datatype attributes extracted from the DBLP xml file and represented in the ontology model as data type properties.

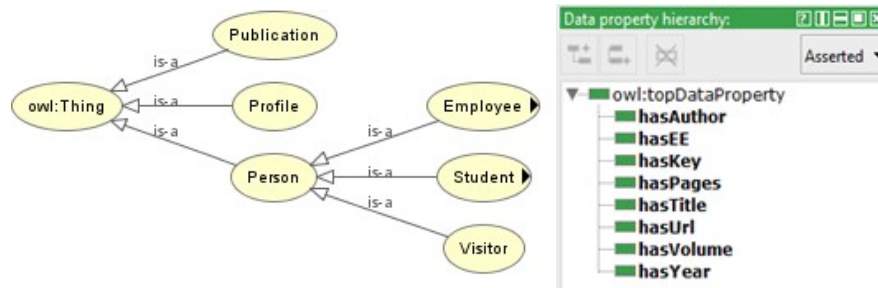


Fig. 4. Publication class hierarchy and data proerties.

5.3 Researcher Profile Ontology

The Researcher Profile Ontology incorporates the conceptualizations from *Person Ontology* and *Publication Class*. From *Person* ontology imports *Professor* personal data, such as full name, and economical number; from the *Publication* Class imports publications organized by year, type of publication, among others. Additionally, incorporates *Department* and *AcademicTitle* concepts. All these conceptualizations are used to complete the definition of a *Researcher Profile*, considering the associated publications, the affiliated department, the academic title obtained, and the rest of personal data.

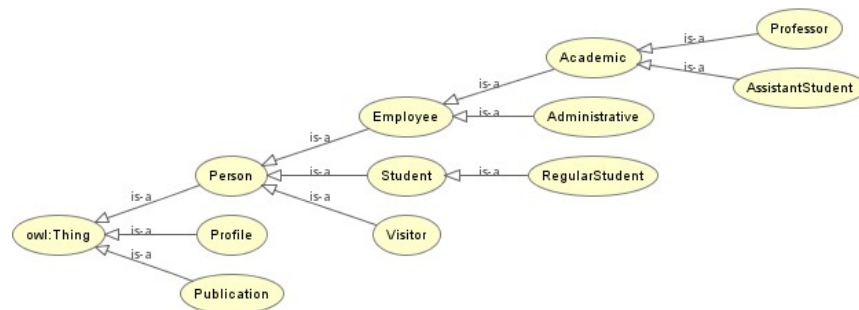


Fig. 5. Publication class hierarchy and data proerties.

6 Ontology Population

Ontology population is the process of adding (instantiating) new individuals in the ontology concepts (classes). Automated ontology population is desirable due to the large amount of data that must be extracted and instantiated in the *Person* and *Publication* concepts.

6.1 Person Ontology Population

For *Person* ontology population, initially the data source came from a set of excel files that the management staff of the university uses for different purposes. These excel files contain the information of all academic staff who is affiliated with the university, such as: professor's full name, gender, department, email, economical number, academic projects, and alias. For the automated population of the *Person* ontology, two Java modules were developed: a module to parse and extract the information from the source files; and another module to interact with the ontologies using the Java OWL Application Programming Interface (OWL API) to load and manipulate ontologies, creating new individuals, instantiating object properties and data properties with individuals, and register them in the ontology. Alias names are important in order to facilitate the semantic association of the researcher individual with all his publications. Figure 6 shows an example of an individual from the *Person* – *Professor* class which was automatically populated.

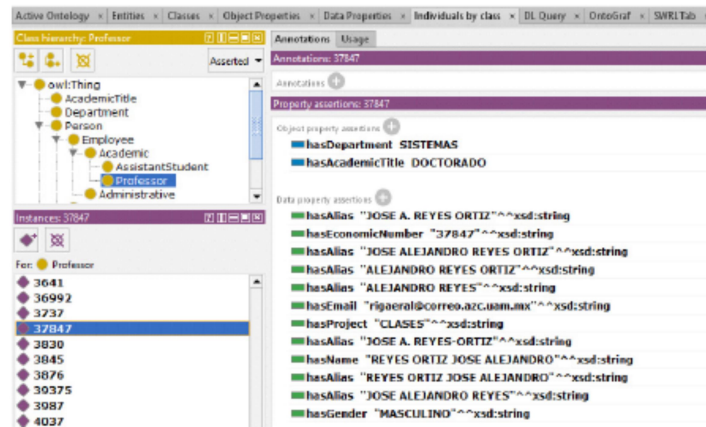


Fig. 6. Person Ontology population.

6.2 Publication Class Population

For *Publication* Class population, the data was extracted from the DBLP XML file, which contains more than a million of Computer Science publications. The XML file from the DBLP contains publication title, author names, publication year, volume, EE (a unique publication identifier), URL, and pages. However, it does not provide the abstract and keywords of publications. A Java module was built to interact with the ontologies using the OWL API to load and manipulate ontologies, creating new individuals, instantiating object properties and data properties. Figure 7 shows an example of a *Publication* Class individual automatically created.

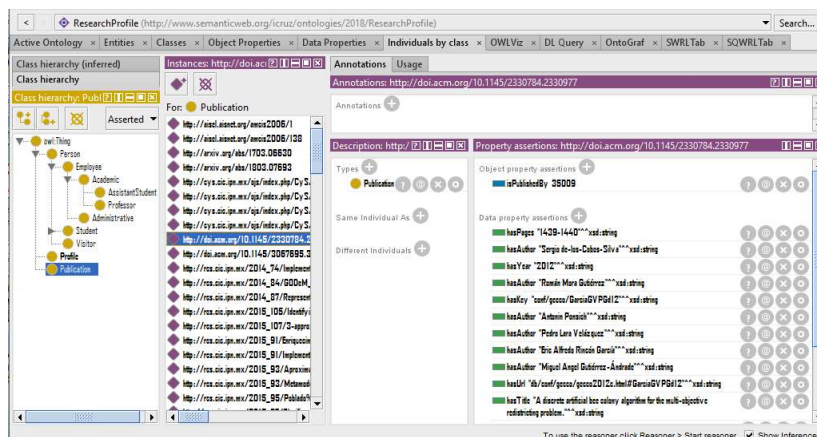


Fig. 7. Publication Class population.

7 Ontology Enrichment

Ontology enrichment is the automatic process of analyzing the population data values and discovering new interesting semantic relations between individuals. Of particular interest in this enrichment process is the automatic discovery of authorships and collaboration relations between authors of publications. For this, the object properties shown in Figure 8 and inference rules were defined.

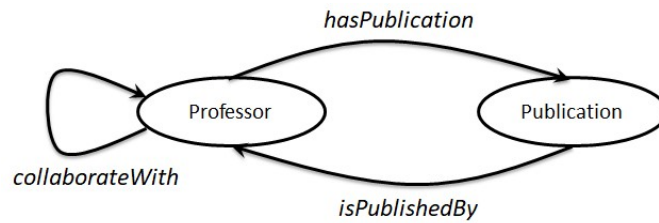


Fig. 8. Object properties utilized for ontology enrichment.

isPublishedBy is an object property with domain **Publication** and range **Professor**. This object property is instantiated during the **Publication** automated population, for each **Publication** instance, the list of authors is extracted and a Java module that recognizes authorships is executed (using the aliases); as an outcome the exact correlation is established. This object property has an inverse definition *hasPublication*, which has as domain the **Professor** class and range the **Publication** class. By executing the reasoner, this inverse semantic relationship is inferred.

collaborateWith is an object property with domain **Person** and range **Person**. This object property was defined to establish semantic relationships between authors of publications, considering that there is the possibility of students, visitors and professors to collaborate. In order to find collaboration relationships, the following SWRL rule is fired resulting in new object property instantiations. Figure 9 shows the result of the ontology enrichment process after the execution of the rule.

```
Publication(?pub) ^ isPublishedBy(?pub, ?prof) ^  
isPublishedBy(?pub, ?prof2) ^ differentFrom(?prof,  
?prof2) ^  
-> collaborateWith(?prof, ?prof2)
```

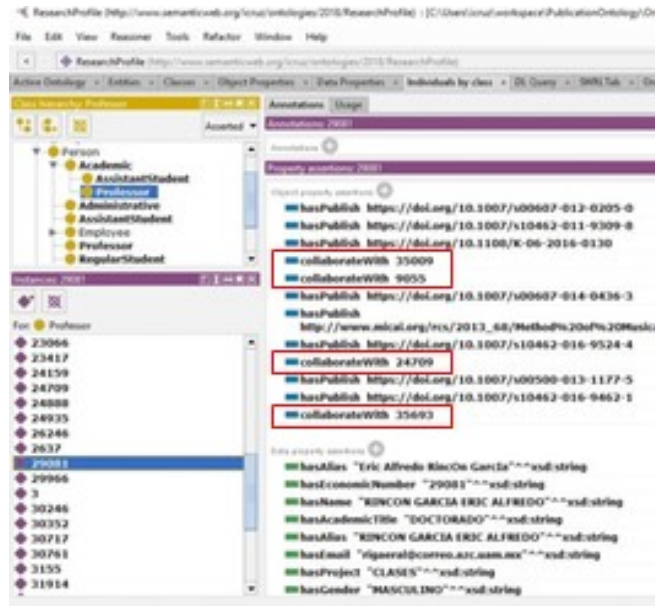



Fig. 9. Object properties utilized for ontology enrichment.

8 Ontology Evaluation

Ontology Evaluation [10] concerns the correct building of the ontology, ensuring that its definitions correctly implement the ontology requirements and competency questions. For evaluation the competence of the ontology was considered, that is, if it is able to respond to a set of competency questions; and the verification of requirements compliance. The following competency questions were coded in SWRL language and their results were correct.

8.1 Researcher Publications

To know how many scientific works does a given researcher has published, the following rule was defined and executed. Figure 9 shows the result of this competency question.

```
Professor(?prof1) ^ hasPublish(?prof1, ?pub) ^
hasEconomicNumber(?prof1, ?e) ^ swrlb:equal(?e, "14233")
-> sqwrl:count(?pub)
```

8.2 Collaboration between Researchers

To find groups of authorship collaborations, the following semantic Web rule is defined. The result is shown in Figure 10.

```
Professor(?p1) ^ collaborateWith(?p1, ?p2) ^  
hasEconomicNumber(?p1, ?e) ^ swrlb:equal(?e, "14233") ->  
sqwrl:select(?p1, ?p2)
```

8.3 Qualified and Specialized Researchers

In order to know the researcher's publications from a specialized topic with high degree studies, the following semantic Web rule was utilized. Figure 11 shows the result of the execution.

```
Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^  
swrlb:contains(?dep, "ELECTRONICA") ^  
hasAcademicTitle(?prof1, ?at) ^ swrlb:equal(?at,  
"DOCTORADO") ^ hasPublish(?prof1, ?pub) ->  
sqwrl:count(?pub)
```

8.4 Publications by Year

The ontology is capable of answering statistical data of publications, for instance: How many publications were there in the year 2017? Figure 12 shows the result of this rule.

```
Publication(?pub) ^ hasYear(?pub, ?y) ^ swrlb:equal(?y,  
"2017") -> sqwrl:count(?pub)
```

8.5 Publications by Department

In order to know the scientific productivity of a given department, the following semantic Web rule is used. Figure 13 shows the result.

```
Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^  
swrlb:contains(?dep, "SISTEMAS") ^ hasPublish(?prof1,  
?pub) -> sqwrl:count(?pub)
```

8.6 Publications by Gender

To know the number of female researchers from a given department that have published scientific works, the following semantic Web rule was used. Figure 14 shows the results.

```
Professor(?prof) ^ hasGender(?prof, ?gen) ^
```

```
swrlb:equal(?gen, "FEMENINO") ^ hasDepartment(?prof,  
?dep) ^ swrlb:equal(?dep, "SISTEMAS") ^ hasPublish(?prof,  
?pub) -> sqwrl:count(?prof)
```

9 Conclusions

In the work reported in this paper, an ontology-based approach was used to populate a *Researcher Profile Ontology*. The researcher profile ontology consists of three main concepts: the *Person Ontology*, the *Publication Class*, and the *Profile Ontology*. For automated population of the *Person Ontology*, a dataset of 373 professors was used, two departments were instantiated in the *Department* class, 50 professor individuals are from the *Systems Department* and 60 from *Electronics Department*. 100% of them were correctly inserted in the Person ontology, specifically in the *Professor* class. For experimentation purposes two ontologies were generated: one was used for professors of the electronics department and another for professors of the systems department. The ontology had a total of 50 researchers in total. Once the universe of professors was divided, the *Publication* ontology was populated using as a data source the DBLP file, which contains approximately one million articles and more than 56 million lines. The result of the population of the *Publication* Class with professors of the Systems department resulted in a total of 135 publications that coincided between the aliases of the professors and the authors indicated within the <author> label of the DBLP. At the same time 116 collaboration relationships were found among the professors from the Systems department.

The ontology of *Professor* from the Electronics department, with a total of 58 professors, was subjected to the same test as the ontology of the Systems department and 22 publications were found from the same DBLP file, that is, all the aliases were compared of the researchers of the Electronics department against the authors of the million articles of the DBLP. In this ontology 4 collaborative relationships were found, that is, in two publications two or more professors from the same department participated. As future work, other sources can be considered to continue enriching the ontologies with more semantic relationships, such as ArnetMiner [4] which contains abstracts and keywords of publications in order to enable the semantic relationship between publications, researchers and topics of interest. In this way the ontology would comply with the characteristic of being scalable and make the profile of each researcher a more complete.

References

1. Maedche, A., Staab, S.: Ontology Learning. In: Handbook on Ontologies (2004)
2. Yao, L., Tang, J., & Li, J. (2007, November). A unified approach to researcher profiling. In Proceedings of the iee/wic/acm international conference on web intelligence (pp. 359-366). IEEE Computer Society.
3. Liu, P., Liu, K., & Liu, J. (2007, September). Ontology-based expertise matching system within academia. In Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on (pp. 5431-5434). IEEE.

4. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008, August). Arnetminer: extraction and mining of academic social networks. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 990-998). ACM.
5. Thiagarajan, R., Manjunath, G., & Stumptner, M. (2008). Finding experts by semantic matching of user profiles (Doctoral dissertation, CEUR-WS).
6. Adnan, S., Tahir, A., Basharat, A., & de Cesare, S. (2009, September). Semantic agent oriented architecture for researcher profiling and association (semora). In Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on (Vol. 3, pp. 559-562). IEEE.
7. Punarnut, R., & Sriharee, G. (2010, January). A researcher expertise search system using ontology-based data mining. In Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling-Volume 110 (pp. 71-78). Australian Computer Society, Inc.
8. Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies.
9. <https://dblp.uni-trier.de/db/about/>
10. Gómez-Pérez, A. Towards a framework to verify knowledge sharing technology. Expert Systems with Applications, 11(4), 519-529, (1996).

Representación de Perfiles de investigación usando ontologías

Isabel Cruz Ruiz¹ y Juan Villegas-Cortez²

¹Universidad Autónoma Metropolitana, Azcapotzalco. Departamento de Sistemas. San Pablo Xalpa No.180, Col Reynosa Tamaulipas, CP 02200, Ciudad de México.

²Universidad Autónoma Metropolitana, Azcapotzalco. Departamento de Electrónica. San Pablo Xalpa No.180, Col Reynosa Tamaulipas, CP 02200, Ciudad de México.
icruzr2014@gmail.com, juanvc@azc.uam.mx

Resumen - En el presente trabajo se muestra una metodología que servirá de guía para el diseño, creación y modelado de una ontología, la cual consiste en lo siguiente: a partir de una base de datos contenida en un archivo XML, donde la información que contiene se encuentra semi-estructurada, pase por un proceso de extracción de etiquetas que representan el perfil académico de un investigador, se hace lectura de algunos registros, pre-procesado de etiquetas XML, conformación de patrón, poblado de ontología y evaluación de la misma a través de preguntas de competencia. Esto tiene como objetivo hacer el poblado automático de la ontología, sin utilizar un recurso humano de supervisión para la organización y estructurado de toda la información.

Palabras clave: Ontología, minería de datos, DBLP.

Abstract— This paper presents a methodology in order to guide the design, creation and modeling of an ontology, which consists of the stages: from an XML file, where the information on it is presented in form semi-structured, pass through a process of extracting labels which represent the academic profile of a researcher, reading some records, pre-processing XML tags, pattern forming, populating ontology and evaluating it through competition questions. This aims to make the automatic population of the ontology, without using a human supervision resource for the organization and structured of all the information.

Keywords: Ontology, data mining, DBLP.

I. INTRODUCCIÓN

En la actualidad la información que se encuentra disponible en la web, además de estar en crecimiento exponencial, no tiene una estructura semántica que permita a un computador procesarla, al contrario, el ser humano debe usar recursos tanto humanos como materiales para lograr este objetivo. Tom Gruber afirma que “una ontología es la especificación explícita de una conceptualización” [1]. Las ontologías han resultado ser útiles para hacer la tarea de a partir de un archivo

no estructurado o semi-estructurado, procesar la información y de alguna manera mantenerla ordenada para que se pueda utilizar en distintas aplicaciones y que sea adaptable a las necesidades de cualquier aplicación. Cabe mencionar que una ontología bien construida, es reutilizable, es decir, se puede “combinar” con una o más ontologías; tal que se utilice en n aplicaciones diferentes; y actualizable, es decir, como consecuencia de que sea reutilizable y usable, la ontología tiene posibilidades de crecer en tamaño, calidad, relaciones y mantenerse vigente.

Para efectos del presente trabajo se creará una ontología de perfiles académicos, considerando a un perfil tal que tiene características de una persona, como gustos, habilidades, profesión, etc. Un perfil académico es aquel que contiene publicaciones, redes sociales académicas, temas científicos de interés, investigadores que trabajan en un tema determinado, etc. Los perfiles académicos y la ontología son expresados en lenguaje OWL (Web Ontology Language).

El DBLP es un archivo semi-estructurado con etiquetas XML que ayudará a extraer la información correspondiente a perfil académico de un investigador, ya que contiene aproximadamente un millón de artículos de ciencias de la computación [2].

Un paso importante para construir un buen diseño es la metodología por seguir para lograr la construcción de una ontología. Algunas de las principales metodologías hasta ahora conocidas para la creación y desarrollo de una ontología se encuentran: Lenat y Guha [8], Uschold y King [9], Grünninger y Fox [10], METHONTOLOGY [11]. La metodología Uschold y King consta de una serie de procesos para la construcción de una ontología y son: identificar el proceso de la ontología, construir la ontología, evaluación y documentación.

La metodología propuesta en este trabajo abarca algunos puntos de estos trabajos, además de complementarse con otros procesos para representar un perfil académico de investigador.

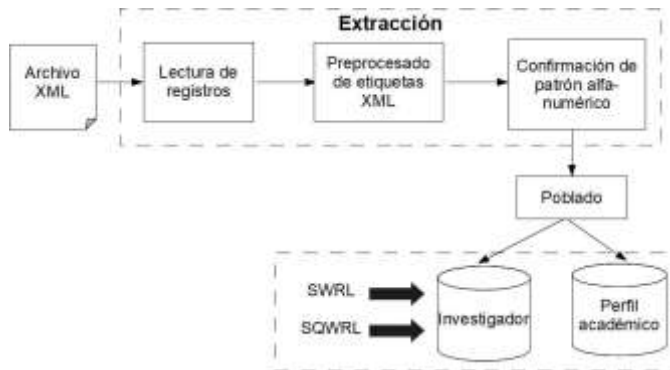


Fig. 1. Metodología propuesta para la construcción de perfiles académicos a partir de un archivo XML.

II. TRABAJOS RELACIONADOS

Dentro de la literatura encontramos algunos trabajos relacionados tanto con perfiles académicos, metodología para la creación, definición, diseño, evaluación y documentación de una ontología.

Una ontología, de acuerdo con Thomas Gruber (1993) define un conjunto de representaciones, las cuales son clases, atributos o propiedades y relaciones con las que se modela un dominio de conocimiento o discurso. En el presente trabajo se utilizarán ontologías ya que se han convertido en una opción muy flexible para dar estructura a la información que se almacena en algunos textos, además de representar el conocimiento [14].

De acuerdo con [6] la ontología es una descripción formal explícita de conceptos (clases) en un dominio con un conjunto de propiedades. Las propiedades de cada concepto describen las características y atributos del concepto. Las ontologías también se pueden definir como un gráfico conceptual para representar las consultas y las descripciones de los recursos.

Como se menciona en los trabajos [3], [6] y [13] han usado como fuente de información el archivo DBLP, una de las razones es porque contiene los datos que se van a utilizar para poblar la ontología, además de contar con información de investigadores y de artículos, los cuales se utilizarán para crear el perfil del investigador.

Para crear una ontología educativa y de investigación se necesita definir publicaciones, con sus tipos y conexiones a otros conceptos. FOAF (por sus siglas en inglés Friend Of a Friend), tiene un concepto llamado documento, y se utiliza como un rango de relación de publicaciones. FOAF no tiene las propiedades que se necesita, como los tipos de publicación o una relación de coautor. Por esta razón, se utiliza la ontología DBLP [3]. El DBLP, junto con CiteSeer y Google Scholar, son plataformas populares, pero no emplean información basada en semántica; este tipo de información proporciona contenido estructurado para facilitar la comprensión y reutilización de la misma.

En los trabajos [1], [3], [4], [6], [7] y [13] utilizan perfiles de investigación, es decir, características académicas de un investigador, como grado académico, publicaciones, co-autores, temas de interés, etc.

Los trabajos [1] y [6] hacen poblado de ontología.

III. ARQUITECTURA DE SOLUCIÓN

Después de analizar el estado del arte, se propone una metodología para resolver el problema de darle estructura semántica a la información que está contenida dentro de las etiquetas XML del archivo que se obtiene del DBLP. En el presente trabajo se utilizará este archivo debido a las características que incluye, como el contener artículos científicos y estar semi-estructurado.

La arquitectura que se utilizará es la mostrada en la Figura 1. Esta arquitectura se compone de las siguientes etapas:

1) *Archivo DBLP.*

Por sus siglas en inglés Digital Bibliography & Library Project. Es un repositorio bibliográfico de artículos científicos, actualmente cuenta con una gran cantidad de artículos y día a día aumenta el número de estos. Cada uno de los artículos tiene un conjunto de etiquetas como autor, título del artículo, año de publicación, título, páginas y URL.

2) *Lectura de registros.*

Un registro representa un objeto único de datos implícitamente estructurados ya sea en una tabla, renglón o conjunto de renglones.

3) *Pre-procesado de etiquetas XML.*

La ontología DBLP crea descripciones OWL de los conceptos DBLP correspondientes. El sitio web del DBLP es una herramienta popular para rastrear el trabajo de los investigadores en ciencias de la computación. DBLP originalmente tiene una API basada en XML, que no es una representación perfecta, es decir, está semi-estructurado. Sus datos son accesibles a través de archivos XML y su aplicación es como un origen de datos.

4) *Poblado.*

En esta fase se implementarán métodos computacionales para la creación de instancias y relaciones en el modelo ontológico a partir de la información académica extraída de los documentos XML como el DBLP [3], por lo que se almacenará dentro de tres diferentes ontologías: investigador, perfil académico y persona.

5) *Reglas SWRL*

Por sus siglas en inglés Semantic Web Rule Language y SQWRL (por sus siglas en inglés Semantic Query Web Rule Language), la función de estas reglas es evaluar la ontología, a través de preguntas de competencia, las cuales son cuestionamientos en lenguaje natural que se utilizarán para verificar que la información que se encuentra almacenada en la ontología está bien estructurada, tiene coherencia y es consistente.

IV. IMPLEMENTACIÓN

Para construir la ontología de perfiles académicos hay diferentes fuentes de información relevante que debe ser recolectada. En [13] se utilizan tres diferentes categorías de información de la publicación de los investigadores, las cuales son: título, resumen (abstract) y palabras clave o keywords.

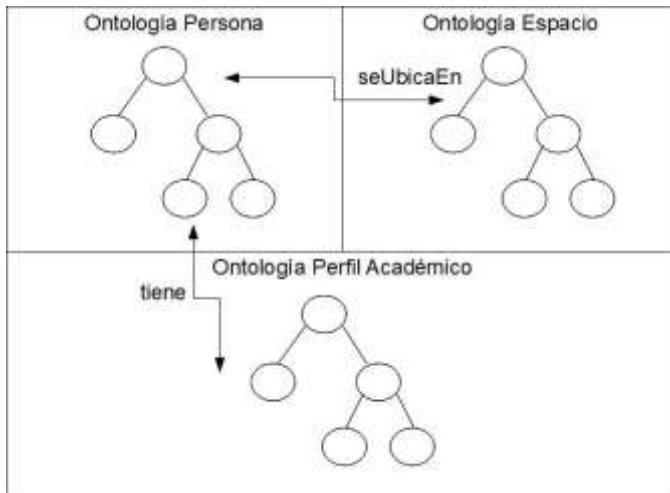


Fig. 2. Esquema general de las ontologías que se propone crear y posteriormente poblar.

En la Figura 2 se muestra un esquema general con las ontologías que resultarán después de ejecutar la metodología.

En esta figura se muestran tres ontologías, persona, espacio y perfil académico. La ontología de persona incluye publicaciones y datos académicos.

El lenguaje de ontología web OWL 2, informalmente OWL 2, es un lenguaje de ontología para la Web semántica con significado definido formalmente. Las ontologías OWL 2 proporcionan clases, propiedades, individuos y valores de datos y se almacenan como documentos web semánticos. Las ontologías OWL 2 se pueden usar junto con la información escrita en RDF, y las ontologías OWL 2 se intercambian principalmente como documentos RDF [15].

La evaluación de la ontología se hace de la siguiente manera: se escriben algunas preguntas de competencia, estas preguntas son expresiones en lenguaje formal e informal. Estas preguntas se traducen a los lenguajes como DL Query o SWRL, para determinar la consistencia de la ontología y ver si cuenta con la estructura para la que fue diseñada, es decir, que cuente con las instancias, relaciones, atributos necesarios para la generación de nuevo conocimiento.

Las reglas SWRL [12] (por sus siglas en inglés Semantic Web Rule Language), consta de dos partes principales, un cuerpo (antecedente) y la cabeza (consecuente), cada una de ellas está formada por un conjunto de átomos, que a su vez están formados por uno o más argumentos (expresión).

SQWRL [16] (por sus siglas en inglés Semantic Query Web Rule Language) es un lenguaje de consultas basado en el lenguaje SWRL, el cual es parecido al lenguaje de consulta de base de datos SQL, y es utilizado para extraer información de una ontología, así como para evaluar la estructura de la misma. En la Tabla I se muestran ejemplos de reglas SQWRL.

V. RESULTADOS

En el presente trabajo se propuso seguir una metodología para la construcción de ontologías para representar perfiles académicos, a partir de la extracción de información de un archivo semi-estructurado de etiquetas XML, con aproximadamente un millón de artículos, y posteriormente hacer un poblado en la ontología. Una vez que se ha simulado

TABLA I

EJEMPLOS DE PREGUNTAS DE COMPETENCIA

| Pregunta de competencia | Regla SQWRL |
|--|---|
| ¿Cuál es la temperatura ideal en el cubículo del profesor Alejandro? | <code>Professor(?p) ^ hasProfessorName(?p, ?n) ^ hasPreferedTemperature(?p, ?t) ^ swrlb:contains(?n, "REYES ORTIZ JOSE ALEJANDRO") -> sqwrl:select(?n, ?t)</code> |
| ¿Cuál es la temperatura del espacio físico E-212 y E-313? | <code>PhysicalSpace(?s) ^ hasNetworkService(?s, ?t) ^ hasNode(?t, ?n) ^ hasEnvironmentSensor(?n, ?e) ^ hasRealTemperature(?e, ?temp) -> sqwrl:select(?s, ?temp)</code> |
| ¿Quién está presente en el espacio E-212? | <code>Professor(?p) ^ isInvolvedInEvent(?p, ?e) ^ isDetectedBy(?e, ?s) ^ isPartOf(?s, ?n) ^ PhysicalSpace(E212) ^ IsLocatedIn(?n, E212) -> sqwrl:select(?p, E212)</code> |

la ejecución de esta metodología el resultado que se obtiene son tres ontologías: perfil académico, investigador y persona con información respectiva estructurada y consistente. Para la simulación de esta metodología, se utilizó una laptop ASUS con procesador Intel Core i7; para las ontologías se utilizó Protegé 5.2.0.

VI. CONCLUSIONES

En este artículo se ha presentado una metodología para la representación de perfiles de investigadores, los cuales se construyen a partir de un archivo XML (semi-estructurado), y que contiene etiquetas con los datos de artículos como nombre, autor, url, año de publicación, etc. Se ha logrado una extracción de la información de cada una de las etiquetas, y finalmente se hace el poblado de la ontología. Para el objetivo del presente artículo las ontologías resultan ser muy útiles para estructurar información sea la fuente que sea, siempre y cuando esta también sea consistente. A través de la metodología se pudieron crear tres diferentes ontologías de acuerdo con el propósito de este trabajo.

REFERENCES

- [1] <http://tomgruber.org/writing/ontology-definition-2007.htm>
- [2] <http://dblp.uni-trier.de/>
- [3] Challenger, M. (2012). The ontology and architecture for an academic social network. *IJCSI International Journal of Computer Science*, (9), 2.
- [4] Gollapalli, S. D., Mitra, P., & Giles, C. L. (2012, June). Similar researcher search in academic environments. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries* (pp. 167-170). ACM.
- [5] Punnarut, R., & Sriharee, G. (2010, January). A researcher expertise search system using ontology-based data mining. In *Proceedings of the Seventh Asia-Pacific Conference on*

Conceptual Modelling-Volume 110 (pp. 71-78). Australian Computer Society, Inc.

[6] Uddin, M. N., Duong, T. H., Oh, K. J., & Jo, G. S. (2011, April). An ontology based model for experts search and ranking. In Asian Conference on Intelligent Information and Database Systems (pp. 150-160). Springer Berlin Heidelberg.

[7] Bibri, S. E. (2015). Context and Context Awareness of Humans and AML Systems: Characteristics and Differences and Technological Challenges and Limitations. In The Human Face of Ambient Intelligence (pp. 67-128). Atlantis Press.

[8] Lenat D.B., Guha R.V., (1990) Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project, Addison-Wesley, Boston.

[9] Uschold M., King M., (1995) Towards a Methodology for Building Ontologies, in: IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal.

[10] Gruninger M., Fox M.S., (1995) Methodology for the design and evaluation of ontologies, in: Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal.

[11] Gomez-Perez A., M. Fernandez-Lopez, A. de Vicente (1996), Towards a Method to Conceptualize Domain Ontologies, in: ECAI96 Workshop on Ontological Engineering, Budapest, pp. 41–51.

[12] Skillen, K. L., Nugent, C., Donnelly, M., Chen, L., & Burns, W. (2015). Using Ontologies for Managing User Profiles in Personalised Mobile Service Delivery. In Health Monitoring and Personalized Feedback using Multimedia Data (pp. 245-264). Springer International Publishing.

[13] Davoodi, E., & Kianmehr, K. (2012, June). A Semantic-Based Social Network of Academic Researchers. In IEA/AIE (pp. 323-332).

[14] Maedche, A.: Ontology learning for the semantic web. Springer Science & Business Media, Massachusetts, USA (2002).

[15] OWL 2 Web Ontology Language. Document Overview (Second Edition). <https://www.w3.org/TR/owl2-overview/>

[16] <https://github.com/protegeproject/swrlapi/wiki/SQWRL>



Ing. Isabel Cruz Ruiz recibió su título de Ingeniero en Computación por la Universidad Autónoma Metropolitana (UAM), unidad Azcapotzalco, en el año 2010, actualmente es estudiante de posgrado de la Maestría en Ciencias de la Computación, de la UAM Azcapotzalco.

Ha trabajado en desarrollos de software con bases de datos para dependencias gubernamentales e internacionales. Sus intereses de investigación son el desarrollo de sistemas de información inteligentes, la minería de datos y el desarrollo de ontologías.



Juan Villegas-Cortez recibió su licenciatura en Matemáticas Aplicadas en Ciencias de la Computación en la Universidad Autónoma Metropolitana, Unidad Iztapalapa (UAM-Iztapalapa) en 1996, obtuvo su Maestría en Ciencias de la Computación en la Universidad Autónoma Metropolitana, Unidad Azcapotzalco (UAM- Azcapotzalco) en

2005, recibió su doctorado en Ciencias de la Computación en el Laboratorio de Reconocimiento de Patrones del Centro de Investigación en Computación, del Instituto Politécnico Nacional en 2009. Ha sido profesor investigador en el Departamento de Electrónica, de la Universidad Autónoma Metropolitana desde 2010. Sus intereses actuales de investigación incluyen Reconocimiento de Patrones, Análisis de Imágenes, Visión por Computadora, Computación Evolutiva, Redes Neuronales Artificiales, Memorias Asociativas y Seguridad en Redes de Cómputo.

Referencias

- [1] Georgios Petasis, Vangelis Karkaletsis, Georgios Paliouras, Anastasia Krithara, and Elias Zavitsanos, “Ontology population and enrichment: State of the art,” <https://pdfs.semanticscholar.org/942e/578f44623dbe60fb15a5a20b1b588eb572fb.pdf>.
- [2] Tom Grubber, “Ontology,” <http://tomgruber.org/writing/ontology-definition-2007.htm>.
- [3] John F. Sowa, “John f. guided tour of ontology,” <http://www.jfsowa.com/ontology/guided.htm>.
- [4] “Ontologías en la biblioteca digital,” http://www.iar.unlp.edu.ar/biblio/htdocs/archi/boeris_texto.htm.
- [5] G. Antoniou and F. van Harmelen, “A researcher expertise search system using ontology-based data mining,” *Kathryn Laskey*, vol. 46, no. 3, 2008.
- [6] María Belén Mora Arciniega y Verónica Segarra Faggioni, “Modelo ontológico para la representación de datos académicos y su publicación con tecnología semántica,” <http://www.redalyc.org/html/310/31048901015/index.html>.
- [7] María Jesús Lamarca Lapuente, “Lenguajes hipertextuales,” http://www.hipertexto.info/documentos/lenguajes_h.htm.
- [8] “¿qué es un documento sgml?” <http://www6.uniovi.es/sgml.html>.
- [9] OWL Working Group, “Owl-semantic web standards,” <https://www.w3.org/OWL/>.
- [10] “Owl,” <https://es.wikipedia.org/wiki/OWL>.

- [11] Javier García Castellano, “Documentos xml,” <http://flanagan.ugr.es/xml/documento.htm>.
- [12] Schloss Dagstuhl, Universität Trier, “dblp - computer science bibliography,” <https://dblp.uni-trier.de/>.
- [13] A. Pretschner and S. Gauch, “Ontology based personalized search,” in *1999. Proceedings. 11th IEEE International Conference*, 1999, pp. 391–398.
- [14] R. Punnarut and G. Sriharee, “A researcher expertise search system using ontology-based data mining,” *Kathryn Laskey*, p. 39, 2017.
- [15] M. Challenger, *The ontology and architecture for an academic social network*. IJCSI International Journal of Computer Science, 2012.
- [16] S. D. Gollapalli, P. Mitra, and C. L. Giles, “Similar researcher search in academic environments,” in *In Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, June 2012, pp. 167–170.
- [17] J. Trajkova and S. Gauch, “Improving ontology-based user profiles,” in *LE CENTRE DE HAUTES ETUDES INTERNATIONALES D’INFORMATIQUE DOCUMENTAIRE*, Apr 2004, pp. 380–390.
- [18] Prof. Manuel Blázquez Ochando, “Técnicas avanzadas de recuperación de información,” <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com/2012/11/frecuencias-y-pesos-de-los-terminos-de.html>.
- [19] M. N. Uddin, T. H. Duong, K. J. Oh, and G. S. Jo, “An ontology based model for experts search and ranking,” in *Springer Berlin Heidelberg*, Apr 2011, pp. 150–160.
- [20] S. E. Bibri, *Context and Context Awareness of Humans and AmI Systems*. Atlantis Press, 2015.
- [21] H. Guermah, T. Fissaa, H. Hafiddi, M. Nassar, and A. Kriouile, *An Ontology Oriented Architecture for Context Aware Services Adaptation*. arXiv preprint arXiv, 2014.

- [22] L. M. McAvoy, L. Chen, and M. P. Donnelly, “A generic upper-level ontological model for context-aware applications within smart environments,” in *UCAmI*, Nov 2013, pp. 270–277.
- [23] K. L. Skillen, L. Chen, C. Nugent, M. Donnelly, W. Burns, and I. Solheim, *Ontological user profile modeling for context-aware application personalization*. Ubiquitous Computing and Ambient Intelligence, 2012.
- [24] H. Chen, T. Finin, and A. Joshi, *An ontology for context-aware pervasive computing environments*. The knowledge engineering review, 2003.
- [25] K. L. Skillen, C. Nugent, M. Donnelly, L. Chen, and W. Burns, *Using Ontologies for Managing User Profiles in Personalised Mobile Service Delivery*. Springer International Publishing, 2015.
- [26] E. Davoodi and K. Kianmehr, “A semantic-based social network of academic researchers,” in *IEA/AIE*, Jun 2012, pp. 323–332.
- [27] B. Sateli and et. al, “ScholarLens: extracting competences from research publications for the automatic generation of semantic user profiles,” *Australian Computer Society, Inc.*, vol. 110, no. 99, pp. 71–78, 2010.
- [28] “El modelo ontológico del museo del prado,” <https://www.museodelprado.es/modelo-semantico-digital/modelo-ontologico>.
- [29] Camilo Alejandro Valencia Martínez, “Definición de un modelo ontológico, que facilite la medición de niveles de usabilidad de herramientas en plataformas lcms mediante sna,” <http://repository.udistrital.edu.com/bitstream/11349/3130/1/ValenciaMart%C3%ADnezCamiloAlejandro2016.pdf>.
- [30] Phillip Cimiano, “*Ontology Learning and Population from Text*,” <https://www.springer.com/us/book/9780387306322>.