

UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN - TACNA

Facultad de Ingeniería

Escuela Académico Profesional de Ingeniería en Informática y Sistemas

**“CARACTERÍSTICAS ONTOLÓGICAS DEL MOTOR DE
BÚSQUEDA SEMÁNTICA “SERON” DEL CONSEJO
DE INVESTIGACIÓN UNJBG, 2014”**

TESIS

PRESENTADA POR:

Bach. Walther Arturo Bueno Mariaca

Para optar el Título Profesional de:

Ingeniero en Informática y Sistemas

TACNA - PERÚ

2015

UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN – TACNA
JURADO CALIFICADOR Y CALIFICACIÓN DE LA SUSTENTACIÓN DE TESIS

TESIS N°

TÍTULO PROFESIONAL DE:

Ingeniero en Informática y Sistemas

La Secretaría Académica de la Facultad de Ingeniería, por Resolución de Facultad N° 02638-2014-FAIN/UNJBG, designó jurado para la sustentación oral de la Tesis titulada: “CARACTERÍSTICAS ONTOLÓGICAS DEL MOTOR DE BÚSQUEDA SEMÁNTICA SERON DEL CONSEJO DE INVESTIGACIÓN UNJBG, 2014”.

El mismo que está conformado por

Presidente : Dr. Julio Miguel Fernández Prado
Secretario : Ing. Edwin Antonio Hinojosa Ramos
Vocal : Ing. Gianfranco Alexey Málaga Tejada

Para calificar la sustentación de la Tesis en acto público el día 31 de diciembre del 2014. Presentado por el Bachiller Walther Arturo Bueno Mariaca, de la Escuela Académico Profesional de Ingeniería en Informática y Sistemas.

El Jurado calificador en forma secreta e individual emitió su opinión sobre el tema de la tesis expuesta y procedió a obtener el promedio que arrojó el calificativo de aprobado con la nota de Diecisiete (17) – promedio sobresaliente.

Para ratificar lo detallado firman:



Dr. Julio Miguel Fernández Prado
Presidente



Ing. Edwin Antonio Hinojosa Ramos
Secretario



Ing. Gianfranco Alexey Málaga Tejada
Vocal

UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN - TACNA

FACULTAD DE INGENIERÍA

ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA EN

INFORMÁTICA Y SISTEMAS

**“CARACTERÍSTICAS ONTOLÓGICAS DEL MOTOR DE BÚSQUEDA
SEMÁNTICA SERON DEL CONSEJO DE INVESTIGACIÓN UNJBG, 2014”**

TESIS PRESENTADA A LA COMISIÓN REVISADORA Y


APROBADA POR EL JURADO CALIFICADOR, INTEGRADO POR:

Presidente:




Dr. Julio Miguel Fernández Prado

Secretario:



Ing. Edwin Antonio Hinojosa Ramos

Vocal:



Ing. Gianfranco Alexey Málaga Tejada

Asesor:



Msc. Edgar Aurelio Taya Acosta

Dedicatoria

A Dios.

Por brindarme las fuerzas necesarias para terminar el presente trabajo de investigación que por muchos años había postergado sin razón alguna.

A mi Familia.

Por apoyarme en todo momento durante el desarrollo de mi vida académica y profesional.

A mis Amigos.

Por brindarme su apoyo en los momentos difíciles y sobre todo por animarme cuando las fuerzas se terminaban.

Agradecimientos

A mi asesor, Msc. Edgar Aurelio Taya Acosta, por sus consejos y sapiencia en la materia de investigación, lo cual fue muy valioso para mi tesis.

A mis profesores, que durante todo el desarrollo de la carrera profesional volcaron valiosos conocimientos que me permiten desenvolverme efectivamente en el campo profesional.

Agradezco también a mi familia y amigos porque sin ellos no hubiera cumplido con el sueño de titularme.

CONTENIDO

ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS	vii
RESUMEN	ix
INTRODUCCIÓN	1
I. PLANTEAMIENTO DE LA INVESTIGACIÓN	4
1.1 <i>Descripción del Problema</i>	4
1.1.1 Antecedentes del Problema	4
1.1.2 Problemática de la Investigación	13
1.2 <i>Formulación del problema</i>	14
1.3 <i>Justificación</i>	14
1.4 <i>Alcances y limitaciones</i>	15
1.5 <i>Objetivos</i>	15
1.5.1 Objetivo General	15
1.5.2 Objetivos Específicos	16
1.6 <i>Hipótesis</i>	16
1.6.1 Hipótesis General	16
1.6.2 Hipótesis Específicas	17
1.7 <i>Variables</i>	18
1.7.1 Identificación de Variables	18
1.7.2 Definición de Variables	19
1.7.3 Operacionalización de Variables	19
1.7.4 Clasificación de las Variables	20
1.8 <i>Diseño de la investigación</i>	20
1.8.1 Diseño Experimental o no Experimental	20
1.8.2 Población y Muestra	20
1.8.3 Técnicas e Instrumentos para recolección de datos	20
1.8.4 Análisis de Datos	21
1.8.5 Selección de Pruebas Estadísticas	21
II. MARCO TEÓRICO	22
2.1 <i>Marco referencial</i>	22
2.1.1 Conceptos Previos	22
2.2 <i>Bases Teóricas respecto al problema</i>	23
2.2.1 Ontologías	23
2.2.2 Ontología en la Informática	27
2.2.3 Componentes de una Ontología	27
2.2.4 Ingeniería Ontológica	29
2.2.5 Principios para el Diseño de Ontologías	30
2.2.6 El Proceso de Desarrollo y Ciclo de Vida Ontológico	32
2.2.7 Datos Enlazados (RDF)	34
2.2.8 RDF Scheme (RDFS)	34

2.2.9	Web Ontology Language (OWL)	35
2.2.10	La Web Semántica	37
2.2.11	Herramientas para Desarrollar Ontologías	39
2.2.12	Protége	41
2.2.13	SparQL	42
2.2.14	Base de Datos	43
2.2.15	Redes	52
2.2.16	Teoría de la Información	54
2.2.17	Methontology	59
III.	DESARROLLO	61
3.1	<i>Especificación</i>	61
3.1.1	Meta	61
3.1.2	Alcance	61
3.1.3	Granularidad	62
3.2	<i>Conceptualización</i>	62
3.2.1	Glosario de Términos	64
3.2.2	Taxonomías	65
3.2.3	Relaciones	67
3.2.4	Diccionario de Conceptos	71
3.2.5	Descripción de Instancias	75
3.2.6	Métricas de la Ontología	81
3.2.7	Diseño Conceptual de la Ontología	82
3.3	<i>Diseño de la Base de Datos</i>	86
IV.	RESULTADOS	87
4.1	<i>Descripción de Características</i>	87
4.1.1	Datos Enlazados (RDF)	87
4.1.2	Vocabularios (OWL)	92
4.1.3	Consultas (SPARQL)	95
V.	DISCUSIONES	99
5.1	<i>Cumplimiento de Objetivos</i>	99
5.2	<i>Primer Objetivo Específico</i>	100
5.3	<i>Segundo Objetivo Específico</i>	102
5.4	<i>Tercer Objetivo Específico</i>	103
5.5	<i>Medición de Indicadores</i>	105
VI.	CONCLUSIONES	106
VII.	RECOMENDACIONES	107
VIII.	REFERENCIAS BIBLIOGRÁFICAS	108

ÍNDICE DE TABLAS

TABLA 1. MEDICIÓN DE INDICADORES	19
TABLA 2. GLOSARIO DE TÉRMINOS	64

ÍNDICE DE FIGURAS

FIGURA 1. VARIABLES DE ESTUDIO	18
FIGURA 2. PROCESO DE DESARROLLO DE ONTOLOGÍAS	33
FIGURA 3. SERIALIZACIÓN DE RDF A XML	34
FIGURA 4. RDFS DE PERSONAS Y GRUPOS DE TRABAJO	35
FIGURA 5. PROTEGE	42
FIGURA 6. MODELADO CONCEPTUAL	60
FIGURA 7. PROCESO DE BÚSQUEDA DE INFORMACIÓN	62
FIGURA 8. FUNCIONES DEL ADMINISTRADOR	63
FIGURA 9. PROCESO DE RETORNO DE INFORMACIÓN	63
FIGURA 10. TAXONOMÍA INVESTIGADORES	66
FIGURA 11. TAXONOMÍA PERSONAL	66
FIGURA 12. TAXONOMÍA PROYECTO	67
FIGURA 13. RELACIÓN PROYECTO - INVESTIGADOR	67
FIGURA 14. RELACIÓN INVESTIGADOR - PROYECTO	68
FIGURA 15. RELACIÓN PROYECTO - FACULTAD	68
FIGURA 16. RELACIÓN REVISTA - PROYECTO	69
FIGURA 17. RELACIÓN REVISTA - PERSONAL	69
FIGURA 18. RELACIÓN REVISTA - PERSONAL	70
FIGURA 19. RELACIÓN REVISTA - IMPRENTA	70
FIGURA 20. RELACIÓN IMPRENTA - UBICACIÓN	71
FIGURA 21. ÁRBOL DE CLASES Y SUBCLASES	72
FIGURA 22. PROPIEDADES DE IMPRENTA	73
FIGURA 23. PROPIEDADES DE FACULTAD	73
FIGURA 24. PROPIEDADES DE INVESTIGADOR	73
FIGURA 25. PROPIEDADES DE PERSONAL	74
FIGURA 26. PROPIEDADES DE PROYECTO	74
FIGURA 27. PROPIEDADES DE UBICACIÓN	74
FIGURA 28. PROPIEDADES DE REVISTA	75
FIGURA 29. INDIVIDUO INVESTIGADOR	76
FIGURA 30. INDIVIDUO PROYECTO	77
FIGURA 31. INDIVIDUO REVISTA	78
FIGURA 32. INDIVIDUO UBICACIÓN	79
FIGURA 33. INDIVIDUO IMPRENTA	80
FIGURA 34. MÉTRICAS DE LA ONTOLOGÍA	81
FIGURA 35. DISEÑO CONCEPTUAL DE LA ONTOLOGÍA	82
FIGURA 36. GRÁFICA DE DOMINIO Y RANGO	83
FIGURA 37. GRÁFICA DE ÁRBOL	84
FIGURA 38. GRÁFICA DE VISTA COMPUESTA	85
FIGURA 39. DISEÑO DE LA BASE DE DATOS	86
FIGURA 40. REGISTRO DEL PROYECTO	86
FIGURA 41: MÉTRICAS DE RDF	89
FIGURA 42: PROPORCIÓN COMPONENTES RDF	90
FIGURA 43: CANTIDAD DE CLASES USADAS EN LA ONTOLOGÍA	90
FIGURA 44: OBJECT PROPERTY EN LA ONTOLOGÍA	91
FIGURA 45: OBJECT PROPERTY ES_INVESTIGADO_POR	92
FIGURA 46: REGLAS PARA LA CLASE UBICACIÓN	93

FIGURA 47: MÉTRICAS PARA OWL	94
FIGURA 48: MÉTRICAS PARA OWL	94
FIGURA 49: TUPLA INVESTIGADOR	97
FIGURA 50: DIAGRAMA INVESTIGADOR	98

RESUMEN

La presente tesis, pretende caracterizar la importancia de los componentes esenciales en los motores de búsqueda semántica, tomando como base sus componentes principales como son: RDF, OWL y SPARQL.

El uso de ontologías permitirá modelar el conocimiento logrando de esta manera crear repositorios que puedan ser consumidos por cualquier aplicación, esto impulsará la filosofía de la reutilización, propiedad que los sistemas ontológicos permiten explotar al máximo.

Sin embargo, es importante brindar las pautas sobre las buenas prácticas en el desarrollo de ontologías, de tal forma que se puedan desarrollar sistemas de calidad basados en estándares mundiales de tal forma que la interoperabilidad sea transparente para las distintas plataformas de desarrollo.

Las aplicaciones ontológicas que basan su diseño en los tres factores básicos (RDF, OWL y Sparql) podrán ser consumidas fácilmente por cualquier aplicación indistintamente de la plataforma tecnológica, lo cual es el objetivo primordial del desarrollo de sistemas ontológicos.

INTRODUCCIÓN

El desarrollo de nuevas metodologías y técnicas para mejorar el desarrollo de Sistemas que administran conocimiento está en una etapa de apogeo. Los Sistemas del futuro podrán convertir la información en conocimiento.

Específicamente en Tacna, se trata de llevar a la práctica los sistemas basados en ontologías, sin embargo, hace falta verificar el uso de las buenas prácticas del diseño ontológico para poder de esta forma describir la robustez de tales sistemas del conocimiento.

La World Wide Web Consortium detalla los requisitos para lograr un sistema ontológico (motor de búsqueda semántica) de calidad para poder funcionar adecuadamente y así contribuir en lo que será la web semántica.

La presente investigación se introducirá en el mundo de la Ingeniería Ontológica para aplicarla a un caso real el cual es: “Características Ontológicas del Motor Semántico SERON del área COIN - UNJBG”, entiéndase por COIN (Consejo de Investigación).

Para solucionar los problemas planteados, esta tesis se organiza de la siguiente manera:

El capítulo 1 se describe los objetivos y alcances de los Sistemas Ontológicos, así como sus características de acuerdo a las recomendaciones del Consorcio de la World Wide Web.

El capítulo 2 presenta el marco referencial que describe los conceptos relacionados con los Sistemas Ontológicos, La Web Semántica, Las Herramientas de Desarrollo y Métricas para Ontologías y las Metodologías para Ontologías. También se hará una somera descripción de los fundamentos de los Sistemas Ontológicos, es decir, Base de Datos, Redes, Teoría de la Información, etc.

El capítulo 3 se centra en las características del motor de búsqueda semántica (Datos Enlazados, Inferencias y Consultas), así como las herramientas utilizadas para tal propósito.

En el capítulo 4 se muestran los resultados obtenidos describiendo cada una de las características fundamentales de los buscadores semánticos.

En el capítulo 5 se comprueba si se cumplieron los objetivos específicos así como las pruebas de hipótesis utilizando Modus Ponendo Ponens para las variables cualitativas del estudio.

El capítulo 6 muestra las conclusiones y recomendaciones para investigaciones futuras que mejoren la medición de los buscadores semánticos.

I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 Descripción del Problema

1.1.1 Antecedentes del Problema

En la investigación realizada por **Graciela Barchini, Margarita Álvarez y Susana Herrera** (Barchini, Alvarez, & Herrera, 2006) cuyo título es: “Sistemas de Información: Nuevos escenarios basados en ontologías”, publicada en la revista “Gestão da Tecnologia e Sistemas de Informação” de la Universidad Nacional de Santiago del Estero – Argentina, se afirma lo siguiente:

- Mientras la ontología sigue siendo un área fecunda de investigación en el campo de la filosofía, las ontologías son actualmente materia de investigación, desarrollo, y aplicación en disciplinas relacionadas con la computación, la información y el conocimiento. Los sistemas de información (SI) son esencialmente artefactos de conocimiento que capturan y representan el conocimiento sobre ciertos dominios.

Considerando que las ontologías generalmente se usan para especificar y comunicar el conocimiento del dominio y que existe un reconocimiento creciente que los principios y conceptos ontológicos pueden aplicarse fructíferamente en el campo de los SI, se puede afirmar que un SI tiene su propia ontología implícita, al atribuir significado a los símbolos usados. Sin embargo, de manera explícita, una ontología puede tener distintos roles en un SI. Los escenarios presentados en este artículo permiten visualizar el rol que las ontologías están teniendo o tendrán en un futuro próximo en los SI. Se ofrece un amplio panorama del uso de las ontologías en diversos escenarios que varían desde el análisis conceptual de técnicas de modelación de los SI hasta su uso como soporte para el diseño, desarrollo y uso de los SI. Se espera de esta manera incrementar nuestro conocimiento acerca del rol de las ontologías en los Sistemas de Información.

Al respecto podemos manifestar que las ontologías permiten construir sistemas robustos basados en una nueva filosofía de desarrollo, en el cual la gestión del conocimiento es el objetivo principal. Sin embargo, se debe tomar en cuenta la necesidad de

medir o evaluar las ontologías para lograr un sistema semántico de calidad.

Por otro lado, **María de los Ángeles Martín (2004)** en su tesis: “Sistema de Catalogación de Métricas e Indicadores con Potencia de Web Semántica”, para optar el grado de magister en Ingeniería de Software en la Universidad de la Plata – Argentina, en el cual concluye a lo siguiente:

- En las últimas décadas han surgido numerosas propuestas que definen conjuntos de métricas de calidad para la evaluación de productos y procesos de software en distintos dominios de aplicación. También existen en la actualidad algunos trabajos que proponen marcos conceptuales para la documentación de dichas métricas y catálogos que brindan información de utilidad en los proyectos de medición y/o evaluación de software. Sin embargo, hemos observado que no existe un consenso general sobre cómo ha de documentarse la información de métricas e indicadores en forma efectiva, de manera de facilitar su explotación, reutilización y aplicación a los procesos de medición y evaluación, de manera consistente. Además, si bien existen

documentos estándares que definen parcialmente la terminología usada para el ámbito de modelos de calidad, métricas, y procesos de medición y evaluación [ISO9126, ISO14598, ISO15939], no existe una terminología universal consensuada entre los investigadores del área, y en consecuencia, la información relacionada proveniente de distintas fuentes no usa un mismo vocabulario, dificultado su aplicación en forma generalizada. En esta tesis se muestra el trabajo realizado para la construcción del “Sistema de Catalogación de Métricas e Indicadores con potencia de Web Semántica”, que intenta dar solución a dichos problemas. Este trabajo proporciona por un lado, una Ontología de Métricas e Indicadores que especifica un vocabulario común, con una semántica bien definida, facilitando la interpretación de la información de métricas e indicadores en forma uniforme, y permitiendo que puedan ser comparados los resultados obtenidos en distintos proyectos en forma consistente. Por otro lado, la implementación de un sistema de catalogación con potencia de web semántica, basado en dicha ontología, que facilita la explotación e intercambio consistente de información de métricas e indicadores tanto

por parte de usuarios como por parte de aplicaciones, agentes y herramientas automáticas en la Web. Dicho sistema de catalogación, fue implementado con tecnologías de web semántica, adecuadas para proporcionar búsqueda y navegación semántica, de modo de proveer “entendimiento” automático de la información.

De la investigación anterior, podemos afirmar que es necesario hacer una evaluación ontológica para asegurar la calidad de los sistemas de información convencionales o semánticos. Es importante medir las características esenciales de todo sistema basado en ontologías para lograr un artefacto semántico de calidad.

También podemos mencionar a **Gutiérrez González D., Roche Rodríguez F. y Marrero Clark M. (2012)** en su trabajo de investigación: “Aplicación informática para gestionar repositorios, establecer semejanzas y caracterizar ontologías” publicado en la Revista Cubana de Ciencias Informáticas en el año 2012, concluyen lo siguiente:

- The research objective is to develop a software application to manage ontologies repositories, characterize and establish

similarities between ontologies representing the knowledge on the web. That's why was performed an analysis a selection of appropriate technologies and tools for ontology management and application development. The application functionality were identified and documented and were made the design, implementation and testing of it. This application will manage the ontologies repositories, obtaining characteristics and similarities between representative ontologies of the knowledge on the web, making easier the processes of storage, search and retrieval of ontologies and reuse of knowledge stored in them.

Según la investigación, se puede verificar la importancia que tiene medir, caracterizar o cuantificar las ontologías o sistemas ontológicos a través de métricas estandarizadas, lo cual permitirá compartir las bases de conocimiento para ser consumidos por otros sistemas de forma transparente.

Lozano Tello A. (2002) en su tesis: “Métrica de idoneidad de Ontologías” para optar el grado de doctor en Ciencias de la Computación en la Universidad Complutense de Madrid, concluye

lo siguiente:

- En los últimos años, la creación de sistemas software basados en ontologías está aumentando considerablemente. Una ontología se define como una especificación explícita y formal de una conceptualización consensuada. Actualmente, los usuarios que buscan ontologías para incorporarlas a sus sistemas se basan únicamente en su experiencia e intuición, y esto hace difícil que puedan justificar las elecciones tomadas. Esto es debido, principalmente, a que no existe ningún método que indique al usuario qué ontologías son las más apropiadas para un nuevo sistema. Para resolver esta carencia, en este trabajo se propone el método OntoMetric, que permite a los usuarios medir la idoneidad de las ontologías existentes respecto a las necesidades de sus sistemas. Sus principales aportaciones son:
 - **Identificación de un marco multinivel de características.**
El marco consiste en un conjunto de 160 características que describen el dominio de las ontologías. Se clasifican en las siguientes dimensiones: el contenido representado en la ontología, el lenguaje en el que se encuentra implementada,

la metodología que se ha seguido para su construcción, los entornos software de desarrollo de ontologías, y los costes de usar las ontologías en el sistema. Este marco (a) proporciona el esquema para representar la información de ontologías existentes, (b) permite comparar las ontologías y (c) elegir las más apropiadas según las necesidades del nuevo sistema.

- **Construcción del modelo conceptual de una ontología en el dominio de las ontologías, Reference Ontology (RO), que está basada en el marco multinivel de características.**

El modelo conceptual de la RO se ha desarrollado siguiendo la metodología METHONTOLOGY y el entorno de desarrollo WebODE. La dimensión “contenido” de esta ontología se ha instanciado con información de ontologías procedentes de servidores de ontologías y disponibles en la web. Las demás dimensiones se han instanciado con información de publicaciones existentes y también analizando directamente los lenguajes, metodologías y entornos de desarrollo de ontologías.

- **Elaboración de un método que mide la idoneidad de un conjunto de ontologías candidatas que pueden ser incorporadas en un nuevo sistema.** El método utiliza: (a) el marco multinivel de características, (b) el modelo conceptual de la RO y sus instancias, y (c) una adaptación del método de las jerarquías analíticas. Para cada una de las ontologías candidatas, el método ofrece una medida cuantitativa de su idoneidad y sirve para decidir, de forma justificada, qué ontologías son las más adecuadas para el sistema que el usuario va a desarrollar.
- **Construcción del soporte tecnológico que asiste al método.** Las aplicaciones software construidas son: a) OntoWrappers: extraen la información de ontologías existentes en servidores de ontologías y páginas web; b) Reference Ontology Instances Selector: selecciona las ontologías que cumplen ciertos requisitos indicados por los usuarios; y c) OntoMetric Tool: calcula, usando una adaptación del método de las jerarquías analíticas, los valores de idoneidad para las ontologías candidatas.

La investigación del Dr. Tello confirma la necesidad de medir y

evaluar las ontologías para lograr un sistema semántico óptimo que tome en consideración las mejores prácticas del campo de la ingeniería ontológica para alcanzar sistemas robustos, estables, altamente escalables e interoperables.

1.1.2 Problemática de la Investigación

El Consejo de Investigación cuenta con un sin número de proyectos que son el producto intelectual de los investigadores Basadrinos. A través de un motor de búsqueda semántica se pueden realizar consultas a todos los trabajos de investigación. Sin embargo, no se tienen razones verificables para calificarlo como un sistema ontológico de calidad, según las características esenciales que promueve la World Wide Web Consortium.

Esta falta de descripción puede generar pérdida de información y bajo rendimiento si es que no se verifica la utilización de los componentes mínimos (Sparql, RDF y OWL) que según el ente mundial W3C son básicas e indispensables.

La principal causa de un sistema ontológico mal diseñado es no considerar los estándares y métricas internacionales que promueven el desarrollo óptimo de estos sistemas.

1.2 Formulación del problema

a. Problema General:

¿Cuáles son las características Ontológicas del motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014?

b. Problemas Específicos:

¿Se utilizan Datos Enlazados (RDF) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014?

¿Se utilizan Vocabularios (OWL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014?

¿Se utilizan consultas (SPARQL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014?

1.3 Justificación

La presente investigación se justifica porque se aplica el método científico para describir un artefacto semántico del cual hasta el momento no tiene un estudio similar. Además es importante porque permite verificar la calidad de los sistemas internos en la Universidad Nacional Jorge Basadre

Grohmann, aplicando el método científico para describir las cualidades de su software.

Además, se justifica porque es totalmente viable, ya que el Consejo de Investigación conocerá con certeza las características ontológicas de su motor de búsqueda semántica, el cual permitirá tener un diagnóstico preliminar sobre sus fortalezas y debilidades, de tal forma que se implementen mejoras en el futuro para beneficio de la población universitaria.

1.4 Alcances y limitaciones

La presente investigación plantea como ente de investigación al motor de búsqueda semántica de COIN, el cual representa el objeto de estudio. Cabe señalar que los permisos para acceder a la estructura del sistema serán solicitados a COIN por lo que representaría una limitación si la respuesta es negativa.

1.5 Objetivos

1.5.1 Objetivo General

Determinar las características ontológicas del buscador semántico SERON del Consejo de Investigación – UNJBG, 2014.

1.5.2 Objetivos Específicos

- Verificar el uso de Datos Enlazados (RDF) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.
- Verificar el uso de Vocabularios (OWL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.
- Verificar el uso de consultas (SPARQL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.

1.6 Hipótesis

1.6.1 Hipótesis General

Si las Ontologías permiten construir artefactos semánticos

Entonces, las características ontológicas son fundamentales para los motores de búsqueda semántica.

1.6.2 Hipótesis Específicas

Si las características ontológicas son fundamentales para los motores de búsqueda semántica.

Entonces, el uso de datos enlazados (RDF) tiene una importancia alta.

Si las características ontológicas son fundamentales para los motores de búsqueda semántica.

Entonces, el uso de vocabularios (OWL) tiene una importancia alta.

Si las características ontológicas son fundamentales para los motores de búsqueda semántica.

Entonces, el uso de consultas (Sparql) tiene una importancia alta.

1.7 Variables

1.7.1 Identificación de Variables

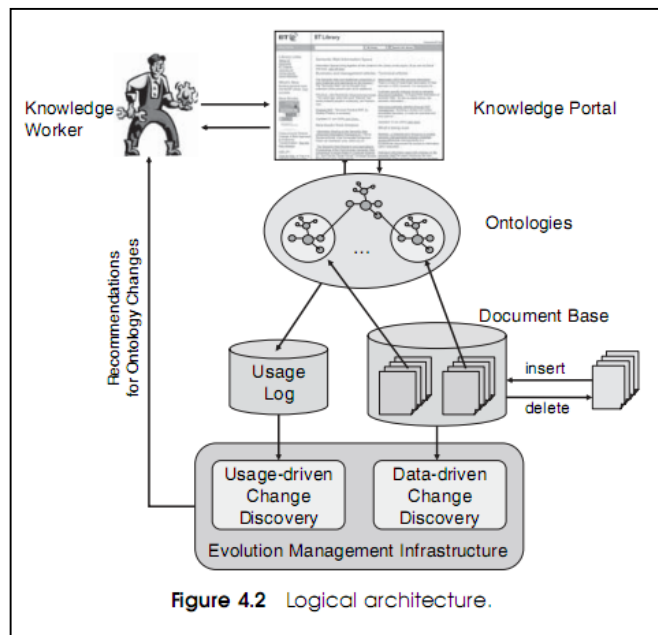


Figura 1. Variables de estudio

Fuente: (Davies, Studer, & Warren, 2006)

a) Variable

Características Ontológicas.

1.7.2 Definición de Variables

Características Ontológicas: Conjunto de propiedades que permiten optimizar el diseño y funcionamiento de los sistemas ontológicos permitiendo un óptimo desempeño de los motores de búsqueda semántica. (W3C, 2014)

1.7.3 Operacionalización de Variables

Indicadores: Características Ontológicas (Consortium, 2014)

- Datos Enlazados (RDF)
- Vocabularios (OWL)
- Consultas (SPARQL)

La medición de los indicadores se describe en la siguiente tabla:

Tabla 1. Medición de Indicadores

Indicador	Si	No
Datos Enlazados		
Vocabularios		
Consultas		
Inferencias		

Fuente: Elaboración propia

1.7.4 Clasificación de las Variables

Variable: Características Ontológicas

Por su nivel de medición: **Nominal**

Por su naturaleza: **Discreta**

Por el método de Estudio: **Cualitativo**

Por los valores que adquieren: **Dicotómica**

1.8 Diseño de la investigación

1.8.1 Diseño Experimental o no Experimental

No experimental (Descriptivo Simple).

M → O.

1.8.2 Población y Muestra

Para el presente proyecto se utilizará como objeto de estudio al motor de búsqueda semántica del Consejo de Investigación – UNJBG, 2014.

1.8.3 Técnicas e Instrumentos para recolección de datos

- La observación – guías de observación

1.8.4 Análisis de Datos

El análisis de datos y estructura para determinar las métricas se realizará con el “Jambalaya”, módulo presente en el software Protége (diseñador de ontologías).

1.8.5 Selección de Pruebas Estadísticas

Modus Ponendo Ponens

II. MARCO TEÓRICO

2.1 Marco referencial

2.1.1 Conceptos Previos

Un **sistema** puede ser definido como un complejo de elementos interactuantes. (Bertalanffy, 1968)

También sistema es un conjunto de elementos organizados que se encuentran en interacción, que buscan alguna meta o metas comunes, operando para ello sobre datos o información sobre energía o materia u organismos en una referencia temporal para producir como salida de información o energía o materia u organismos.(Murdick, 1988)

La **Búsqueda Semántica de los Proyectos de COIN**, es un mecanismo a nivel informático para compartir e informar de los distintos trabajos de investigación a la comunidad universitaria. Específicamente, la Exploración Semántica se refiere al mapeo de bases de conocimiento sobre un dominio específico, en este caso el

dominio de investigación son los proyectos de COIN, que por muchos años tuvieron un pobrísimo acceso.

2.2 Bases Teóricas respecto al problema

2.2.1 Ontologías

Concepto

Una ontología es una descripción formal de los conceptos y las relaciones entre dichos conceptos (Gruber, 1993).

El sentido filosófico del término ontología hace referencia a la esencia misma del ser, a su existencia [onto: ser]. Para los sistemas de Inteligencia Artificial, lo que "existe" es lo que puede representarse (Gruber, 1993)

Entre las definiciones más conocidas y utilizadas destaca la aportada por Thomas R. Gruber, afirma que una ontología constituye una especificación explícita y formal de abstracciones mentales, las cuales se conforman mediante un acuerdo de la comunidad experta en un dominio y en un diseño para un propósito específico (Gruber, 1993). Esta definición conlleva implícito multitud de aspectos que contribuyen a clarificar dicho concepto. Por ello, al

analizarlo de forma pormenorizada, se extrae que una ontología proporciona la estructura para un vocabulario controlado de manera similar a un diccionario o a un tesoro. Dicho vocabulario, consensuado por la comunidad, se expresa a través de conceptos, conformando así las abstracciones mentales del susodicho dominio (Bishr, 1998; Harvey et al., 1999; Sheth, 1999).

Por otro lado, en cuanto a la especificación formal de un vocabulario, esta se podría dar de diversas formas, tales como una lista plana de palabras, un diccionario, una taxonomía, un diagrama Entidad – Relación, un modelo UML (Unified Markup Language), un esquema de XML y muchos otros posibles (Batini et al., 1992).

Hoy en día la clave del enfoque para tratar con léxicos complejos son las ontologías, por el que nos referimos a una caracterización (entendido a la vez por personas y procesado por máquinas), del significado conceptual de las unidades léxicas y de sus relaciones con otros términos.

Spyns (2002), afirma que las ontologías en el actual lenguaje de ciencias de la computación son computadoras basadas en recursos que representan un acuerdo de dominios semánticos. A diferencia de los modelos de datos, el principal valor de las ontologías es su relativa independencia de aplicaciones particulares, es decir, una ontología consiste en conocimientos relativamente genéricos que pueden ser rehusados por diferentes tipos de aplicaciones.

Las ontologías, a pesar de que contengan instancia de datos, es principalmente una estructura de metadatos. Una vista tiene su propio esquema de base de datos (definición de tablas), y esta siempre es “materializada”, ya que esta es almacenada en un servidor de base de datos. Para las ontologías, este es un hecho importante, ya que el mapeo dinámico puede ser establecido sin la necesidad de un meta - nivel materializado (ontología).

Informalmente, la ontología de un cierto dominio está sobre esta terminología (vocabulario del dominio), todos los conceptos esenciales en el dominio, su clasificación, su taxonomía, sus relaciones (incluyendo todas las restricciones y jerarquías importantes) y el dominio de axiomas. Formalmente, para quien

quiera estudiar temas en un dominio D con un lenguaje L, una ontología provee una lista de tipos de cosas asumidos para existir en D; los tipos y la ontología son representados in términos de conceptos, relaciones y predicados de L.

Ambas, formal e informalmente, la ontología es una parte extremadamente importante del conocimiento sobre algún dominio, además la ontología es la parte fundamental del conocimiento, y todos los otros conocimientos podrían confiar o referirse a ésta.

Una de las razones clave para usar ontologías en los sistemas de recuperación de información, es que ellos permiten la representación del conocimiento de fondo sobre un dominio en una máquina de formato comprensible. Los humanos usan el conocimiento de fondo comúnmente en las tareas de recuperación de información. Por ejemplo, si una persona está buscando documentos sobre Europa usará su conocimiento a fondo sobre países de Europa en la tarea de búsqueda. Encontrará un documento sobre Alemania, muy pertinente si la palabra “Europa” no ha sido mencionada. Con la ayuda de una apropiada ontología geográfica y un sistema de recuperación de información se puede hacer fácilmente inferencias.

2.2.2 Ontología en la Informática

Hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados.

Facilita la comunicación y el compartimiento de información entre diferentes sistemas y entidades.

Se pueden utilizar para una variedad de propósitos: razonamiento inductivo, clasificación, técnicas de resolución de problemas, etc.

2.2.3 Componentes de una Ontología

Una ontología consta de un conjunto no vacío de conceptos identificados como entidades relevantes en el dominio a modelar, un conjunto de atributos que describen los conceptos y que pueden ser propios o heredados en una especialización, un conjunto de relaciones entre dichos conceptos, un conjunto de funciones (que son un caso especial de relaciones), un conjunto de axiomas que vinculan elementos de la ontología en condiciones que deben cumplirse

siempre y un conjunto de instancias. A continuación se describirá cada uno de estos elementos:

- Conceptos.- un concepto puede ser cualquier cosa acerca del cual se pueda aseverar algo, y por lo tanto puede ser un objeto físico (tangible), u objetos intangibles como por ejemplo la descripción de una tarea, función, acción, estrategia, entre otros. Cada concepto tiene un término asociado como nombre y un conjunto de atributos que lo identifican.
- Relaciones.- las relaciones representan el tipo de interacción entre los conceptos de un dominio, son formalmente definidas como subconjuntos del producto cartesiano de “n” conjuntos, esto es $R \subset C_1 \times C_2 \times C_3 \times \dots \times C_n$.
- Funciones.- son un caso especial de relaciones donde el enésimo elemento de la relación es único para los n-1 anteriores.
- Axiomas.- los axiomas se usan para modelar verdades que se cumplen siempre en la realidad modelada. Los axiomas definidos en una ontología pueden ser estructurales o no estructurales.

Un axioma estructural establece condiciones relacionadas a las jerarquías de la ontología, conceptos y atributos definidos. Un ejemplo de axioma estructural puede ser: “Una métrica directa no puede ser a la vez métrica indirecta”. Es decir:

$$MD \cap MI = \Phi.$$

- Instancias.- se usan para representar elementos específicos del dominio de la ontología

2.2.4 Ingeniería Ontológica

El campo de la Ingeniería Ontológica ha sido tema considerable para el estudio y la investigación durante la última década. La Ingeniería Ontológica se refiere al conjunto de actividades que concierne el proceso de desarrollo de ontologías, el ciclo de vida de la ontología, los principios, métodos y metodologías para construir ontologías y la suite de herramientas y lenguajes que nos permiten trabajar con las ontologías. La noción de Ingeniería Ontológica en red ha entrado en juego por la necesidad de la web semántica, donde una de las más relevantes hipótesis es que las ontologías son distribuidas a través de diferentes servidores web y repositorios de

ontologías y pueden tener superposición de representaciones de dominios iguales o diferentes.

En relación con los métodos y metodologías, varias propuestas han sido reportadas para el desarrollo de ontologías. En 1990, Lenat y Guha publicaron “the general steps and some interesting points about the Cyc development”. Unos años más tarde, en 1995, se desarrolla una ontología empresarial denominada TOVE (Toronto Virtual Enterprise). En la 12th Conferencia Europea de Inteligencia Artificial (ECAI’ 96), Bernaras y colegas, presentan un método usado para construir una ontología en el dominio de redes eléctricas como parte del proyecto “ESPRIT KAPTUS”. La metodología METHONTOLOGY apareció al mismo tiempo y fue extendida en los últimos papers. Ésta fue propuesta para la construcción de ontologías por la “Foundation for Intelligent Physical Agents (FIPA)”, el cual promueve la interoperabilidad a través de aplicaciones basadas en agentes.

2.2.5 Principios para el Diseño de Ontologías

Se han identificado cinco principios principales:

- Claridad.- se define en los siguientes términos: “Una ontología debería comunicar efectivamente el significado pretendido de los términos definidos. Las definiciones deberían ser objetivas. Las definiciones pueden señalarse sobre axiomas formales y una definición completa (definido por condiciones necesarias y suficientes). Todas las definiciones deberían ser documentadas con lenguaje natural”.
- Mínima Codificación de Prejuicios.- el cual significa que: “La conceptualización debería ser especificado sobre el nivel de conocimiento sin depender de una codificación particular de nivel simbólico”.
- Extensión.- se dice que: “Uno debe poder definir nuevos términos para usos especiales basado en el vocabulario existente, en una forma que no requiere la revisión de las definiciones existentes”.
- Coherencia.- se define como sigue: “Una ontología debe ser coherente: esto es, debería sancionar inferencias que son consistentes con las definiciones. Si una oración que se puede inferir de los axiomas contradictorios, una definición o una

ejemplo dado informalmente, entonces la ontología es incoherente”.

- **Mínimo Compromiso Ontológico.**- se describe en los siguientes términos: “Desde que el compromiso ontológico está basado sobre el uso consistente del vocabulario, el compromiso ontológico puede ser minimizado especificando las teorías más débiles y definiendo sólo esos términos que son esenciales para la comunicación del conocimiento consistente con la teoría”.

De acuerdo a este último principio, no deberíamos comprometernos a un formato específico de fechas, monedas, etc., cuando diseñamos nuestras ontologías, ya que tales detalles podrían ser diferentes en diferentes sistemas.

2.2.6 El Proceso de Desarrollo y Ciclo de Vida Ontológico

En 1997, el proceso de desarrollo de ontologías fue identificado en el framework de la Metodología METHONTOLOGY para la construcción de ontologías. Tal propuesta fue basada en el estándar IEEE para el desarrollo de software. El proceso de desarrollo de ontologías se refiere a las

actividades que deben ser terminadas cuando se construyen ontologías. Éstas pueden ser divididas en tres categorías:

- Administración de Actividades
- Desarrollo Orientado a Actividades
- Soporte de Actividades

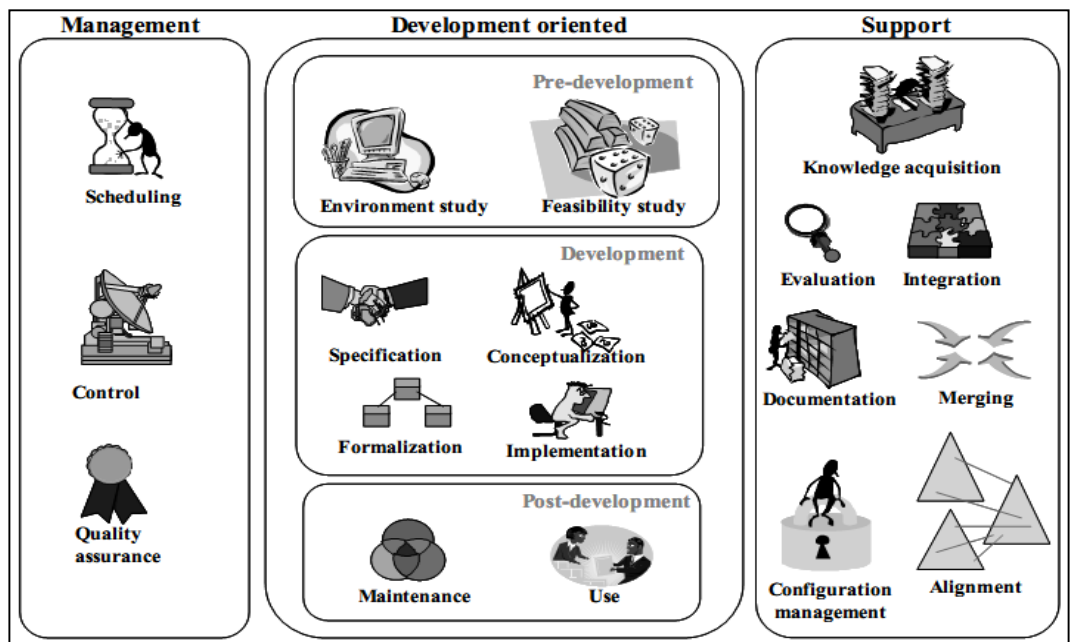


Figura 2. Proceso de Desarrollo de Ontologías

Fuente: Coral Calero, Ruiz Francisco, Piattini Mario. "Ontologies for Software Engineering and Software Technology", Berlin, 2006.

2.2.7 Datos Enlazados (RDF)

El Resource Description Framework (RDF) es el primer lenguaje desarrollado especialmente para la Web Semántica. RDF fue desarrollado como un lenguaje para añadir legitimidad de metadatos para datos existentes en la web. RDF usa XML para su serialización.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://example.org/#"
  xml:base="http://example.org/" >
  <rdf:Description about="#john" >
    <ex:hasName rdf:parseType="Resource" >
      <ex:firstName>John</ex:firstName>
      <ex:lastName>Smith</ex:lastName>
    </ex:hasName>
  </rdf:Description>
</rdf:RDF>
```

Figura 3. Serialización de RDF a XML

Fuente: Dieter Fensel, Holger Lausen, Axel Polleres Jos de Bruijin, Michael Stollberg, Dumitru Roman Jhon Domingue. "Enabling Semantic Web Services – The Web Service Modeling Ontology", Berlin, 2007. Página 32.

2.2.8 RDF Scheme (RDFS)

RDF Scheme (RDFS) es un ligero lenguaje para ontologías que permite definir vocabularios para RDF. A diferencia del esquema XML, el cual prescribe el orden y combinación de etiquetas o tags (la estructura) en un documento XML, el esquema RDF sólo provee

información sobre la interpretación de las sentencias dadas en un modelo de datos RDF. El esquema RDF no explica nada sobre la apariencia sintáctica de la descripción RDF. RDFS puede en realidad ser visto como una extensión de RDF con un vocabulario para definir clases, jerarquía de clases, propiedades (relaciones binarias), jerarquía de propiedades y restricciones de las propiedades. Las clases y propiedades de RDFS pueden ser instanciadas en un RDF.

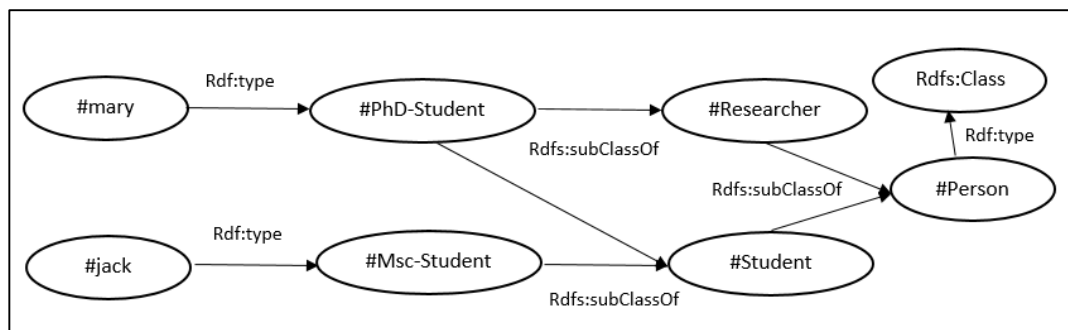


Figura 4. RDFS de personas y grupos de trabajo

Fuente: Dieter Fensel, Holger Lausen, Axel Polleres Jos de Bruijin, Michael Stollberg, Dumitru Roman Jhon Domingue. “Enabling Semantic Web Services – The Web Service Modeling Ontology”, Berlin, 2007. Página 32.

2.2.9 Web Ontology Language (OWL)

El lenguaje para la web ontológica (OWL), es un lenguaje expresivo para ontologías que extiende el RDFS.

OWL tiene sus raíces en sus propios lenguajes predecesores (ejemplo: XML y RDF), y en una representación tradicional del

conocimiento formal que tienen históricamente incertidumbre no considerada.

OWL en sí consiste de tres tipos diferenciados por el incremento de expresividad en cada uno. Entre ellos tenemos:

- OWL Lite.- es el menos expresivo de los OWL. Comparado con RDFS, éste agrega un rango de restricciones locales, restricciones existenciales, restricciones de cardinalidad simple, igualdades y varios tipos de propiedades (inversas, transitivas y simétricas).
- OWL DL.- comparado con OWL Lite, OWL DL agrega todo el soporte para: negación, disyunción, restricciones cardinales, enumeraciones y restricciones de valor. El elemento “DL” proviene del parecido con un lenguaje expresivo para la descripción lógica.
- OWL Full.- considerando que OWL Lite y OWL DL imponen restricciones en el uso del vocabulario y las sentencias RDF, OWL Full no tiene tales restricciones. Por tanto, OWL FULL

permite simultáneamente la especificación de classes-as-
instances y el uso de un lenguaje construido en el mismo
lenguaje, lo cual modifica el lenguaje.

2.2.10 La Web Semántica

El principal problema de la web actual, en esta etapa, es que el significado del contenido no es accesible a las máquinas, la información retornada y las herramientas de procesamiento de texto son ampliamente usadas, pero estos aún presentan dificultades cuando interpretan oraciones o extraen información útil para los usuarios. El desarrollo de la Web Semántica con máquinas capaces de entender el contenido, tiene el potencial para revolucionar la actual web.

“La Web semántica no constituye una web independiente, sino una ampliación de la actual; en la cual la información está dotada de significados bien definidos, con el fin de permitir un mejor trabajo en la cooperación entre humanos y computadoras” (Berners Lee, 1998).

La web semántica y las tecnologías para la web semántica nos ofrecen un nuevo enfoque para administrar y procesar la información, principio fundamental del cual es la creación y uso de los metadatos semánticos.

El problema que la web semántica ha de resolver, consiste en proporcionar un lenguaje capaz de dar expresión tanto a datos como a las reglas para razonar sobre datos, y que permita, además, la exportación a la web de las reglas de inferencia de cualesquiera sistemas de representación de conocimientos que ya existan.

Ya están disponibles tres importantes tecnologías para el desarrollo de la web semántica: el lenguaje XML (eXtensible Markup Language, Lenguaje de Marcado Extensible), el lenguaje RDF (Resource Description Framework, Marco de Descripción de Recursos) y el lenguaje OWL (Ontology Web Language, Lenguaje de Ontologías en la Web).

2.2.11 Herramientas para Desarrollar Ontologías

Al igual que sucedía con las definiciones y las tipologías, también son varias las herramientas disponibles para la definición y gestión de ontologías. Desde las iniciales y más rudimentarias (Ontolingua Server, Ontosaurus) a las más recientes y avanzadas (Protégé, WebODE, y Ontoedit). Gómez (2002) plantea la siguiente tipología:

- Herramientas de desarrollo de ontologías: este grupo incluye las herramientas que sirven para la construcción de nuevas ontologías o bien para la reutilización de las existentes. Destacan entre sus funcionalidades la edición y la consulta, así como la exportación e importación de ontologías, la visualización en diversos formatos gráficos, etc.
- Herramientas de fusión y de integración de ontologías: pretenden solucionar el problema de la combinación e integración de diversas ontologías del mismo dominio, lo que ocurre cuando se unen dos organizaciones diferenciadas, o cuando se pretende obtener una ontología de calidad, a partir de las ya existentes.

- Herramientas de evaluación de ontologías: aparecen como instrumentos de apoyo que deben asegurar que tanto las ontologías como las tecnologías relacionadas tengan un nivel mínimo de calidad. Para el futuro, este esfuerzo puede conducir a las certificaciones estandarizadas.
- Herramientas basadas en la anotación: estas herramientas se han diseñado para permitir a usuarios que inserten informaciones y datos. La mayoría de estas herramientas han aparecido recientemente, junto con la idea de Web Semántica.
- Herramientas de almacenaje y de preguntas: son instrumentos que se han creado para permitir usar fácilmente las ontologías. La clave está en el intento de que la web se convierta en una auténtica plataforma para transmitir conocimiento.
- Herramientas de aprendizaje: se utilizan semi-automáticamente para construir ontologías a partir del lenguaje natural.

Junto a las herramientas o editores de ontologías hay que destacar la existencia de diferentes lenguajes que permiten la

creación y/o el desarrollo de las mismas. La mayoría están contruidos por sintaxis de XML o RDF(S) como OIL (Ontology Inference Layer), DAML (DARPA Agent Mark-up Language) + OIL y OWL (Web Ontology Language). En cualquier caso, estos lenguajes de ontologías están aún en desarrollo, aunque gracias al trabajo y apoyo del World Wide Web Consortium (W3C) parece que OWL se consolida adquiriendo una progresión clara y evidente, por lo que lleva camino de convertirse en estándar de facto.

2.2.12 Protége

Protege es una herramienta libre basado en Java. Es un editor de ontologías con un framework basado en el conocimiento.

La Plataforma Protege soporta dos formas principales de modelamiento de ontologías vía el editor Protege – Frames y Protege OWL. Las ontologías de Protége pueden ser exportadas a una variedad de formatos incluyendo RDF(S), OWL y XML Scheme.

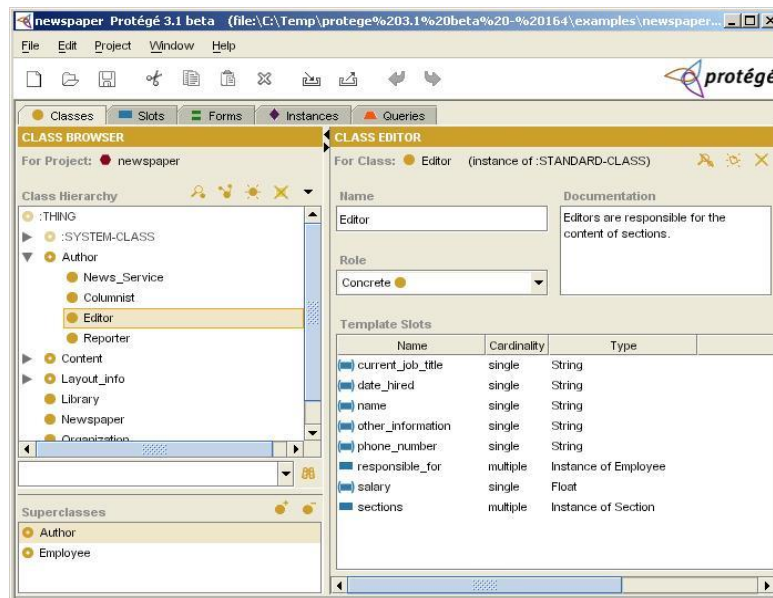


Figura 5. Protege

Fuente: “<http://protege.stanford.edu/overview/pf-screenshots.html>”

2.2.13 SparQL

Lenguaje de consulta para RDF, está diseñado para trabajar con casos de uso y requerimientos identificados por el RDF Data Access Working Group.

SPARQL es un acrónimo recursivo del inglés SPARQL Protocol and RDF Query Language. Se trata de una recomendación para crear un lenguaje de consulta dentro de la Web semántica que, aunque en 2005 está en proceso de estandarización por el DAWG del W3C, está ya implementada en muchos lenguajes y bases de datos.

El 15 de enero de 2008 el W3C publicó como recomendación la especificación de SPARQL.

El lenguaje de recuperación SPARQL aparece especificado en los siguientes documentos:

- El Protocolo SPARQL para RDF [SPROT] especificación que define los protocolos de acceso remoto mediante el uso de consultas SPARQL y los datos que devuelve.
- Los Resultados en formato XML de la Consulta SPARQL [RESULTS] especificación que define el formato del documento XML para representar los resultados de las consultas SELECT y ASK de SPARQL.

2.2.14 Base de Datos

Una base de datos es un conjunto de datos relacionados entre sí. Por datos entendemos hechos conocidos que pueden registrarse y que tienen un significado implícito. Por ejemplo, consideremos los nombres, números telefónicos y direcciones de personas que conocemos. Tal vez hayamos registrado estos datos en una libreta de direcciones indizada, o quizá lo hayamos hecho en un disquete,

empleando un computador personal y software del tipo de DBASE IV o V, PARADOX o EXCEL. Se trata de un conjunto de datos relacionados entre sí y que tienen un significado implícito; por tanto, constituyen una base de datos.

Las bases de datos y su tecnología están teniendo un impacto decisivo sobre el creciente uso de las computadoras.

Una base de datos o banco de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. En este sentido, una biblioteca puede considerarse una base de datos compuesta en su mayoría por documentos y textos impresos en papel e indexados para su consulta. En la actualidad, y debido al desarrollo tecnológico de campos como la informática y la electrónica, la mayoría de las bases de datos están en formato digital (electrónico), que ofrece un amplio rango de soluciones al problema de almacenar datos.

Existen programas denominados sistemas gestores de bases de datos, abreviados SGBD, que permiten almacenar y posteriormente acceder a los datos de forma rápida y estructurada. Las propiedades

de estos SGBD, así como su utilización y administración, se estudian dentro del ámbito de la informática.

Las aplicaciones más usuales son para la gestión de empresas e instituciones públicas. También son ampliamente utilizadas en entornos científicos con el objeto de almacenar la información experimental.

Aunque las bases de datos pueden contener muchos tipos de datos, algunos de ellos se encuentran protegidos por las leyes de varios países. Por ejemplo, en España los datos personales se encuentran protegidos por la Ley Orgánica de Protección de Datos de Carácter Personal (LOPD).

Modelos de Bases de datos

Además de la clasificación por la función de las bases de datos, éstas también se pueden clasificar de acuerdo a su modelo de administración de datos.

Un modelo de datos es básicamente una "descripción" de algo conocido como contenedor de datos (algo en donde se guarda la

información), así como de los métodos para almacenar y recuperar información de esos contenedores. Los modelos de datos no son cosas físicas: son abstracciones que permiten la implementación de un sistema eficiente de base de datos; por lo general se refieren a algoritmos, y conceptos matemáticos.

Algunos modelos con frecuencia utilizados en las bases de datos:

- Base de Datos Jerárquicas
- Base de Datos de Red
- Base de Datos Transaccionales
- Base de Datos Relacionales
- Base de Datos Multidimensionales
- Base de Datos Orientada a Objetos
- Base de Datos Deductivas
- Base de Datos Federadas
- Base de Datos Distribuidas

Álgebra Relacional

El álgebra relacional es un conjunto de operaciones que describen paso a paso como computar una respuesta sobre las relaciones, tal y como éstas son definidas en el modelo relacional. Denominada de tipo procedimental, a diferencia del Cálculo relacional que es de tipo declarativo.

Describe el aspecto de la manipulación de datos. Estas operaciones se usan como una representación intermedia de una consulta a una base de datos y, debido a sus propiedades algebraicas, sirven para obtener una versión más optimizada y eficiente de dicha consulta.

Operaciones Básicas

- Selección (σ).- Permite seleccionar un subconjunto de tuplas de una relación (R), todas aquellas que cumplan la(s) condición(es) P, esto es:

$$\sigma_P (R)$$

Ejemplo:

$$\sigma_{\text{Apellido}=\text{Gomez}}(\text{Alumnos})$$

Selecciona todas las tuplas que contengan Gómez como apellido en la relación Alumnos.

Una condición puede ser una combinación booleana, donde se pueden usar operadores como: \wedge , \vee , combinándolos con operadores $<$, $>$, \leq , \geq , $=$, \neq

- Proyección (Π)

Permite extraer columnas (atributos) de una relación, dando como resultado un *subconjunto vertical* de atributos de la relación, esto es:

$$\Pi_{A_1, A_2, \dots, A_n}(R)$$

Donde A_1, A_2, \dots, A_n son atributos de la relación **R**.

Ejemplo:

$$\Pi_{\text{Apellido}, \text{Semestre}, \text{NumeroControl}}(\text{Alumnos})$$

Selecciona los atributos Apellido, Semestre y NumeroControl de la relación Alumnos, mostrados como un subconjunto de la relación Alumnos.

- Producto cartesiano (\times)

El producto cartesiano de *dos relaciones* se escribe como:

$R \times S$

Y entrega una relación, cuyo *esquema* corresponde a una combinación de todas las tuplas de **R** con cada una de las tuplas de **S**, y sus atributos corresponden a los de **R** seguidos por los de **S**.

Ejemplo:

Alumnos \times Maestros

Muestra una nueva relación, cuyo esquema contiene cada una de las tuplas de la relación Alumnos junto con las tuplas de la relación Maestros, mostrando primero los atributos de la relación Alumnos seguidos por las tuplas de la relación Maestros.

- Unión (U)

La operación $R \cup S$ retorna el conjunto de tuplas que están en R, o en S, o en ambas. R y S deben ser *uniones compatibles*.

- Diferencia (-)

La diferencia de dos relaciones, R y S denotada por:

$$R - S$$

Entrega todas aquellas tuplas que están en R, pero **no** en S. R y S deben ser *uniones compatibles*.

Estas operaciones son fundamentales en el sentido en que (1) todas las demás operaciones pueden ser expresadas como una combinación de éstas y (2) ninguna de estas operaciones pueden ser omitidas sin que con ello se pierda información.

Operaciones No Básicas

Entre los operadores no básicos tenemos:

- Intersección (\cap)

La intersección de dos relaciones se puede especificar en función de otros operadores básicos:

$$R \cap S = R - (R - S)$$

La intersección, como en Teoría de conjuntos, corresponde al conjunto de todas las tuplas que están en R y en S, siendo R y S *uniones compatibles*.

- Reunión natural (\bowtie) (Natural Join)

La operación Reunión natural en el álgebra relacional es la que permite reconstruir las tablas originales previas al proceso de normalización. Consiste en combinar las proyección, selección y producto cartesiano en una sola operación, donde la condición θ es la igualdad Clave Primaria = Clave Externa (o Foránea), y la proyección elimina la columna duplicada (clave externa).

Expresada en las operaciones básicas, queda:

$$R \bowtie S = \Pi_{A_1, A_2, \dots, A_n}(\sigma_\theta(R \times S))$$

Una reunión zeta (θ -Join) de dos relaciones es equivalente a:

$$R \bowtie_\theta S = \sigma_\theta(R \times S)$$

Donde la condición θ es libre. Si la condición θ es una igualdad se denomina EquiJoin.

- División (/)

Supongamos que tenemos dos relaciones $A(x, y)$ y $B(y)$ donde el dominio de y en A y B , es el mismo. El operador división A / B retorna todos los distintos valores de x tales que para todo valor y en B existe una tupla $\langle x, y \rangle$ en A .

2.2.15 Redes

Una **red de computadoras**, también llamada **red de ordenadores** o **red informática**, es un conjunto de equipos (computadoras y/o dispositivos) conectados por medio de cables,

señales, ondas o cualquier otro método de transporte de datos, que comparten información (archivos), recursos (CD-ROM, impresoras, etc.), servicios (acceso a internet, e-mail, chat, juegos), etc.

Una red de comunicaciones es un conjunto de medios técnicos que permiten la comunicación a distancia entre equipos autónomos (no jerárquica -master/slave-). Normalmente se trata de transmitir datos, audio y vídeo por ondas electromagnéticas a través de diversos medios (aire, vacío, cable de cobre, cable de fibra óptica, etc.).

Para simplificar la comunicación entre programas (aplicaciones) de distintos equipos, se definió el Modelo OSI por la ISO, el cual especifica 7 distintas capas de abstracción. Con ello, cada capa desarrolla una función específica con un alcance definido.

2.2.16 Teoría de la Información

Introducción

El concepto de información supone dos objetos:

- La fuente de información (Tx).
- El destinatario de la información (RX)

Los tipos de problemas que surgen en el estudio de la información son de alguno de estos tres tipos:

- **Técnicos:** Relativos a la exactitud, confiabilidad, velocidad de transmisión, etc.
- **Semánticos:** Están dirigidos a las investigaciones que pretenden resolver el problema de cómo es posible transmitir exactamente el significado del mensaje con ayuda de textos.
- **Pragmáticos:** Analizan el grado de eficiencia en que influye la información sobre el comportamiento del destinatario.

Campo de Ocupación

De la descripción matemática y evaluación de los métodos y transmisión, conservación, extracción, clasificación y medida de la información.

Teoría Matemática de las Comunicaciones

Se planteó el siguiente problema: Dado un conjunto de posibles mensajes que una fuente puede producir, no de nuestra elección, ¿Cómo podemos representar de la mejor forma el mensaje para llevar la información a través de un sistema dado con sus limitaciones físicas inherentes?. La consecución de respuestas a esta pregunta le llevo a consolidar su teoría matemática de las comunicaciones con plena validez hoy.

El nombre de Shannon se asocia a dos teoremas que tuvieron una grandísima importancia en el desarrollo de la ciencia de la computación y en las comunicaciones digitales. El primero señala que el número de bits necesarios para describir unívocamente una fuente de información puede aproximarse al correspondiente contenido de información tanto como se desee (teorema de codificación de la fuente). El segundo teorema declara que el ratio de errores de los datos transmitidos en un canal confinado y con ruido puede reducirse a una cantidad arbitrariamente pequeña si la velocidad de transmisión es menor que la capacidad del canal (teorema de la codificación del canal). Los dos se dan como resultado del estudio de los trabajos de Hartley y Nyquist. Una

introducción a su trabajo se publicó en 1948 en un artículo de la revista Bell System Technical Journal, con un título prometedor: "Una Teoría Matemática de la Comunicación".

Conceptos Básicos

Los principales conceptos a desarrollar en la Teoría de la Información son:

- La medida de la información.
- La capacidad de un canal de comunicación para transferir la información.
- La codificación como medio de utilizar el canal a plena capacidad.

Estos tres conceptos básicos se encuentran reunidos en lo que puede llamarse *“Teorema fundamental de la teoría de la información”*.

La codificación se emplea para adaptar la fuente al canal para una transferencia de información con un máximo de confiabilidad.

Medida de la Información

Sea un mensaje “A” de probabilidad de ocurrencia “ $P(A)$ ”, la información que conlleva ese mensaje definida como auto información es:

$$I_A = f(P(A)) \dots\dots\dots (I)$$

Donde $f(P(A))$ debe ser determinada.

Para encontrar $f(P(A))$ se suponen los siguientes requerimientos:

1. $f(P(A)) \geq 0$ donde $0 \leq P(A) \leq 1$
2. $\lim_{P(A) \rightarrow 1} f(P(A)) = 0$
3. $f(P(A)) > f(P(B))$ Para $P(A) < P(B)$
4. $f(P(A) \cdot P(B)) = f(P(A)) + f(P(B)) = f(P(C))$

Es decir: $I_C = I_A + I_B$

La única función que satisface estas cuatro condiciones es:

$$f(P(A)) = -K \log_b P_A = K \log_b \frac{1}{P_A} \dots\dots\dots(2)$$

Donde K es una constante positiva y b la base.

Hartley demostró que K es 1.

Al especificar la base se selecciona la unidad de información así:

- $I_A = \log_r \frac{1}{P(A)}$ Unidades de orden r
- $I_A = \log_2 \frac{1}{P(A)}$ Shannon o bit
- $I_A = \log_{10} \frac{1}{P(A)}$ Hartley
- $I_A = \ln_r \frac{1}{P(A)}$ Natural

2.2.17 Methontology

Methontology es una metodología creada en el Laboratorio de Inteligencia Artificial de la Universidad Técnica de Madrid. La creación de la ontología puede empezar desde cero o en base a la reutilización de otras existentes. Methontology incluye la identificación del proceso de desarrollo de la ontología (calendario, control, aseguramiento de calidad, adquisición de conocimiento), un ciclo de vida basado en la evolución de prototipos, para lo cual se sigue -los pasos definidos en el estándar IEEE 1074 de desarrollo de software (Corcho, 2001) que son:

- Especificación.- Definir el alcance y granularidad de la ontología.
- Conceptualización.- Permite organizar y estructurar el conocimiento adquirido mediante tablas, lenguaje UML, jerarquías etc.
- Implementación.- Representa la formalización de la ontología; es decir pasar la conceptualización de la ontología a un lenguaje como RDF, OWL, etc.
- Evaluación.- Comprobar el funcionamiento de la ontología.

La parte principal del desarrollo de ontologías usando Methontology es la etapa del modelado conceptual que comprende distintas tareas.

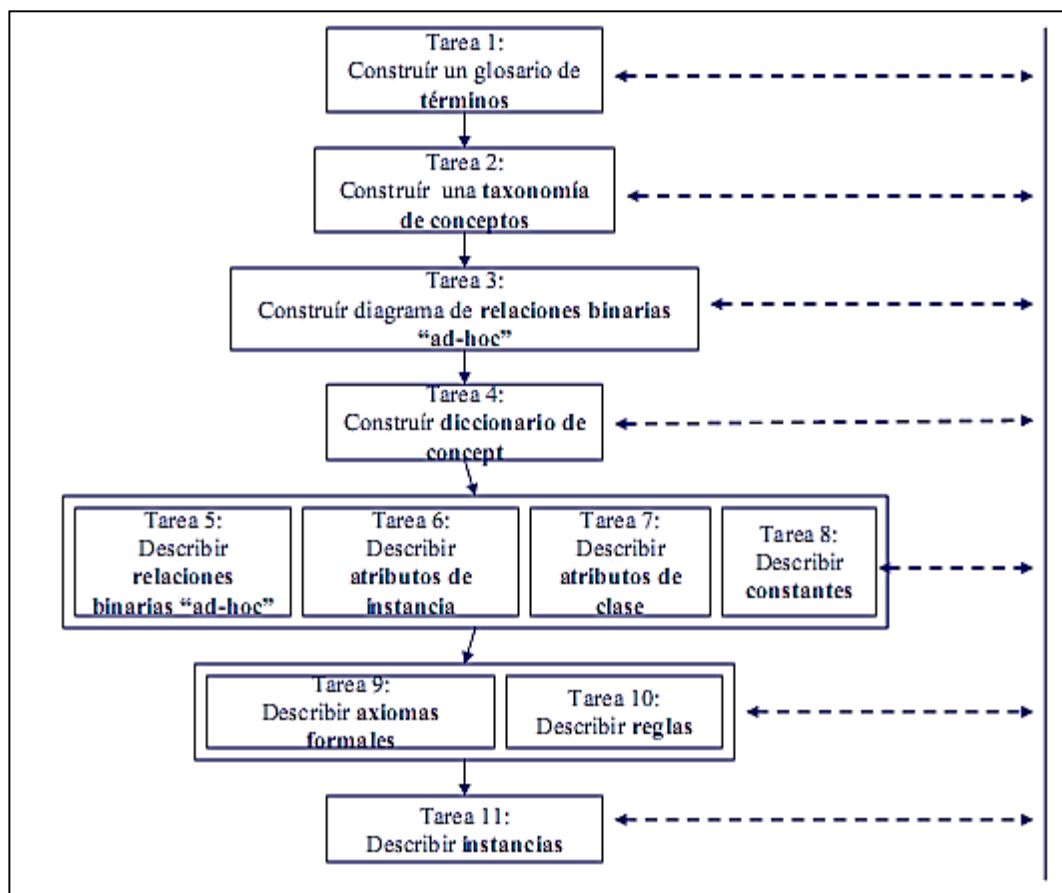


Figura 6. Modelado Conceptual
Fuente: *diseño propio.*

III. DESARROLLO

Para determinar las características internas de acuerdo a los indicadores para la presente investigación se debe empezar por explorar cada uno de los componentes internos desde el diseño de la Ontología para lo cual se usará la metodología Methontology,

3.1 Especificación

3.1.1 Meta

Determinar las características internas del buscador ontológico basado en RDF, OWL y SPARQL.

3.1.2 Alcance

Se analizará el motor de búsqueda semántica de la ontología de COIN la cual almacena todos los proyectos publicados en COIN desde el año 2003.

3.1.3 Granularidad

El nivel de detalle utilizado será máximo para representar y especificar cada uno de los diccionarios para el almacenamiento de información; así como la especificación de relaciones y consultas semánticas.

3.2 Conceptualización

Las actividades principales dentro de la búsqueda de proyectos en COIN se muestran a continuación:

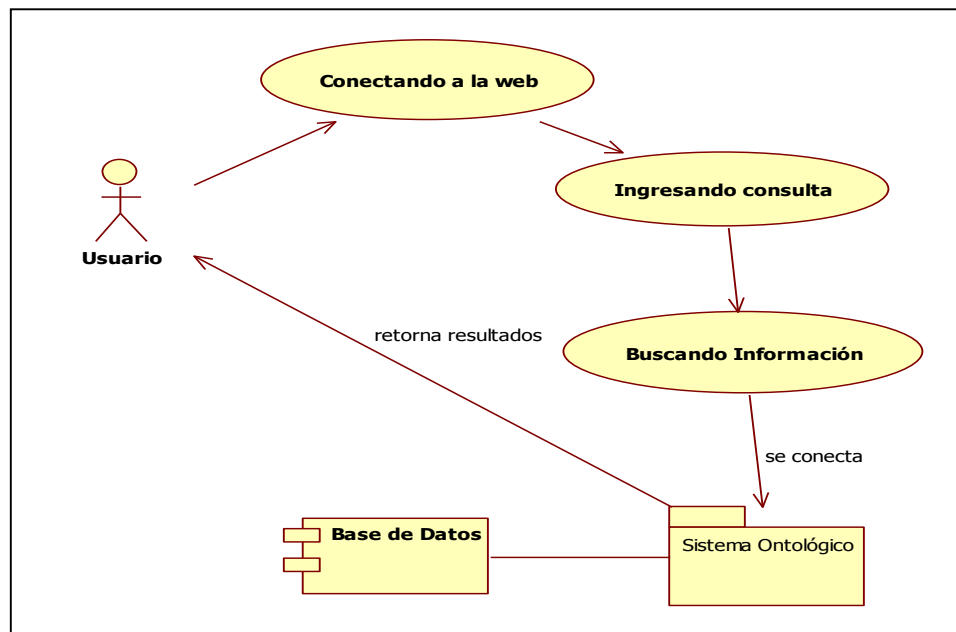


Figura 7. Proceso de Búsqueda de Información

Fuente: *Elaboración propia*

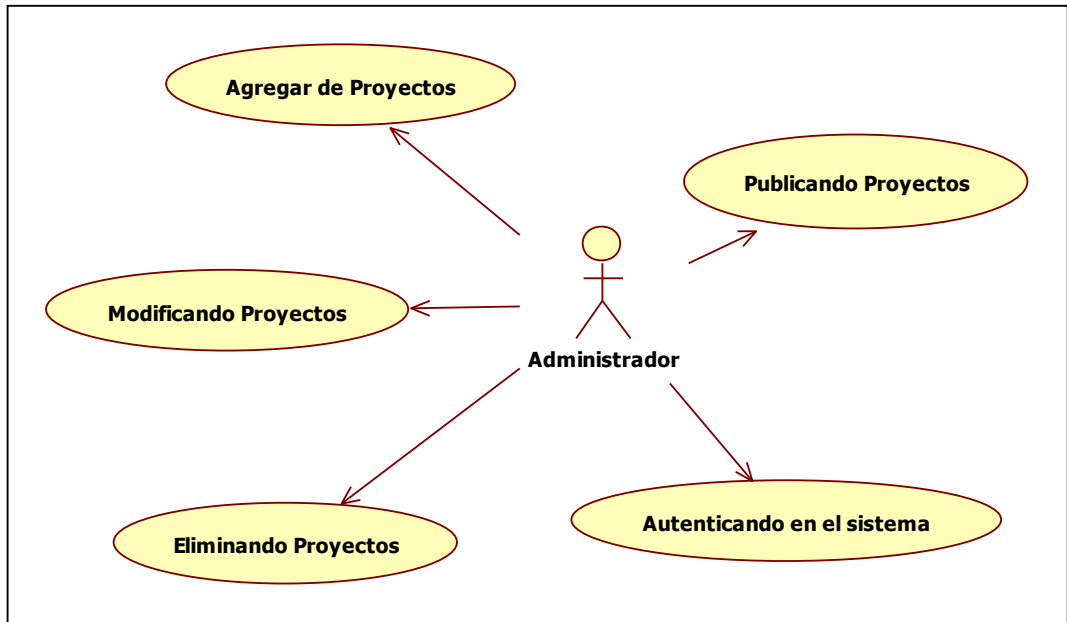


Figura 8. Funciones del Administrador
 Fuente: *Elaboración propia*

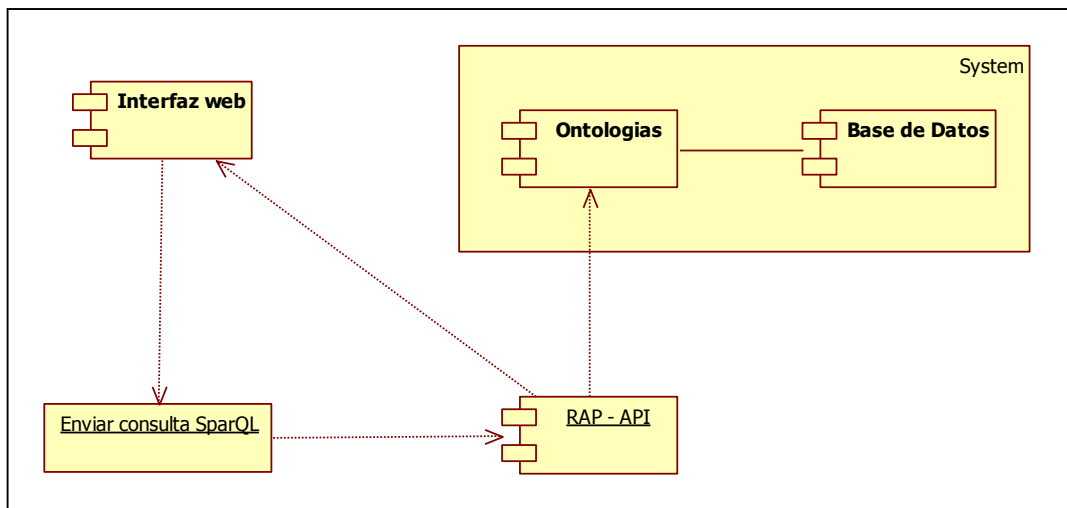


Figura 9. Proceso de Retorno de Información
 Fuente: *Elaboración propia*

3.2.1 Glosario de Términos

Aquí se detallarán los conceptos que se utilizaron en el diseño de la ontología. Específicamente, se define el vocabulario de los términos que se usaron en COIN cuando se refieren a los proyectos.

Tabla 2. Glosario de Términos

Nombre	Sinónimo	Acrónimo	Definición	Tipo
Proyecto	Plan, Planificación.	PY	Un proyecto es una planificación que consiste en un conjunto de actividades que se encuentran interrelacionadas y coordinadas.	Concepto
Investigación	Indagación, Exploración.		Búsqueda intencionada de conocimientos o de soluciones a problemas de carácter científico.	Concepto
Facultad			Infraestructura que contiene a una o más escuelas.	Concepto

Responsable	Autor, Investigador.		Persona que hace un trabajo de investigación.	Concepto
Resolución de Facultad			Documento emitido por las facultades.	Concepto
Miembros	Apoyo, Colaboradores o Ayudantes.		Personas que apoyan en la investigación a los investigadores.	Concepto
Escuela			Órgano educativo dentro de una facultad.	Concepto
Grado Académico	Grado, nivel.		Grados obtenidos luego de haber concluido la etapa de pre grado.	Concepto

Fuente: *Elaboración propia*

3.2.2 Taxonomías

El resultado de esta tarea es una o más taxonomías donde los conceptos estén clasificados.

- Para Investigadores

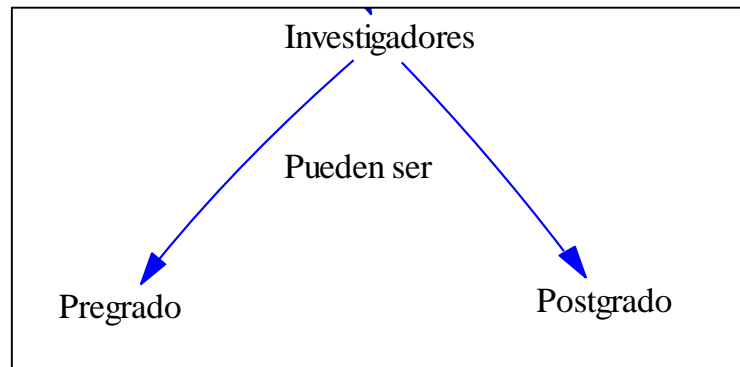


Figura 10. Taxonomía Investigadores

Fuente: *Elaboración propia*

- Para Personal

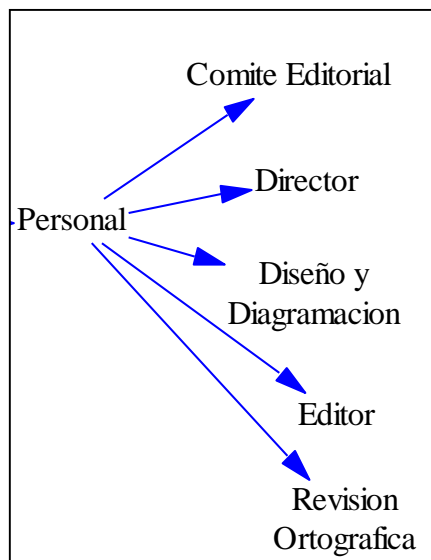


Figura 11. Taxonomía Personal

Fuente: *Elaboración propia*

- Para Proyecto

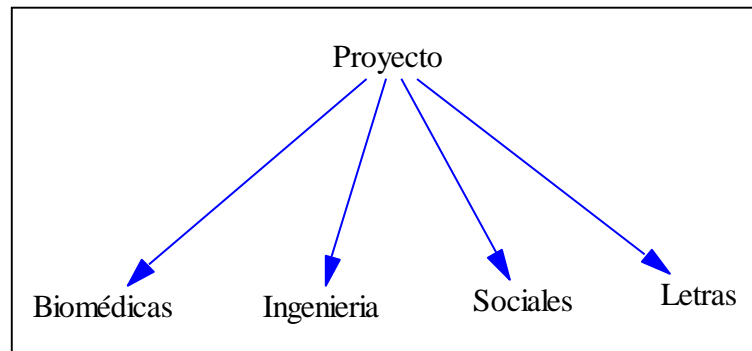


Figura 12. Taxonomía Proyecto

Fuente: *Elaboración propia*

3.2.3 Relaciones

En esta tarea, se debe definir las relaciones entre conceptos presentes en el dominio de estudio.

- Proyecto - Investigador

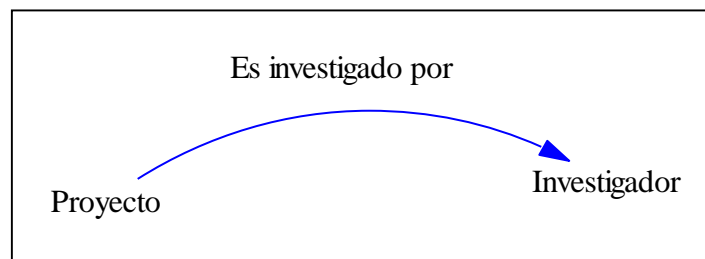


Figura 13. Relación Proyecto - Investigador

Fuente: *Elaboración propia*

- Investigador - Proyecto

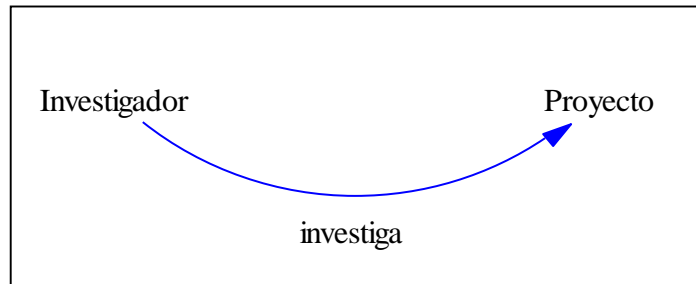


Figura 14. Relación Investigador - Proyecto

Fuente: Elaboración propia

- Proyecto - Facultad

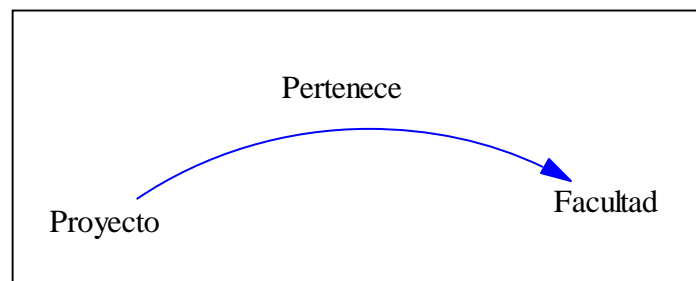


Figura 15. Relación Proyecto - Facultad

Fuente: Elaboración propia

- Revista - Proyecto

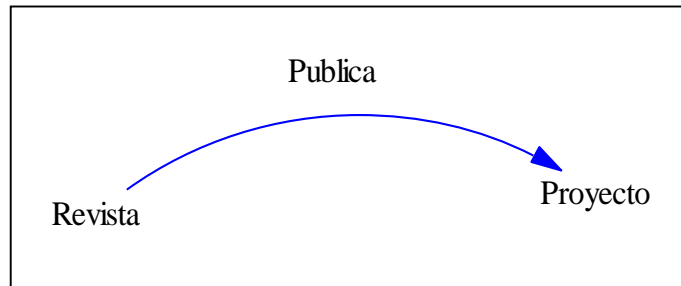


Figura 16. Relación Revista - Proyecto

Fuente: Elaboración propia

- Revista - Personal

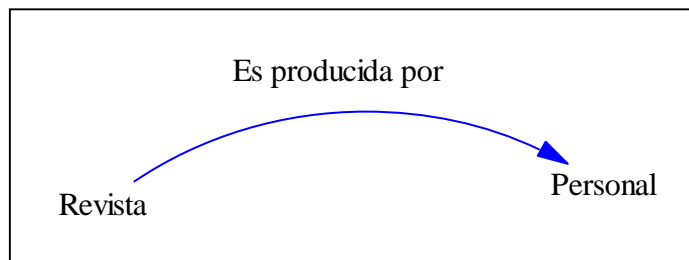


Figura 17. Relación Revista - Personal

Fuente: Elaboración propia

- Personal - Revista

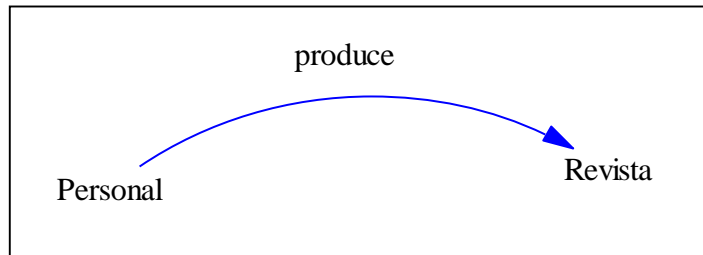


Figura 18. Relación Revista - Personal

Fuente: *Elaboración propia*

- Revista - Imprenta

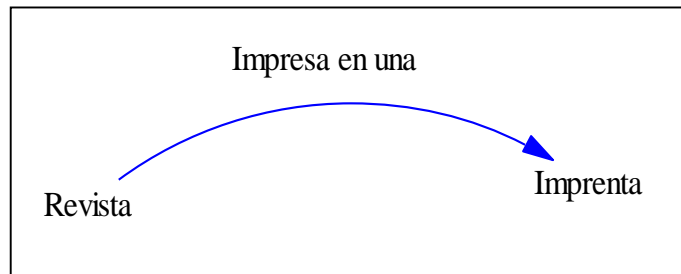


Figura 19. Relación Revista - Imprenta

Fuente: *Elaboración propia*

- Imprenta - Ubicación

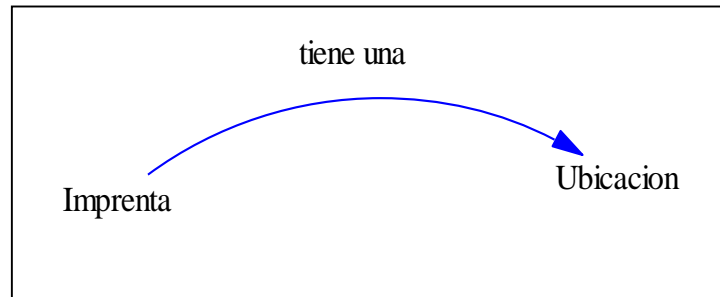


Figura 20. Relación Imprenta - Ubicación

Fuente: *Elaboración propia*

3.2.4 Diccionario de Conceptos

Ahora se debe definir las clases y subclases, así como también sus propiedades.

El árbol de clases y subclases de la ontología de acuerdo a las relaciones planteadas anteriormente se muestra a continuación:



Figura 21. Árbol de Clases y Subclases

Fuente: *Elaboración propia*

Los atributos y relaciones de cada clase, se muestran a continuación:

- Clase Imprenta

imprime	(multiple Revista)
nombre	(multiple string)
se_ubica_en	(multiple Ubicacion)

Figura 22. Propiedades de Imprenta

Fuente: *Elaboración propia*

- Clase Facultad

Escuelas	(multiple string)
nombre	(multiple string)

Figura 23. Propiedades de Facultad

Fuente: *Elaboración propia*

- Clase Investigador

Email	(multiple string)
investiga	(multiple Proyecto)
nombre	(multiple string)

Figura 24. Propiedades de Investigador

Fuente: *Elaboración propia*

- Clase Personal

■	Email	(multiple string)
■	Grado_Academico	(multiple string)
■	nombre	(multiple string)
■	produce	(multiple Revista)

Figura 25. Propiedades de Personal

Fuente: *Elaboración propia*

- Clase Proyecto

■	es_investigado_por	(multiple Investigador)
■	es_publicado_en	(multiple Revista)
■	pertenece	(multiple Facultad)
■	Titulo	(multiple string)
■	url	(multiple string)

Figura 26. Propiedades de Proyecto

Fuente: *Elaboración propia*

- Clase Ubicación

■	Ciudad	(multiple string)
■	Departamento	(multiple string)
■	direccion	(multiple string)

Figura 27. Propiedades de Ubicación

Fuente: *Elaboración propia*

- Clase Revista

es_producida	(multiple Personal)
fecha_publicacion	(multiple string)
nombre	(multiple string)
Portada	(multiple string)
publica	(multiple Proyecto)
se_imprime_en	(multiple Imprenta)

Figura 28. Propiedades de Revista

Fuente: *Elaboración propia*

3.2.5 Descripción de Instancias

En esta parte se describen cada una de las instancias en la ontología (población de datos).

- Individuo Investigador

INDIVIDUAL EDITOR for Postgrado_101 (instance of Postgrado) + - F T

For Individual: http://www.owl-ontologies.com/onto_coin.owl#Postgrado_101

Annotations

Property	Value	Lang
<input checked="" type="checkbox"/> rdfs:comment	Cientifico, Inventor, Intelectual, Analista, Researcher, Ricercatore	

Email

Value	Lang
maria@hotmail.com	

nombre

Value	Lang
Maria Dalila Salas Romero	

Grado_Academico

Value	Lang
Magister	

investiga

◆ Proyecto_102	
----------------	--

Figura 29. Individuo Investigador

Fuente: *Elaboración propia*

- Individuo Investigador

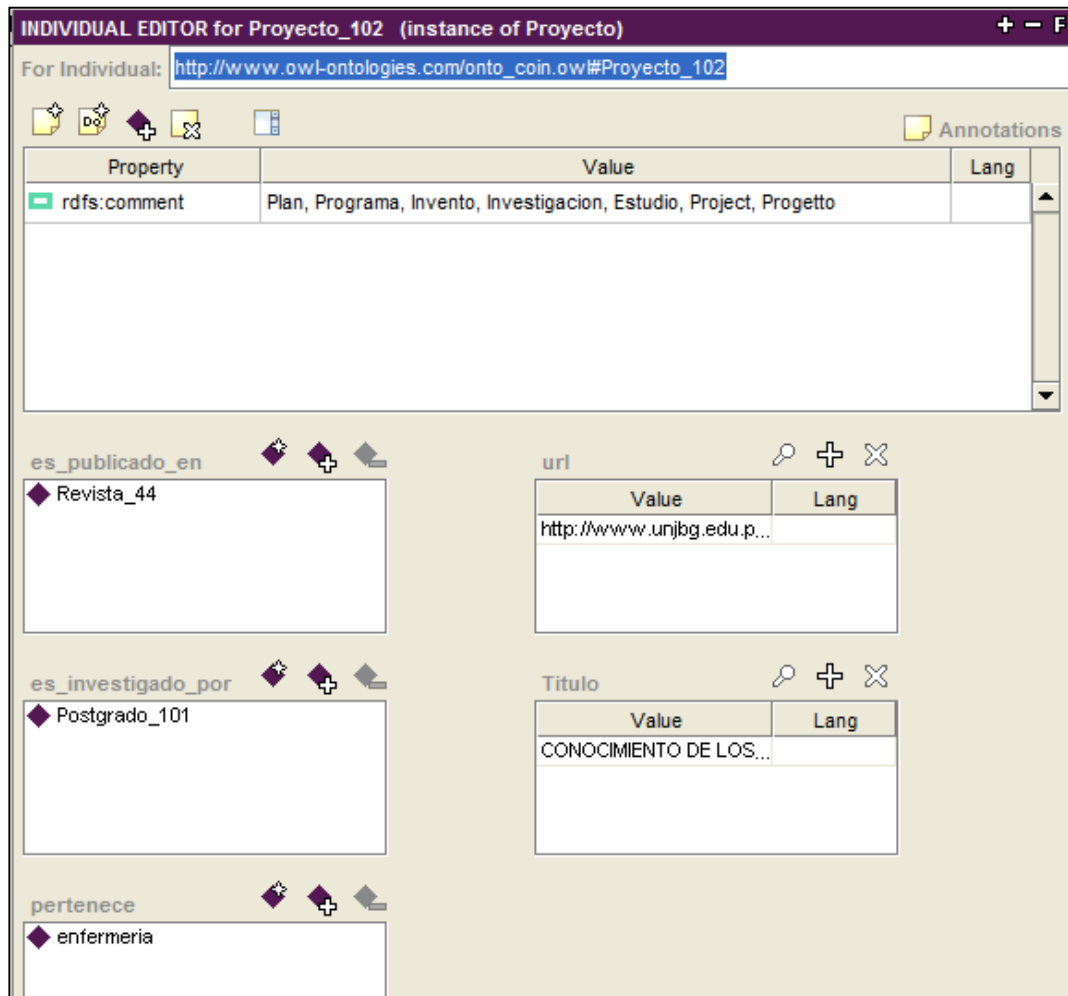


Figura 30. Individuo Proyecto

Fuente: *Elaboración propia*

- Individuo Revista

INDIVIDUAL EDITOR for Revista_44 (instance of Revista)

For Individual: http://www.owl-ontologies.com/onto_coin.owl#Revista_44

Annotations

Property	Value	Lang
rdfs:comment	Publicacion, Periodico, Semanario, Boletin, Informativo, Magazine, Periodical, Journal	

fecha_publicacion

Value	Lang
2009	

Portada

Value	Lang
Revista 11	

publica

- ◆ Proyecto_100
- ◆ Proyecto_102
- ◆ Proyecto_104
- ◆ Proyecto_106
- ◆ Proyecto_108
- ◆ Proyecto_110

nombre

Value	Lang
Ciencia y Desarrollo	

es_producida

- ◆ Comite_Editorial_67
- ◆ Director_73
- ◆ Personal_42
- ◆ Revision_Ortografica_72

se_imprime_en

- ◆ Imprenta_40

Figura 31. Individuo Revista

Fuente: *Elaboración propia*

- Individuo Ubicación

INDIVIDUAL EDITOR for Ubicacion_45 (instance of Ubicacion) + - F T

For Individual: http://www.owl-ontologies.com/onto_coin.owl#Ubicacion_45

Annotations

Property	Value	Lang
<input checked="" type="checkbox"/> rdfs:comment	Direccion, Lugar	

Ciudad	Value	Lang
	Tacna	

direccion	Value	Lang
	Dos de Mayo #345	

Departamento	Value	Lang
	Tacna	

Figura 32. Individuo Ubicación

Fuente: *Elaboración propia*

- Individuo Imprenta

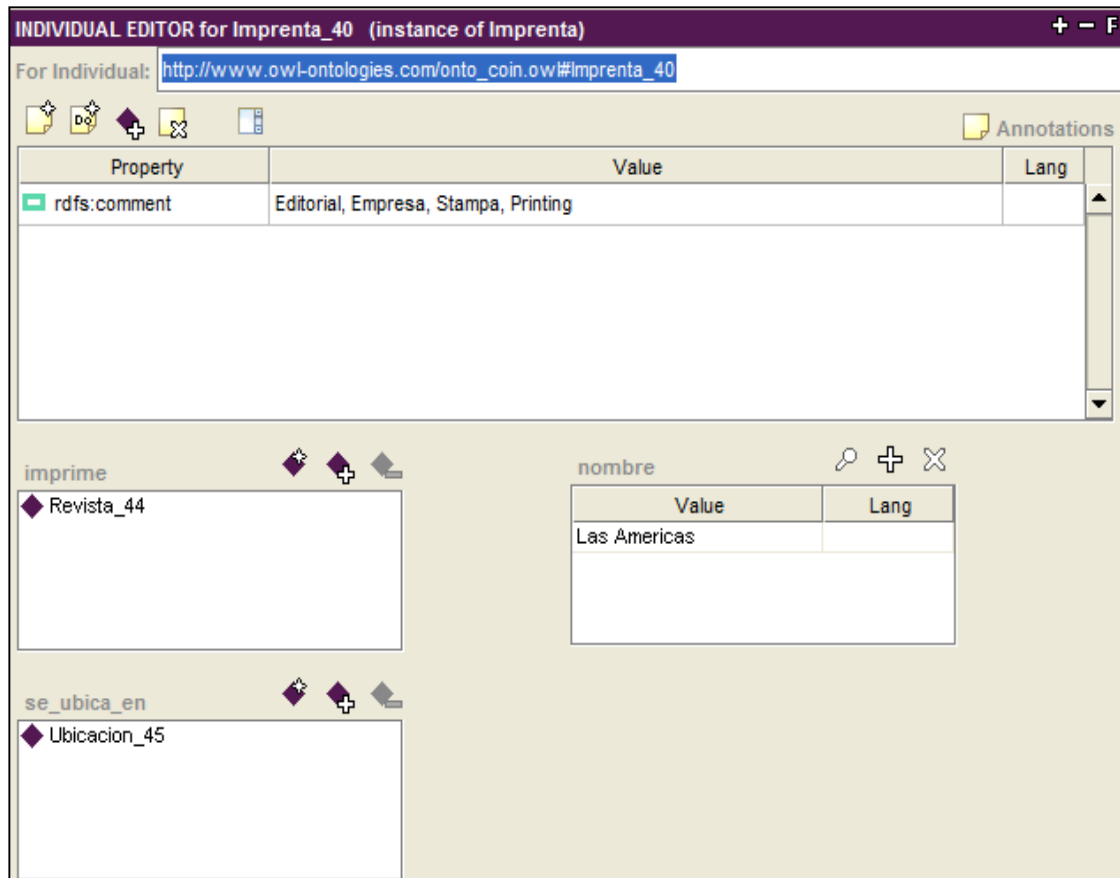


Figura 33. Individuo Imprenta

Fuente: *Elaboración propia*

3.2.6 Métricas de la Ontología

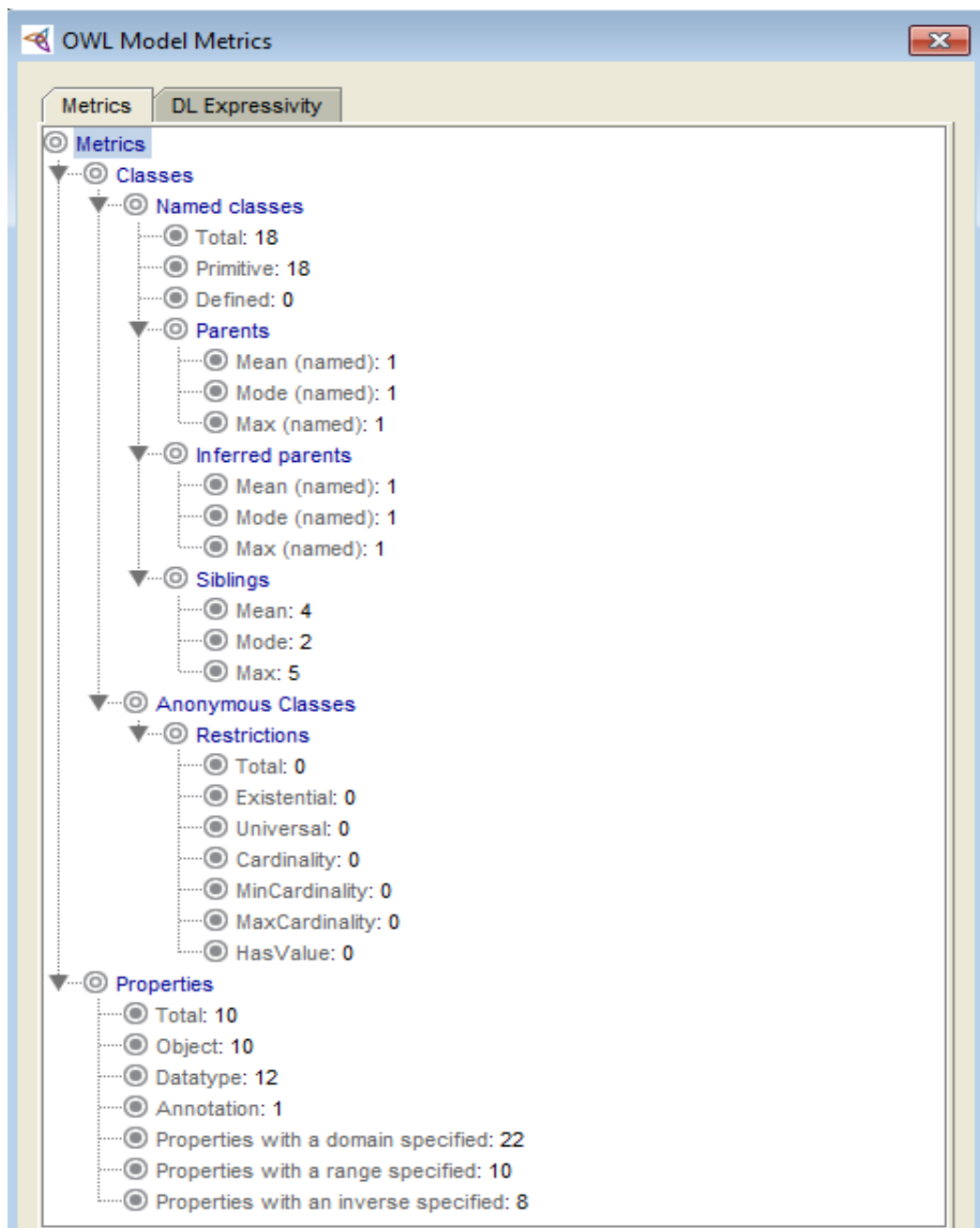


Figura 34. Métricas de la Ontología

Fuente: *Elaboración propia*

3.2.7 Diseño Conceptual de la Ontología

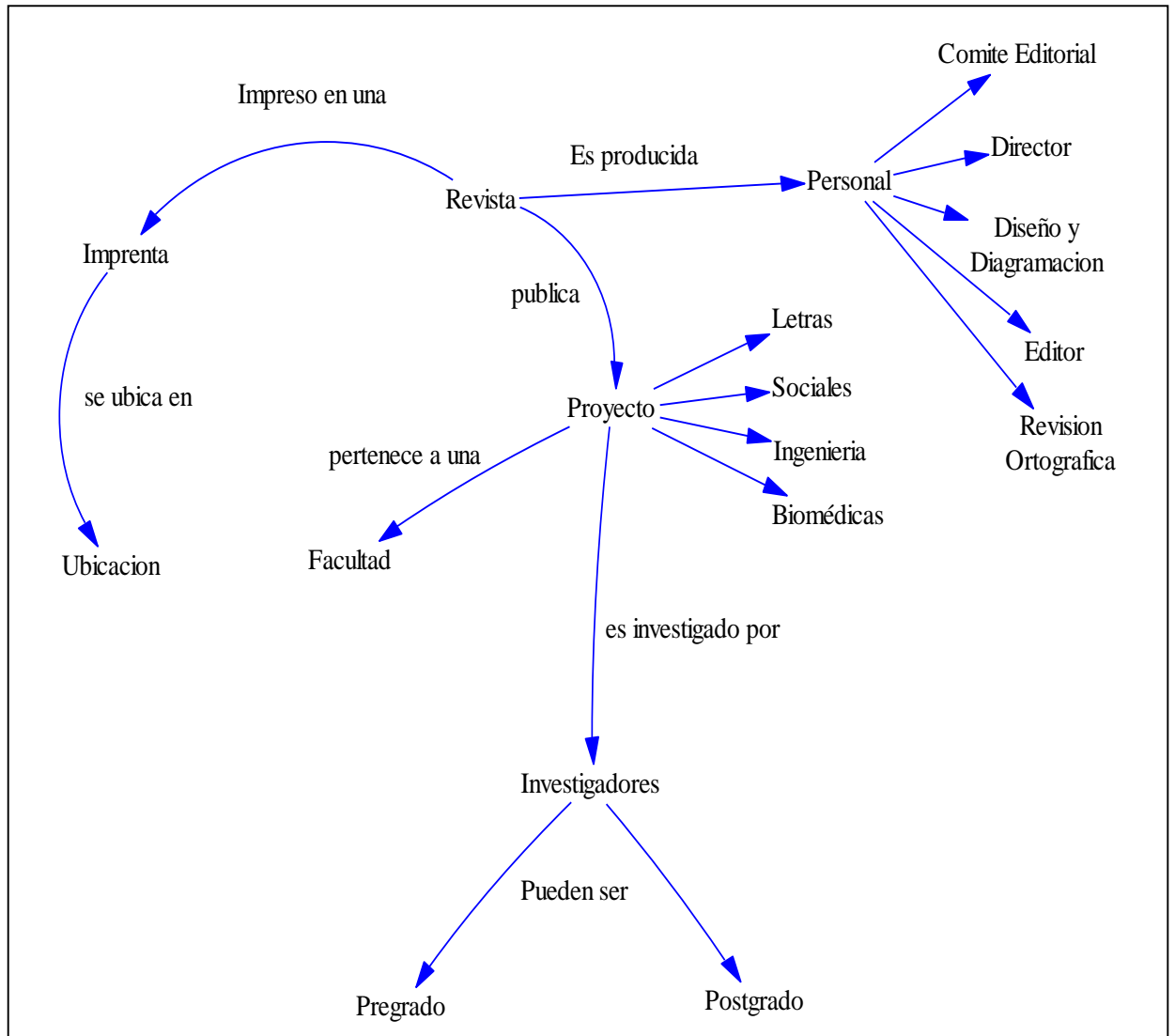


Figura 35. Diseño Conceptual de la Ontología

Fuente: Elaboración propia

Gráficas de Jambalaya (Analizador gráfico de ontologías)

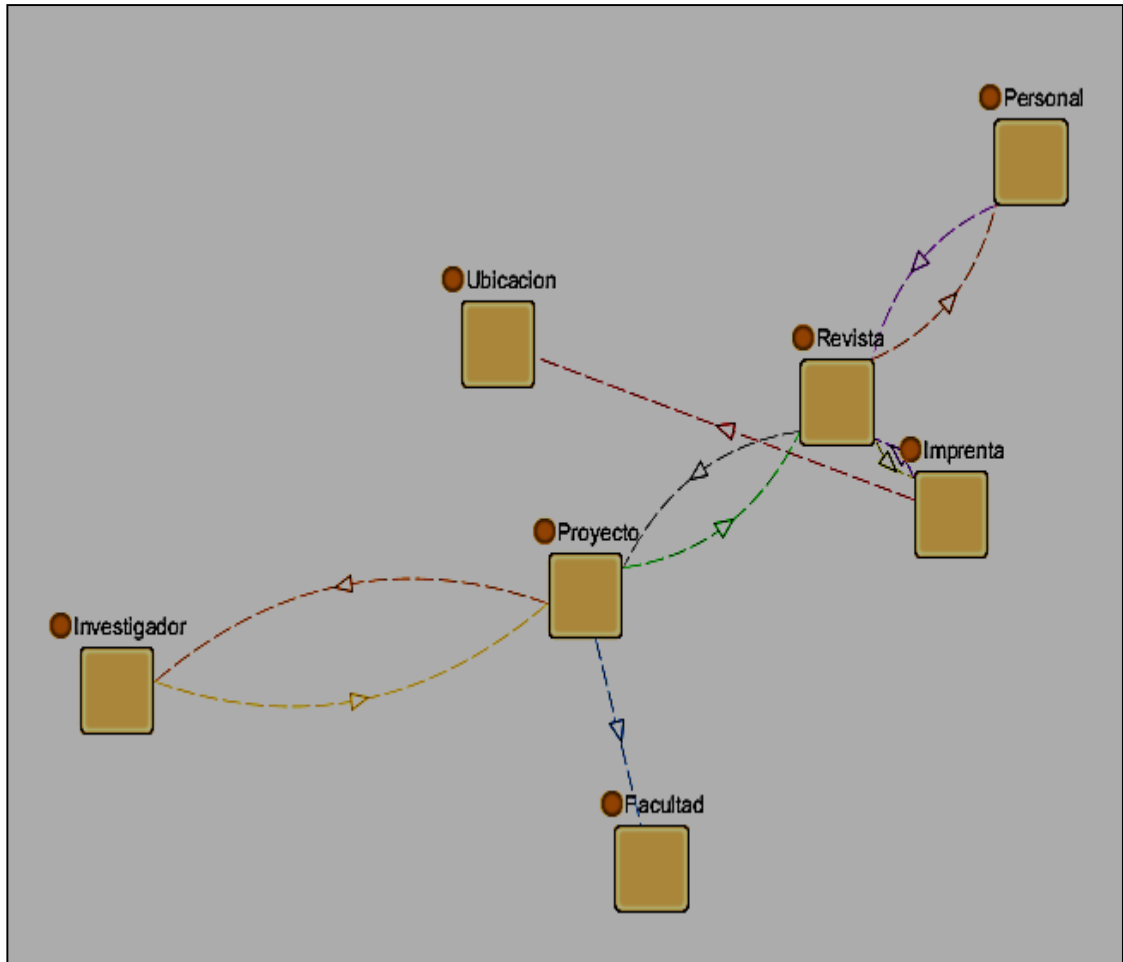


Figura 36. Gráfica de Dominio y Rango

Fuente: *Elaboración propia*

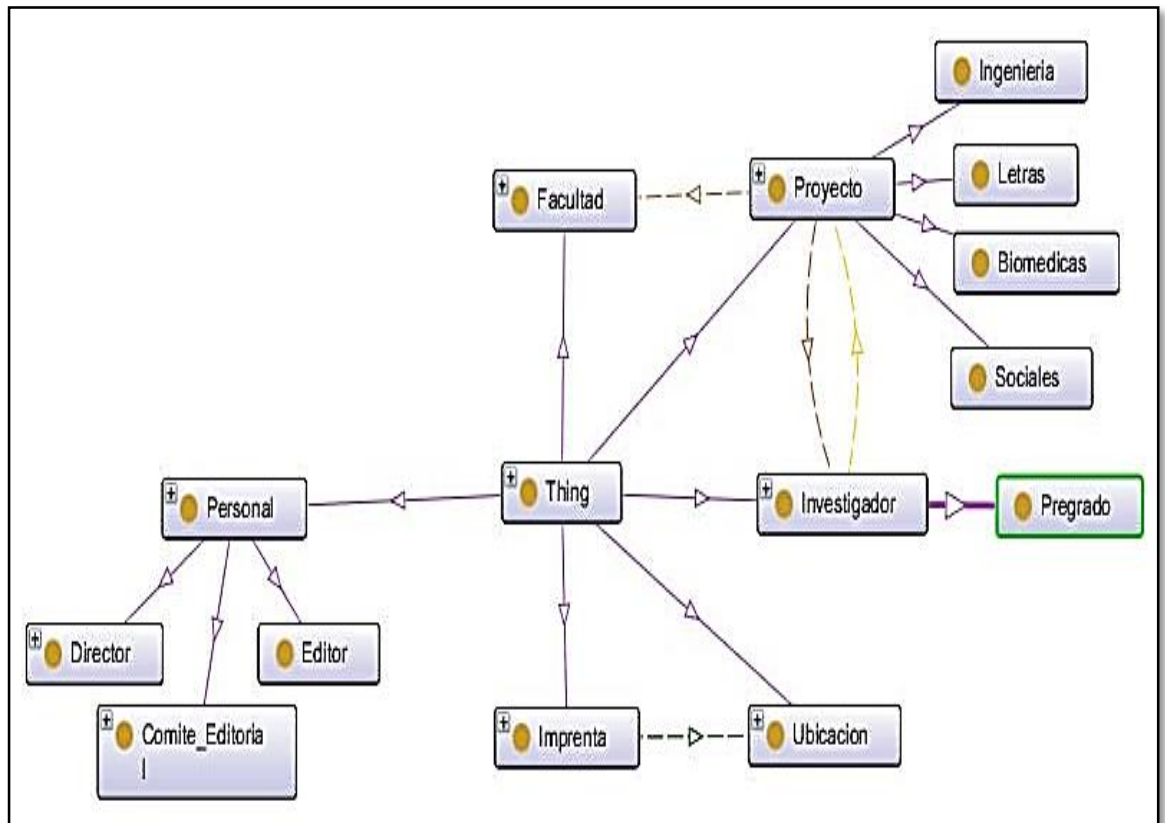


Figura 37. Gráfica de Árbol

Fuente: *Elaboración propia*

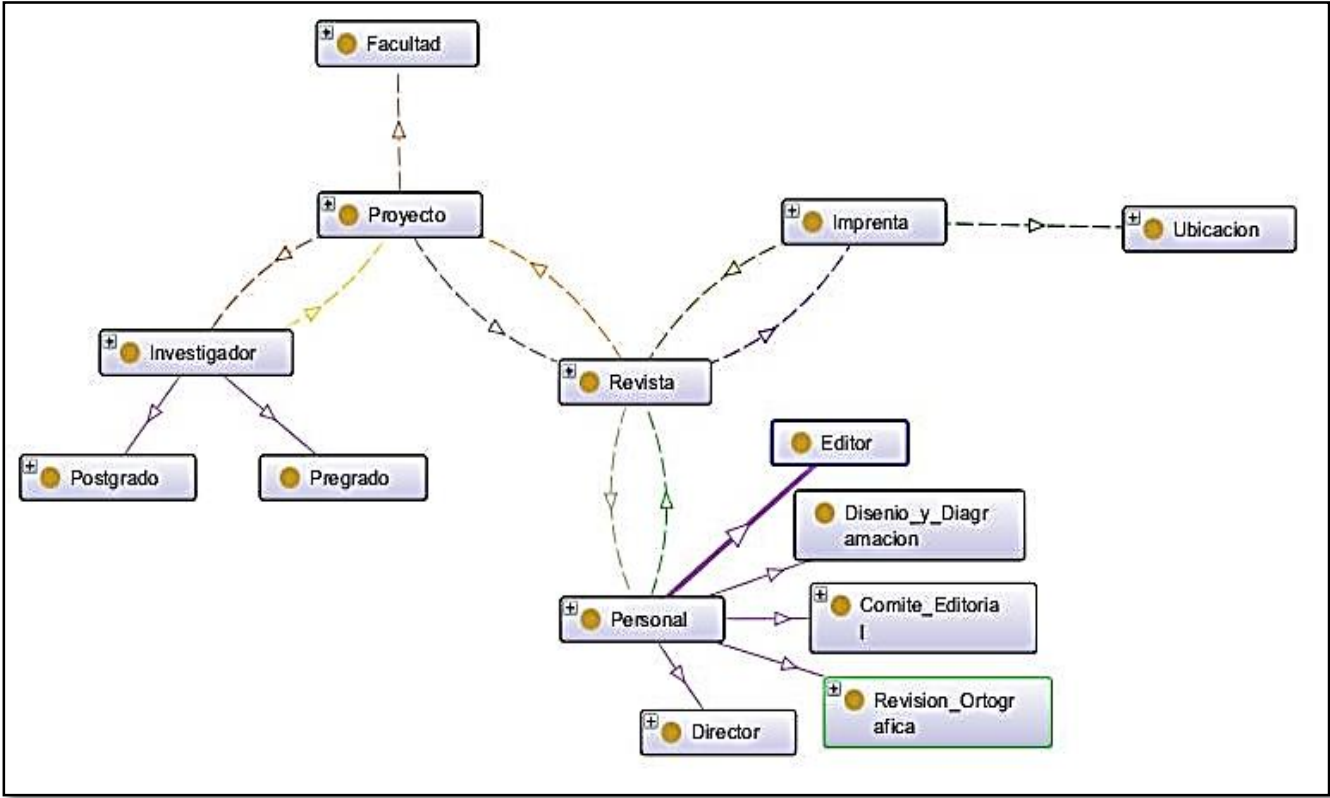


Figura 38. Gráfica de Vista Compuesta

Fuente: Elaboración propia

3.3 Diseño de la Base de Datos

Se analizó la composición de las tablas usadas en la base de datos que consume la ontología.

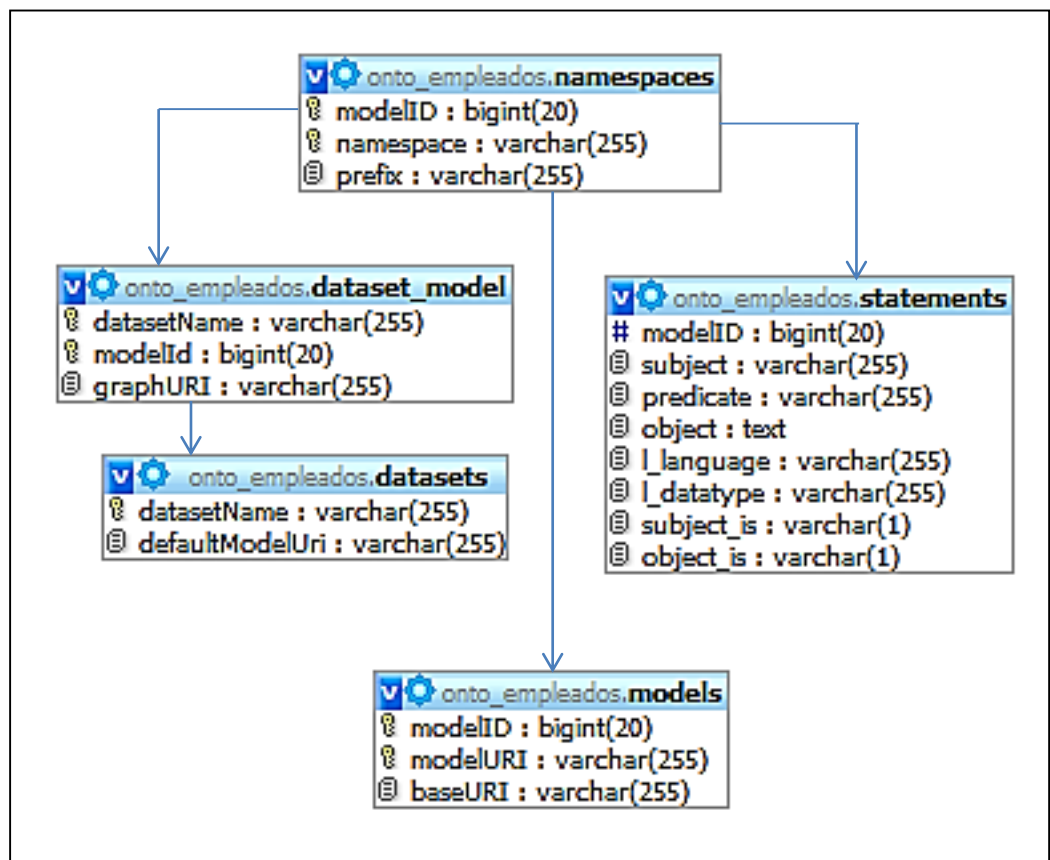


Figura 39. Diseño de la Base de Datos
Fuente: *Elaboración Propia*

IV. RESULTADOS

4.1 Descripción de Características

Tomando en consideración el análisis realizado de la estructura del motor de búsqueda de la ontología se puede definir la medición realizada para cada uno de los objetivos planteados. Para ello se utilizará el modelo Ontometric incluido dentro del analizador visual de Protége.

4.1.1 Datos Enlazados (RDF)

El análisis de los diccionarios semánticos utilizados para la ontología se desprende el uso de RDF para la descripción de las entidades, clases y subclases propias del motor de búsqueda semántica.

Para describir una instancia “Proyecto_86” del ente **proyecto**

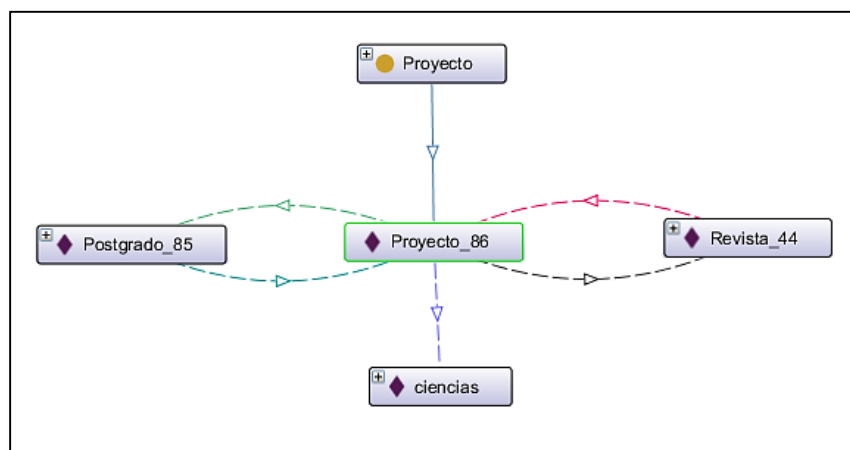


Figura 40. Registro del Proyecto 86
Fuente: *Elaboración Propia*

La codificación RDF es como sigue:

```
<Proyecto RDF:ID="Proyecto_86">
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Plan, Programa, Invento, Investigacion, Estudio, Project,
  Progetto</rdfs:comment>
  <pertenece>
    <Facultad rdf:ID="ciencias">
      <Escuelas rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Biologia y Microbiologia</Escuelas>
      <Escuelas rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Ingenieria en Informatica y Sistemas</Escuelas>
      <Escuelas rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Ingenieria Quimica</Escuelas>
      <Escuelas rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Fisica Aplicada</Escuelas>
      <nombre rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Ciencias</nombre>
    </Facultad>
  </pertenece>
  <es_publicado_en rdf:resource="#Revista_44"/>
  <es_investigado_por>
    <Postgrado rdf:ID="Postgrado_85">
      <nombre rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Daladier Castillo Cotrina </nombre>
      <nombre rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Cesar Huanacuni Lupaca</nombre>
      <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Cientifico, Inventor, Intelectual, Analista, Researcher,
      Ricercatore</rdfs:comment>
```

```

<investiga rdf:resource="#Proyecto_86"/>
  <Grado_Academico
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Maestro en Ciencias</Grado_Academico>
  <Grado_Academico
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Microbiologo</Grado_Academico>
  </Postgrado>
</es_investigado_por>
<url rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >http://www.unjbg.edu.pe/coin/pdf/07-2009.pdf</url>
</Proyecto>

```

A través del analizador de Protégé se determinaron las métricas para el primer componente RDF.

El analizador semántico presenta un total de 18 clases para toda la ontología.

Metrics	
Axiom	697
Logical axiom count	504
Class count	18
Object property count	11
Data property count	12
Individual count	84
DL expressivity	ALCI(D)

Figura 41: Métricas de RDF

Fuente: *Elaboración propia*

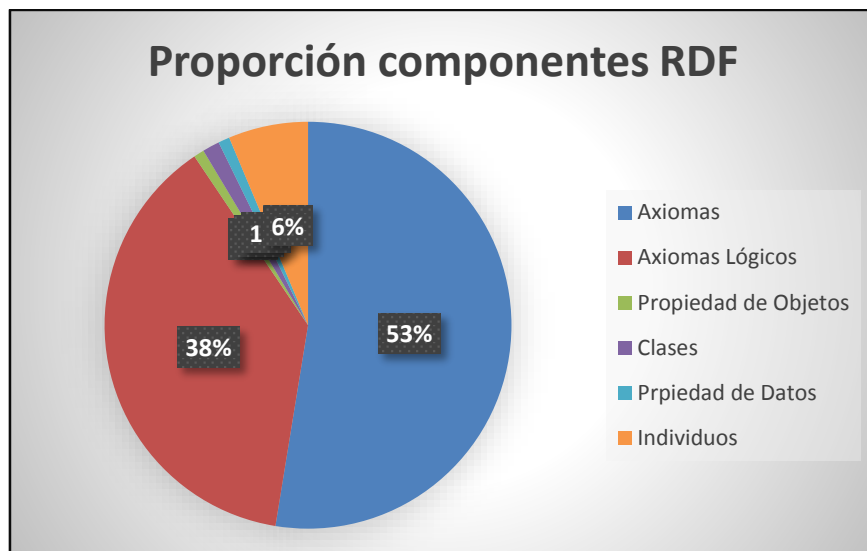


Figura 42: Proporción Componentes RDF

Fuente: *Elaboración propia*

Se muestran cada una de las 18 clases que se implementaron en la ontología:

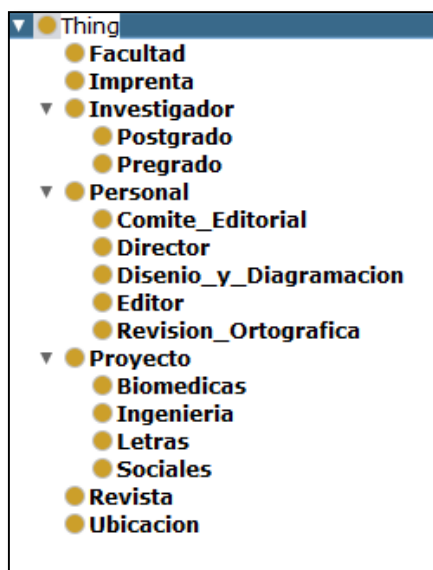


Figura 43: Cantidad de Clases usadas en la Ontología

Fuente: *Elaboración propia*

Además se utilizaron 11 Object Property



Figura 44: Object Property en la Ontología
Fuente: *Elaboración propia*

El primer Object Property es codificado en RDF según el estándar recomendado y es como sigue:

```
<!-- http://www.owl-  
ontologies.com/onto_coin.owl#es_investigado_por --  
>  
  
<owl:ObjectProperty rdf:about="http://www.owl-  
ontologies.com/onto_coin.owl#es_investigado_por">  
  <rdfs:range rdf:resource="http://www.owl-  
ontologies.com/onto_coin.owl#Investigador"/>  
  <rdfs:domain rdf:resource="http://www.owl-  
ontologies.com/onto_coin.owl#Proyecto"/>  
  <owl:inverseOf rdf:resource="http://www.owl-  
ontologies.com/onto_coin.owl#investiga"/>  
</owl:ObjectProperty>
```

En el código anterior se puede verificar que el Object Property nace a partir de las clases Investigador y Proyecto y además tiene una

contra relación “Investiga” como se puede observar en el siguiente gráfico:

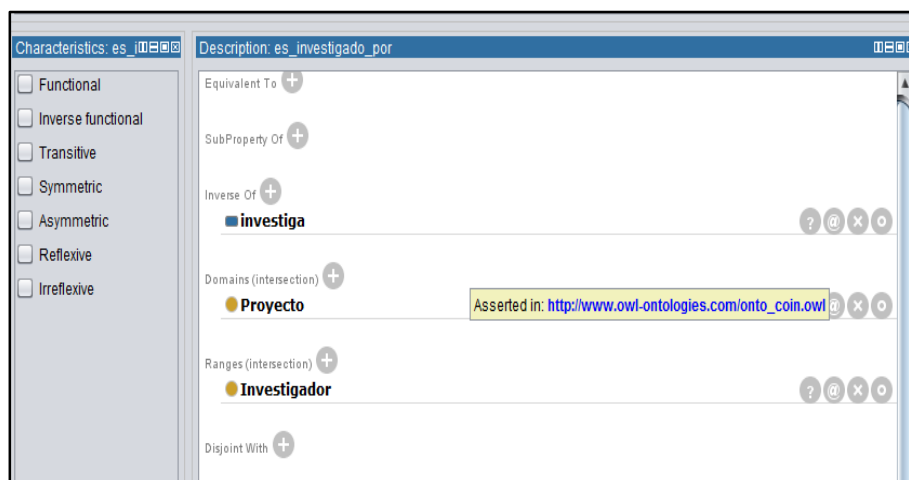


Figura 45: Object Property es_investigado_por
Fuente: *Elaboración propia*

Esto permite comprobar el uso de RDF en cada una de las 18 clases de la ontología.

4.1.2 Vocabularios (OWL)

En el análisis también se pudo comprobar el uso de OWL para definir las reglas entre instancias así como las relaciones semánticas entre ellas.

A continuación la definición de reglas para la clase **Ubicación**

```

<!-- http://www.owl-
ontologies.com/onto_coin.owl#Ubicacion -->

    <owl:Class rdf:about="http://www.owl-
ontologies.com/onto_coin.owl#Ubicacion">
        <sinonimo rdf:resource="http://www.owl-
ontologies.com/onto_coin.owl#Ubicacion_45"/>
<sinonimo rdf:resource="http://www.owl-
ontologies.com/onto_coin.owl#Ubicacion_47"/>
<sinonimo rdf:resource="http://www.owl-
ontologies.com/onto_coin.owl#Ubicacion_48"/>
<sinonimo rdf:resource="http://www.owl-
ontologies.com/onto_coin.owl#Ubicacion_49"/>
    </owl:Class>

```

En el analizador OWL de Protégé se pudo determinar la gráfica de lo anteriormente codificado:

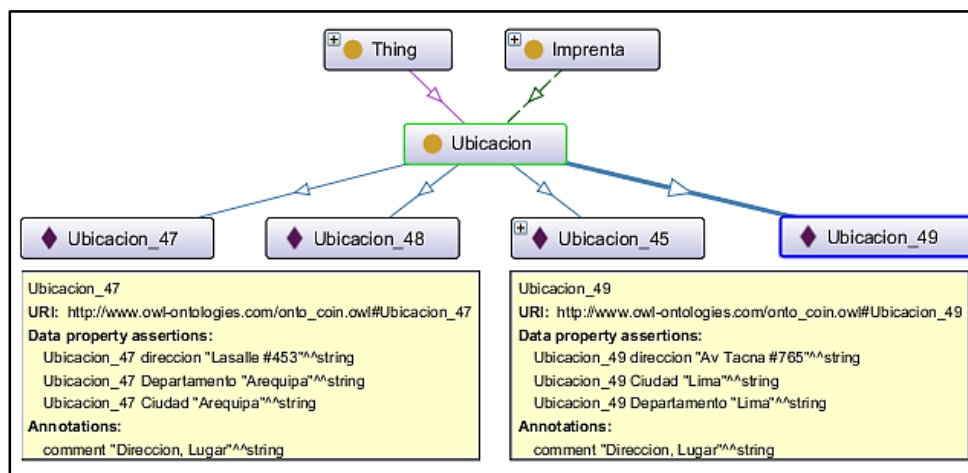


Figura 46: Reglas para la clase ubicación

Fuente: *Elaboración propia*

Además se determinaron las métricas para el componente OWL y es como sigue:

SubClassOf axioms count	11
EquivalentClasses axioms count	0
DisjointClasses axioms count	20
GCI count	0
Hidden GCI Count	0

Figura 47: Métricas para OWL
Fuente: *Elaboración propia*

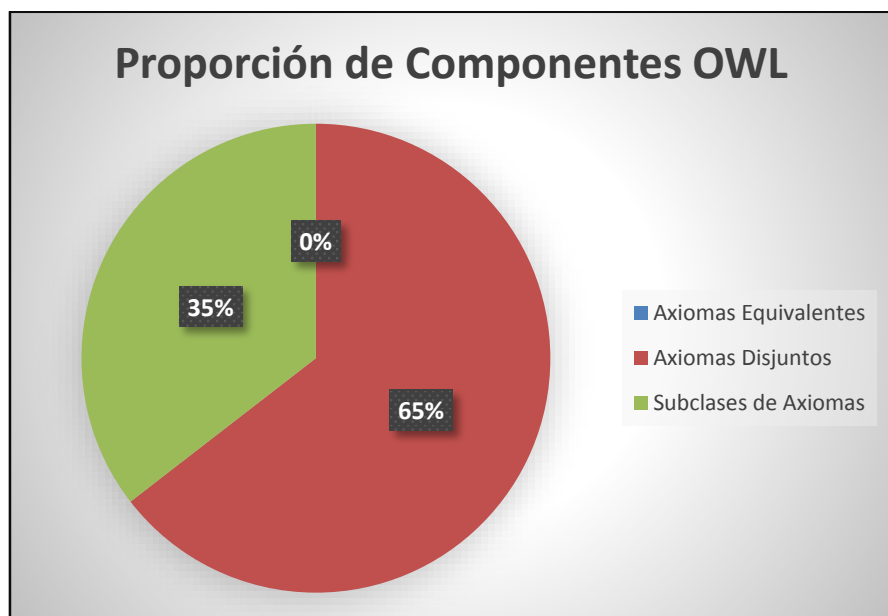


Figura 48: Métricas para OWL
Fuente: *Elaboración propia*

Cada uno de los axiomas es codificado en OWL según el estándar recomendado, lo cual permite comprobar el uso de OWL en los 31 axiomas de la ontología.

4.1.3 Consultas (SPARQL)

Las consultas SPARQL son fundamentales para consumir las ontologías, lo cual se puede verificar su uso en el sistema ontológico de COIN.

Las consultas Sparql se basan en tuplas que están conformadas por sujeto, predicado y objeto, esta composición de tres elementos permite explorar el conocimiento en una ontología

La Consulta SPARQL para la búsqueda de proyectos es como sigue:

```
$sparql_query = '
    SELECT ?title, ?autor, ?grado, ?descargar, ?facu
    WHERE (?x, vpro:Titulo, ?title),
           (?x, vpro:pertenece, ?fa),
           (?fa, vpro:nombre, ?facu),
           (?x, vpro:url, ?descargar),
           (?x, vpro:es_investigado_por ?y),
           (?y, vpro:Grado_Academico ?grado),
           (?y, vpro:nombre, ?autor)

           AND ?title ~~ '@' . $split . '@' | ?autor ~~ '@' .
ucfirst(strtolower($split)) . '@' \
| ?facu ~~ '@' . ucfirst(strtolower($split)) . '@' \ | ?grado
~~ '@' . ucfirst(strtolower($split)) . '@' \

    USING vpro FOR <http://www.owl-
ontologies.com/onto_coin.owl#>';
```

157	http://www.owl-ontologies.com/onto_coin.owl#Comite_Editorial_67	http://www.owl-ontologies.com/onto_coin.owl#nombre	"Edgar Faustino Taya Osorio"^^http://www.w3.org/2001/XMLSchema#string
158	http://www.owl-ontologies.com/onto_coin.owl#Comite_Editorial_67	http://www.owl-ontologies.com/onto_coin.owl#Grado_Academico	"Ingeniero"^^http://www.w3.org/2001/XMLSchema#string
159	http://www.owl-ontologies.com/onto_coin.owl#Comite_Editorial_67	http://www.owl-ontologies.com/onto_coin.owl#Email	"taya@hotmail.com"^^http://www.w3.org/2001/XMLSchema#string
480	http://www.owl-ontologies.com/onto_coin.owl#Proyecto_120	http://www.owl-ontologies.com/onto_coin.owl#Titulo	"ESTUDIO DE LA CONTAMINACION POR METALES PESADOS EN LAS COSTAS DE TACNA - PERU"^^http://www.w3.org/2001/XMLSchema#string
482	http://www.owl-ontologies.com/onto_coin.owl#Proyecto_120	http://www.owl-ontologies.com/onto_coin.owl#url	"http://www.unjbg.edu.pe/coin2/pdf/024-2009.pdf"^^http://www.w3.org/2001/XMLSchema#string

Para la tupla del registro 157 se mapeó el resultado de la siguiente manera:

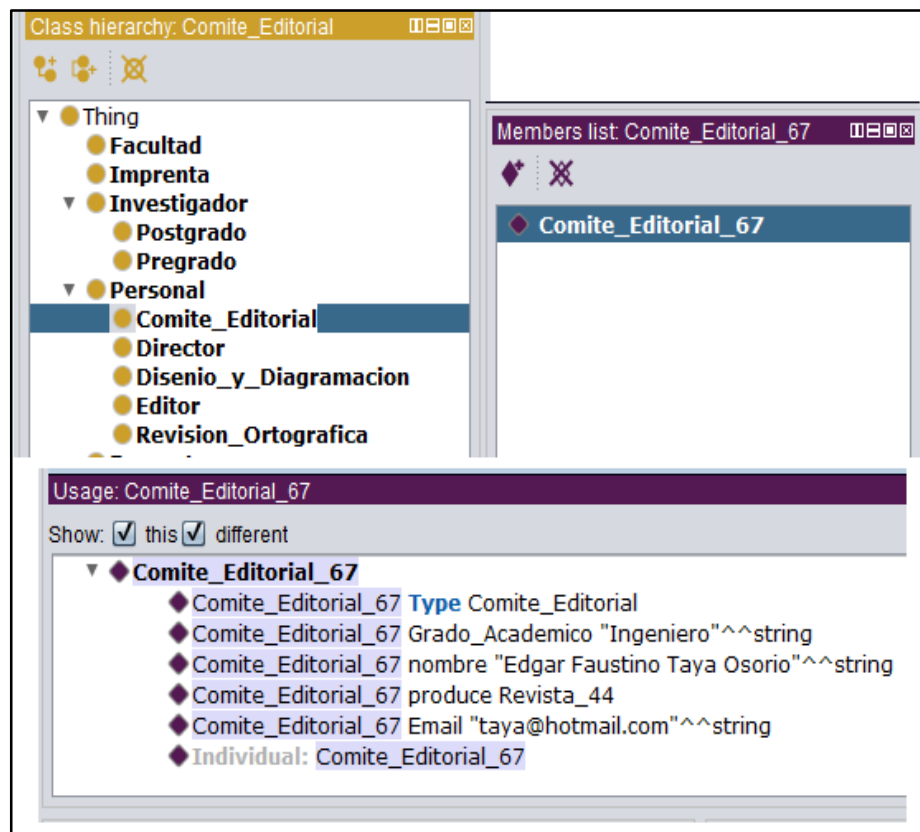


Figura 49: Tupla Investigador
Fuente: *Elaboración propia*

Lo cual se desprende del siguiente diagrama:

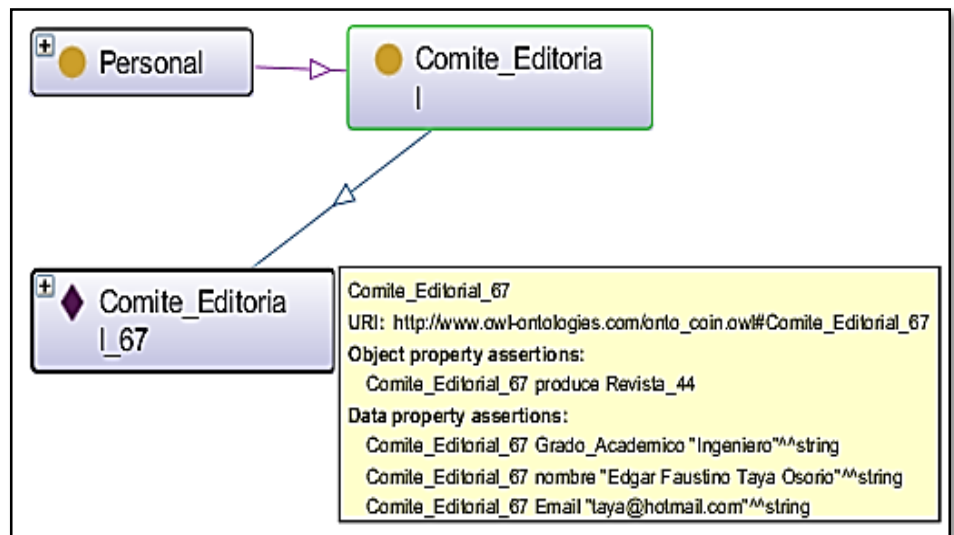


Figura 50: Diagrama Investigador
Fuente: *Elaboración propia*

Además, considerar que se tienen en total 723 tuplas debidamente validadas a través del **Validation Service** de la W3C.

Por lo tanto se puede comprobar el uso correcto del Sparql para la consulta de datos enlazados dentro de la ontología de acuerdo a las recomendaciones y los estándares para la web semántica.

V. DISCUSIONES

Las demostraciones de las hipótesis se harán bajo proposiciones recurrentes, es decir, proposiciones que afirman afirmando (Modus Ponendo Ponens).

Si en una demostración figura $(p \rightarrow q)$ y en la misma demostración también figura P , entonces, en dicha demostración también se puede concluir Q . Esta regla universal se conoce con el nombre del Modus Ponendo Pones ó Modus Ponens.

La regla MODUS PONENS (MP) se sintetiza como

$$\frac{p \rightarrow q}{p} \\ q$$

5.1 Cumplimiento de Objetivos

Se expresa con total seguridad que los objetivos se alcanzaron plenamente durante el desarrollo de la presente investigación.

El objetivo general fue el siguiente:

Determinar las características ontológicas del buscador semántico SERON del Consejo de Investigación – UNJBG, 2014.

Los objetivos específicos son tres, cuyo cumplimiento a continuación pasamos a expresar.

5.2 Primer Objetivo Específico

Verificar el uso de Datos Enlazados (RDF) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.

Este objetivo fue alcanzado, porque los diccionarios semánticos utilizados para la ontología usan RDF para la descripción de las entidades, clases y subclases propias del motor de búsqueda semántica.

Prueba de Hipótesis – Primer Sub hipótesis

Si las características ontológicas son fundamentales para los motores de búsqueda semántica.

Entonces, el uso de datos enlazados (RDF) tiene una importancia alta.

Para verificar la validez de la hipótesis procedemos de la siguiente forma:

Expresamos la hipótesis ($p \rightarrow q$):

Si las características ontológicas son fundamentales para los motores de búsqueda semántica. Entonces, el uso de datos enlazados (RDF) tiene una importancia alta.

Identificamos las premisas:

- p: Las características ontológicas son fundamentales para los motores de búsqueda semántica.
- q: El uso de datos enlazados tiene una importancia alta.

Por Modus Ponens podemos decir que aprobamos p. Por lo tanto se aprueba la sub hipótesis alternativa.

5.3 Segundo Objetivo Específico

Verificar el uso de Vocabularios (OWL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.

Este objetivo fue alcanzado porque se pudo comprobar el uso de OWL para definir las reglas entre instancias así como las relaciones semánticas entre ellas.

Prueba de Hipótesis – Segunda Sub hipótesis

Si las características ontológicas son fundamentales para los motores de búsqueda semántica.

Entonces, el uso de vocabularios (OWL) tiene una importancia alta.

Para verificar la validez de la hipótesis procedemos de la siguiente forma:

Expresamos la hipótesis (p->q):

Si las características ontológicas son fundamentales para los motores de búsqueda semántica. Entonces, el uso de vocabularios (OWL) tiene una importancia alta.

Identificamos las premisas:

- p: Las características ontológicas son fundamentales para los motores de búsqueda semántica.
- q: El uso de vocabularios (OWL) tiene una importancia alta.

Por Modus Ponens podemos decir que aprobamos p. Por lo tanto se aprueba la sub hipótesis alternativa.

5.4 Tercer Objetivo Específico

Verificar el uso de consultas (SPARQL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.

Las consultas SPARQL son fundamentales para consumir las ontologías, lo cual se puede verificar su uso en el sistema ontológico de COIN. Todo esto fue desarrollado en el capítulo III.

Prueba de Hipótesis – Tercera Sub hipótesis

Si las características ontológicas son fundamentales para los motores de búsqueda semántica.

Entonces, el uso de consultas (Sparql) tiene una importancia alta.

Para verificar la validez de la hipótesis procedemos de la siguiente forma:

Expresamos la hipótesis ($p \rightarrow q$):

Si las características ontológicas son fundamentales para los motores de búsqueda semántica. Entonces, el uso de consultas (SPARQL) tiene una importancia alta.

Identificamos las premisas:

- p: Las características ontológicas son fundamentales para los motores de búsqueda semántica.
- q: El uso de consultas (SPARQL) tiene una importancia alta.

Por Modus Ponens podemos decir que aprobamos p. Por lo tanto se acepta la sub hipótesis alternativa.

5.5 Medición de Indicadores

- Datos Enlazados (RDF)
- Vocabularios (OWL)
- Consultas (SPARQL)

Indicador	Si	No
Datos Enlazados	x	
Vocabularios	x	
Consultas	x	

VI. CONCLUSIONES

PRIMERA

En el análisis de la estructura del sistema ontológico de COIN se verificó el uso de Datos Enlazados (RDF) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.

SEGUNDA

Se pudo constatar el uso de Vocabularios (OWL) en el motor de búsqueda semántica SERON del Consejo de Investigación – UNJBG, 2014.

TERCERA

Se afirma que el sistema ontológico hace uso de consultas (SPARQL) en el motor de búsqueda semántica SERON del Consejo de Investigación UNJBG, 2014.

CUARTA

Se determinaron las características ontológicas del buscador semántico SERON del Consejo de Investigación – UNJBG, 2014.

VII. RECOMENDACIONES

- Evaluar el uso de RDFS extendido para poder determinar la potencia en la definición de clases del motor de búsqueda ontológico.
- Categorizar el uso de OWL en sus tres tipos desde lite a full para medir la complejidad del buscador semántico.
- Medir el tiempo de ejecución de las consultas SPARQL en producción para proponer mejoras que optimicen el tiempo de respuesta.

VIII. REFERENCIAS BIBLIOGRÁFICAS

- Barchini, G., Alvarez, M., & Herrera, S. (2006). Sistemas de Información: Nuevos Escenarios Basados en Ontologías. *Revista de Gestão da Tecnologia e Sistemas de Informação*, 2 - 18.
- Bennett, B., & Fellbaum, C. (2006). *Formal Ontology in .* Netherlands: IOS Press.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web. Scientific American, Volumen 284*, Páginas 34-43.
- Calero, C., Ruiz, F., & Piattini, M. (2006). *Ontologies for Software Engineering and Software Technology*. Berlin: Springer-Verlag.
- Colomb, R. (2007). *Ontology and the Semantic Web*. Amsterdam: IOS Press.
- Corcho, O., & Gómez-Pérez, A. (2000). *A Roadmap to Ontology Specification Languages*. Springer Verlag: EKAW.
- Davies, J., Studer, R., & Warren, P. (2006). *Semantic Web Technologies Trends and Research in Ontology Based Systems*. Londres: Wiley.
- Elmasri, R., & Navathe, S. (2013). *Fundamentos de Sistemas de Base de Datos .* Texas: Pearson.
- Fensel, D., Lausen, H., Jos de Bruijin, A., & Stollberg, M. (2007). *Enabling Semantic Web Services – The Web Service Modeling Ontology*. New York: Springer.
- Gasevic, D., Djuric, D., & Devedzic, V. (2006). *Model Driven Architecture and Ontology Development*. Berlin: Springer.
- Gomez Perez, A., Fernandez Lopez, M., & Corcho, O. (2010). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. London: Springer.

- Gómez-Pérez, A., Fernández López, M., & Corcho, O. (2004). *Ontological Engineering with examples from the areas of knowledge management, e-commerce and the semantic web*. London: Springer.
- Gruber, & Gruber, T. (1993). *A Translation Approach to Portable Ontology Specifications*. Stanford: Stanford University.
- Gruber, T. (1993). *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. California: Stanford University.
- Kompatsiaris, Y., & Hobson, P. (2008). *Semantic Multimedia and Ontologies – Theory and Applications*. London: Springer-Verlag.
- Nigro, H., González Císaro, S., & Xodo, D. (2008). *Data Mining with Ontologies: Implementations, Findings and Frameworks*. London: IGI Global.
- Rebstock, M., Fengel, J., & Paulheim, H. (2008). *Ontologies – Based Business Integration*. Berlin: Springer-Verlag.
- Studer, S., & Benjamins, R. (1998). Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, 161-197.
- Taniar, D., & Rahayu, J. (2006). *Web Semantics and Ontology*. Hershey: Idea Group.
- W3C. (2014). *World Wide Web Consortium*. Recuperado el 17 de Mayo de 2014, de <http://www.w3c.org>
- Zongmin, M. (2006). *Soft Computing in Ontologies and Semantic WebV*. Netherlands: Springer-Verlag.

Recursos de Internet

1. Site de COIN.

<http://www.unjbg.edu.pe/coin/>

2. Site de la World Wide Web Consortium

www.w3c.com

3. Web de Mysql

www.mysql.com

4. Web de Protégé

www.protege.stanford.edu