

	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA			
	<u>Documento</u>	<u>Código</u>	<u>Fecha</u>	<u>Revisión</u>
	FORMATO HOJA DE RESUMEN PARA TRABAJO DE GRADO	F-AC-DBL-007	10-04-2012	A
	<u>Dependencia</u>	<u>Aprobado</u>		<u>Pág.</u>
DIVISIÓN DE BIBLIOTECA	SUBDIRECTOR ACADEMICO		1(89)	

RESUMEN - TESIS DE GRADO

AUTORES	JOEL ALBERTO ACUÑA BON CARLOS MIGUEL VERA BAENE
FACULTAD	DE INGENIERÍAS
PLAN DE ESTUDIOS	INGENIERÍA DE SISTEMAS
DIRECTOR	ANDRES MAURICIO PUENTES VELASQUEZ
TÍTULO DE LA TESIS	PROTOTIPO DE SISTEMA INTELIGENTE DE CONSULTA DE OBJETOS VIRTUALES DE APRENDIZAJE (OVAS) DE LA ESPECIALIZACIÓN EN AUDITORIA DE SISTEMAS DE LA UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA (UFPSO)

RESUMEN (70 palabras aproximadamente)

EL PRESENTE TRABAJO PROPONE UN PROTOTIPO PARA EL MANEJO DE LA INFORMACIÓN VIRTUAL (OVAS) EN LA UFPSO, ESPECÍFICAMENTE EN LA ESPECIALIZACIÓN EN AUDITORÍA DE SISTEMAS. SE TRATA DE UNA INVESTIGACIÓN DE APROXIMADAMENTE DOS AÑOS ACERCA DE TEMAS RELACIONADOS CON LA WEB SEMÁNTICA. SE EXPONE LA FORMA EN LA QUE SE DISEÑA LA APLICACIÓN “ONTOBUS”, UN BUSCADOR SEMÁNTICO DE OBJETOS VIRTUALES DE APRENDIZAJE

CARACTERÍSTICAS

PÁGINAS. 89	PLANOS.	ILUSTRACIONES. 25	CD-ROM. 1
--------------------	----------------	--------------------------	------------------



**PROTOTIPO DE SISTEMA INTELIGENTE DE CONSULTA DE OBJETOS
VIRTUALES DE APRENDIZAJE (OVAS) DE LA ESPECIALIZACIÓN EN
AUDITORIA DE SISTEMAS DE LA UNIVERSIDAD FRANCISCO DE PAULA
SANTANDER OCAÑA (UFPSO)**

**JOEL ALBERTO ACUÑA BON
CARLOS MIGUEL VERA BAENE**

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA
FACULTAD DE INGENIERÍAS
INGENIERÍA DE SISTEMAS
OCAÑA
2016**

**PROTOTIPO DE SISTEMA INTELIGENTE DE CONSULTA DE OBJETOS
VIRTUALES DE APRENDIZAJE (OVAS) DE LA ESPECIALIZACIÓN EN
AUDITORIA DE SISTEMAS DE LA UNIVERSIDAD FRANCISCO DE PAULA
SANTANDER OCAÑA (UFPSO)**

**JOEL ALBERTO ACUÑA BON
CARLOS MIGUEL VERA BAENE**

**Proyecto de grado presentado como requisito para optar al título de
Ingeniero de Sistemas**

**Director
ANDRES MAURICIO PUENTES VELASQUEZ
Ingeniero de Sistemas**

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA
FACULTAD DE INGENIERÍAS
INGENIERÍA DE SISTEMAS
OCAÑA
2016**

CONTENIDO

	pág.
<u>INTRODUCCIÓN</u>	12
<u>1. PROTOTIPO DE SISTEMA INTELIGENTE DE CONSULTA DE OBJETOS VIRTUALES DE APRENDIZAJE (OVAS) DE LA ESPECIALIZACIÓN EN AUDITORIA DE SISTEMAS DE LA UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA (UFPSO)</u>	13
1.1 <u>PLANTEAMIENTO DEL PROBLEMA</u>	13
1.2 <u>FORMULACIÓN DEL PROBLEMA</u>	14
1.3 <u>OBJETIVOS</u>	14
1.3.1 Objetivo general	14
1.3.2 Objetivos específicos	14
1.4 <u>JUSTIFICACIÓN</u>	14
1.5 <u>DELIMITACIONES</u>	15
1.5.1 Conceptual	15
1.5.2 Operativa	16
1.5.3 Temporal.	16
1.5.4 Geográfica	16
<u>2. MARCO REFERENCIAL</u>	17
2.1 <u>MARCO HISTÓRICO</u>	17
2.1.1 Evolución de la web	17
2.1.1.1 Web 1.0.	17
2.1.1.2 Web 3.0	18
2.1.2 Historia de los buscadores semánticos	19
2.1.3 Generaciones de buscadores	19
2.1.4 Antecedentes en Colombia	20
2.2 <u>MARCO TEÓRICO</u>	20
2.2.1 Procesamiento de lenguaje natural (PLN).	21
2.2.2 Des-ambigüedad	23
2.2.3 Red semántica	24
2.2.4 Ontología. .	24
2.2.5 Buscador semántico	25
2.2.6 Inteligencia artificial. .	26
2.3 <u>MARCO CONCEPTUAL</u>	29
2.3.1 Lenguaje para desarrollo de ontologías.	29
2.3.2 JAVA/JAVASCRIPT	31
2.3.3 Frameworks	33
2.3.4 Herramientas de trabajo o IDE's.	35
2.3.5 Servidores	36
2.4 <u>MARCO LEGAL</u>	36

3. <u>DISEÑO METODOLÓGICO</u>	39
3.1 <u>TIPO DE INVESTIGACIÓN</u>	39
3.2 <u>POBLACIÓN Y MUESTRA</u>	39
3.3 <u>RECOLECCIÓN DE DATOS</u>	40
3.4 <u>ANÁLISIS DE LA INFORMACIÓN</u>	41
3.5 <u>METODOLOGÍA DEL DESARROLLO DE UN BUSCADOR SEMÁNTICO COMO UN SISTEMA INTELIGENTE PARA LAS CONSULTAS DE OVAS</u>	50
3.5.1 Investigación de teoría y conceptos.	50
3.5.2 Modelamiento del dominio del conocimiento.	50
3.6 <u>SEGUIMIENTO METODOLÓGICO DEL PROYECTO</u>	50
4. <u>RESULTADOS</u>	52
4.1 <u>MODELO DE ESPECIFICACIÓN DE REQUISITOS</u>	52
4.1.1 Requisitos funcionales y no funcionales del sistema	52
4.1.2 Requisitos no funcionales.	53
4.2 <u>DIAGRAMA DE COMPONENTES</u>	55
4.3 <u>FUNCIONALIDADES DEL SISTEMA</u>	56
4.3.1 Casos de uso	58
4.3.2 Diagrama de Secuencias.	59
5. <u>CONCLUSIONES</u>	62
6. <u>RECOMENDACIONES</u>	63
<u>BIBLIOGRAFÍA</u>	66
<u>ANEXOS</u>	69

LISTA DE ILUSTRACIONES

	pág.
Ilustración 1. Modelo de tripletas Sujeto, Propiedad, Objeto	25
Ilustración 2: Tabla de elementos de los metadatos	28
Ilustración 3: Constante "K" y Nivel de confianza	40
Ilustración 4: Grafica a la pregunta 1 de encuesta	42
Ilustración 5: Grafica a la pregunta 2 de encuesta	42
Ilustración 6: Grafica a la pregunta 3 de encuesta	43
Ilustración 7: Grafica a la pregunta 4 de encuesta	43
Ilustración 8: Grafica a la pregunta 5 de encuesta	44
Ilustración 9: Grafica a la pregunta 6 de encuesta	44
Ilustración 10: Grafica a la pregunta 7 de encuesta	45
Ilustración 11: Grafica a la pregunta 8 de encuesta	45
Ilustración 12: Grafica a la pregunta 9 de encuesta	46
Ilustración 13: Grafica a la pregunta 10 de encuesta	46
Ilustración 14: Grafica a la pregunta 11 de encuesta	47
Ilustración 15: Grafica a la pregunta 12 de encuesta	47
Ilustración 16: Grafica a la pregunta 13 de encuesta	48
Ilustración 17: Grafica a la pregunta 14 de encuesta	48
Ilustración 18: Grafica a la pregunta 15 de encuesta	49
Ilustración 19: Grafica a la pregunta 16 de encuesta	49
Ilustración 20: Diagrama de Componentes	56
Ilustración 21: Actores	58
Ilustración 22: Diagrama de Casos de Uso – Usuario Registrado	58
Ilustración 23: Diagrama de Casos de Uso – Sistema	59
Ilustración 24: Diagrama de secuencias de una búsqueda.	59
Ilustración 25: Arquitectura de la aplicación.	65

LISTA DE CUADROS

	pág.
Cuadro 1: Seguimiento Metodológico	49
Cuadro 2: Requisitos funcionales del sistema	51
Cuadro 3: Requisitos no funcionales del sistema	52

LISTA DE ANEXOS

	pág.
Anexo A. Encuesta dirigida a estudiantes, docentes y administrativos del área de postgrados de la Universidad Francisco de Paula Santander Ocaña.	69
Anexo B. Entrevista dirigida al director del proyecto	72
Anexo C. Diseño del componente de persistencia: “Base de Conocimiento”	74
Anexo D. Documento de especificación de Casos de Uso	82

RESUMEN

El presente trabajo propone un prototipo para el manejo de la información virtual (OVAS) en la UFPSO, específicamente en la especialización en auditoría de sistemas. Se trata de una investigación de aproximadamente dos años acerca de temas relacionados con la web semántica. Se expone la forma en la que se diseña la aplicación “Ontobus”, Un buscador semántico de objetos virtuales de aprendizaje. Se describe el funcionamiento y gestión de bases de conocimiento como ontologías interactuando con bases de datos y algoritmos para el procesamiento de lenguaje natural apoyados en la base de datos léxica “WordNet”. Además, se emulan las condiciones en una plataforma con un servidor de prueba con el fin de demostrar que los algoritmos funcionen y con la visión de usarla en todas las áreas de estudio. Los programas usados para la realización de la aplicación fueron: Netbeans, Protégé, DreamWeaver CS5, Apache TomCat y MySQL.

INTRODUCCIÓN

Durante los últimos años el crecimiento y avance tecnológico es exponencial, haciendo que las aplicaciones se vuelvan obsoletas más rápido y creando la necesidad de evolucionar constantemente. En los sistemas de información, el manejo de datos es un reto, la forma en la que se accede a la información y los resultados no siempre son los esperados e incluso las consultas se deben hacer muy cuidadosamente. A esto apunta esta investigación, las tecnologías de la web 3.0 y los diferentes algoritmos para el procesamiento de las consultas en lenguaje natural proveen un mejor servicio pues hacen más simple el trabajo. Como resultado: una aplicación inteligente que es capaz de distinguir el contexto en el que se escribe, buscando las relaciones de los datos, analizando sinónimos y corrigiendo palabras dándole significado a la web.

Los buscadores semánticos, encuentran resultados en función del contexto, información más exacta acerca de lo que se busca, ofreciendo una cantidad de resultados más sesgada, facilitando la labor de filtrar los resultados por parte del usuario. Desde el punto de vista investigativo, la semántica aplicada a sistemas de búsqueda de información es un tema que a nivel mundial ha madurado, debido a la cantidad de beneficios que promete y a su creciente acogida en las funcionalidades de la web, se considera pues el estudio de esta temática de gran interés y utilidad por parte de las nuevas generaciones de ingenieros que en gran medida se postulan como los desarrolladores de las tecnologías venideras. En este proyecto se desarrollará un prototipo de buscador semántico, partiendo de los fundamentos teóricos existentes y, del análisis que se llevará a cabo acerca de las tecnologías que se involucran, como son los agentes inteligentes de software, las ontologías implementadas en lenguajes como RDF y OWL, y demás herramientas de desarrollo.

1. PROTOTIPO DE SISTEMA INTELIGENTE DE CONSULTA DE OBJETOS VIRTUALES DE APRENDIZAJE (OVAS) DE LA ESPECIALIZACIÓN EN AUDITORIA DE SISTEMAS DE LA UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA (UFPSO)

1.1 PLANTEAMIENTO DEL PROBLEMA

La evolución de la web exige a las diferentes tecnologías, aplicaciones, sistemas y/o dispositivos un cambio o mejoramiento con la visión de optimizar los procesos que se efectúan, haciéndolos más simples y con mejores resultados, con menor uso de recursos. Con la llegada de la Web Semántica (Web 3.0) llegan inmersos los Buscadores Semánticos, en ellos los datos no son solo instancias individuales que son llamados directamente por medio de palabras clave, se usan además de bases de datos, bases de conocimiento como ontologías que incluyen de manera explícita el significado de los contenidos, facilitando la posterior gestión de los datos.

Los buscadores tradicionales y sus coincidencias directas no permiten incluir información sobre el significado de la búsqueda, lo que somete al usuario a un esfuerzo adicional al hacer “filtros” desechando los resultados menos relevantes. Con los años se han optimizado el tiempo de respuesta y la cantidad de resultados, sin embargo, no han logrado optimizar la calidad de los resultados; también se han desarrollado nuevas formas de representar los datos, es aquí donde la web semántica juega un papel preponderante como una solución tecnológica a algunos problemas relacionados con la ineficacia de resultados en las búsquedas.¹

Este trabajo se desarrolla como una solución para la recuperación de información de la especialización en auditoría de sistemas de la Universidad Francisco de Paula Santander Ocaña, el cual es uno de los programas académicos que más utiliza recursos virtuales para el aprendizaje y en diferentes plataformas (LMS, U-virtual, directorios compartidos en la nube, correos de difusión); conllevando a un crecimiento en la cantidad de estos recursos disponibles para todos los módulos del programa. Este crecimiento de recursos que deben estar disponibles en la web, debe acompañarse del desarrollo de herramientas tecnológicas que permitan gestionar y buscar eficientemente los contenidos que allí se representan.

Este es el punto de partida, pues eso es lo que se pretende realizar: un Buscador Semántico² donde la interacción del usuario con los contenidos virtuales de apoyo académico sea rápida y la información recolectada sea consistente. En la investigación nos encontramos con un trabajo realizado en la universidad de Cartagena que es similar, ya que este trabajo consiste en desarrollar una plataforma para la gestión de OVAS para la facultad de

¹ CODINA, Iluís. La web semántica: una visión crítica (online). Disponible en <http://www.elprofesionaldelainformacion.com/contenidos/2003/marzo/16.pdf>

² PUELLO, Plinio. Desarrollo de una plataforma para la gestión de objetos virtuales de aprendizaje para la facultad de odontología en la universidad de Cartagena. (online). Disponible en: <http://190.25.234.130:8080/jspui/handle/11227/42>

odontología.

1.2 FORMULACIÓN DEL PROBLEMA

¿El diseño de una aplicación web semántica servirá como alternativa para gestionar los objetos virtuales de aprendizaje de la especialización en auditoría de sistemas de la UFPSO?

1.3 OBJETIVOS

1.3.1 Objetivo general. Desarrollar un buscador semántico para gestionar la consulta de los objetos virtuales de aprendizaje de la especialización en auditoría de sistemas de la UFPSO.

1.3.2 Objetivos específicos.

Abstraer la información y el conocimiento del dominio a través de artefactos de ingeniería de software, modelos de especificación de requisitos e interfaces de usuario.

Diseñar los componentes de persistencia, tales como bases de Datos y bases de conocimiento.

Implementar las funcionalidades del buscador empleando metodologías de sistemas basados en conocimiento en un servidor de prueba, comprobando los resultados de las consultas y el acceso a los recursos virtuales de aprendizaje disponibles en la especialización en auditoría de sistemas.

1.4 JUSTIFICACIÓN

La web semántica se prepara para dar solución a los problemas de conceptualización de información de la web actual. El solo uso de los servicios web trae grandes ventajas y la capacidad de poder combinar varios servicios, es uno de los aspectos clave para maximizar las potencialidades individuales de las organizaciones a través de la integración de datos y de aplicaciones, lo cual es algo que en general se conoce como composición o colaboración de servicios. Como antecedentes de Buscadores semánticos encontramos: Hakia, Swoogle, SenseBot.

Los buscadores proveen lo necesario para solucionar la problemática. Tal como su nombre lo indica el Buscador Semántico recurren a ciertas particularidades vinculadas a la sintaxis, ontología, relación de las palabras o definiciones empleadas presentando una lista de respuestas acorde a la inquietud presentada en estos buscadores. Pero surge un problema con los motores de búsqueda, y está en la manera de recuperar la información, sin embargo ahí está el centro de esta investigación. Para comprender de una mejor manera el significado de los Buscadores Semánticos se procede a mencionar algunas palabras claves:

Ontología, Sintaxis, Relación, Vínculo, Significado, Clase, Instancia, Persona, Etiqueta.³

El presente trabajo constituye un aporte en materia de integración y uso de tecnologías emergentes enfocadas a optimizar el acceso a los recursos educativos digitales; esto se consigue con la creación y reúso de ontologías (para marcar eficientemente los recursos) y con el desarrollo de aplicaciones especializadas en consulta y recuperación de la información.

En ese punto, teniendo en cuenta que se busca la competitividad de la universidad en un mundo de cambios y mejoras constantes, idealmente tratando de estar a la vanguardia de la gestión datos, se pretende construir una aplicación web semántica basada en ontologías funcionando en un servidor de prueba para verificar la eficacia de las consultas de los materiales de la especialización en auditoría de sistemas y que sirva como base para futuras integraciones con sistemas activos de la universidad.

La diferencia que tiene este proyecto con respecto a otros que se han desarrollado en la universidad usando la tecnología de las ontologías, es que además de bases de datos y de conocimiento, se incluye: Programación web en HTML y JSP usando Servlets para la interconexión con formularios escritos en código JAVA, JAVASCRIPT para gestionar dinámicamente la interacción del usuario con los formularios HTML y lo más importante, en lo que hemos investigado, JENA, un paquete de librerías o código fuente que contiene clases y métodos que permiten la gestión de Tripletas RDF/OWL desde java, librerías que añadiremos a nuestro IDE NetBeans que será el lienzo en el que se trabaje. Su Arquitectura incluye una API para trabajar (Leer, procesar, escribir) ontologías RDF y OWL, un motor de inferencia para razonar sobre las ontologías, estrategias de almacenamiento flexible para guardar tripletas en memoria o ficheros y un motor de “queries” compatible con especificación SPARQL.

De esta manera los resultados de esta propuesta de investigación van a ser satisfactorios permitiendo alcanzar logros tecnológicos a futuro, quizás usando MICRO JENA que se usa para trabajar sobre dispositivos móviles, es ahí donde la institución puede obtener mejores niveles de competencia frente a otras. Destacando el aumento de la frecuencia de uso de los materiales académicos, incentivando a las personas a la investigación y desarrollo de sus trabajos, proyectos y tareas específicas dentro del campus universitario, aprovechando los recursos propios de la universidad.

1.5 DELIMITACIONES

1.5.1 Conceptual. La articulación del proyecto requiere de la aplicación teórica en aspectos relacionados con: Inteligencia artificial, Objetos virtuales de aprendizaje (OVA), Buscador semántico, Ontologías, Web 3.0, Procesamiento de Lenguaje Natural (PLN) y desarrollo software orientado a la web.

³ MIRA jose, Aspectos básicos de la inteligencia artificial. 1ª edición. REBIUN. 1995. p. 201.

1.5.2 Operativa. Los recursos virtuales de aprendizaje y los metadatos asociados a ellos son administrados por las personas encargadas de la especialización en auditoría de sistemas con el soporte efectivo de la unidad virtual, por esta razón, se harán requerimientos formales a estas dos dependencias, buscando la información como el insumo más importante para el desarrollo de este sistema.

El proyecto se divide en 4 principales tareas: La construcción de la(s) Ontología(s) Necesaria(s) usando Protégé, el modelado de la Base de Datos usando MySQL, la creación de los métodos y clases que permitirán el funcionamiento de la aplicación en JAVA y JAVASCRIPT, implementando las librerías de JENA y JWNL y por último, el diseño y la implementación de las interfaces sobre Formularios Web.

1.5.3 Temporal. Se estima un periodo máximo de 8 meses para su realización.

1.5.4 Geográfica. El proyecto de investigación se llevará a cabo en la ciudad de Ocaña, Norte de Santander, se emplearán los laboratorios del Tecnoparque SENA nodo Ocaña, que facilita el espacio y los equipos adecuados para la realización.

2. MARCO REFERENCIAL

2.1 MARCO HISTÓRICO

2.1.1 Evolución de la web

En 1957 se crea ARPA, Agencia de Proyectos de Investigaciones Avanzadas.

En 1969 se crea ArpaNet, partiendo de los avances en las tecnologías en redes.

En 1990 se desmantela ArpaNet y aparecen, Milnet de carácter militar e Internet de naturaleza pública, en el sector industrial, académico e investigativo.

En 1991 Tim Berners Lee crea la World Wide Web (WWW) y propone un sistema de hipertexto para compartir documentos (HTML).⁴

2.1.1.1 Web 1.0. No es que se trate de una versión de producto (1.0 o 2.0) pero la Web es ahora denominada así por las evoluciones a través del tiempo. La Web 1.0 empezó en los años 60's junto al Internet, de la forma más básica que existe, con navegadores de solo texto, como ELISA, era bastante rápido pero muy simple, luego Tim Berners crea el HTML (Hyper Text Markup Language) donde los primeros navegadores visuales fueron Netscape e Internet Explorer.

La característica de esta Web es que era de solo lectura, ya que el usuario no podía interactuar con el contenido de la página (ningún comentario, respuesta, etc).

Funciona por hipertexto y gráficas e incluye efectos multimedios.

Es considerado el acceso más sencillo y comprensible a la información disponible en internet.

Enlaza páginas o documentos localizados en la red sin importar su ubicación física o geográfica.

Protocolos: Http, Protocolo de transferencia de hipertexto. Html. Formato hipertextual e hipergráfico para publicar documentos en red, creado para codificarlos y visualizarlos, con formatos, layout y estructura de un documento web.

2.1.1.2 Web 2.0. La Web 2.0, es un concepto que se originó en una sesión de brainstorming (o lluvia de ideas) entre Dale Dougherty de O'Really y Craig Cline de MediaLive International, a partir de ideas sobre evolucionar en la web, en una conferencia en octubre de 2004, caracterizan a la Web 2.0 como una nueva actitud o evolución de la Internet, que se resumía en tres principios básicos: la web como plataforma, la inteligencia colectiva y la

⁴ Disponible en: http://julionica.udem.edu.ni/wp-content/uploads/2014/01/Evolucion_Web.pdf

arquitectura de la participación.

Dicho de otro modo, este término propuesto por Tim O'Reilly se refiere a una segunda generación en la historia del desarrollo de tecnología Web, basada en comunidades de usuarios y en una amplia gama de herramientas, como las redes sociales, blogs, wikis y otros, que fomentan la colaboración y el intercambio de información.

La Web 2.0 es también llamada Web social por el enfoque colaborativo y de construcción social de esta herramienta, donde el usuario es el ente más importante de información. Es una web basada en comunidades de usuarios, se pasa de una web informativa, creada por expertos a una web social, donde cualquiera participa, aparecen aplicaciones web muy potentes y sencillas de manejar enfocadas al usuario final y basa su desarrollo en Sistemas de Gestión de Contenidos, los cuales permiten la creación y administración de contenidos en páginas web principalmente.

El usuario es el protagonista, crea y comparte información.

El conocimiento compartido en base a la suma de esfuerzos individuales.

No son necesario muchos conocimientos para crear espacios en internet.

No es ninguna tecnología ni lenguaje de programación.

Es una técnica de desarrollo web que combina tecnologías para conseguir una navegación ágil y rápida.

Tecnologías: XHTML (HTML) y CSS para la presentación de datos, Document Object Model (DOM) para interactuar con la información dinámicamente, XML y XSLT para intercambiar y manipular datos con servidores web, XMLHttpRequest para la recuperación y envío de datos de modo asíncrono y JavaScript como nexo de unión.

2.1.1.2 Web 3.0. Hoy en día apenas estamos desglosando la Web 3.0 y ya se habla hasta de la 4.0. Si bien estas no están tan difundidas como la Web 2.0, representan una evolución considerable.

La Web 3.0 es considerada como una extensión de Web definida por el premio Nobel Sir Timothy “Tim” John Berners – Lee. Y trata de poder enriquecer la comunicación mediante metadatos semánticos (ontologías) que aportan un valor añadido a la información, la diferencian y la hacen más inteligente. Por lo que los contenidos ahora ya no son tratados por su sintáctica sino por su semántica.

La web actual es útil porque se puede comprender, principalmente está compuesta por documentos HTML que son comprendidos por humanos, pero la idea es integrar la información o encontrarla acertadamente, de ahí surge la idea de la Web Semántica.

Entonces, mientras en la Web Sintáctica nos encontrábamos con recursos enlazados entre sí que no aseguran que se enlacen con la totalidad de las páginas que existen, con escasa precisión de los resultados y alta sensibilidad al vocabulario usado, en la Web Semántica se le añade lo que le falta para convertirse en un entorno donde se accede a la información, facilitando el procesado de datos y resolviendo problemas de interoperabilidad entre aplicaciones.

La Web mantiene los principios de: Descentralización, Compartición, Compatibilidad, Facilidad de Acceso y Contribución y, además, rescata la noción del concepto de ontología del campo de Inteligencia Artificial que será expuesta más adelante en este documento.

Y por último cabe mencionar que la Web 4.0 es un término que se va acuñando recientemente y trata de movernos hacia una Web Ubicua donde el objetivo primordial será el de unir las inteligencias, para que tanto las personas como las cosas se comuniquen entre sí para generar la toma de decisiones.⁵

2.1.2 Historia de los buscadores semánticos. Sobre el año 2007 surgieron algunos buscadores semánticos que acabaron en fracaso, pues la falta de estructura, de adaptación a formatos y a anotaciones semánticas en los documentos en la web dificultó el desarrollo de estas búsquedas inteligentes. A pesar de ello, actualmente, hay algunos de estos buscadores semánticos funcionando, como es el caso de Hakia y Evri. La búsqueda semántica despierta interés, debido a que tiene un gran vínculo con las ideas iniciales de los creadores del Internet.

En el año 2008 el buscador Google ganó el premio Príncipe de Asturias de Comunicación. El jurado había considerado que Google hizo posible una gigantesca revolución cultural y ha propiciado el acceso generalizado al conocimiento.

En el 2010 Google compra la compañía Metaweb, líder en web semántica. Esta compañía mantiene una inmensa base de datos abierta llamada Freebase, que estructura información proveniente de todas partes del mundo, y que Google ya venía estudiando en servicios como su Sitio de Noticias. A raíz de dicha compra, se crea la aplicación llamada Knowledge Graph, que cuenta con 500 millones de datos y 3,5 mil millones de vínculos entre estos datos, el cual acerca al buscador Google a la web semántica y el procesamiento del lenguaje natural, describiendo significados, contextos y relaciones en los contenidos hallados.⁶

2.1.3 Generaciones de buscadores. Los buscadores se dividen en tres generaciones, esta son:

⁵ CODINA, Iluis. ¿Web 2.0, Web 3.0 o Web Semántica? : El impacto en los sistemas de información de la Web (online) UPF, 2009 [citado 1 Nov 2014]

⁶ MERO, David. Buscadores Semánticos, 2da edición, Universidad Jaen, Octubre 1999. cap 14.

Buscadores de primera generación: se basa en palabras clave de búsqueda para mostrar los resultados. También llamados buscadores “keywords” (palabras claves).

Buscadores de segunda generación: Se basa en palabras clave y otras variables agregadas para poder generar la búsqueda y así mostrar los resultados al usuario. Podemos encontrar dentro de esta segunda generación a los metabuscadores como MetaCrawler (siendo el primero metabuscador) y los buscadores verticales, también se encuentran nuevos sistemas que obtienen información de fuentes externas a los documentos. Otro cambio importante es que ya no se asume la honestidad de los autores de los documentos.

Buscadores de tercera generación: Son los llamados buscadores semánticos. Estos extraen el significado escondido en (bases de datos, ficheros de e-mail, grabaciones, imágenes o vídeos) o se encuentra en varios idiomas. Estos buscadores pueden contestar a cualquier pregunta que se le formule, incluso si se realizase en lenguaje natural, por esta razón son llamados también buscadores inteligentes, y se basan en poder contextualizar y “entender” la información que indexan.⁷

2.1.4 Antecedentes en Colombia. Actualmente en Colombia no hay muchos trabajos enfocados a la web semántica, específicamente en buscadores semánticos que fortalezcan la consulta de los datos y el acceso a la información académica. Hay un trabajo muy interesante sobre un buscador semántico orientado a páginas de turismo, en el año 2009 en Barranquilla. Pero cabe rescatar un buscador colombiano que pretende competirle a google.

Se trata de Gyffu, un buscador desarrollado por la iniciativa de 5 colombianos y que reúne características importantes como la consulta anónima. Gyffu actúa como un buscador tradicional pero se trabaja en implementar Gyffusion que es la evolución de este buscador con tecnologías semánticas y de inteligencia artificial. Está en línea hace muy poco tiempo y cuenta con más de 45.000 usuarios. En su página se puede encontrar más información: www.gyffu.com.⁸

2.2 MARCO TEÓRICO

Debido a la rápida evolución de la web (desde la primera generación o web 1.0 pasando por la web 2.0 y llegando a la web 3.0 o web semántica) y el gran incremento de contenidos presentes en Internet como red global, cada vez se hace más necesario tener métodos eficientes de recuperación de información.

La recuperación de información consiste en encontrar el material (normalmente documentos) de entre grandes colecciones de datos para satisfacer la necesidad de un

⁷ LOZADA, Pablo. Evolución de la Web (Online), 2014 [citado 28 abril 2014]. Disponible en http://julionica.udem.edu.ni/wp-content/uploads/2014/01/Evolucion_Web.pdf

⁸ Disponible en <http://www.eltiempo.com/tecnosfera/gyffu-el-buscador-colombiano-que-quiere-ser-mejor-que-google/14795836>

usuario. El principal objetivo de un sistema de recuperación de la información debe ser obtener los documentos más relevantes posibles en relación a una consulta particular.

Hoy en día es imprescindible el uso de motores de búsqueda para realizar las consultas en Internet y es posible que un mismo motor proporcione respuestas diferentes para diferentes versiones de una misma pregunta. Estos resultados pueden depender de las palabras clave utilizadas y no siempre son correctos. Estas son las principales motivaciones para el desarrollo de métodos de búsqueda semántica, aprovechar las propiedades de la semántica (como el estudio del significado de las palabras) para orientar la búsqueda y así intentar obtener resultados óptimos.

La dificultad de este tipo de búsqueda recae en que para los seres humanos es fácil establecer equivalencias semánticas entre diferentes expresiones pero este proceso no es evidente para los sistemas automatizados. Un sistema de búsqueda semántica ideal tendría que emular un hipotético sistema de búsqueda humano con una memoria suficientemente grande para recordar y relacionar todas las preguntas y respuestas anteriormente consultadas. Es cierto que diferentes personas pueden dar diferentes respuestas a una misma pregunta pero por mucho que re-formulemos la consulta la respuesta será similar ya que semánticamente serán consultas equivalentes.

Finalmente el objetivo definitivo para un sistema artificial de búsqueda semántica será obtener los mismos resultados y en el mismo orden de relevancia respecto a diferentes consultas semánticamente equivalentes.

2.2.1 Procesamiento de lenguaje natural (PLN). Es un campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia las interacciones entre las computadoras y el lenguaje humano.

El PLN se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio de lenguajes naturales. El PLN no trata de la comunicación por medio de lenguajes naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente (que se puedan realizar por medio de programas que ejecuten o simulen la comunicación). Los modelos aplicados se enfocan no sólo a la comprensión del lenguaje de por sí, sino a aspectos generales cognitivos humanos y a la organización de la memoria. El lenguaje natural sirve sólo de medio para estudiar estos fenómenos. Hasta la década de 1980, la mayoría de los sistemas de PNL se basaban en un complejo conjunto de reglas diseñadas a mano. A partir de finales de 1980, sin embargo, hubo una revolución en PNL con la introducción de algoritmos de aprendizaje automático para el procesamiento del lenguaje.⁹

Dificultades en el procesamiento de lenguajes naturales

⁹ CARBONELL, Jaime. El procesamiento del lenguaje natural, tecnología en transición. (online) 2006. Disponible en <http://www.sc.edu.es/sbweb/fisica/cursoJava/fundamentos/colecciones/stringtokenizer.htm>

Ambigüedad. El lenguaje natural es inherentemente ambiguo a diferentes niveles:

A nivel léxico, una misma palabra puede tener varios significados, y la selección del apropiado se debe deducir a partir del contexto oracional o conocimiento básico. Muchas investigaciones en el campo del procesamiento de lenguajes naturales han estudiado métodos de resolver las ambigüedades léxicas mediante diccionarios, gramáticas, bases de conocimiento y correlaciones estadísticas.

A nivel referencial, la resolución de anáforas y catáforas implica determinar la entidad lingüística previa o posterior a que hacen referencia.

A nivel estructural, se requiere de la semántica para desambiguar la dependencia de los sintagmas preposicionales que conducen a la construcción de distintos árboles sintácticos. Por ejemplo, en la frase Rompió el dibujo de un ataque de nervios.

A nivel pragmático, una oración, a menudo, no significa lo que realmente se está diciendo. Elementos tales como la ironía tienen un papel importante en la interpretación del mensaje.

Para resolver estos tipos de ambigüedades y otros, el problema central en el PLN es la traducción de entradas en lenguaje natural a una representación interna sin ambigüedad, como árboles de análisis.

Detección de separación entre las palabras. En la lengua hablada no se suelen hacer pausas entre palabra y palabra. El lugar en el que se debe separar las palabras a menudo depende de cuál es la posibilidad que mantenga un sentido lógico tanto gramatical como contextual. En la lengua escrita, idiomas como el chino mandarín tampoco tienen separaciones entre las palabras.

Recepción imperfecta de datos. Acentos extranjeros, regionalismos o dificultades en la producción del habla, errores de mecanografiado o expresiones no gramaticales, errores en la lectura de textos mediante OCR.

Componentes

Análisis morfológico. El análisis de las palabras para extraer raíces, rasgos flexivos, unidades léxicas compuestas y otros fenómenos.

Análisis sintáctico. El análisis de la estructura sintáctica de la frase mediante una gramática de la lengua en cuestión.

Análisis semántico. La extracción del significado de la frase, y la resolución de ambigüedades léxicas y estructurales.

Análisis pragmático. El análisis del texto más allá de los límites de la frase, por ejemplo, para determinar los antecedentes referenciales de los pronombres.

Planificación de la frase. Estructurar cada frase del texto con el fin de expresar el significado adecuado.

Generación de la frase. La generación de la cadena lineal de palabras a partir de la estructura general de la frase, con sus correspondientes flexiones, concordancias y restantes fenómenos sintácticos y morfológicos.

2.2.2 Des-ambigüedad. Típicamente el caso que suele presentarse es el de un usuario con una necesidad de información más o menos concreta que propone una consulta a un motor de búsqueda, esta consulta contiene palabras clave que el usuario considera necesarias o correctas para obtener la información deseada. Entonces éste convierte en metadatos (crea una representación) las palabras clave utilizadas en la consulta y realiza la búsqueda en su BD.

Esta búsqueda contiene la relación de metadatos con todos los documentos que conoce y devuelve una lista de resultados en función de la relevancia establecida por el orden de clasificación. Este sistema tiene dos limitaciones principales: a veces el usuario no es capaz de definir correctamente su objetivo mediante palabras clave además de que los motores de búsqueda no entienden el lenguaje natural.

El lenguaje natural es muy complejo debido, en gran parte, al gran número de sinónimos y palabras polisémicas que contiene. En este punto entra en juego la importancia de la aplicación de sistemas de búsqueda semántica en los motores de búsqueda. En general el proceso de búsqueda semántica es:

Interpretar la pregunta del usuario extrayendo los conceptos más relevantes de la frase.

Utilizar este grupo de conceptos para crear una consulta y utilizarla contra la ontología del sistema.

Presentar los resultados al usuario.

Con tal de entender que es lo que el usuario está buscando (punto A del proceso), se debe desambiguar el significado de las palabras clave utilizadas en la pregunta. Se considera que un término es ambiguo cuando este puede tener un considerable número de significados posibles, por ejemplo la palabra hoja como "la hoja de un árbol", "una hoja de papel" o "una hoja de afeitar". Gracias a los procesos de desambiguación se elige el significado más probable de entre todos los posibles.

Estos procesos tienen en cuenta el significado del resto de palabras presentes en la frase y el resto del texto de las web's. La determinación de cada significado influye en la desambiguación de los demás hasta llegar a una situación de máxima verosimilitud y coherencia para la frase inicial consultada. Toda la información fundamental para el proceso, es decir, todo el conocimiento utilizado por el sistema, se ve representada en forma

de una red semántica organizada alrededor de un núcleo conceptual.¹⁰

2.2.3 Red semántica

El principal objetivo de la investigación de redes semánticas es el desarrollo de una serie de lenguajes y la tecnología necesaria para expresar información semántica que pueda ser entendida y procesada por las computadoras para poder aplicarlo al entorno del trabajo en red. Una estructura de este tipo pretende representar el conocimiento lingüístico mostrando las interrelaciones entre conceptos.

Cada concepto léxico coincide con el nodo de una red semántica y está conectado con otros por relaciones semánticas específicas en una estructura jerárquica y hereditaria. De esta forma, cada concepto enriquece con sus características y su significado a los nodos cercanos.

Cada nodo de la red agrupa un conjunto de sinónimos que representan el mismo concepto léxico y pueden contener:

Lemas simples ('asiento', 'vacaciones', 'trabajo', 'rápido', 'más', etc.).

Compuestos ('guardaespaldas', 'pararrayos', 'aguardiente', etc.).

Colocaciones ('plan de choque', 'paquete bomba', 'llevar a cabo', 'bajo consumo', etc.).

Los enlaces que identifican las relaciones semánticas entre los conjuntos de sinónimos son las directrices a seguir para la organización de la red semántica de conceptos.¹¹

2.2.4 Ontología. Antes se ha mencionado el concepto de ontología que aparece en la parte del proceso de la búsqueda semántica en que ya se ha desambiguado el mensaje inicial del usuario y se pretende realizar la consulta definitiva.

Ontología es un término originalmente utilizado en filosofía y ahora relacionado con la informática como un conjunto de términos jerárquicamente estructurado para describir un dominio que puede ser utilizado como núcleo de una base de conocimientos. Con tal de que las ontologías puedan ser interpretadas por los ordenadores, los conceptos, propiedades y sus relaciones, restricciones y normas se describen en un lenguaje formal. Por lo tanto la ontología de un dominio de conocimiento proporciona un vocabulario con el que se representa el conocimiento de ese dominio de conocimiento y el conjunto de relaciones que mantienen los términos utilizados en el vocabulario.

La definición de ontologías es un proceso básico para el desarrollo de búsquedas semánticas ya que con ellas se obtienen muchos beneficios. Definen la terminología de un

¹⁰ Disponible en <http://www.sc.ehu.es/sbweb/fisica/cursoJava/fundamentos/colecciones/stringtokenizer.htm>

¹¹ VERA José. PIMENTEL Carlos. Redes semánticas: Aspectos teóricos, técnicos, metodológicos, y analíticos 1ª Edición, Mc Graw Hill Agosto 2009. p. 104.

dominio de búsqueda de forma detallada con lo que los conceptos semánticos quedan establecidos de forma no ambigua además de obtener una relación directa entre conceptos semánticos similares.¹²

2.2.5 Buscador semántico. Es aquel que realiza el rastreo atendiendo al significado del grupo de palabras que se escribe y no basándose en las actuales etiquetas. En pocas palabras, Un buscador inteligente.

Funcionamiento. El buscador semántico extrae información de los datos etiquetados (cabeceras, RDF, microformatos). Evidentemente si nuestra web no contiene estas etiquetas ser hace más difícil que sea entendida por los algoritmos de búsqueda.

Los algoritmos de estos buscadores generan las ontologías que definen las relaciones entre las entidades mediante el Lenguaje de Ontología Web (OWL), turtle, N3, N-Triples. Una ontología define cosas como clases, superclases, subclases, propiedades y las relaciones entre ellos. Por ejemplo, podría especificar que “si una persona trabaja para una compañía, entonces la compañía emplea a la persona” (una relación inversa).

Ilustración 1. Modelo de tripletas Sujeto, Propiedad, Objeto



Fuente. Autores del proyecto

Las ontologías son normalmente representadas con tripletas que tienen un sujeto, un predicado y un objeto. El giro real aquí es que los predicados también se pueden utilizar como el sujeto u objeto de cualquier otra declaración en la ontología. Por ejemplo, puede definir lo que la “inversa” o “reflexiva” significa, usando el mismo lenguaje que se utiliza para definir que: X persona trabaja para la compañía Y.

Otro punto importante en los buscadores semánticos son los razonadores que pueden procesar ontologías y los datos para crear nuevo conocimiento que antes no existía. Por ejemplo, saber todas las personas que trabajan para las filiales de una empresa sin que estas personas hayan estado etiquetadas antes. Los buscadores semánticos también usan herramientas de consulta como SPARQL, que se utiliza para consultar este conocimiento

¹² ALCINA Amparo, VALERO Esperanza. RAMBLA Elena. Terminología y Sociedad del Conocimiento. 1ª Edición. PETERLANG, 2009. p. 255

almacenado y generado.¹³

2.2.6 Inteligencia artificial. Es un área multidisciplinaria que, a través de ciencias como las ciencias de la computación, la lógica y la filosofía, estudia la creación y diseño de entidades capaces de resolver cuestiones por sí mismas utilizando como paradigma la inteligencia humana.

General y amplio como eso, reúne a amplios campos, los cuales tienen en común la creación de máquinas capaces de pensar. En ciencias de la computación se denomina inteligencia artificial a la capacidad de razonar de un agente no vivo. John McCarthy acuñó la expresión «inteligencia artificial» en 1956, y la definió así: “Es la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes”.¹⁴

Búsqueda del estado requerido en el conjunto de los estados producidos por las acciones posibles.

Algoritmos genéticos (análogo al proceso de evolución de las cadenas de ADN).

Redes neuronales artificiales (análogo al funcionamiento físico del cerebro de animales y humanos).

Razonamiento mediante una lógica formal análogo al pensamiento abstracto humano.

También existen distintos tipos de percepciones y acciones, que pueden ser obtenidas y producidas, respectivamente, por sensores físicos y sensores mecánicos en máquinas, pulsos eléctricos u ópticos en computadoras, tanto como por entradas y salidas de bits de un software y su entorno software. Varios ejemplos se encuentran en el área de control de sistemas, planificación automática, la habilidad de responder a diagnósticos y a consultas de los consumidores, reconocimiento de escritura, reconocimiento del habla y reconocimiento de patrones.

Los sistemas de IA actualmente son parte de la rutina en campos como economía, medicina, ingeniería y la milicia, y se ha usado en gran variedad de aplicaciones de software, juegos de estrategia, como ajedrez de computador, y otros videojuegos.

2.2.7 Objetos virtuales de aprendizaje (OVAS). Un Objeto Virtual de Aprendizaje, comúnmente llamado OVA, también se conoce en algunos contextos como OA, que significa Objeto de Aprendizaje, tomado de OL (en Inglés) Object Learning. Aunque existen diferentes definiciones de un Objeto de Aprendizaje, en Colombia, expertos de diferentes universidades, por solicitud del Ministerio de Educación, construyeron un concepto propio para nuestro país el cual establece que es “Un conjunto de recursos

¹³ MEIR Louis. Semantic Web 1ª Edición. UHA, 2008. P. 111

¹⁴ Tomado de: http://elpais.com/diario/2011/10/27/necrologicas/1319666402_850215.html

digitales, que pueden ser utilizados en diversos contextos, con un propósito educativo y constituido por al menos tres componentes internos: contenidos, actividades de aprendizaje y elementos de contextualización. Además, el Objeto de Aprendizaje, debe tener una estructura de información externa (metadato), para facilitar su almacenamiento, identificación y recuperación”.¹⁵

A su vez, se define un Objeto Informativo como: “Un conjunto de recursos digitales, que pueden ser utilizados en diversos contextos educativos, y que posee una estructura de información externa (metadato), para facilitar su almacenamiento, identificación y recuperación”.¹⁶ Y, un recurso digital: “Un recurso digital es cualquier tipo de información, que se encuentra almacenada en formato digital”.¹⁷

Componentes de un objeto de aprendizaje. Así como la definición construida por el Equipo del Ministerio de Educación Nacional, en Colombia existen otras definiciones, como la del Instituto de Ingenieros Eléctricos y Electrónicos, IEEE: “Un objeto es cualquier entidad digital o no digital, que puede ser usada, re-usada o referenciada, para el aprendizaje soportado en tecnología”. Aunque existe cierta similitud entre ambas definiciones, la diferencia sustancial radica, en que en el ámbito nacional, se ha enfocado a los Objetos Digitales o Virtuales. Además, se destacan los componentes que deben contener.

Componentes internos de un OA.

Contenidos: Se da a conocer el tema, utilizando diferentes estrategias, con el fin de capturar la atención del estudiante, puede ser a través de aplicaciones multimedia, donde se involucre texto, imágenes, animaciones, audio, etc. Todo esto, con el fin de contribuir con la comprensión del tema, por parte de los estudiantes

Actividades de aprendizaje: son actividades que debe desarrollar el estudiante, ya sean directamente en el software, o a través de otros mecanismos.

Elementos de contextualización: Esta información, conocida como metadatos, hace referencia los datos que describen el objeto, como: título, idioma, la versión, la información relacionada con los derechos de autor. Esta información, permitirá ubicar fácilmente el objeto, desde diferentes sistemas, así como su reutilización en otros escenarios.

Estructura externa de un OA. La estructura externa de un Objeto de Aprendizaje, está conformada por los metadatos, los cuales describen múltiples atributos de los Objetos de Aprendizaje e Informativos. Esta información se encuentra clasificada en las siguientes categorías, según la especificación LOM¹⁸ de la IEEE:

¹⁵ Tomado de: <http://www.colombiaaprende.edu.co/objetos/>

¹⁶ Tomado de: <http://www.colombiaaprende.edu.co/objetos/>

¹⁷ Tomado de: <http://www.colombiaaprende.edu.co/html/directivos/1598/article-117373.html>

¹⁸ IEEE P1484.12.2 - Metadatos para Objetos de Aprendizaje - <http://ltsc.ieee.org/wg12/materials.html>

Ilustración 2: Tabla de elementos de los metadatos

CATEGORÍA METADATO	ELEMENTOS QUE LA CONFORMAN
General:	Título, idioma, descripción, palabras clave.
Ciclo de Vida:	Versión, autor(es), entidad, fecha
Técnico:	Formato, tamaño, ubicación, requerimientos, instrucciones de instalación.
Derechos:	Costo, derechos de autor y otras restricciones
Anotación:	Uso educativo.
Clasificación:	Fuente de clasificación, ruta taxonómica.

Fuente. Autores del proyecto

Tales metadatos, con la información descriptiva de los recursos, son los que facilitan el almacenamiento, organización e identificación de los Objetos de Aprendizaje e Informativos.

Funciones de los objetos de aprendizaje

Favorecer la generación, integración y reutilización de Objetos de Aprendizaje.

Posibilitar el acceso remoto a la información y contenidos de aprendizaje.

Posibilitar la integración de diferentes elementos multimedia, a través de una interfaz gráfica.

Contribuir a la actualización permanente de profesores y alumnos.

Estructuración de la información en formato hipertextual.

Facilitar la interacción de diferentes niveles de usuarios. (Administrador, diseñador y alumno).

Características de un objeto de aprendizaje

Reusable: Es decir, que pueda ser utilizable en diferentes contextos, con fines educativos.

Interoperable: Capacidad de integrarse en diferentes plataformas de aprendizaje.

Escalable: Permite integración con estructuras más complejas.

Interactivo: Capacidad de generar actividades y comunicación entre sujetos involucrados.

Autocontenido: El contenido debe ser lo suficientemente completo, como para el tema que se pretende enseñar.

2.3 MARCO CONCEPTUAL

2.3.1 Lenguaje para desarrollo de ontologías. Para el desarrollo de las ontologías existen diferentes lenguajes que se pueden utilizar en diferentes aplicaciones de la base de conocimientos, entre esas:

Repositorios para la organización del conocimiento

Servir de herramienta para la adquisición de información

Servir de herramientas de referencia en la construcción de sistemas de bases de conocimiento que aporten consistencia, fiabilidad y falta de ambigüedad a la hora de recuperar información

Normalizar los atributos de los metadatos aplicables a los documentos

Crear una red de relaciones que aporte especificación y fiabilidad

Permitir compartir conocimiento

Posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre organizaciones, comunidades científicas, etc.

Permitir la integración de diferentes perspectivas de usuarios

Permitir el tratamiento ponderado del conocimiento para recuperar información de forma automatizada

Permitir la construcción automatizada de mapas conceptuales y mapas temáticos

Permitir la reutilización del conocimiento existente en nuevos sistemas

Permitir la inter operatividad entre sistemas distintos

Establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo entre diferentes contextos

Servir de base para la construcción de lenguajes de representación del conocimiento
Para llevar a cabo esto es necesario tener claro los conceptos de los lenguajes de programación que se usan.

A continuación se describen los lenguajes de programación utilizados en el Proyecto.

RDF/OWL

RDF: Es un modelo estándar para el intercambio de datos en la Web. RDF tiene características que facilitan la fusión de datos, incluso si los esquemas subyacentes difieren, y soporta específicamente la evolución de esquemas en el tiempo sin necesidad de todos los consumidores de datos que pueden cambiar.

RDF amplía la estructura de enlaces de la Web para utilizar URI para nombrar la relación entre las cosas, así como los dos extremos de la conexión (esto se refiere generalmente como un "triple"). El uso de este modelo simple, que permite que los datos estructurados y semi-estructurados para ser mezclados, expusieron y compartieron a través de diferentes aplicaciones.

Esta estructura forma una vinculación dirigida, gráfica marcada, donde los bordes representan el enlace nombrado entre dos recursos, representados por los nodos del gráfico. Este punto de vista gráfico es el modelo mental más fácil posible para RDF y se utiliza a menudo en las explicaciones visuales fáciles de entender.¹⁹

OWL: El Lenguaje de Ontologías Web (OWL) está diseñado para ser usado en aplicaciones que necesitan procesar el contenido de la información en lugar de únicamente representar información para los humanos. OWL facilita un mejor mecanismo de interpretabilidad de contenido Web que los mecanismos admitidos por XML, RDF, y esquema RDF (RDF-S) proporcionando vocabulario adicional junto con una semántica formal. OWL tiene tres sublenguajes, con un nivel de expresividad creciente: OWL Lite, OWL DL, y OWL Full.²⁰

Clases: Una clase define un grupo de individuos que pertenecen a la misma porque comparten algunas propiedades. Por ejemplo, Diego y Carlos son miembros de la clase Persona. Las clases pueden organizarse en una jerarquía de especialización usando *subClassOf*. Se puede encontrar una clase general llamada Thing que es una clase de todos los individuos y es una superclase de todas las clases de OWL. También se puede encontrar una clase general llamada Nothing que es la clase que no tiene instancias y es una subclase de todas las clases de OWL.

Instancias o individuales: Los individuos son instancias de clases, y las propiedades pueden ser usadas para relacionar un individuo con otro. Por ejemplo, un individuo llamado Diego puede ser descrito como una instancia de la clase Persona y la propiedad *Es_Estudiante_de* puede ser usada para relacionar el individuo Diego con el individuo *Ingenieria_de_Sistemas*.

RDF:Property: las propiedades pueden utilizarse para establecer relaciones de individuos a valores de datos. Ejemplos de propiedades son *Creada_por*, *Es_dictada_por*, *Asistida_por*,

¹⁹ Disponible en: <http://www.w3.org/RDF/>

²⁰ Disponible en: <https://ozone-development.github.io/ozone-website/owf.html>

y `Nacido_En`. Los tres primeros pueden utilizarse para relacionar una instancia de la clase `Persona` con otra instancia de la clase `Persona` (siendo casos de `ObjectProperty`), y el último (`Nacido_En`) puede ser usado para relacionar una instancia de la clase `Persona` con una instancia del tipo de datos `String` (siendo un caso de `DatatypeProperty`). Ambas, `owl:ObjectProperty` y `owl:DatatypeProperty`, son subclases de la clase de RDF `rdf:Property`

RDFS: domain: Un dominio de propiedad reduce los individuos a los que puede aplicarse la propiedad. Si una propiedad relaciona un individuo con otro individuo, y la propiedad tiene una clase como uno de sus dominios, entonces el individuo debe pertenecer a esa clase. Por ejemplo, puede establecerse que la propiedad `Es_La_Nacionalidad_De` tenga como dominio la clase `PAIS`. De esto, un razonador puede deducir que "Diego `Es_La_Nacionalidad_De` COLOMBIA". Obsérvese que *rdfs: domain* se denomina restricción global debido a que la restricción se refiere a la propiedad y no sólo a la propiedad cuando está asociada con una clase en concreto.

RDFS: range: El rango de una propiedad reduce los individuos que una propiedad puede tener como su valor. Si una propiedad relaciona a un individuo con otro individuo, y esta como rango a una clase, entonces el otro individuo debe pertenecer a dicha clase. Por ejemplo, la propiedad `Es_Asistida_Por` debe establecerse que tiene como rango la clase `MATERIA`. A partir de aquí, un razonador puede deducir que si `DIEGO` está relacionada con `COMPILADORES` mediante la propiedad `Es_Asistida_Por`, (por ejemplo, `Compiladores` es Asistida por `Diego`), entonces `Diego` es un `ALUMNO`. El rango es, al igual que el dominio, una restricción global.

Anotaciones: Las Anotaciones son las que facilitan la extracción y recuperación de la información, permite la reutilización de los corpus, a la vez que es multifuncional. Por ejemplo, la información sobre la clase gramatical de las palabras nos puede servir para la lexicografía, el análisis sintáctico, para elaborar listas de frecuencia y otras muchas aplicaciones. En el caso de este proyecto las anotaciones son utilizadas para la incursión de sinónimos.

OntGraph: Es utilizado para mostrar por medio de una gráfica la ontología con sus respectivas relaciones.

2.3.2 JAVA/JAVASCRIPT

JAVA: Java es un lenguaje de programación con el que podemos realizar cualquier tipo de programa. En la actualidad es un lenguaje muy extendido y cada vez cobra más importancia tanto en el ámbito de Internet como en la informática en general. Está desarrollado por la compañía Sun Microsystems con gran dedicación y siempre enfocado a cubrir las necesidades tecnológicas más punteras.

Una de las principales características por las que Java se ha hecho muy famoso es que es un lenguaje independiente de la plataforma. Eso quiere decir que si hacemos un programa en Java podrá funcionar en cualquier ordenador del mercado. Es una ventaja significativa para los desarrolladores de software, pues antes tenían que hacer un programa para cada sistema

operativo, por ejemplo Windows, Linux, Apple, etc. Esto lo consigue porque se ha creado una Máquina de Java para cada sistema que hace de puente entre el sistema operativo y el programa de Java y posibilita que este último se entienda perfectamente.

La independencia de plataforma es una de las razones por las que Java es interesante para Internet, ya que muchas personas deben tener acceso con ordenadores distintos. Pero no se queda ahí, Java está desarrollándose incluso para distintos tipos de dispositivos además del ordenador como móviles, agendas y en general para cualquier cosa que se le ocurra a la industria.

JAVASCRIPT: JavaScript es un lenguaje de scripting multiplataforma y orientado a objetos. Es un lenguaje pequeño y liviano. Dentro de un ambiente de host, JavaScript puede conectarse a los objetos de su ambiente y proporcionar control programático sobre ellos.

JavaScript contiene una librería estándar de objetos, tales como Array, Date, y Math, y un conjunto central de elementos del lenguaje, tales como operadores, estructuras de control, y sentencias. El núcleo de JavaScript puede extenderse para varios propósitos, complementándolo con objetos adicionales, por ejemplo:

Client-side JavaScript extiende el núcleo del lenguaje proporcionando objetos para controlar un navegador y su modelo de objetos (o DOM, por las iniciales de Document Object Model). Por ejemplo, las extensiones del lado del cliente permiten que una aplicación coloque elementos en un formulario HTML y responda a eventos del usuario, tales como clicks del ratón, ingreso de datos al formulario y navegación de páginas.

Server-side JavaScript extiende el núcleo del lenguaje proporcionando objetos relevantes a la ejecución de JavaScript en un servidor. Por ejemplo, las extensiones del lado del servidor permiten que una aplicación se comunique con una base de datos, proporcionar continuidad de la información de una invocación de la aplicación a otra, o efectuar manipulación de archivos en un servidor

JavaScript y Java son similares en algunos aspectos, pero fundamentalmente diferentes en otros. El lenguaje JavaScript se parece a Java pero no tiene el tipo estático (static) de Java, ni tiene un chequeo de tipos fuerte. JavaScript usa la mayoría de la sintaxis de expresiones de Java, convenciones de nombrado, y las construcciones básicas de control de flujo, razón por la cual se le cambió el nombre del original LiveScript a JavaScript.

En contraste con el sistema de clases construidas por declaraciones que se usa en tiempo de compilación de Java, JavaScript soporta un sistema de tiempo de ejecución basado en un pequeño número de tipos de datos que representan valores numéricos, lógicos, y de cadena de caracteres (string). JavaScript tiene un modelo de objetos basado en prototipos en lugar del modelo de objetos basado en clases, que es más común. El modelo basado en prototipo proporciona herencia dinámica; esto es, que lo que se hereda puede variar entre objetos individuales. JavaScript también soporta funciones sin ningún requerimiento declarativo especial. Las funciones pueden ser propiedades de los objetos, ejecutándose como métodos

levemente tipados.

Comparado con Java, JavaScript es un lenguaje muy libre de forma. No hay que declarar todas las variables, clases, y métodos. No hay que preocuparse de si los métodos son públicos, privados, o protegidos, y no hay que implementar interfases. Las variables, parámetros, y retornos de funciones no tienen que declararse explícitamente de un tipo dado.

Java es un lenguaje de programación basado en clases, diseñado para lograr seguridad de tipos, y ejecución rápida. Seguridad de tipos significa, por ejemplo, que no es posible hacer una conversión de tipos a un Integer de Java para obtener una referencia a un objeto, o acceder a memoria protegida mediante la alteración del bytecode de una clase. El modelo basado en clases de Java implica en que los programas consisten exclusivamente en clases y sus métodos. La herencia de clases de Java y el tipado fuerte generalmente requiere que jerarquías de objetos fuertemente acopladas.

En contraposición, JavaScript proviene en espíritu de una línea de lenguajes de programación más pequeños, con tipado dinámico, como HyperTalk, y dBASE. Estos lenguajes de scripting hacen accesibles las herramientas de programación a audiencias mucho más amplias. Esto se debe a su sintaxis más indulgente, funcionalidad especializada ya incluida, y mínimos requisitos para la creación de objetos.²¹

2.3.3 Frameworks

Además de los lenguajes de programación también necesitamos de los entornos de desarrollo y las librerías que se describen a continuación.

JENA. Es un framework Java para construir aplicaciones basadas en ontologías.

Su Arquitectura incluye:

API para trabajar (leer, procesar, escribir) ontologías RDF y OWL.

Motor de inferencia para razonar sobre ontologías RDF y OWL.

Estrategias de almacenamiento flexible para almacenar tripletas RDF en memoria o fichero.

Motor de queries compatible con especificación SPARQL²²

²¹ Disponible en: https://www.java.com/es/download/faq/whatis_java.xml

²² Disponible en: <https://jena.apache.org/>

SPARQL. Es un lenguaje de consulta para RDF, y actualmente está siendo considerada como una recomendación del W3C para crear un lenguaje de consulta dentro de la Web semántica que está ya implementada en varios lenguajes y bases de datos. Desde 2005 está en proceso de estandarización por el RDF Data Access Working Group (DAWG) del W3C; y en junio del 2007 se anunció el paso de su especificación a Candidate Recommendation.²³

HTML/JSP

HTML: HTML es la sigla de HiperText Markup Language (Lenguaje de Marcación de Hipertexto) es un lenguaje que se utiliza comúnmente para establecer la estructura y contenido de un sitio web, tanto de texto, objetos e imágenes. Los archivos desarrollados en HTML usan la extensión .htm o .html.

JSP: JSP es un acrónimo de Java Server Pages, que en castellano vendría a decir algo como Páginas de Servidor Java. Es, pues, una tecnología orientada a crear páginas web con programación en Java.

Motor JSP. El motor de las páginas JSP está basado en los servlets de Java -programas en Java destinados a ejecutarse en el servidor-, aunque el número de desarrolladores que pueden afrontar la programación de JSP es mucho mayor, dado que resulta mucho más sencillo aprender que los servlets.

En JSP creamos páginas de manera parecida a como se crean en ASP o PHP. Generamos archivos con extensión .jsp que incluyen, dentro de la estructura de etiquetas HTML, las sentencias Java a ejecutar en el servidor. Antes de que sean funcionales los archivos, el motor JSP lleva a cabo una fase de traducción de esa página en un servlet, implementado en un archivo class (Byte codes de Java). Esta fase de traducción se lleva a cabo habitualmente cuando se recibe la primera solicitud de la página .jsp, aunque existe la opción de precompilar en código para evitar ese tiempo de espera la primera vez que un cliente solicita la página.²⁴

API's y librerías. (Application Program Interface). Conjunto de convenciones internacionales que definen cómo debe invocarse una determinada función de un programa desde una aplicación. Cuando se intenta estandarizar una plataforma, se estipulan unos APIs comunes a los que deben ajustarse todos los desarrolladores de aplicaciones. Herramientas de programación para rutinas, protocolos y software.

SLF4J: Sirve como una fachada simple o abstracción para diversos marcos de registro (por ejemplo java.util.logging, logback, log4j) que permite al usuario final conectarse en el marco de registro deseado.

²³ Disponible en: <http://www.w3.org/TR/rdf-sparql-query/>

²⁴ <http://www.it.uc3m.es/labttlat/2007-08/material/jsp/>

JWNL: JWNL es una API de Java para acceder al diccionario relacional WordNet. WordNet es ampliamente utilizado para el desarrollo de aplicaciones de PLN.

MySQL Connector: Es la librería de java que permite la conexión con MySQL para la gestión de base de datos por medio de métodos propios de la librería que permiten consultar, actualizar y eliminar registros de una base de datos.

JSON: JSON (JavaScript Object Notation - Notación de Objetos de JavaScript) es un formato ligero de intercambio de datos. Leerlo y escribirlo es simple para humanos, mientras que para las máquinas es simple interpretarlo y generarlo. Está basado en un subconjunto del Lenguaje de Programación JavaScript, Standard ECMA-262 3rd Edition - Diciembre 1999.

JSON es un formato de texto que es completamente independiente del lenguaje pero utiliza convenciones que son ampliamente conocidos por los programadores de la familia de lenguajes C, incluyendo C, C++, C#, Java, JavaScript, Perl, Python, y muchos otros. Estas propiedades hacen que JSON sea un lenguaje ideal para el intercambio de datos.

Motor de búsqueda web. Un motor de búsqueda (también llamado Searchbot) es una herramienta hardware y software que indexa páginas Web para que se puedan buscar a través de palabras claves en un formulario de búsqueda.

Sistema de gestión de base de datos. En sistema de gestión de bases de datos relacionales (RDBMS) es un programa que te permite crear, actualizar y administrar una base de datos relacional. La mayoría de los RDBMS comerciales utilizan el lenguaje de consultas estructuradas (SQL) para acceder a labase de datos, aunque SQL fue inventado después del desarrollo del modelo relacional y no es necesario para su uso.²⁵

Sistema de gestión de base de conocimientos. Un sistema de gestión de base de conocimientos es una aplicación que permite la creación e interacción con archivos en formatos como RDF-OWL que contienen las ontologías, permitiendo la gestión de la información por medio de consultas a las tripletas o metadatos a través del lenguaje de consulta SparQL.

2.3.4 Herramientas de trabajo o IDE's. Los programas que se usaron para el desarrollo del Proyecto fueron los siguientes:

NetBeans. Es un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. Existe además un número importante de módulos para extenderlo. NetBeans IDE es un producto libre y gratuito sin restricciones de uso.

²⁵ Disponible en: <http://es.ccm.net/contents/830-web-motor-de-busqueda>

Protégé. Es un editor de código abierto usado para construir Ontologías y un marco general para representar el conocimiento.σ Está escrito en Java que es un lenguaje de programación orientado a objetos.σ Se usa para construir aplicaciones para la Web Semántica, para hacer una descripción semántica de la información.σ Sus archivos se hacen en el lenguaje OWL.

MySql. Es un sistema de gestión de bases de datos relacional, multi-hilo y multiusuario con más de seis millones de instalaciones.

Wordnet (Base de datos léxica). Es una base de datos léxica del Idioma Inglés. Agrupa palabras en inglés en conjuntos de sinónimos llamados synsets, proporcionando definiciones cortas y generales, y almacena las relaciones semánticas entre los conjuntos de sinónimos.²⁶

2.3.5 Servidores. Los servidores son programas de computadora en ejecución que atienden las peticiones de otros programas, los clientes. Por tanto, el servidor realiza otras tareas para beneficio de los clientes. Ofrece a los clientes la posibilidad de compartir datos, información y recursos de hardware y software.

Apache TomCat. Apache Tomcat funciona como un contenedor de servlets desarrollado bajo el proyecto Jakarta en la Apache Software Foundation. Tomcat implementa las especificaciones de los servlets y de JavaServer Pages de Sun Microsystems.²⁷

2.4 MARCO LEGAL

Este trabajo se realiza bajo la licencia Creative Commons. Reconocimiento no comercial.

Compartir bajo la misma licencia 2.5 Colombia, usted es libre de copiar, distribuir y comunicar públicamente la obra, o hacer otros trabajos derivados bajo las siguientes condiciones:

Reconocimiento: Debe reconocerse los créditos de la manera especificada por los autores. (Sin insinuar que tienen su apoyo o apoyan el uso de su obra).

No comercial: No se puede utilizar esta obra bajo fines comerciales.

Compartir bajo la misma licencia: Si se altera o se transforma esta obra, o genera una obra derivada, solo se puede distribuir la obra generada bajo una licencia idéntica a esta. Al reutilizar o distribuir la obra, se debe dejar claro los términos de la licencia de esta obra. Algunas de las condiciones pueden no aplicarse si hay el permiso de titular de los derechos de autor. Nada en esta licencia menoscaba o restringe los derechos morales del autor.

Lo anterior fue un resumen de fácil entendimiento del texto legal. Para ver la licencia

²⁶ Disponible en: <http://www.batanga.com/tech/13220/los-mejores-ides-y-editores-para-programadores>

²⁷ Disponible en: <http://tomcat.apache.org/>

completa se puede visitar:

<http://creativecommons.org/licenses/by-ncsa/2.5/co/legalcode>

El Software como producto final, está licenciado bajo GNU GPL v3 y por ende es software libre, puede ser restringido y/o modificado bajo los términos de la GNU General Public License como fue publicada por la Free Software Foundation, versión 3.

Una copia de la licencia puede ser vista en:

<http://www.gnu.org/licenses/>

El Software como producto final hace uso de trabajo de terceros bajo diferentes licencias, en el documento se expone la lista de librerías usadas o trabajos de otras personas.

Además se tendrán en cuenta los artículos de la Ley 1607 Reforma Tributaria 2012, que tienen incidencia con la Industria Software en Colombia:

Artículo 161. Prorrogase la vigencia del artículo 207-2 numeral 8 del Estatuto Tributario, respecto de la producción de software nacional, por el término de cinco (5) años, contados a partir del primero de enero de 2013. El certificado exigido en dicha norma será expedido por el Consejo Nacional de Beneficios Tributarios en Ciencia, Tecnología e Innovación, creado por el artículo 34 de la Ley 1450 de 2011, el cual también estará integrado por el Ministro de las Tecnologías de la Información y las Comunicaciones o su representante. El Gobierno reglamentará el funcionamiento del Consejo.

Artículo 481. Bienes exentos con derecho a devolución bimestral. Para efectos del Impuesto sobre las ventas, únicamente conservarán la calidad de bienes y servicios exentos con derecho a devolución bimestral:

Los servicios que sean prestados en el país y se utilicen exclusivamente en el exterior por empresas o personas sin negocios o actividades en Colombia, de acuerdo con los requisitos que señale el reglamento. Quienes exporten servicios deberán conservar los documentos que acrediten debidamente la existencia de la operación. El Gobierno Nacional reglamentará la materia.

Los servicios de conexión y acceso a internet desde redes fijas de los suscriptores residenciales de los estratos 1 y 2.

Parágrafo. Sin perjuicio de lo establecido en el literal c de este artículo, se entenderá que existe una exportación de servicios en los casos de servicios relacionados con la producción de cine y televisión y con el desarrollo de software, que estén protegidos por el derecho de autor, y que una vez exportados sean difundidos desde el exterior por el beneficiario de los mismos en el mercado internacional y a ellos se pueda acceder desde Colombia por cualquier medio tecnológico.

Artículo 50. 15. Los servicios de conexión y acceso a internet de los usuarios residenciales del estrato 3.²⁸

²⁸ Disponible en: <http://co.creativecommons.org/>

3. DISEÑO METODOLÓGICO

3.1 TIPO DE INVESTIGACIÓN

Es una Investigación cuantitativa porque se usan algunos conceptos de semántica computacional y bases de conocimiento que pueden ser medidos precisamente para describirlos, ubicando propiedades importantes y delimitando los hechos que conforman el problema a solucionar, con información que será analizada en cuanto a la consecución de la población y las características que se exponen de la misma, para probar que la hipótesis de un Sistema Inteligente de Consulta de OVAS, propuesta por el grupo de investigadores, es la mejor solución para encontrar las respuestas producidas de las consultas de las personas que solo usan el lenguaje natural.

3.2 POBLACIÓN Y MUESTRA

La población que se tendrá en cuenta está conformada por las personas del área de Postgrados de la UFPSO, distribuidos en las categorías: Estudiantes, Docentes y Administrativos.

Estudiantes

Docentes

Administrativos

Un total de 36 personas. En algunos casos, un estudiante es también un administrativo, sin embargo aunque se tiene en cuenta su respuesta, puede darse que estos administrativos no necesariamente sean del área de postgrados.

La muestra tomará el valor de acuerdo a una ecuación. Debido a que la población es pequeña y que en algunos casos se rehúsan de responder la encuesta por motivos como falta de conocimiento del tema o simplemente porque es difícil acceder a ellos por su horario laboral, para determinar con más exactitud los datos recolectados por las encuestas y entrevistas, hemos estimado el error de muestra del 10.5%.

A continuación detallamos las variables y las fórmulas que consideramos para recolectar los datos:

Población: Representa el total de personas que hacen parte (N) = 36 Personas.

Nivel de confianza: Porcentaje de la certeza de los resultados de la investigación. (C) = 90%.

Constante dependiente del nivel de confianza de las respuestas: (K) = 1.65.

A continuación una tabla representativa para hallar la constante (K):

Ilustración 3: Constante "K" y Nivel de confianza

K	1,15	1,28	1,44	1,65	1,96	2	2,58
Nivel de confianza	75%	80%	85%	90%	95%	95,5%	99%

Fuente. Autores del proyecto

Error de muestra: Diferencia entre los resultados obtenidos y los posibles si el total de la población resuelve la encuesta (E) = 10.5%.

Proporción de individuos que poseen las características de estudio: (P) = 50%.

Proporción de individuos que no poseen las características de estudio: (Q) = 50%.

Para hallar la muestra se usa la siguiente ecuación apoyada en los valores antes expuestos:

$$M = \frac{K^2 * P * Q * N}{(E^2 * (N - 1)) + K^2 * P + Q}$$

Muestra: (M) = 23 Personas.

3.3 RECOLECCIÓN DE DATOS

Para la recolección de la información, las técnicas usadas fueron la encuesta y la entrevista, además de una revisión de la documentación.

La encuesta fue diseñada con un total de 16 Preguntas, que reunirán datos estadísticos representados en gráficas para su posterior análisis.

La encuesta va dirigida a gente con al menos conocimientos básicos de Ingeniería de Sistemas y Diseño de Software, abarcando temas como: Buscadores Semánticos, Procesamiento del Lenguaje Natural, Gestión de la Información (OVAS en este caso), para observar el conocimiento de los usuarios sobre el tema, conocer el nivel de aceptación de la aplicación y evidenciar si hay antecedentes de sistemas inteligentes de manejo de la información.

Para conocer más detalles antes de la aplicación de la encuesta se realizó una entrevista al Ing. Andrés Mauricio Puentes Velásquez, que cuenta con conocimientos de Inteligencia Computacional y construcción de Ontologías, además de la gestión de la información apuntando a la web semántica.

El modelo de entrevista tiene 3 apartados, el primer apartado consta de 6 preguntas y cada uno de los otros dos apartados de 5 preguntas:

Apartado 1: Preguntas abiertas para fortalecer el tema. Estas preguntas permitirán conocer

aspectos o pensamientos un poco más personales del experto, opiniones o pensamientos propios que servirán para la comprensión de sus respuestas.

Apartado 2: Preguntas cerradas con dos opciones de respuesta (**SI/NO**). Estas preguntas encaminarán la propuesta sobre dos alternativas que confirmarán o rechazarán la pregunta de acuerdo a las experiencias del experto.

Apartado 3: Preguntas cerradas con más de dos opciones de respuesta. Estas preguntas ayudarán a la categorización de ciertos aspectos de la propuesta y de la situación actual de la investigación.

Esto tiene como propósito obtener en los resultados:

Mejora de la calidad y precisión de los resultados de las búsquedas de material educativo y ahorro de tiempo para el usuario.

Fácil Interacción de los contenidos y la aplicación con diseño llamativo que genere interés al usuario por el uso de los materiales bibliográficos.

La revisión de la documentación existente fue enfocada a la descarga de los archivos y configuración de las interfaces, IDE, variables de entorno y de las aplicaciones necesarias para el funcionamiento sincronizado para el desarrollo del Sistema Inteligente como producto final.

3.4 ANÁLISIS DE LA INFORMACIÓN

Con el propósito de mostrar de manera cuantitativa los resultados obtenidos de esta investigación, para demostrar estadísticamente la necesidad que se presente con miras a ser pioneros de la gestión de la información se realizó un análisis que se llevó a cabo en 3 fases:

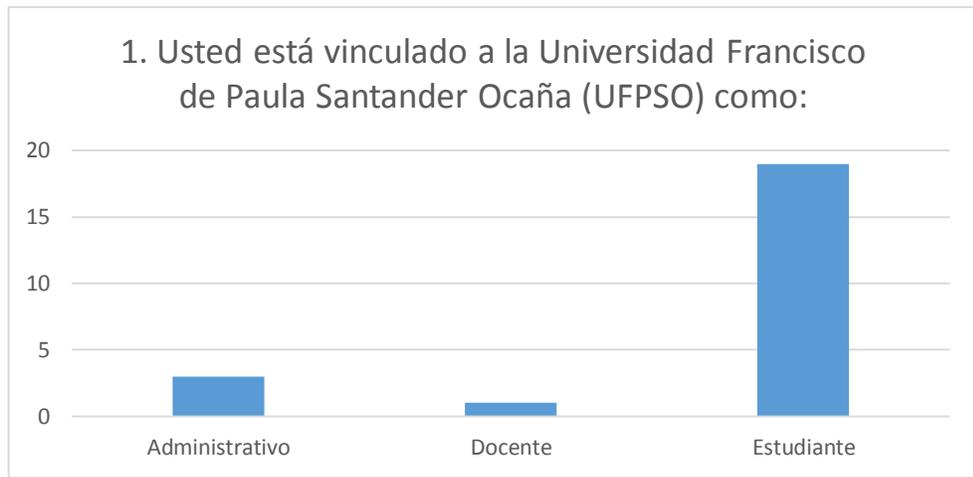
Recolección y distribución de los datos en tablas.

Representación gráfica de la información obtenida.

Conclusiones e interpretación de los datos obtenidos en las gráficas y las tablas.

A continuación se muestran los resultados en las tablas con sus respectivas gráficas, organizados por preguntas:

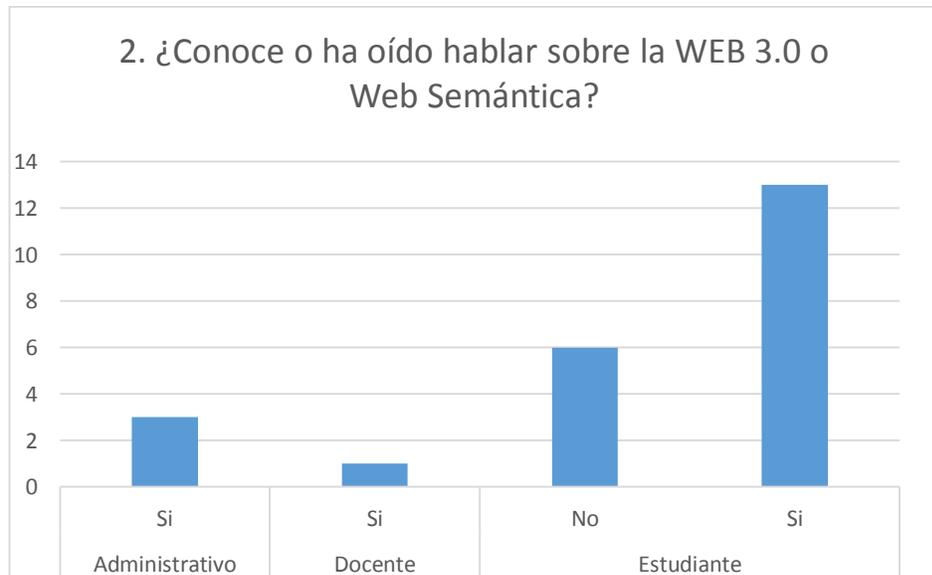
Ilustración 4: Grafica a la pregunta 1 de encuesta



Fuente. Autores del proyecto

Conclusión: En el área de postgrados la mayoría de personas son estudiantes.

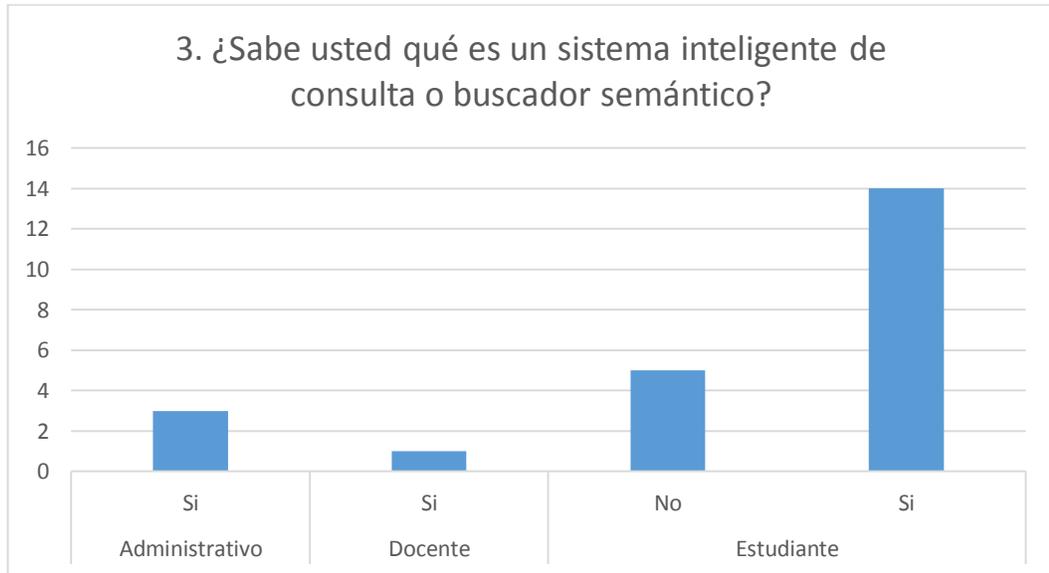
Ilustración 5: Grafica a la pregunta 2 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos tiene conocimiento acerca de la Web Semántica.

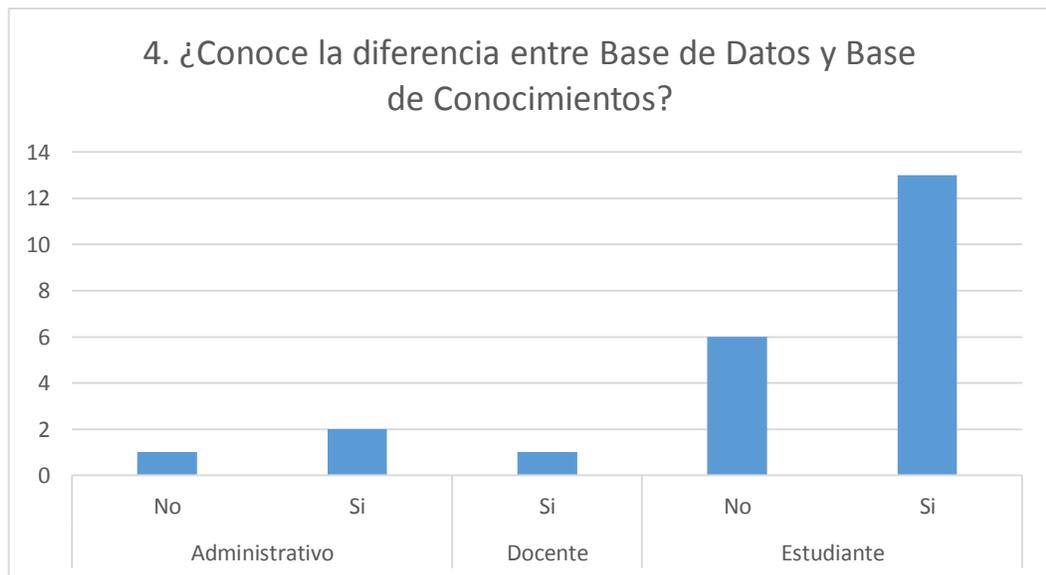
Ilustración 6: Grafica a la pregunta 3 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos tienen conocimientos sobre que es un buscador semántico.

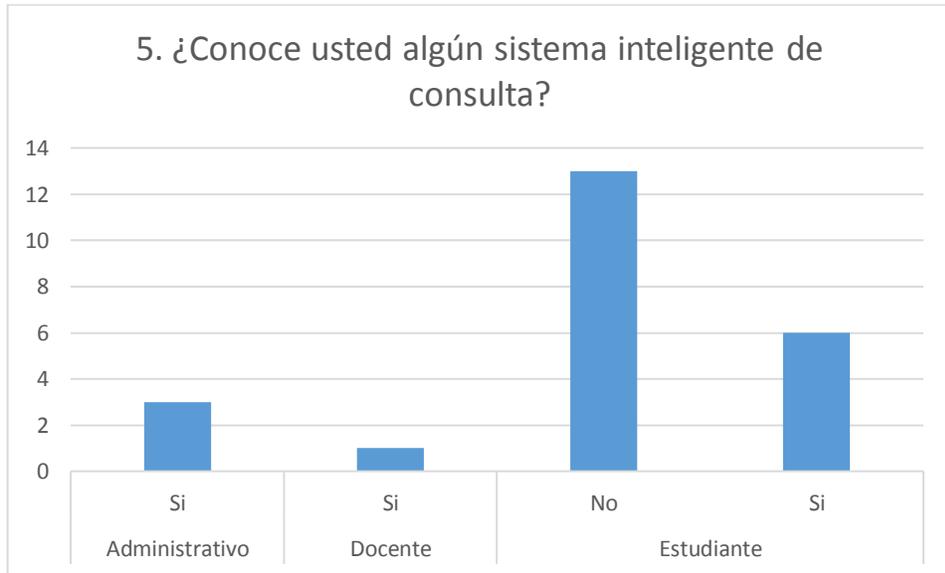
Ilustración 7: Grafica a la pregunta 4 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos saben la diferencia entre base de datos y base de conocimientos.

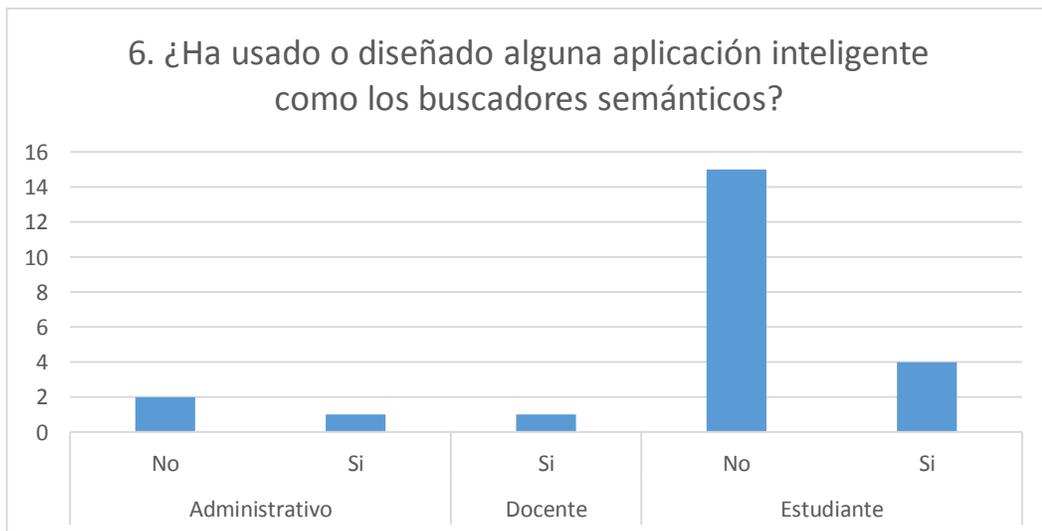
Ilustración 8: Grafica a la pregunta 5 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos no conocen ningún sistema inteligente de consultas.

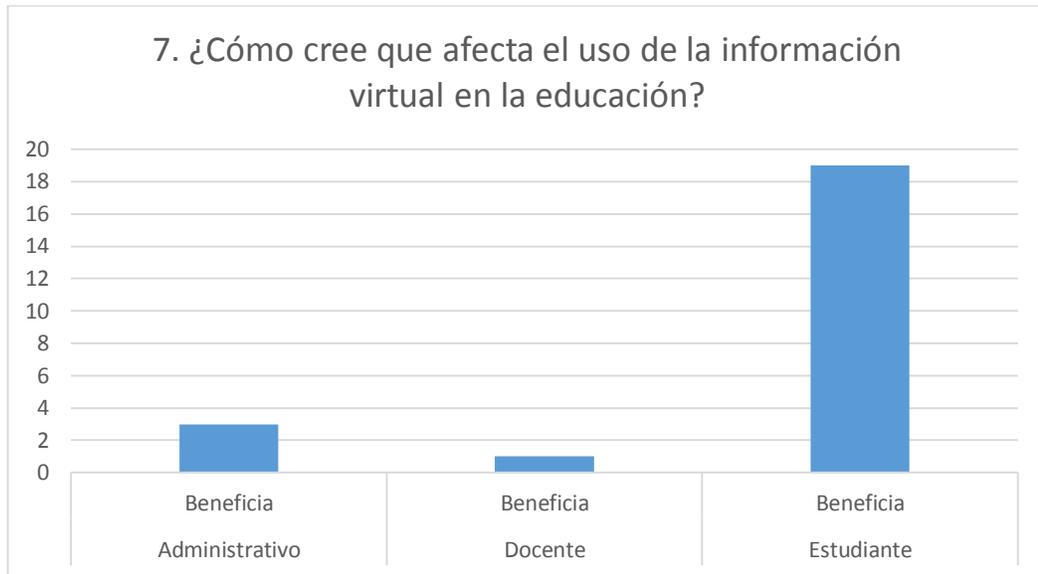
Ilustración 9: Grafica a la pregunta 6 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos no ha usado ni diseñado ninguna aplicación inteligente parecida a un buscador semántico.

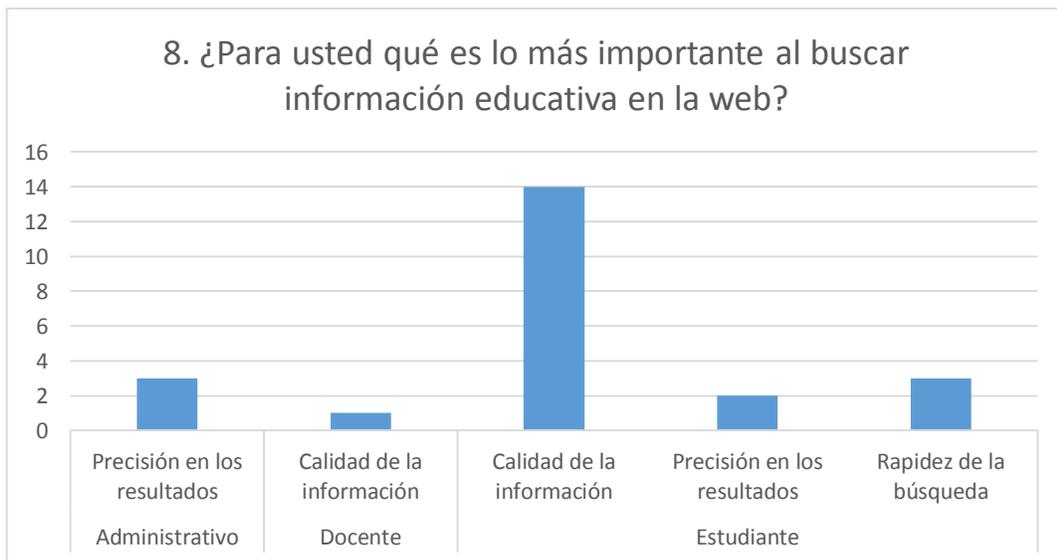
Ilustración 10: Grafica a la pregunta 7 de encuesta



Fuente. Autores del proyecto

Conclusión: Todas las personas entre estudiantes, docentes y administrativos consideran benéfica la incursión de la información virtual en la educación.

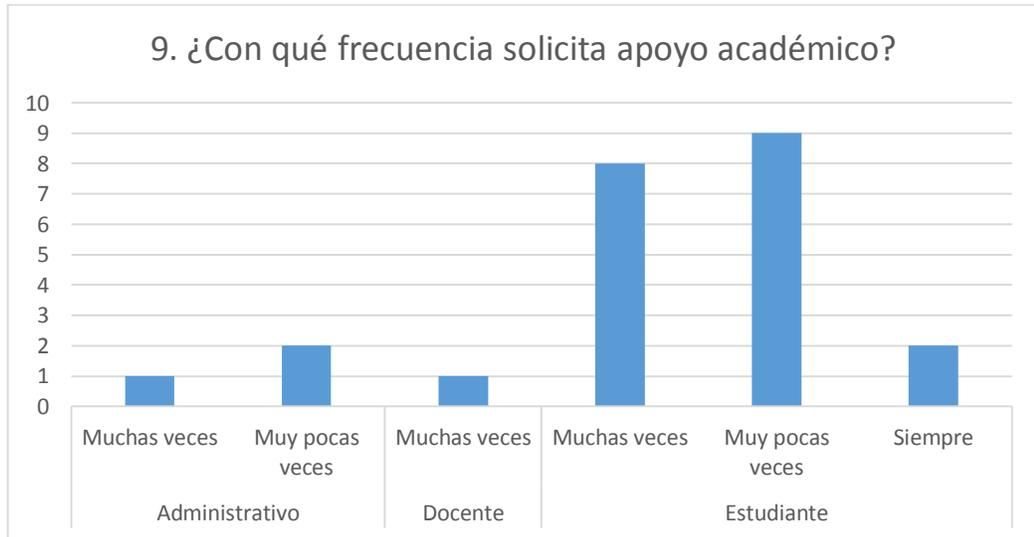
Ilustración 11: Grafica a la pregunta 8 de encuesta



Fuente. Autores del proyecto

Conclusión: Se puede observar que los factores importantes que consideran los encuestados son la calidad de la información y la precisión de los resultados, además que la rapidez de la búsqueda también es importante al buscar información educativa en la web.

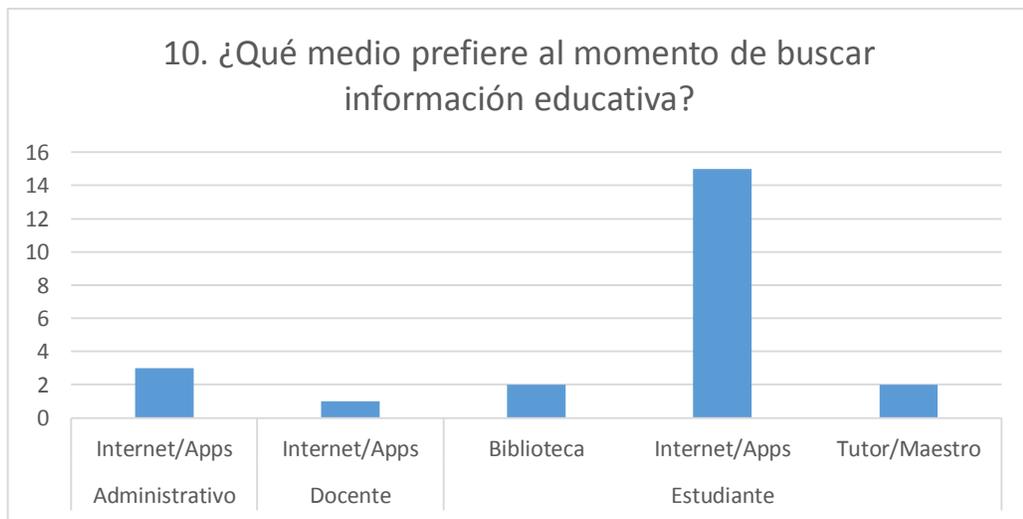
Ilustración 12: Grafica a la pregunta 9 de encuesta



Fuente. Autores del proyecto

Conclusión: De esta pregunta se puede analizar que la mayoría de los administrativos encuestados muy pocas veces solicitan apoyo académico, lo que refleja que ellos no tienen esa necesidad debido a su cargo. Los docentes solicitan muchas veces apoyo académico lo cual es obvio, ya que son ellos los que ofrecen la educación. Por otro lado, de los estudiantes no se puede decir mucho porque la diferencia no es muy marcada, esto permite concluir que en ellos depende es el interés de la persona al solicitar los recursos académicos.

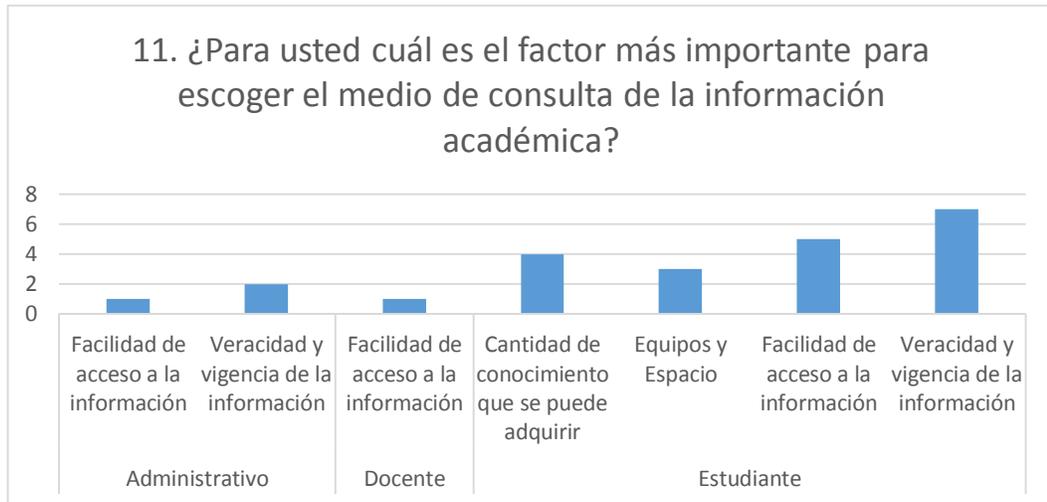
Ilustración 13: Grafica a la pregunta 10 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos prefieren usar la Internet y aplicaciones a la hora de buscar información educativa.

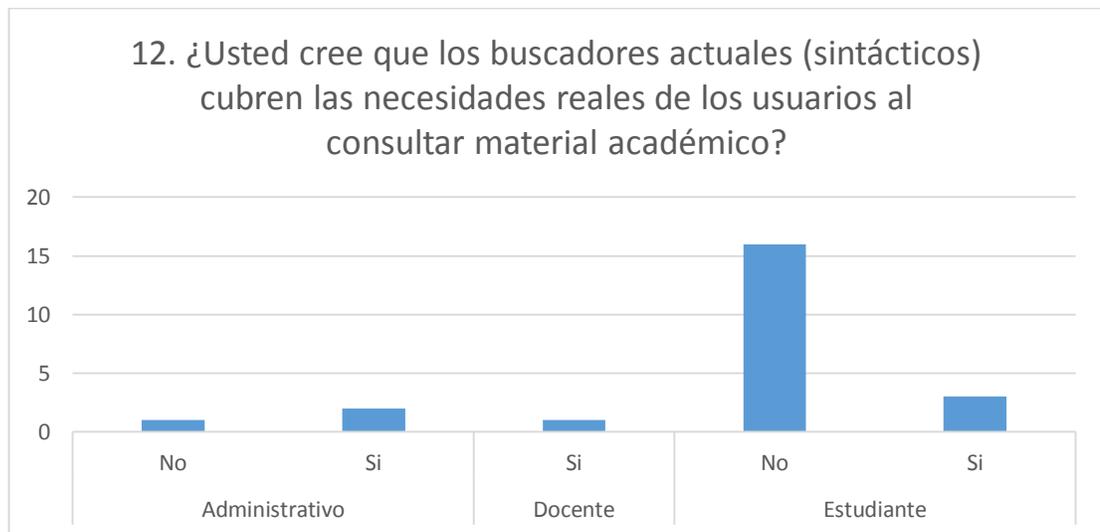
Ilustración 14: Grafica a la pregunta 11 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos considera que los factores más importantes a la hora que escoger un medio de consulta académica son la veracidad y vigencia de la información además de la facilidad de acceso a ella.

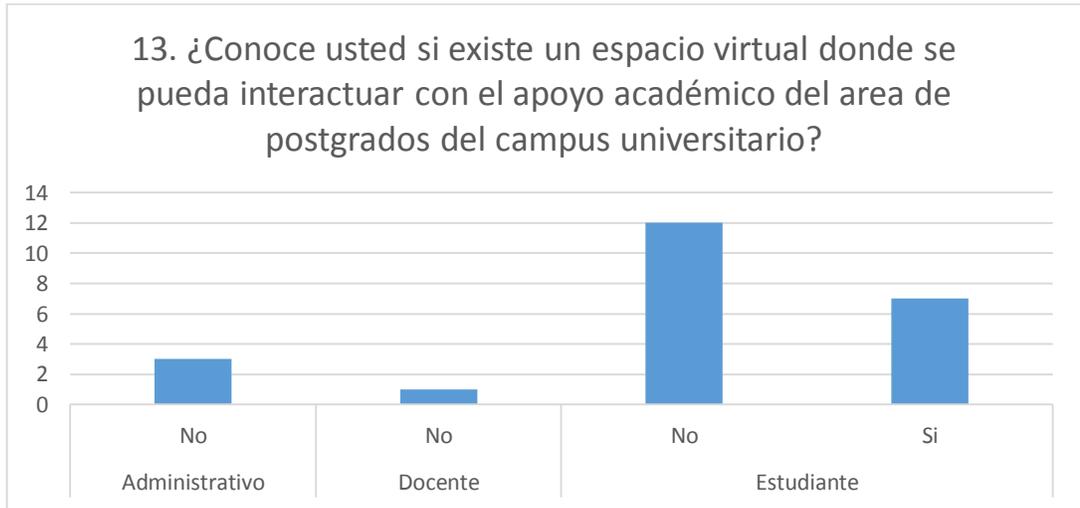
Ilustración 15: Grafica a la pregunta 12 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos piensan que los buscadores sintácticos actuales no cubren sus necesidades a la hora de consultar material académico.

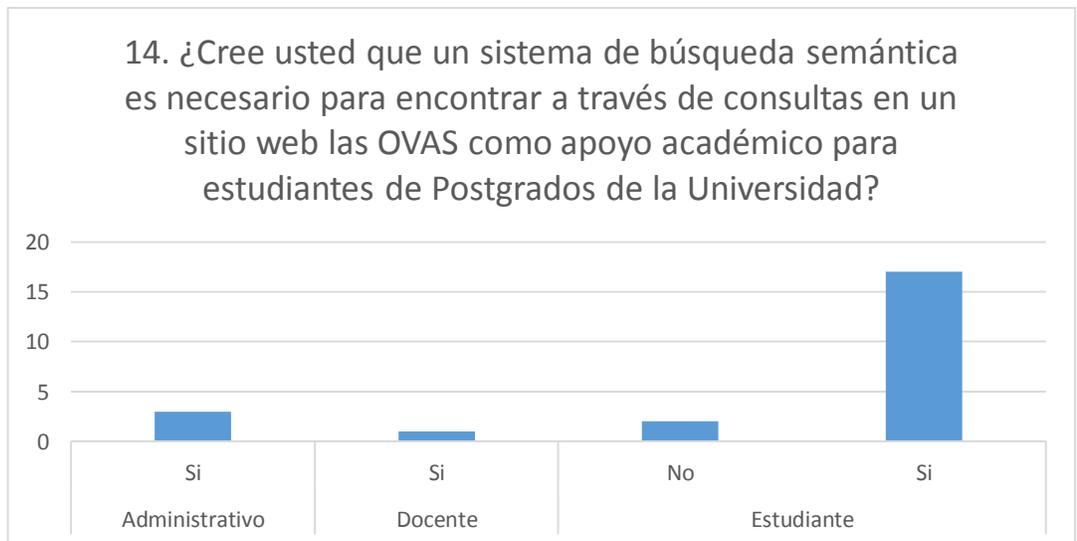
Ilustración 16: Grafica a la pregunta 13 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos no conocen un espacio virtual donde se pueda interactuar con apoyo académico del área de postgrados dentro de la universidad.

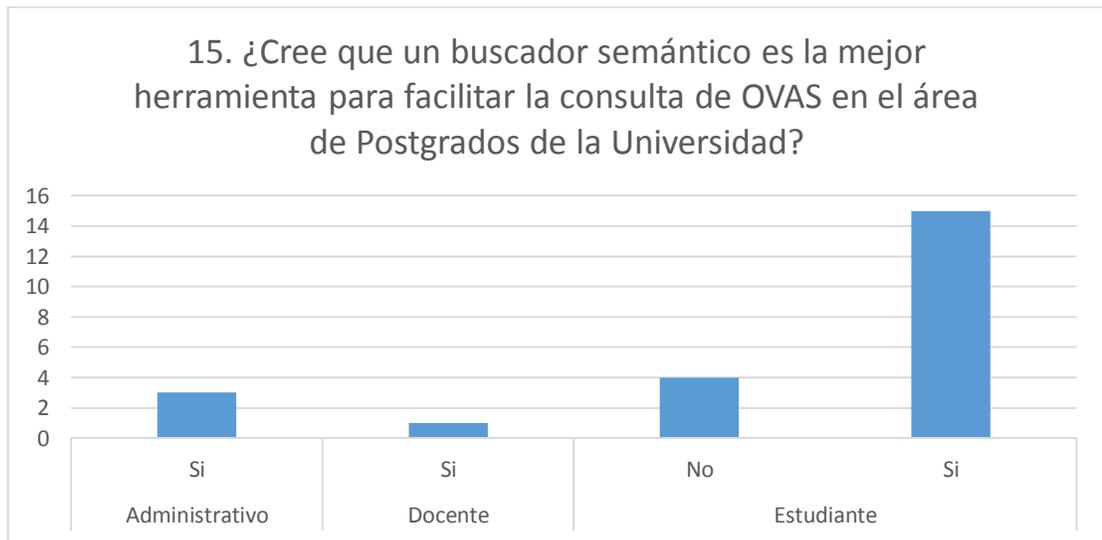
Ilustración 17: Grafica a la pregunta 14 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos considera necesario un sistema de búsquedas para encontrar OVAS a través de un sitio web como apoyo académico para estudiantes de postgrados de la Universidad.

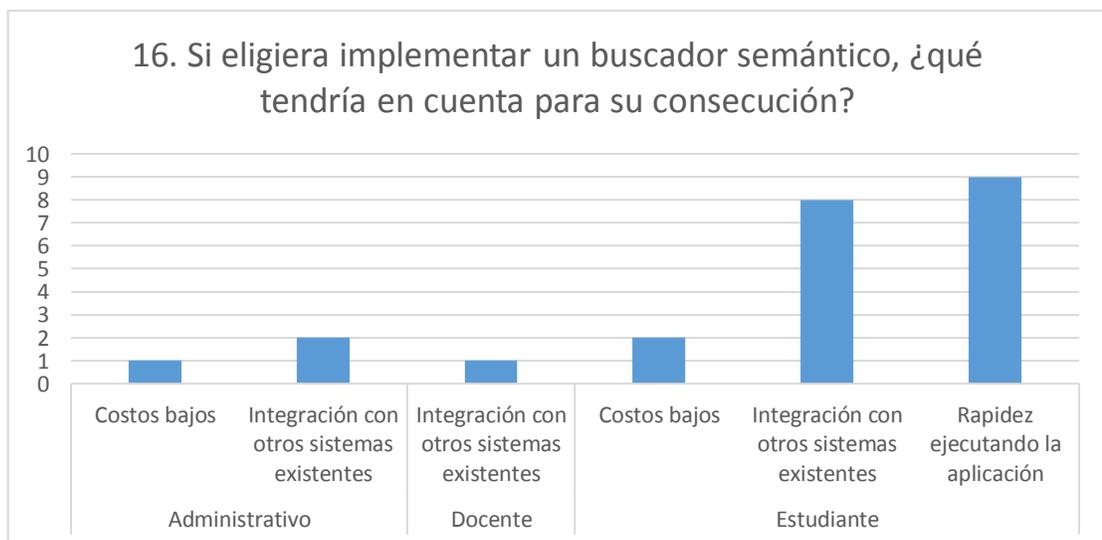
Ilustración 18: Grafica a la pregunta 15 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos cree que un buscador semántico es la mejor herramienta para facilitar la consulta de OVAS en el área de postgrados de la Universidad.

Ilustración 19: Grafica a la pregunta 16 de encuesta



Fuente. Autores del proyecto

Conclusión: La mayoría de personas entre estudiantes, docentes y administrativos a la hora de implementar un buscador semántico lo que más tendrían en cuenta sería la rapidez de ejecución de la aplicación y la integración con otros sistemas.

3.5 METODOLOGÍA DEL DESARROLLO DE UN BUSCADOR SEMÁNTICO COMO UN SISTEMA INTELIGENTE PARA LAS CONSULTAS DE OVAS

3.5.1 Investigación de teoría y conceptos. Como primer paso de la investigación se necesita un fundamento teórico suficiente sobre el cual sostener el desarrollo de la misma, la investigación inicio con la adquisición de información teórica sobre la web semántica y las tecnologías que la conforman, como son las ontologías y sus formas de construcción. Al mismo tiempo se va depositando la información adquirida en los marcos teóricos y conceptuales donde se explican y definen tanto los fundamentos teóricos del proyecto como los conceptos involucrados en este.

3.5.2 Modelamiento del dominio del conocimiento. La base del desarrollo de un sistema que pretende incluir cierto grado de semántica es un correcto moldeamiento del dominio del conocimiento, esto se logra a través de la recolección de documentos que contengan los conceptos relacionados al tema, realizar una extracción manual de estos y encontrar también las formas que estos se relacionan entre sí. El resultado de este proceso es una ontología rudimentaria que poco a poco se ira refinando.

3.6 SEGUIMIENTO METODOLÓGICO DEL PROYECTO

Cuadro 20: Seguimiento Metodológico

Desarrollar una aplicación inteligente que permita gestionar los objetos digitales de aprendizaje de la especialización en auditoría de sistemas de la UFPSO, optimizando las consultas			
OBJETIVO ESPECÍFICO	NOMBRE	ACTIVIDAD	INDICADOR
Abstraer la información y el conocimiento del dominio a través de artefactos de ingeniería de software, modelos de especificación de requisitos e interfaces de usuario.	Fundamentación Teórica y Metodología (Identificación)	Reunión Con los Expertos	-Marco Referencial.
		Determinación del Objeto de Estudio	-Doc. Anexo: Encuesta.
		Recolección de datos e información	-Doc. Anexo: Entrevista.
	Estructuración de los Conceptos en base al proyecto (Conceptualización)	Revisión y Elaboración del Marco Teórico, Marco Legal y	-Doc. Anexo: Construcción de Componente de Persistencia “Base de Conocimiento”.
			-Doc. Anexo: Especificación de

Cuadro 1. (Continuación)

	n)	Estado del Arte Realización de encuesta y entrevista a la población y a los expertos Construcción de un Glosario de términos	Caso de Uso. - Diagramas incluidos en este documento.
Diseñar los componentes de persistencia, tales como bases de Datos y bases de conocimiento.	Diseño y elaboración de prototipos y proyecto total (Construcción)	Construcción de Ontologías Elaboración de Base de Datos Construcción de Interfaz y código fuente Mejoramiento de detalles estéticos	-Archivo de las Ontologías como Base de Conocimientos. -Cód. SQL de la Base de Datos -Cód. Fuente de la aplicación.
Implementar las funcionalidades del sistema empleando metodologías de sistemas basados en conocimiento, probando el acceso a los recursos virtuales de aprendizaje disponibles en la especialización en auditoría de sistemas.	Experimentación y Evaluación del Producto (Pruebas)	Procesamiento de los datos recolectados y producción de hipótesis para los problemas que surgen Prueba de errores y optimización del producto Elaboración de un informe final con los resultados obtenidos	-Aplicación Desarrollada. -Doc. Anexo: Informe de los resultados.

4. RESULTADOS

Con el fin de lograr los objetivos planteados al inicio de este proyecto, luego de tabular la información recolectada por las encuestas y la entrevista, se hizo un análisis de los resultados de la aplicación. Parece factible que se generen páginas semánticas sin demasiado esfuerzo cuya información formalmente anotada pueda ser explotada por las aplicaciones semánticas. Esto es lo que hemos procurado visionar.

Se han analizado las debilidades, amenazas, fortalezas y oportunidades de la Web Semántica respecto de la Web actual. La oportunidad principal para implantar la Web Semántica es, en realidad, la debilidad de la Web actual, pero indicábamos que esto no sería posible si no se disminuye la complejidad de la tecnología, que es la principal debilidad de la Web Semántica para que los usuarios activos pudieran adaptarse a ella.

En este último capítulo vamos resumir las aportaciones para, a continuación, dar una valoración de las mismas y establecer las conclusiones correspondientes y además dar una propuesta para trabajos futuros que se pueden realizar en la UFPSO que surgen a partir de nuestra investigación.

4.1 MODELO DE ESPECIFICACIÓN DE REQUISITOS

De nuestra investigación construimos un documento con la especificación de las funcionalidades del sistema, los actores que interactúan y su jerarquía, y tablas de los requerimientos funcionales y no funcionales del software.

Esto nos permitió acceder más rápido a la información necesaria para la construcción de la Ontología, ya que la relación de los datos al momento de crear esta base de conocimientos debía suministrarse cuidadosamente para que no hubiese incoherencia en los resultados.

4.1.1 Requisitos funcionales y no funcionales del sistema. A continuación hay algunas tablas que muestran la información de requisitos funcionales y no funcionales obtenidos mediante el análisis de nuestro sistema.

Requisitos funcionales. En la siguiente tabla se exponen los requisitos funcionales para nuestro sistema:

Cuadro 21: Requisitos funcionales del sistema

RFUN01	
NOMBRE	Ingreso
DESCRIPCIÓN	El usuario puede ingresar al sistema con los datos académicos, debe estar previamente registrado como estudiante, profesor o administrativo. La autenticación se hará mediante la introducción de un nombre de usuario (Ejemplo: el código del estudiante), el documento y una contraseña.

Cuadro 2. (Continuación)

RFUN02	
NOMBRE	Salida
DESCRIPCIÓN	El usuario podrá salir de la sesión de búsqueda en cualquier momento o volver a formularios anteriores.
RFUN03	
NOMBRE	Consulta y PLN
DESCRIPCIÓN	El usuario podrá realizar una consulta en lenguaje natural, el sistema hace el procesamiento de frases necesario para interpretar la consulta para al final implementar las técnicas de búsqueda semántica. Las búsquedas a la Base de Datos se hacen automáticas para adherir más detalles.
RFUN04	
NOMBRE	Recuperación
DESCRIPCIÓN	Se deberá mostrar o visualizar los detalles de las OVAS consultadas en un formulario JSP donde se encuentren las coincidencias semánticas pertinentes.
RFUN05	
NOMBRE	Servidor
DESCRIPCIÓN	El servidor usado en esta aplicación inteligente es TomCat Apache.
RFUN06	
NOMBRE	Alimentación de Información
DESCRIPCIÓN	El administrador debe llenar las ontologías directamente desde un software de gestión de bases de conocimiento como Protégé. La gestión de base de datos (en especial la tabla "Usuario") también se hace desde MySQL.
RFUN07	
NOMBRE	Consulta de Sinónimos a WordNet
DESCRIPCIÓN	Se hace una consulta invisible a WordNet para solucionar ambigüedad de palabras.
RFUN08	
NOMBRE	Filtrado
DESCRIPCIÓN	Se borran palabras vacías que no afectan de manera significativa a la consulta. (Ejemplo: de, la, los, etc.)

Fuente. Autores del proyecto

4.1.2 Requisitos no funcionales. A continuación se exponen, en forma de tabla, los requisitos no funcionales para nuestro sistema:

Cuadro 22: Requisitos no funcionales del sistema

RNFUN01	
NOMBRE	Idiomas
DESCRIPCIÓN	El lenguaje usado para el uso de la aplicación será el español (Colombia). Sin embargo, para las consultas a WordNet para los sinónimos de las palabras, se hace una traducción automática al

Cuadro 3. (Continuación)

	inglés que el usuario no va a notar.
RNFUN02	
NOMBRE	Lenguaje de Programación
DESCRIPCIÓN	El Sistema se desarrolla en formularios JSP usando JavaScript, bajo controladores Servlets para el paquete Web. Para el almacenamiento de la información se usan los archivos en formato OWL/RDF que es el usado para las ontologías y para las bases de datos se usa MySQL. Por otra parte se usa JAVA para las siguientes funciones: con JENA para las consultas a las ontologías y con JWNL para las consultas a WordNet.
RNFUN03	
NOMBRE	Intuitivo
DESCRIPCIÓN	El sistema desarrollado ha de ser mostrado al usuario para que lo use de la manera más simple posible. En este caso solo basta con ingresar y hacer la consulta para recibir los resultados que se quieren.
RNFUN04	
NOMBRE	Escalabilidad
DESCRIPCIÓN	El sistema es de fácil modificación para administradores con conocimiento básico de bases de conocimiento y permite la ampliación de sus funciones para futuras mejoras.
RNFUN05	
NOMBRE	Modularidad
DESCRIPCIÓN	El sistema se construye por módulos o partes que interactúan entre sí y cumplen un objetivo en común: Consulta Inteligente. (Ver Diagrama de Componentes).
RNFUN06	
NOMBRE	Algoritmos
DESCRIPCIÓN	Los algoritmos implementados para el Procesamiento del Lenguaje Natural y las consultas a las ontologías son diseños realizados por parte de los autores de este proyecto como propuesta para el mejoramiento de las consultas de material académico en la UFPSO.
RFUN07	
NOMBRE	Rendimiento
DESCRIPCIÓN	Los requerimientos en esta categoría están relacionados con tiempos de respuesta estimados, requeridos y esperados para la ejecución en línea de procesos del sistema, teniendo como base la plataforma tecnológica y escenarios específicos a los que en teoría el sistema estará expuesto y frente a los que deberá responder.
RFUN08	
NOMBRE	Fiabilidad
DESCRIPCIÓN	Estos requerimientos están relacionados con la capacidad del usuario para confiar en las respuestas del sistema, en un sentido técnico, es decir, que la funcionalidad del sistema no se vea afectada por factores ajenos al sistema como los son los factores técnicos.

Cuadro 3. (Continuación)

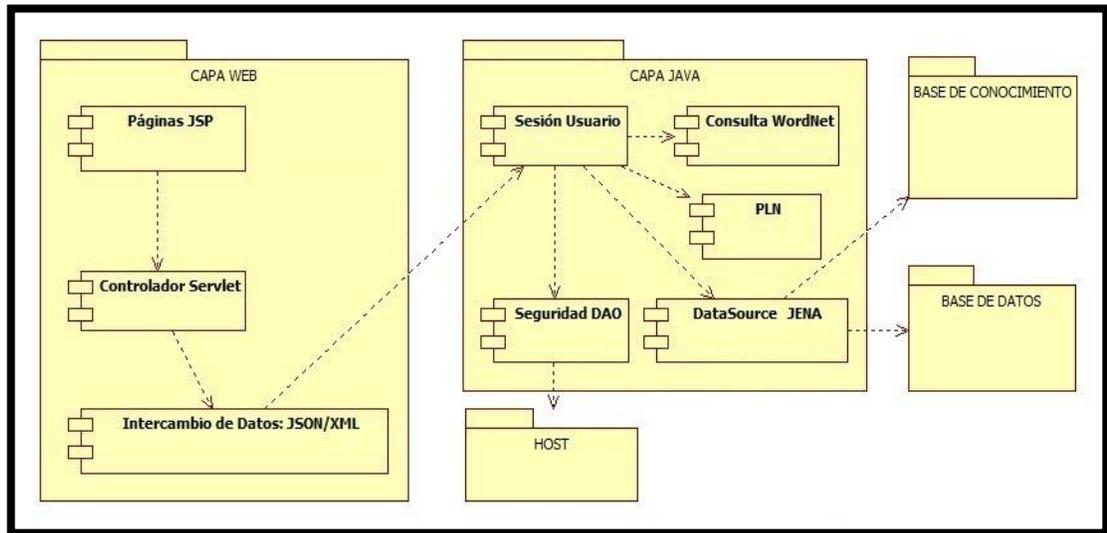
RFUN09	
NOMBRE	Seguridad
DESCRIPCIÓN	Requerimientos relacionados con la confidencialidad de los datos en la transmisión y en el almacenamiento, junto con las necesidades del sistema para evitar intrusiones no autorizadas al mismo y la capacidad para seguir eventos que comprometan esta seguridad a través del tiempo.
RFUN10	
NOMBRE	Portabilidad
DESCRIPCIÓN	El sistema cuenta con la capacidad para migrar de una plataforma hardware a otra sin que esto represente problemas al cliente, teniendo en cuenta los requisitos técnicos presentados y las generalidades naturales de configuración del sistema.
RFUN11	
NOMBRE	Mantenimiento
DESCRIPCIÓN	El sistema o los componentes de persistencia pueden ser modificados sobre su funcionalidad sin que se haga una inversión de recursos exagerada.
RFUN12	
NOMBRE	Reusabilidad
DESCRIPCIÓN	El sistema funciona solo usando librerías adicionales como JENA Y JWNL para determinadas funciones, aunque es compatible y puede usarse si se desea incluir dentro de otro sistema web más grande.
RFUN13	
NOMBRE	Interfaces
DESCRIPCIÓN	Los formularios que se muestran al usuario no llevan mucho trabajo pues la idea es mostrar resultados de manera simple o resumida, la interfaz planteada es básica, con texto, casillas de entrada, botones y con pocas imágenes o iconos para representar las OVAS, debido a que el procesamiento de las consultas es lo primordial en el sistema.

Fuente. Autores del proyecto

4.2 DIAGRAMA DE COMPONENTES

Los conocimientos sobre el diseño de Bases de Datos eran amplios debido al trabajo realizado a lo largo de la carrera y no duró mucho tiempo su conclusión. Creamos un modelo relacional muy simple para almacenar los detalles de cada consulta y gestionar los usuarios, por otra parte, para la construcción de la ontología se llevó más tiempo porque el conocimiento del tema era muy limitado. Para ver el diseño de este componente (Ontología) revisar el Anexo C.

Ilustración 23: Diagrama de Componentes



Fuente. Autores del proyecto

Por un lado, la ontología guarda toda la información de las ovas, los objetos (individuales) se relacionan por medio de las propiedades de objeto y propiedades de dato. Toda esta información es categorizada gracias a las clases y como un factor adicional podemos ubicar gráficamente estos elementos dentro de una herramienta llamada OntGraph de Protégé. Mientras que de esta forma gracias a la base de datos podemos almacenar información temporal para ubicar indicadores relevantes de la consulta que posteriormente serán condiciones para la búsqueda en la ontología. También nos sirve para el análisis léxico, sintáctico y semántico de las consultas pues se almacenan aproximadamente unas 1000 palabras con su traducción para el procesamiento de las frases digitadas en las búsquedas y por último guardar los enlaces dónde encontramos los archivos fuera de la aplicación.

El resultado: Resultados en un formulario JSP con información que se busca en dos componentes de persistencia que interactúan para poder evitar la redundancia de la información y obtener los detalles específicos necesarios.

4.3 FUNCIONALIDADES DEL SISTEMA

Anteriormente vimos cómo se diseñaron los componentes de persistencia o almacenamiento de información y el diagrama de componentes del sistema, a continuación apoyado de un diagrama de casos de uso, se explica el proceso de toda la aplicación y se obtiene una conclusión de las fallas o inconvenientes identificados y cómo el software provee la solución adecuada.

El primer aporte de este desarrollo es de tipo metodológico. Se creó un procedimiento o algoritmo para contribuir en la extensión del uso de ontologías, para almacenar y buscar

información detallada, facilitando a un usuario activo consultar los materiales académicos de la especialización gracias al etiquetado semántico de la información que se gestiona.

En nuestro trabajo se tiene muy en cuenta la posibilidad de que el contenido pueda hacer referencia a diferentes temas o pueda interpretarse desde diferentes puntos de vista, por eso al terminar la ontología queremos proponer como recomendación que el proceso hecho por nosotros se puede usar a futuro con la idea de “poblar” diferentes ontologías desde el mismo contenido, lo que se conoce como “vistas semánticas”.

Por el momento la aplicación es un prototipo, es decir que se ha planteado un escenario simplificado que recrea los elementos fundamentales, para proponer una estrategia de migración o transformación hacia la Web Semántica y analizar los resultados. Las conclusiones alcanzadas son el resultado de un proceso de autocorrección experimental.

Para realizar esta transformación o migración, se ha implementado una herramienta prototipo que automatiza el proceso de consulta que hemos venido presentando. Nuestra herramienta se ha ejecutado desde un servidor de prueba y corre sin ningún inconveniente en la anotación con la ontología usada en la tesis, pero es fácilmente adaptable para soportar otras. Para probar el proceso, hemos realizado nuestro propio módulo de PLN y nuestra interfaz desde JAVA y usando páginas JSP que permite mostrar la viabilidad para usuarios activos.

Finalmente hemos construido un prototipo de buscador semántico (OntoBus), cuyas búsquedas se basan en la consulta de ontologías desde JENA y cuyos resultados siguen siendo en el formato web al que el usuario está acostumbrado. De manera, que un usuario de un buscador semántico no ve la Web Semántica pero percibe sus efectos en una mejora del interfaz, en cuanto la expresión natural de las preguntas al buscador, y de la calidad de la información recibida, mucho más precisa, más ajustada como respuesta a cuestión planteada.

Es difícil pensar en una evaluación colectiva y colaborativa, ya que, necesitaríamos usuarios que probasen nuestra implementación y posteriormente realizar una encuesta sobre su experiencia. Sin embargo ese es el siguiente paso, implementarlo en la universidad en un futuro para el manejo de la información académica o bibliográfica de cualquier carrera. Pero es cierto también, que actualmente un usuario activo, como un WebMaster no puede crear un sitio web semántico sin esfuerzo. De forma, que con una herramienta como la diseñada en el prototipo de esta tesis, podría transformar su web en muy poco tiempo. Los resultados obtenidos, aunque modestos, son gratificantes, ya que, hemos podido automatizar el proceso completo de consulta, que era el objetivo principal de la tesis. En nuestras pruebas, hemos logrado acceder a mucha más información de forma automática y con poco esfuerzo del usuario.

La calidad de los resultados obtenidos depende de varios factores; como son la propia calidad de la ontología (afinidad, precisión, estandarización, completitud, etc), la claridad del contenido y la capacidad de extracción y análisis, condicionada, en gran medida, al

procesado de lenguaje natural (PLN) o incluso desde la misma consulta con frases o palabras mal escritas.

Nuestra propuesta representa una solución de esas debilidades y amenazas de la web actual que se convierten en fortalezas en una Web Semántica basada en vistas semánticas.

4.3.1 Casos de uso. En este apartado se describen los distintos casos de uso del sistema. Cada caso de uso se centra en describir como alcanzar una única meta o tarea y ayudan a describir qué es lo que el sistema debe hacer.

Lo primero que haremos será definir a los actores que interactuaran con el sistema dentro de los diferentes casos de uso, En nuestro caso tendremos un actor, el usuario registrado tras una sesión de usuario iniciada, un administrador para gestionar los componentes de persistencia y el sistema.

El usuario registrado es el que va a usar la aplicación y el sistema es la aplicación desarrollada funcionando, con todos sus módulos interactuando.

Ahora analizaremos el diagrama de casos de uso, para ver el documento de especificación de casos de uso ver Anexo D.

Ilustración 24: Actores



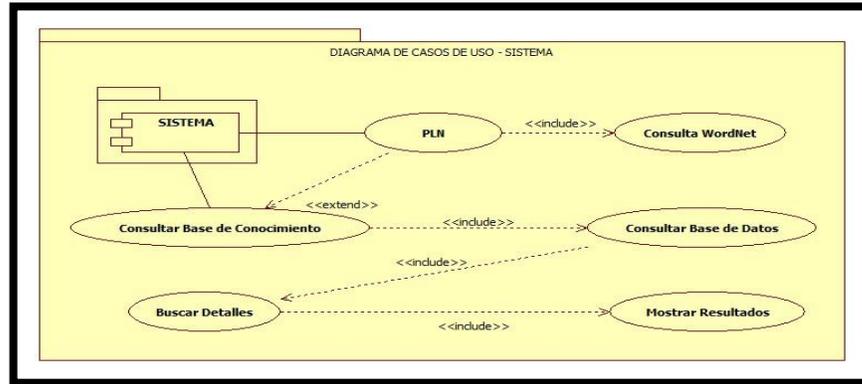
Fuente. Autores del proyecto

Ilustración 25: Diagrama de Casos de Uso – Usuario Registrado



Fuente. Autores del proyecto

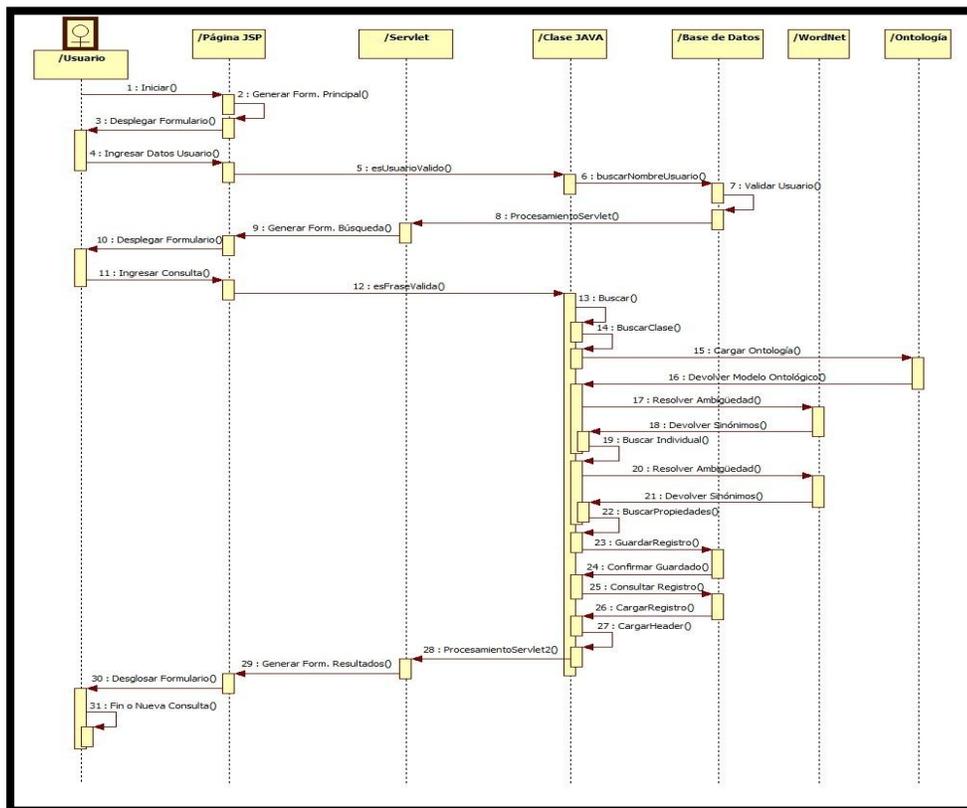
Ilustración 26: Diagrama de Casos de Uso – Sistema



Fuente. Autores del proyecto

4.3.2 Diagrama de Secuencias. A continuación se muestra la forma en que interactúa el usuario con los módulos del sistema en una consulta:

Ilustración 27: Diagrama de secuencias de una búsqueda.



Fuente. Autores del proyecto

Para entender los procesos que se llevan a cabo, se describe cada uno:

Iniciar (): El usuario entra al formulario principal e inicia el sistema.

Generar Form. Principal (): Se carga el formulario principal donde el usuario puede ingresar los datos de autenticación.

Desplegar Formulario (): Se muestra el formulario generado en la página JSP.

Ingresar Datos Usuario (): El usuario rellena las cajas de texto para validar e ingresar al menú de consulta.

esUsuarioValido(): Método en java que verifica si el usuario existe en la base de datos.

buscarNombreUsuario(): Método en java que hace una consulta a la base de datos y se obtiene el nombre del usuario.

Validar Usuario (): El sistema verifica los datos y pasa la información necesaria al controlador del Servlet.

ProcesamientoServlet (): Se pasan los datos cargados de la base de datos para crear el formulario de búsqueda.

Generar Form. Búsqueda (): Se carga el formulario de búsqueda donde el usuario puede ingresar la frase para la consulta.

Ingresar Consulta (): El usuario rellena las cajas de texto para realizar la consulta.

esFraseValida(): Método en java que verifica si las palabras de la frase existen en la base de datos y así saber que la frase está bien escrita.

Buscar (): Método en java que separa la frase por palabras e inicia la búsqueda.

BuscarClase (): Método en java que busca si la palabra de la frase es una clase de la ontología.

Cargar Ontología (): Se carga la ontología desde java gracias a los métodos que brinda JENA.

Devolver Modelo Ontológico (): Se devuelve el modelo por el cual se aplicaran los métodos de búsqueda.

Resolver Ambigüedad (): El sistema revisa las palabras seleccionadas y si no encuentra similitudes busca los sinónimos e hiperónimos.

Devolver Sinónimos (): Si la palabra no está en la ontología busca los sinónimos e hiperónimos y los cambia para encontrar similitudes.

BuscarIndividual (): Método en java que busca si la palabra de la frase es un individual de la ontología.

BuscarPropiedades (): Método en java que busca las propiedades de objeto y de dato (relaciones) de un individual previamente seleccionado.

GuardarRegistro (): Método en java que guarda temporalmente los resultados de la consulta a la ontología para posteriormente buscar la información sin redundar.

Confirmar Guardado (): La base de datos confirma si se guarda o no el registro de la consulta a la ontología.

Consultar Registró (): Se verifica en la base de datos que los registros hayan sido guardados.

CargarRegistro(): Método en java que carga los registros guardados temporalmente de la consulta a la ontología para posteriormente cargarlos en el formulario de resultados.

CargarHeader(): Método en java que recoge los datos que tienen en común los registros guardados para crear un encabezado y categorizar los resultados.

ProcesamientoServlet2 (): Se pasan los registros cargados de la base de datos para crear el formulario de resultados.

Generar Form. Resultados (): Se carga el formulario de resultados donde el usuario obtiene las salidas de todo el proceso de la consulta.

Fin o Nueva Consulta (): El usuario cierra la sesión o vuelve a iniciar una búsqueda repitiendo el proceso descrito en el diagrama de secuencias.

5. CONCLUSIONES

Reunir la información mediante artefactos de ingeniería de software permitió definir los requerimientos necesarios para la construcción de la aplicación web. Entre los requerimientos más importantes se destacan: El filtrado de las consultas y el procesamiento del lenguaje natural.

La investigación contribuye a una innovación a nivel departamental, ya que en Norte de Santander no existe una aplicación similar y en Colombia la cantidad de proyectos relacionados es baja. De los trabajos que implementan ontologías se destaca: “Sistema de información documental sobre el desplazamiento forzado en Colombia basado en ontologías” disponible en el sistema de revistas de la Universidad de Antioquia.

El diseño de los mecanismos de persistencia, usando ontologías apoyadas en bases de datos, facilita el acceso a más información con búsquedas más simples, evidenciando que para generar resultados eficientes en la gestión de los datos, los sistemas de información deben usar tecnologías orientadas a la Web 3.0.

En la Universidad Francisco de Paula Santander Ocaña existen estudiantes de ingeniería que han desarrollado proyectos similares con logros significativos en buscadores sintácticos para acceder a cursos de la plataforma Uni-Virtual y documentos disponibles en la biblioteca. Este proyecto mantiene un proceso evolutivo en los buscadores y en los sistemas de información gracias al procesamiento de frases en lenguaje natural, búsquedas semánticas con sinónimos apoyados en una base de datos léxica llamada WordNet, algoritmos para autocompletar palabras, gestión de una ontología desde un entorno web o local, sugerencias de búsquedas y la idea de acceder solo con los datos del SIA.

6. RECOMENDACIONES

Detallamos en primer lugar, las líneas futuras de investigación que presumiblemente requieren varios años de dedicación y, en segundo lugar, señalamos posibles trabajos de innovación, que aunque requieren implementación de software, se estima que el tiempo será del orden de meses. Finalmente nos hemos aventurado a imaginar algunas oportunidades que presenta este nuevo paradigma al mundo empresarial partiendo de las posibilidades que brinda nuestro proyecto:

La generación de páginas semánticas implica varios procesos como son: identificación, extracción e interpretación. El resultado de la ejecución de estos procesos es la transformación de un sitio web en un sitio web semántico, generando páginas semánticas basadas en vistas semánticas, o lo que es lo mismo, páginas que “entienden” las “máquinas”.

Optimización de los procesos de transformación basados en PLN. Esta investigación debería abarcar desde el proceso de identificación hasta la generación y actualización de vistas semánticas mediante métodos morfológicos. Como se ha descrito en esta tesis, el método de identificación de las palabras ha dado resultados muy positivos. Este método tiene por objetivo localizar los lexemas de los vocablos para identificar otras palabras relacionadas.

Consideramos una línea de gran interés incorporar en el proceso de identificación alguna técnica ligera de PLN. Consideramos que WordNet con su versión europea conocida como EuroWordNet (que incluye castellano) son buenas opciones a tener en cuenta. De todas formas, no debemos olvidar herramientas profesionales (no gratuitas) como DataLexica.

Contenido Semántico de naturaleza Dinámica. Como mencionamos antes, las vistas semánticas se han definido como una estructura que soporta contenido de texto estático y dinámico, pero es evidente que el contenido dinámico es muy complejo de tratar, ya que las páginas web dinámicas no existen en realidad y se generan con una consulta a la Base de Datos (BBDD) en tiempo real; es por esto que se intuye que generar vistas semánticas sincronizadas con contenido dinámico representa uno de los principales problemas de la Web Semántica. En esta investigación habrá que definir un procedimiento para permitir generar y modificar vistas semánticas en tiempo real. Probablemente incluso se tenga que proponer especificaciones de cómo hacer esto. Como referencia para comenzar esta investigación sería muy útil examinar propuestas como R2O y D2R. En concreto la tesis de Jesús Barrasa (enero 2007), que propone el lenguaje R2O y el procesador ODEMapster, que parece que puede servir de base para plantear soluciones al problema de la generación de vistas semánticas a partir de contenido dinámico (información en BBDD).

Algoritmos para la extracción de identificadores y traducción al idioma del contenido. La tarea de asociar una o varias ontologías con un contenido en lenguaje natural requiere diseñar un proceso que permita buscar, dentro de la ontología que se desea comparar, todas las palabras “claves” que la forman en el lenguaje natural de la ontología, esto es, los

identificadores que se utilizaron para nombrar las clases y propiedades. En nuestro proyecto tuvimos ese inconveniente y alertamos que se debe tener en cuenta que la búsqueda proporciona identificadores que deberán ser tratados adecuadamente, por ejemplo, la mayoría de los identificadores estarán formados por palabras concatenadas que habrá que separar. Una vez que se logra identificar y separar palabras, habrá que traducirlas al idioma natural que se use en el contenido, es decir, el idioma que el webmaster utiliza en sus páginas con el objetivo de poder comprobar si estas palabras traducidas aparecen en el contenido del webmaster calculando así la frecuencia de aparición. A mayor frecuencia de coincidencia mayor probabilidad que el contenido se asocie con una determinada ontología. Esta descripción del procedimiento deberá repetirse para cada ontología con la que se desee comparar el contenido. Lo que conduce a plantear también la coherencia y sincronismo de estos procesos con las ontologías originales de las que se parte.

Selección automática de contextos para aplicaciones semánticas El problema del contexto en aplicaciones semánticas concretas, como es el caso de los buscadores semánticos determinará, en cierta medida, la calidad de la búsqueda y la sencillez para el usuario. Es fundamental desarrollar estrategias para seleccionar el “contexto” de forma automática. De esta forma, el buscador semántico podrá conocer el dominio de conocimiento que debe usar y por consiguiente las ontologías implicadas para sus consultas lógicas.

Trabajos de Innovación Tecnológica relevantes a este proyecto aplicado en la UFPSO.

Se recomienda para proyectos de fin de carrera aquellos que permitan la optimización del Buscador Semántico: ONTOBUS. No se debe olvidar que un buscador semántico deberá incorporar todas las prestaciones actuales mejorándolo con las ventajas semánticas. En esta tesis se ha construido un prototipo de buscador semántico que puede ser mejorado en varios aspectos:

Mejora del interface de PLN

Mejora de funciones lógicas

Redes neuronales para conocer las preferencias del usuario

Incorporación de más ontologías para el uso de vistas semánticas

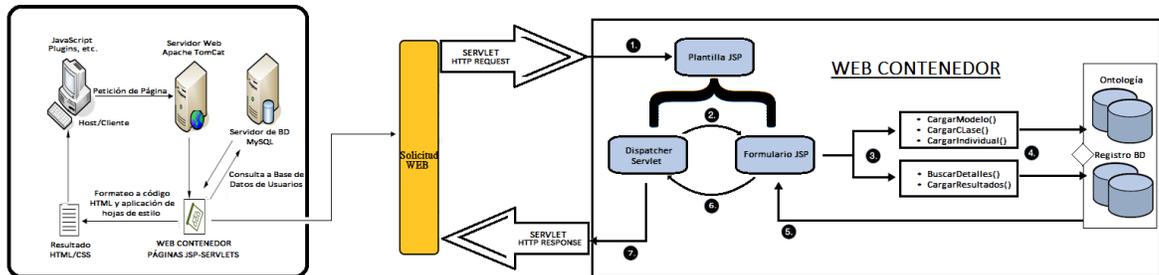
Concluimos que el almacenamiento de la información con estas nuevas metodologías es consistente, los detalles obtenidos relacionados a lo que queremos buscar son relevantes, las consultas permiten un gran manejo de información por muy simples que sean, a lo anterior le sumamos que nos brinda soluciones alternas por si lo que se encuentra no es suficiente. Sin embargo se identifica la necesidad de un procesador del lenguaje natural para reemplazar la base de datos léxica de WordNet haciendo que el programa sea automanejable y autosostenible (como un sub-proyecto para optimizar los resultados).

Al comparar resultados, captar la diferencia fue inmediato. La aplicación muestra satisfactoriamente la información extraída de un componente de almacenamiento orientado a la Web 3.0, una Base de Conocimientos, una Ontología con los detalles que se pueden obtener de ella, sus relaciones y demás consultas. A futuro se recomienda trabajar en un modelo para las consultas SPARQL también, esto da más opciones de resultados.

Con el cumplimiento de los objetivos proyectados inicialmente se cumple con el objetivo general: **Desarrollar un buscador semántico para gestionar la consulta de los objetos virtuales de aprendizaje de la especialización en auditoría de sistemas de la UFPSO.**

A continuación se muestra la arquitectura de la aplicación desarrollada:

Ilustración 28: Arquitectura de la aplicación.



1. Petición web, Carga plantillas.
2. Disparador del Servlet para llamar formularios JSP.
3. Cargar mecanismos persistentes por medio de métodos específicos.
4. Buscar información en la Ontología o en la Base de Datos.
5. Cargar datos al formulario JSP.
6. Llamar al disparador del servlet para mostrar el formulario.
7. Respuesta Web, Muestra Resultados.

Fuente. Autores del proyecto

BIBLIOGRAFÍA

PUELLO, Plinio. Desarrollo de una plataforma para la gestión de objetos virtuales de aprendizaje para la facultad de odontología en la universidad de Cartagena. (online). Disponible en: <http://190.25.234.130:8080/jspui/handle/11227/421>

CODINA, lluis. Web semántica y sistemas de información documental (online) UPF, 2009 [citado 1 Nov 2014]. Disponible en <http://www.lluiscodina.com/wp-content/uploads/webSemantica2009.pdf>

MAYER, miguel. Concepto y aplicaciones de la WEB 3.0 Una introducción para médicos (online) [España]. GRIP, 2009 [citado 5 mayo 2010]. Disponible en <http://www.sciencedirect.com/science/article/pii/S0212656709005083>

HOSEIN, fatima. Los buscadores semánticos y sus efectos en los(as) usuarios(as) y en las empresas (online) [Venezuela], 2012 [citado 2 febrero 2012]. Disponible en <http://fatimakorisha.blogspot.com/2012/02/los-buscadores-semanticos-y-sus-efectos.html>

LOZADA, Pablo. Evolución de la Web (Online), 2014 [citado 28 abril 2014]. Disponible en http://julionica.udem.edu.ni/wp-content/uploads/2014/01/Evolucion_Web.pdf

MERO, David. Buscadores Semánticos, 2da edición, Universidad Jaen, Octubre 1999. cap 14.

CODINA, lluis. ¿Web 2.0, Web 3.0 o Web Semántica? : El impacto en los sistemas de información de la Web (online) UPF, 2009 [citado 1 Nov 2014]. Disponible en http://www.lluiscodina.com/wp-content/uploads/Web20_WebSemantica2009_Nov2009.pdf

CARBONELL, Jaime. El procesamiento del lenguaje natural, tecnología en transición. (online) 2006. Disponible en <http://www.sc.ehu.es/sbweb/fisica/cursoJava/fundamentos/colecciones/stringtokenizer.htm>

CODINA, lluis. ¿Web 2.0, Web 3.0 o Web Semántica? : El impacto en los sistemas de información de la Web (online) UPF, 2009 [citado 1 Nov 2014]. Disponible en http://www.lluiscodina.com/wp-content/uploads/Web20_WebSemantica2009_Nov2009.pdf

VERA José. PIMENTEL Carlos. Redes semanticas: Aspectos teoricos, técnicos, metodológicos, y analíticos 1a Edicion, Mc Graw Hill Agosto 2009. p. 104.

ALCINA Amparo, VALERO Esperanza. RAMBLA Elena. Terminología y Sociedad del Conocimiento. 1ª Edición. PETERLANG, 2009. p. 255

MEIR Louis. Semantic Web 1ª Edicion. UHA, 2008. P. 111

MIRA jose, Aspectos básicos de la inteligencia artificial. 1ª edición. REBIUN. 1995. p.

201.

COLOMBIA APRENDE, Objetos Virtuales de Aprendizaje. (online) 2010. Disponible en <http://www.colombiaaprende.edu.co/objetos/>

COLOMBIA APRENDE, Objetos Virtuales de Aprendizaje: Educación innovadora (online) 2010. Disponible en: <http://www.colombiaaprende.edu.co/html/directivos/1598/article-117373.html>

LAMARCA LAPUENTE, Maria Jesùs, Ontologías: El nuevo concepto de documento en la cultura de la imagen (online). Disponible en: <http://www.hipertexto.info/documentos/ontologias.htm>

RDF Working Group. Resource Description Framework. (online). 2014. Disponible en: <http://www.w3.org/RDF/>

McGUINNESS, Deborah. Lenguaje de Ontologías Web (OWL). Amsterdam. (online). 2004. Disponible en: <http://www.w3.org/2007/09/OWL-Overview-es.html>

LUJAN MORA, Sergio. Programación en internet: Clientes Web. 1a Edición. (online) 2001. Disponible en: <http://www.definicionabc.com/tecnologia/html.php>.

ROUSE, Margaret. Sistema de gestión de base de datos relacionales (RDBMS). (online). 2015. Disponible en: <http://searchdatacenter.techtarget.com/es/definicion/Sistema-de-gestion-de-bases-de-datos-relacionales-RDBMS>.

ANEXOS

Anexo A. Encuesta dirigida a estudiantes, docentes y administrativos del área de postgrados de la Universidad Francisco de Paula Santander Ocaña.

**UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA
PLAN DE ESTUDIOS DE INGENIERÍAS
PROGRAMA DE INGENIERÍA DE SISTEMAS
ENCUESTA DIRIGIDA AL AREA DE POSTGRADOS DE LA UFPSO**

Objetivo: recolectar información sobre la aceptación de un buscador semántico de OVAS en el área de Postgrados de la Universidad Francisco de Paula Santander Ocaña, se ha diseñado la siguiente encuesta hacia la población que es objeto de estudio, de la cual se pretende organizar los datos estadísticamente para demostrar la viabilidad del proyecto.

La WEB 3.0 es la web del futuro, y el futuro es ahora. La rápida evolución exige mejores sistemas adaptados a las necesidades reales y, en el proceso educativo y la gestión de información los buscadores semánticos dan solución a problemas actuales. Un buscador orientado a la WEB 3.0, es un sistema inteligente que usa la información relacionada como bases de conocimientos (en este caso ontologías) que reemplazan u optimizan las bases de datos, que a su vez permite más solidez para recuperar o consultar los datos, dando precisión y eficacia en los resultados, además de usar el procesamiento del lenguaje natural, es decir: que “la máquina comprenda al humano” sin problemas de ambigüación en lo que se dice. Antes de empezar definimos OVA como cada objeto virtual (Imagen, video, documento, etc.) de aprendizaje o educativo. Marque con una “X” La respuesta escogida (Solo una):

1. Usted está vinculado a la Universidad Francisco de Paula Santander Ocaña (UFPSO) como:

- a. Estudiante
- b. Docente
- c. Administrativo

2. ¿Conoce o ha oído hablar sobre la WEB 3.0 o Web Semántica?

- a. Si
- b. No

3. ¿Sabe usted qué es un sistema inteligente de consulta o buscador semántico?

- a. Si
- b. No

4. ¿Conoce la diferencia entre Base de Datos y Base de Conocimientos?

- a. Si
- b. No

5. ¿Conoce usted algún sistema inteligente de consulta? Si su respuesta es afirmativa, por favor escriba cuál.

- a. Si _____
- b. No _____

6. ¿Ha usado o diseñado alguna aplicación inteligente como los buscadores semánticos?

- a. Si _____
- b. No _____

7. ¿Cómo cree que afecta el uso de la información virtual en la educación?

- a. Beneficia _____
- b. Empeora _____
- c. No afecta _____

8. ¿Para usted qué es lo más importante al buscar información educativa en la web?

- a. Cantidad de los resultados _____
- b. Precisión en los resultados _____
- c. Calidad de la información _____
- d. Rapidez de la búsqueda _____

9. ¿Con qué frecuencia solicita apoyo académico?

- a. Nunca _____
- b. Muy Pocas veces _____
- c. Muchas veces _____
- d. Siempre _____

10. ¿Qué medio prefiere al momento de buscar información educativa?

- a. Biblioteca _____
 - b. Internet/Apps. _____
 - c. Tutor/Maestro _____
 - d. Otro _____
- Cuál: _____

11. ¿Para usted cuál es el factor más importante para escoger el medio de consulta de la información académica?

- a. Equipos y Espacio _____
- b. Facilidad de acceso a la información _____
- c. Cantidad de conocimiento que se puede adquirir _____

- d. Veracidad y vigencia de la información _____
12. ¿Usted cree que los buscadores actuales (sintácticos) cubren las necesidades reales de los usuarios al consultar material académico?
- a. Si _____
- b. No _____
13. ¿Conoce usted si existe un espacio virtual donde se pueda interactuar con apoyo académico del área de postgrados del campus universitario?
- a. Si _____
- b. No _____
14. ¿Cree usted que un sistema de búsqueda semántica es necesario para encontrar a través de consultas en un sitio web las OVAS como apoyo académico para estudiantes de Postgrados de la Universidad?
- a. Si _____
- b. No _____
15. ¿Cree que un buscador semántico es la mejor herramienta para facilitar la consulta de OVAS en el área de Postgrados de la Universidad?
- a. Si _____
- b. No _____
16. Si eligiera implementar un buscador semántico, ¿qué tendría en cuenta para su consecución?
- a. Instalación práctica del sistema _____
- b. Rapidez ejecutando la aplicación _____
- c. Costos bajos _____
- d. Integración con otros sistemas existentes _____

Anexo B. Entrevista dirigida al director del proyecto

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER OCAÑA PLAN DE ESTUDIOS DE INGENIERÍAS PROGRAMA DE INGENIERÍA DE SISTEMAS ENTREVISTA DIRIGIDA A: ING. ANDRÉS MAURICIO PUENTES VELÁSQUEZ

Objetivo: El objetivo de esta entrevista es realizar una investigación sobre el uso de las tecnologías en el proyecto, conociendo opiniones personales de nuestro director distribuidas en 3 apartados. A continuación el modelo de entrevista:

Apartado 1: Preguntas abiertas para fortalecer el tema. Estas preguntas permitirán conocer aspectos o pensamientos un poco más personales del experto, opiniones o pensamientos propios de él que servirán para la comprensión de sus respuestas.

Pregunta 1: Antes de hablar sobre buscadores semánticos, sería bueno hablar sobre la W3C (Web semántica). Según su opinión, ¿Cómo definiría la Web semántica y qué retos plantea?

Pregunta 2: ¿Qué iniciativas interesantes de Buscadores Semánticos conoce?, ¿Conoce alguno en Colombia?

Pregunta 3: Según su opinión, ¿qué tan rentable y/o beneficioso pueden ser los Buscadores Semánticos?

Pregunta 4: Sobre las Ontologías, ¿son ellas el mejor esquema conceptual para trabajar con Buscadores Semánticos?, ¿existe otra manera de desarrollar búsquedas semánticas?

Pregunta 5: De forma general, ¿podría explicar el funcionamiento de un buscador semántico cualquiera?

Pregunta 6: ¿Qué cree que vendrá después de la Web Semántica y los Buscadores Semánticos?, ¿cree que será este el fin o el tope de la evolución informática?

Apartado 2: Preguntas cerradas con dos opciones de respuesta (**SI/NO**). Estas preguntas encaminarán la propuesta sobre dos alternativas que confirmarán o rechazarán la pregunta de acuerdo a las experiencias del experto.

Pregunta 1: ¿Ha trabajado, utilizado o desarrollado algún Buscador Semántico?

Pregunta 2: ¿Los buscadores habituales usados generalmente brindan eficacia en tiempo de respuesta y resultados?

Pregunta 3: ¿Los buscadores actuales brindan los resultados correctos en las consultas realizadas?

Pregunta 4: Específicamente en la UFPSO, ¿sería beneficioso crear un Buscador Semántico?

Pregunta 5: ¿Cree que un Buscador Semántico sería una motivación al uso apoyo académico por parte de la comunidad universitaria?

Apartado 3: Preguntas cerradas con más de dos opciones de respuesta. Estas preguntas ayudarán a la categorización de ciertos aspectos de la propuesta y de la situación actual de la investigación.

✓ **Pregunta 1:** A la hora de buscar apoyo académico o bibliográfico usted prefiere:

- A) Internet
- B) Tutor
- C) Biblioteca
- D) Otro: _____

✓ **Pregunta 2:** De acuerdo a la pregunta anterior, ¿qué factor es el que determina la elección de su respuesta?

- A) Facilidad de adquirir información adecuada
- B) Velocidad para encontrar lo buscado
- C) Coincidencias en los resultados
- D) Veracidad de la información
- E) Otro: _____

✓ **Pregunta 3:** ¿Con qué frecuencia visita usted algún Sistema de Consulta de Información fuera o dentro de la UFPSO?

- A) Siempre
- B) Algunas veces
- C) Muy pocas veces
- D) Nunca

✓ **Pregunta 4:** ¿Cómo clasificaría los Sistemas de Búsqueda Inteligente de la Información?

- A) Excelentes
- B) Buenos
- C) Regulares
- D) Malos

✓ **Pregunta 5:** ¿Cómo califica el actual uso de OVAS en los espacios virtuales de aprendizaje de la UFPSO?

- A) Excelente
- B) Bueno
- C) Regular
- D) Malo

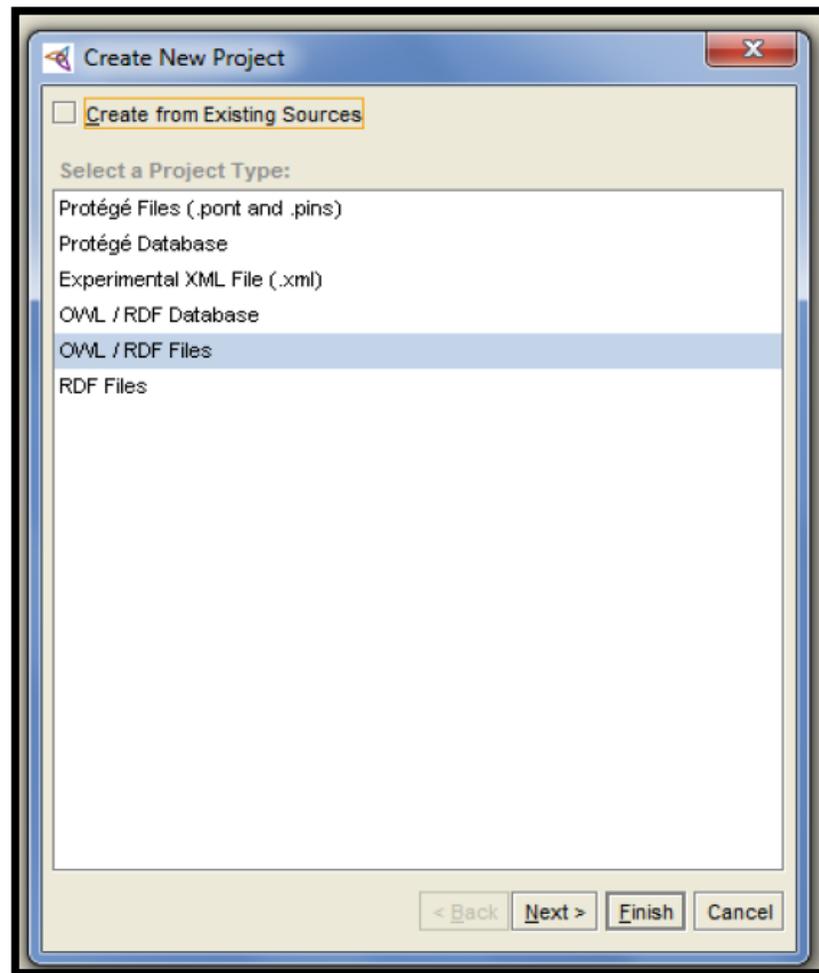
Anexo C. Diseño del componente de persistencia: “Base de Conocimiento”

CREACIÓN DE LA ONTOLOGÍA

A continuación se explica la manera en que se crea un nuevo archivo OWL/RDF con sus clases, individuales y relaciones. Cabe aclarar que los objetos instanciados pueden variar un poco en algunos detalles e incluso agregarse nuevos objetos para las pruebas de usabilidad y exposición del trabajo final.

Creación de un nuevo proyecto OWL/RDF

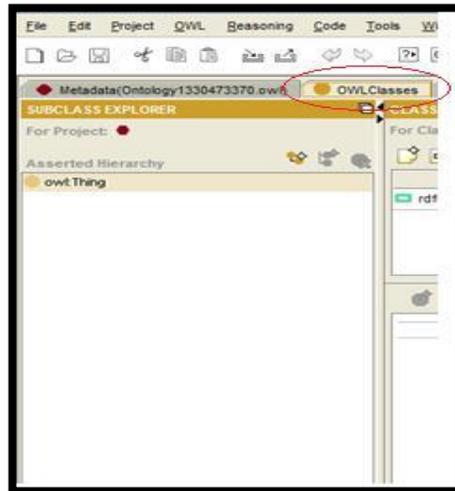
Se selecciona el lenguaje sobre el que se va a soportar la ontología, para el caso en cuestión se selecciona OWL/RDF Files. Se deja el URI por defecto ya que la ontología no es necesario que tenga un alcance web. Por la naturaleza de la ontología se selecciona RDF schema and OWL como perfil de lenguaje. Finalmente se selecciona Logical View.



Fuente. Autores del proyecto

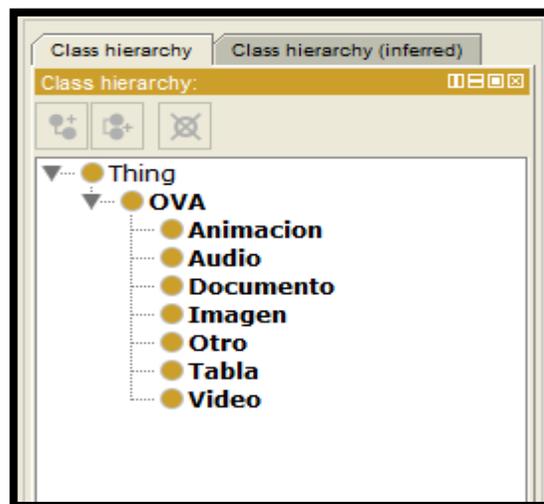
Creación de clases

Para construir los conceptos o clases, de la ontología, se ingresa a la pestaña OWLClases, donde se empiezan a crear clases y subclases de acuerdo al dominio, donde se parte de los conceptos más generales hasta llegar a los particulares.



Fuente. Autores del proyecto

Finalmente se ilustra el resultado de las clases desarrolladas.

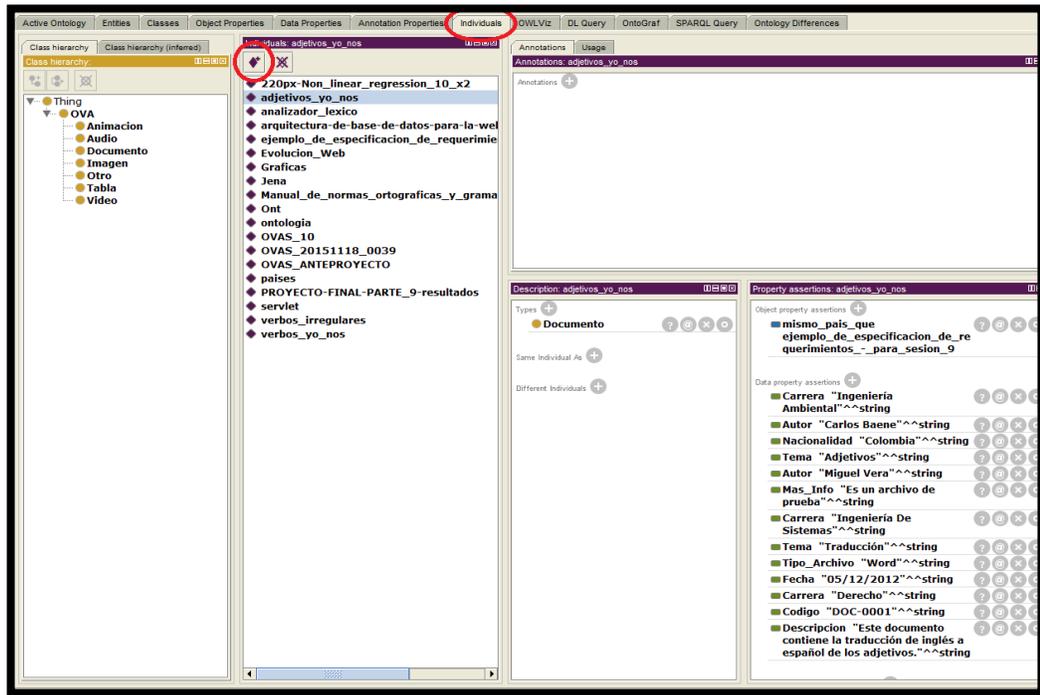


Fuente. Autores del proyecto

Creación de Instancias

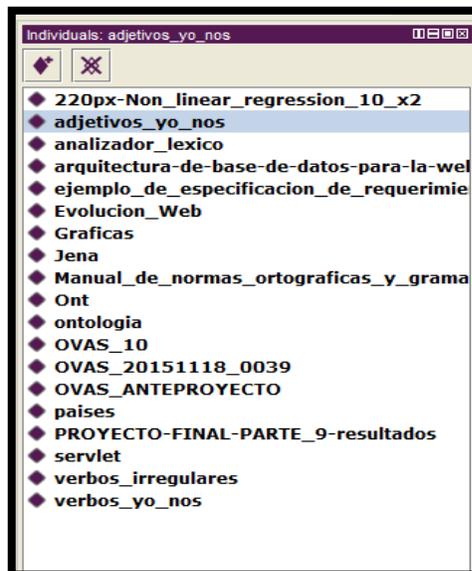
El siguiente paso es agregar instancias, es decir objetos que prueben la validez de nuestra ontología, para ello nos desplazamos a la pestaña de **Individuals**, para agregar un individuo

o instancia, usamos el botón de agregar individuos.



Fuente. Autores del proyecto

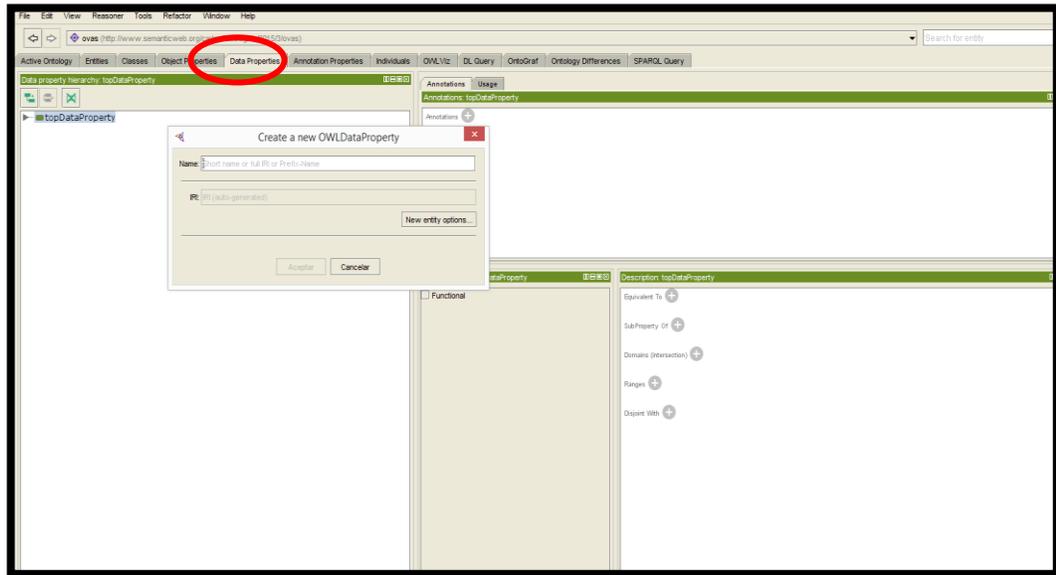
Como mencionamos antes, los objetos pueden variar a lo largo del proceso de nuestro proyecto, esto realmente no afecta en la estructura de la aplicación pues son los archivos de prueba para reconocer el funcionamiento de la aplicación. Por el momento manejamos los siguientes individuos:



Fuente. Autores del proyecto

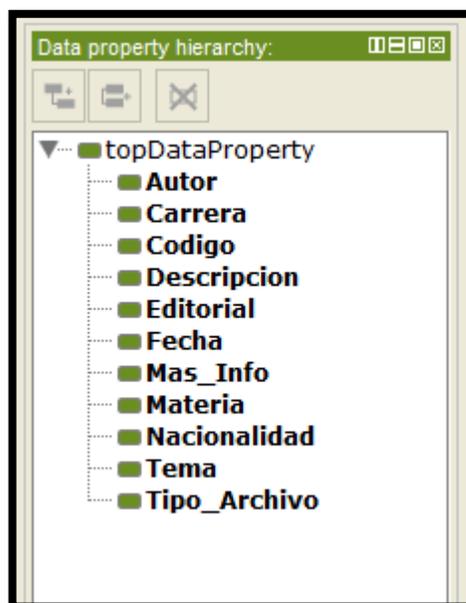
Creación de propiedades de datos

A continuación se crean las propiedades de los datos, esto es el tipo de información que se les asignaran a las clases. Para ello nos desplazamos a la pestaña Data Properties e insertamos dichas propiedades.



Fuente. Autores del proyecto

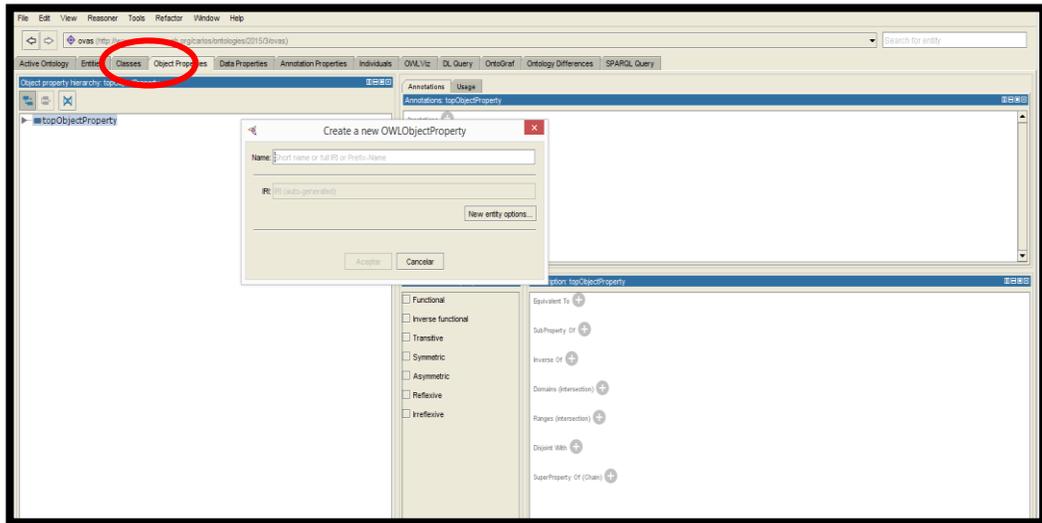
A continuación todas las propiedades creadas.



Fuente. Autores del proyecto

Creación de propiedades de objetos

Lo siguiente a realizar son los objetos de las propiedades, esta información irá asociada tanto a las clases como a los individuales. Para ello nos desplazamos en la pestaña **Object Propeties** e insertamos los valores.



Fuente. Autores del proyecto

El resultado final es el siguiente:

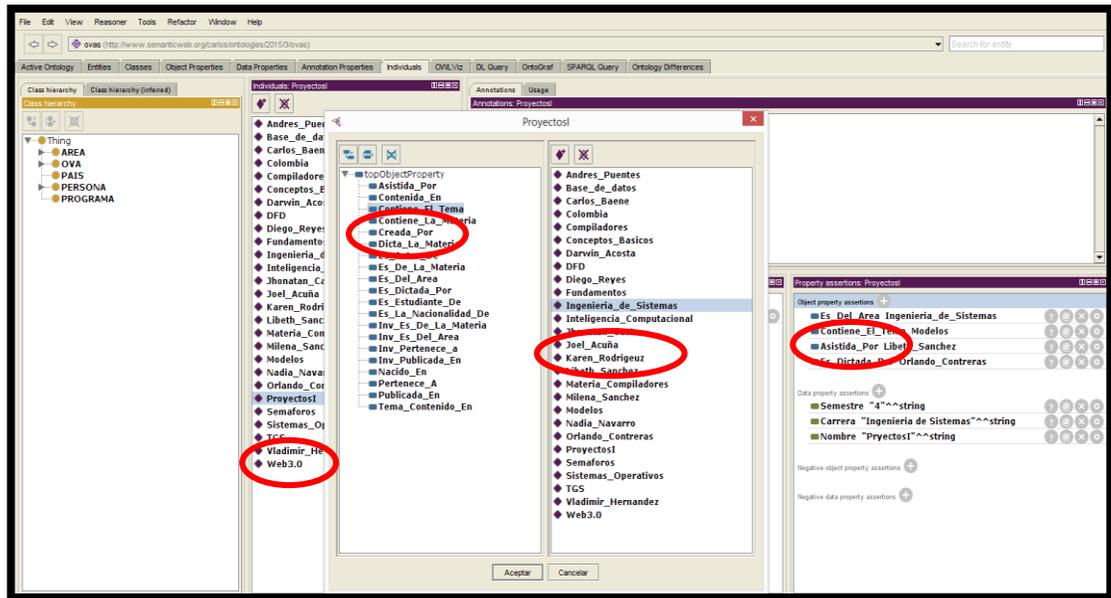


Fuente. Autores del proyecto

Creación de Relaciones

Para la creación de relaciones se selecciona la pestaña *individuals* y se va seleccionando cada uno de ellos, como ejemplo tomaremos un proyecto que hemos hecho paralelo al expuesto en este documento. Esta ontología además de la información de los recursos académicos también permite recuperar información específica de personas, países, programas, entre otros. Se seleccionará Proyectos1 y en la parte derecha vamos a la opción Object property assertions, esto para asociar dicha instancia.

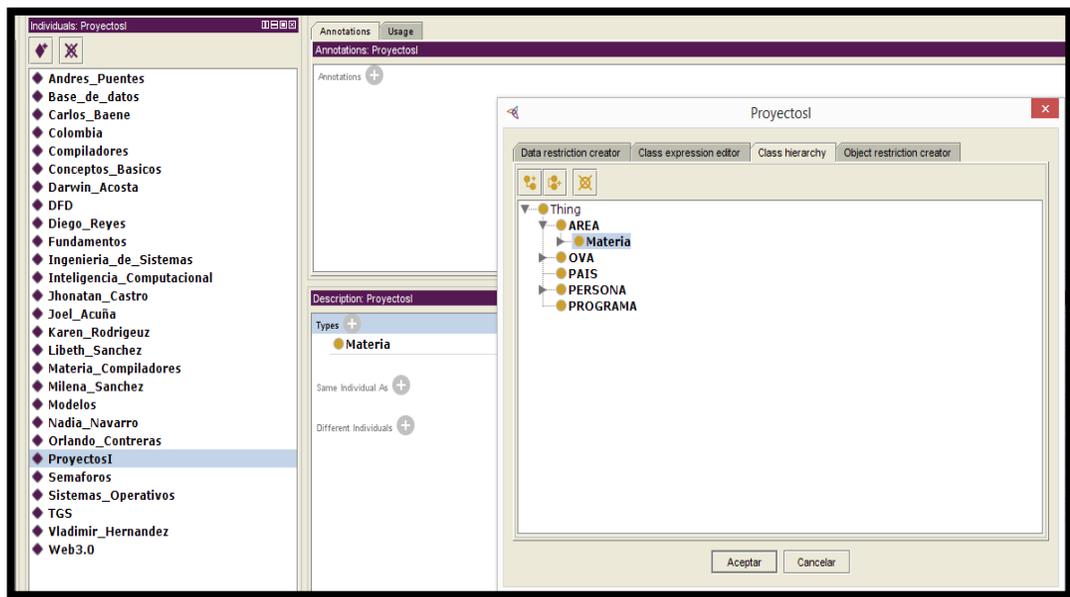
Como el ejemplo es de la materia proyectos1 podemos decir que Es del área de ingeniería de sistemas, contiene el tema modelos, Asistida por Libeth Sánchez, Es dictada por Orlando Contreras. De igual forma se crea las relaciones de clases y se llenan los datos de las propiedades.



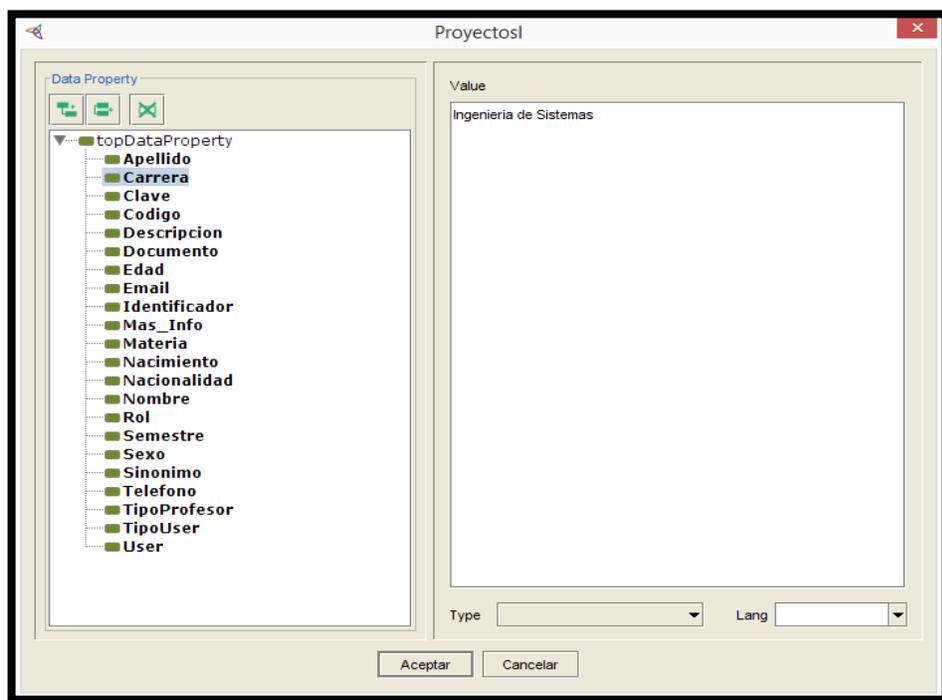
Fuente. Autores del proyecto



Fuente. Autores del proyecto

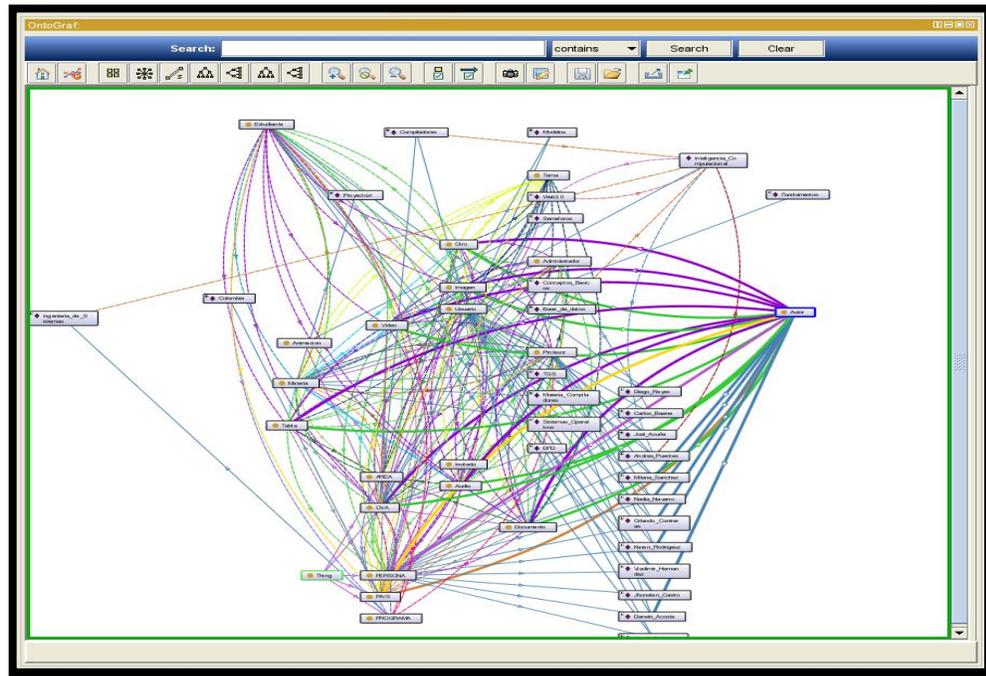


Fuente. Autores del proyecto



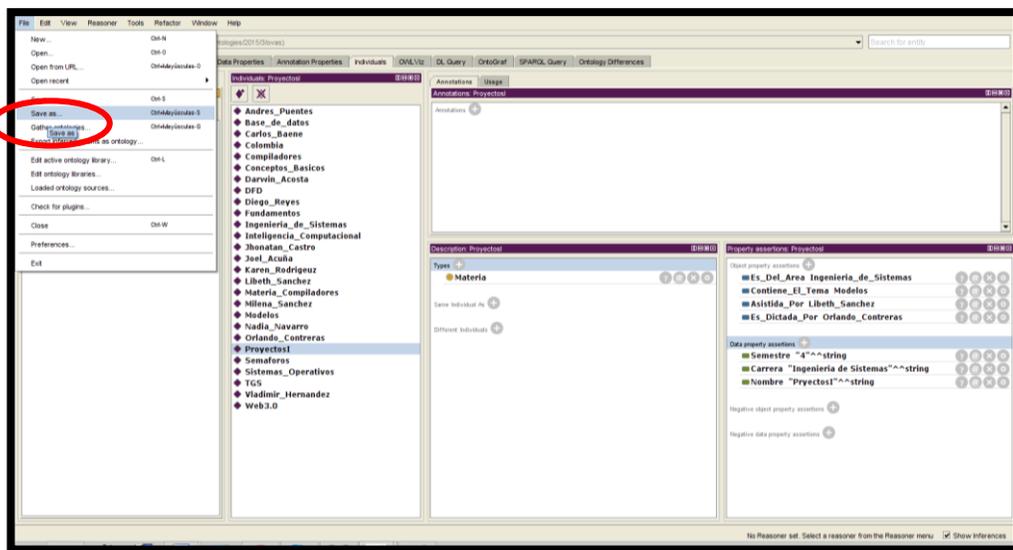
Fuente. Autores del proyecto

Grafica de la ontología.



Fuente. Autores del proyecto

Extracción del proyecto .OWL, Para guardar el proyecto .owl nos vamos a la opción file, save as, y a continuación se selecciona el formato en el que se va a extraer. En este caso es .OWL/XML.



Fuente. Autores del proyecto

Anexo D. Documento de especificación de Casos de Uso

CASOS DE USO DEL USUARIO

NOMBRE: INGRESAR		
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon	Fecha: 28/11/2015
Resumen		
Este caso de uso permite ingresar a la ventana principal del software y describir el formulario en el que se trabaja.		
Actor Principal		
El actor principal de este caso de uso es el Usuario Registrado.		
Precondiciones		
1. Estar registrado previamente como un estudiante, profesor o administrativo de la UFPSO en la base de datos.		
Flujo Básico de Eventos		
Número de Evento	Descripción	
1	El Sistema muestra el formulario principal que contiene en la parte superior el logo del software "OntoBus", debajo una etiqueta "Ingresar" y un poco más abajo dos cajas de texto que tienen dos etiquetas al lado: "Usuario" y "Contraseña", donde el usuario debe rellenar con la información correspondiente para que el sistema valide los datos, justo debajo un botón con la opción "Ingresar". Por último los créditos del creador del Software y del diseño del formulario.	
2	El Usuario Registrado llena las cajas de texto con información correcta y da clic al botón "Ingresar" o presiona la tecla "Enter".	
2.1	El Sistema valida los datos, el formulario principal se cierra y se ejecuta el caso de uso: Realizar Consulta.	
Flujos Alternativos		
Número de Evento	Número	Descripción
2	1	El Usuario Registrado llena las cajas de texto con información incorrecta y da clic al botón "Ingresar" o presiona la tecla "Enter".
2	1.1	El Sistema muestra un mensaje de error advirtiendo que los datos suministrados son erróneos. Se carga de nuevo el formulario principal.
Post-Condiciones		
Al poder ingresar al sistema, el Usuario Registrado puede realizar consultas a las OVAS disponibles en el momento.		
Validaciones		
Ninguna.		
Suposiciones		
1. Los Usuarios Registrados están almacenados previamente en la base de datos de la UFPSO como estudiantes, profesores o administrativos y serán gestionados fuera de la aplicación.		

Fuente. Autores del proyecto

NOMBRE: REALIZAR CONSULTA		
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon	Fecha: 28/11/2015
Resumen		
Este caso de uso permite digitar la consulta y enviarla para que el sistema inteligente haga el procesamiento necesario y la búsqueda solicitada a partir de otros casos de uso que se llaman desde el formulario de búsqueda.		
Actor Principal		
El actor principal de este caso de uso es el Usuario Registrado.		
Precondiciones		
1. Haber ingresado al sistema y por ende validado los datos de usuario.		
Flujo Básico de Eventos		
Número de Evento	Descripción	
1	El Sistema despliega el formulario de búsqueda que contiene en la parte superior el logo del software "OntoBus", debajo una etiqueta que dice "Escriba su consulta" y debajo de esta una caja de texto que será la que permita ingresar la consulta, al lado un botón que dice "Buscar" y un poco más a la derecha un botón que dice "Cerrar Sesión". Más abajo hay una descripción del software con instrucciones de su uso y algunas sugerencias de búsquedas. Por último los créditos del creador del Software y del diseño del formulario.	
2	El Usuario Registrado llena la caja de texto con una consulta con alguna o todas sus palabras correctas (Es decir: palabras bien escritas y existentes, también palabras con poca mala ortografía a las que se puedan aplicar los métodos de desambiguación y de transformación) y da clic al botón "Buscar" o presiona la tecla "Enter".	
2.1	Se ejecuta el caso de uso: PLN.	
2.2	Se ejecuta el caso de uso: Consultar Base de Conocimiento.	
2.3	Se ejecuta el caso de uso: Consultar Base de Datos.	
Flujos Alternativos		
Número de Evento	Número	Descripción
2	1	El Usuario Registrado llena la caja de texto con una consulta con todas sus palabras incorrectas (Es decir: palabras con muy mala ortografía, inexistentes o símbolos) y da clic al botón "Buscar" o presiona la tecla "Enter".
2	1.1	El Sistema muestra un mensaje de error advirtiendo que los datos suministrados son erróneos o la información no es suficiente para ser válida. Se carga de nuevo el formulario de búsqueda.
Post-Condiciones		
Si la consulta es válida, se hace el procesamiento necesario y se procede a realizar la búsqueda de elementos.		
Validaciones		
Ninguna.		
Suposiciones		
1. La información de las OVAS está previamente cargada en la base de conocimientos. 2. Se usa la base de datos para guardar registros temporales que serán consultados por SQL para no redundar información.		

Fuente. Autores del proyecto

CASOS DE USO DEL SISTEMA

NOMBRE: PLN		
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon	Fecha: 29/11/2015
Resumen		
Este caso de uso permite transformar una frase (String que contiene la consulta) en una lista de palabras, realizando con ellas el procesamiento de lenguaje natural mientras interactúa con WordNet. Al final, la información es insertada como un registro temporal en la base de datos.		
Actor Principal		
El actor principal de este caso de uso es el Sistema.		
Precondiciones		
1. Recibir una frase válida (Es decir: palabras bien escritas y existentes, también palabras con poca mala ortografía a las que se puedan aplicar los métodos de desambiguación y de transformación).		
Flujo Básico de Eventos		
Número de Evento	Descripción	
1	El Sistema recibe una frase digitada por el Usuario Registrado, separa la frase por palabras y revisa que están mal escritas.	
2	El Sistema realiza el proceso de transformación con ayuda de una consulta a la base de datos y revisa que las palabras si son válidas pero no existen en la base de conocimiento.	
3	Se ejecuta el caso de uso: Consulta WordNet.	
Flujos Alternativos		
Número de Evento	Número	Descripción
1	1	El Sistema recibe una frase digitada por el Usuario Registrado, separa la frase por palabras y revisa que si son válidas para su posterior consulta en la base de conocimientos.
1	1.1	Se ejecuta el caso de uso: Consultar Base de Conocimiento.
Post-Condiciones		
Si la consulta es válida, se hace el procesamiento necesario y se procede a realizar la búsqueda de elementos. De lo contrario se procede a realizar la desambiguación a través del caso de uso: Consulta WordNet.		
Validaciones		
Ninguna.		
Suposiciones		
1. Interacción dinámica entre este caso de uso y el caso de uso: Consulta WordNet.		

Fuente. Autores del proyecto

NOMBRE: CONSULTA WORDNET	
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon
Fecha: 29/11/2015	
Resumen	
Este caso de uso permite la desambiguación de las palabras y la búsqueda de sinónimos e hiperónimos para reemplazar las palabras en la consulta.	
Actor Principal	
El actor principal de este caso de uso es el Sistema.	
Precondiciones	
1. Este caso de uso se incluye (<<include>>) en el caso de uso: PLN.	
Flujo Básico de Eventos	
Número de Evento	Descripción
1	El Sistema consulta la base de datos léxica de WordNet y devuelve los sinónimos e hiperónimos de la palabra procesada.
Flujos Alternativos	
Ninguno.	
Post-Condiciones	
Se almacena en el sistema una lista de las posibles similitudes de las palabras con el fin de encontrar (si existe) la información consultada.	
Validaciones	
No es necesario instalar la aplicación de WordNet porque las librerías importan la base de datos léxica.	
Suposiciones	
1. La información de las OVAS está previamente cargada en la base de conocimientos. 2. Se usa la base de datos para guardar registros temporales que serán consultados por SQL para no redundar información. 3. Interacción dinámica entre este caso de uso y el caso de uso: PLN.	

Fuente. Autores del proyecto

NOMBRE: CONSULTAR BASE DE CONOCIMIENTO		
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon	Fecha: 30/11/2015
Resumen		
Este caso de uso permite consultar y cargar la base de conocimiento a la aplicación.		
Actor Principal		
El actor principal de este caso de uso es el Sistema.		
Precondiciones		
1. Debe existir y no estar vacía la ontología en formato OWL/RDF para ser cargada. 2. Este caso de uso se extiende (<<extend>>) del caso de uso: PLN.		
Flujo Básico de Eventos		
Número de Evento	Descripción	
1	El Sistema crea un modelo ontológico vacío mediante cláusulas internas.	
2	El Sistema carga satisfactoriamente el modelo ontológico de la base de conocimiento mediante cláusulas internas.	
2.1	El Sistema importa toda la información de la ontología que se encuentra en el archivo de formato OWL/RDF para su posterior uso dentro de la aplicación. Se ubican las clases, individuales y propiedades para compararlos con la frase de consulta. Con las similitudes cargadas en arreglos se guardan los registros temporales en la base de datos.	
2.2	Se ejecuta el caso de uso: Consultar Base de Datos.	
Flujos Alternativos		
Número de Evento	Número	Descripción
2	1	El Sistema no carga el modelo ontológico de la base de conocimiento debido a falla de conexión y/o de interoperabilidad o porque el archivo está dañado.
2	1.1	El Sistema muestra un mensaje de advertencia informando el error al cargar el archivo de la base de conocimientos.
Post-Condiciones		
La información de la base de conocimiento se carga en la aplicación para realizar las consultas.		
Validaciones		
La base de datos usada debe interactuar de manera correcta permitiendo la compatibilidad de los métodos desarrollados en la aplicación.		
Suposiciones		
1. Interacción dinámica entre este caso de uso y el caso de uso: Consulta WordNet.		

Fuente. Autores del proyecto

NOMBRE: CONSULTAR BASE DE DATOS	
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon
Fecha: 30/11/2015	
Resumen	
Este caso de uso permite la consulta (por medio de lenguajes SQL) de los registros temporales almacenados en la base de datos para que posteriormente se pueda realizar una búsqueda detallada y mostrar los resultados en el formulario de resultados.	
Actor Principal	
El actor principal de este caso de uso es el Sistema.	
Precondiciones	
1. En la base de datos deben existir y en caso de consulta no estar vacía. 2. Este caso de uso se incluye (<<include>>) en el caso de uso: Consultar Base de Conocimiento.	
Flujo Básico de Eventos	
Número de Evento	Descripción
1	Se consultan los registros guardados temporalmente para extraer la información necesaria y crear arreglos con todos los datos que coinciden con la consulta. Se cargan los datos y se presentan al controlador del procesamiento de servlet para reenviarlos al formulario de resultados.
2	Se ejecuta el caso de uso: Buscar Detalles.
Flujos Alternativos	
Ninguno.	
Post-Condiciones	
Extraer la información filtrada de los registros que se mostrara en los resultados del formulario de resultados.	
Validaciones	
La base de datos usada debe interactuar de manera correcta permitiendo la compatibilidad de los métodos desarrollados en la aplicación.	
Suposiciones	
1. Interacción dinámica entre este caso de uso y el caso de uso: Consultar Base de Conocimiento.	

Fuente. Autores del proyecto

NOMBRE: BUSCAR DETALLES	
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon
Fecha: 30/11/2015	
Resumen	
Este caso de uso la búsqueda de detalles adicionales (como el encabezado) para la finalización del proceso de búsqueda.	
Actor Principal	
El actor principal de este caso de uso es el Sistema.	
Precondiciones	
1. Este caso de uso se incluye (<<include>>) en el caso de uso: Consultar Base de Datos.	
Flujo Básico de Eventos	
Número de Evento	Descripción
1	El Sistema verifica la información suministrada y filtrada, luego carga el encabezado para cada ítem de la lista de resultados.
2	Se ejecuta el caso de uso: Mostrar Resultados.
Flujos Alternativos	
Ninguno.	
Post-Condiciones	
Se añaden detalles adicionales a los resultados.	
Validaciones	
Ninguna.	
Suposiciones	
Ninguna.	

Fuente. Autores del proyecto

NOMBRE: MOSTRAR RESULTADOS		
Versión: 0.1	Autores: Carlos Miguel Vera Baene Joel Alberto Acuña Bon	Fecha: 30/11/2015
Resumen		
Este caso de uso permite describir el formulario de resultados y cargar todas las salidas de la consulta en él.		
Actor Principal		
El actor principal de este caso de uso es el Sistema.		
Precondiciones		
1. Este caso de uso se incluye (<<include>>) en el caso de uso: Buscar Detalles.		
Flujo Básico de Eventos		
Número de Evento	Descripción	
1	El Sistema carga toda la información obtenida de los procesos de transformación, desambiguación, inserción y selección de registros, consulta de ontología y consulta de base de datos, para diseñar el formulario de resultados.	
2	El Sistema muestra el formulario de resultados que contiene en la parte superior el logo del software “OntoBus”, debajo una etiqueta “Resultados” y un poco más abajo una lista de los resultados obtenidos. Para cada ítem resultante se muestra una imagen del tipo de OVA que se encontró, una lista con las propiedades de objeto y otra con las propiedades de dato. Más abajo un botón que dice “Volver” y otro que dice “Salir”. Por último los créditos del creador del Software y del diseño del formulario.	
2.1	El Usuario Registrado revisa la información resultante y selecciona la opción “Volver” del formulario de resultados.	
2.2	Se ejecuta el caso de uso: Volver.	
Flujos Alternativos		
Número de Evento	Número	Descripción
1	1	El Sistema carga toda la información obtenida de los procesos de transformación, desambiguación, inserción y selección de registros, consulta de ontología y consulta de base de datos, para diseñar el formulario de resultados, pero esta es nula o vacía.
1	1.1	El Sistema muestra el formulario de resultados que contiene en la parte superior el logo del software “OntoBus”, debajo una etiqueta “Resultados” y un poco más abajo una lista de los resultados obtenidos. Se muestra un mensaje informando que no hay resultados para mostrar. Más abajo un botón que dice “Volver” y otro que dice “Salir”. Por último los créditos del creador del Software y del diseño del formulario.
2.1	2	El Usuario Registrado revisa la información resultante y selecciona la opción “Salir” del formulario de resultados.
2.1	2.1	Se ejecuta el caso de uso: Salir.
Post-Condiciones		
Se finaliza el proceso de búsqueda de OVAS y se carga el formulario de resultados con todas las coincidencias semánticas para la consulta de los objetos. Cada objeto va detallado y categorizado por tipo de OVA.		
Validaciones		
Los archivos y registros temporales son borrados para evitar desborde de memoria o de almacenamiento.		
Suposiciones		
Ninguna.		

Fuente. Autores del proyecto