TESIS TESIS TESIS TESIS TESIS

CENTRO DE CIENCIAS BÁSICAS

DEPARTAMENTO DE SISTEMAS DE INFORMACIÓN

TESIS

CONSTRUCCIÓN DE UN BUSCADOR SEMÁNTICO PARA UN SISTEMA WEB DE APOYO PARA EL APRENDIZAJE DE LA PROGRAMACIÓN: UN ESTUDIO DESCRIPTIVO.

PRESENTA
Julio René López Guerrero

PARA OBTENER EL GRADO DE MAESTRÍA EN INFORMÁTICA Y TECNOLOGÍAS COMPUTACIONALES

TUTOR
Dr. Carlos Argelio Arévalo Mercado

COMITÉ TUTORIAL
Dra. Estela Lizbeth Muñoz Andrade
Dra. Loecelia Guadalupe Ruvalcaba Sánchez

Aguascalientes, Ags., 10 de Junio de 2014

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

UNIVERSIDAD AUTONOMA DE AGUASCALIENTES

Centro de Ciencias Básicas

I.S.C. JULIO RENÉ LÓPEZ GUERRERO
ALUMNO (A) DE LA MAESTRIA EN INFORMÁTICA
Y TECNOLOGÍAS COMPUTACIONALES
P R E S E N T E.

Estimado (a) alumno (a) López:

Por medio de este conducto me permito comunicar a Usted que habiendo recibido los votos aprobatorios de los revisores de su trabajo de tesis y/o caso práctico titulado: **"CONSTRUCCIÓN DE UN BUSCADOR SEMÁNTICO PARA UN SISTEMA WEB DE APOYO PARA EL APRENDIZAJE DE LA PROGRAMACIÓN: UN ESTUDIO DESCRIPTIVO",** hago de su conocimiento que puede imprimir dicho documento y continuar con los trámites para la presentación de su examen de grado.

Sin otro particular me permito saludarle muy afectuosamente.

A T E N T A M E N T E
Aguascalientes, Ags., 10 de junio de 2014
'SE LUMEN PROFERRE'
EL DECANO

M. en C. JOSÉ DE JESÚS RUIZ GALLEGOS

UNIVERSIDAD AUTONOMA DE AGUASCALIENTES
CENTRO DE CIENCIAS BÁSICAS

c.c.p.- Archivo.
JJRG,mjda

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

M. EN C. JOSÉ DE JESÚS RUIZ GALLEGOS
DECANO DEL CENTRO DE CIENCIAS BÁSICAS
PRESENTE

Por medio de la presente, como Director de Tesis designado del estudiante **ISC. JULIO RENE LOPEZ GUERRERO** con ID 44439 quien realizó el trabajo de Tesis titulado: **CONSTRUCCIÓN DE UN BUSCADOR SEMÁNTICO PARA UN SISTEMA WEB DE APOYO PARA EL APRENDIZAJE DE LA PROGRAMACIÓN: UN ESTUDIO DESCRIPTIVO** de la **Maestría en Informática y Tecnologías Computacionales**, y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia, me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla y continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

A T E N T A M E N T E
"Se Lumen Proferre"
Aguascalientes, Ags., a 10 de Junio del 2014.

Dr. Carlos Argelio Arévalo Mercado
Director de Tesis

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

M. EN C. JOSÉ DE JESÚS RUIZ GALLEGOS
DECANO DEL CENTRO DE CIENCIAS BÁSICAS
PRESENTE

Por medio de la presente, como Integrante del Comité Tutoral designado del estudiante **JULIO RENE LOPEZ GUERRERO** con ID 44439 quien realizó el trabajo de Tesis titulado: **CONSTRUCCIÓN DE UN BUSCADOR SEMÁNTICO PARA UN SISTEMA WEB DE APOYO PARA EL APRENDIZAJE DE LA PROGRAMACIÓN: UN ESTUDIO DESCRIPTIVO** de la **Maestría en Informática y Tecnologías Computacionales**, y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia, me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla y continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

A T E N T A M E N T E
"Se Lumen Proferre"
Aguascalientes, Ags., a 10 de Junio del 2014.

Dra. Estela Lizbeth Muñoz Andrade
Integrante del Comité Tutoral

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

M. EN C. JOSÉ DE JESÚS RUIZ GALLEGOS
DECANO DEL CENTRO DE CIENCIAS BÁSICAS
PRESENTE

Por medio de la presente, como Integrante del Comité Tutoral designado del estudiante **JULIO RENE LOPEZ GUERRERO** con ID 44439 quien realizó el trabajo de Tesis titulado: **CONSTRUCCIÓN DE UN BUSCADOR SEMÁNTICO PARA UN SISTEMA WEB DE APOYO PARA EL APRENDIZAJE DE LA PROGRAMACIÓN: UN ESTUDIO DESCRIPTIVO** de la **Maestría en Informática y Tecnologías Computacionales**, y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia, me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla y continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

A T E N T A M E N T E
"Se Lumen Proferre"
Aguascalientes, Ags., a 10 de Junio del 2014.

Dra. Loecelia Guadalupe Ruvalcaba Sánchez
Integrante del Comité Tutoral

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Agradecimientos

Me gustaría aprovechar estas líneas para agradecer en particular a mi director de tesis que me asesoró en todo el proceso, siempre provocando nuevos ánimos por seguir adelante y mejorando el trabajo, de igual manera agradecer a mi comité tutoral por sus oportunas observaciones y a todas las personas que me apoyaron en la elaboración del trabajo.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Dedicatorias

Esta tesis la dedico a mi familia que me escuchó y ayudó en el proceso, en especial a mi novia Carolina porque cree en mi trabajo y me apoyó en todo momento, dándome siempre motivos para seguir adelante.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Índice General

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Índice de Tablas

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Índice de Figuras

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Resumen

Este trabajo de tesis se realizó con el objetivo de describir y evaluar las implicaciones técnicas y medir las ventajas en cuanto a la usabilidad y esfuerzo, de la migración de una aplicación bajo un modelo relacional, hacia una representación semántica basada en ontologías.

Para cumplir el objetivo, se propuso una metodología basada en la construcción de un prototipo de buscador semántico. La metodología está compuesta de tres fases principales, la primera fase consiste en la transformación de la base de datos relacional a su representación semántica construyendo un repositorio RDF utilizando la librería EasyRDF y durante el proceso medir el esfuerzo en horas invertidas en la transformación empleando las prácticas de PSP (Personal Software Process), teniendo un total de 91hrs invertidas en la conversión.

La segunda fase consiste de la implementación de una herramienta como ARC2 que permite procesar consultas semánticas en el repositorio RDF, las consultas deben ser construidas en base a expresiones en lenguaje natural, por lo cual la tercera fase es basada en la creación de una interfaz para el buscador con ayuda de un analizador semántico como NplTools que ayude a identificar conceptos clave en la expresión.

Finalmente se realizó un test de usabilidad en base a cinco tareas aplicadas a un grupo de seis personas, con el objetivo de medir indicadores como: tiempo de ejecución promedio, en donde se obtuvo 1.46 min y una tasa de éxito del 56.66%, tiempo entre eventos promedio 19.89 seg y cantidad de clics empleados en cada tarea 9.23 clics.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Abstract

This thesis was conducted with the objective of describe and evaluate the technical implications and measure the usability and effort advantages to migrate an application under a relational model, to a semantic representation based on ontologies.

In order to meet the objective, the proposal was a methodology based in the construction of a semantic search engine prototype. The methodology is composed of three main phases, the first phase consist in the transformation of the relational data base to its semantic representation through the building of a RDF repository using the Easy RDF library and during the process, measure the effort in hours dedicated in the transformation using the PSP (Personal Software Process) practices, having a total of 91hrs inverted in conversion.

The second phase involves the implementation of a tool like ARC2 which can process semantic queries in the RDF repository, queries must be built based on natural language expressions, and whence the third phase is based on the creation of an interface for the search engine using a semantic analyzer as NplTools in order to help to identify key concepts in the expression.

Finally, a usability test was performed base on five tasks applied to a group of six persons, in order to measure indicators such as average execution time, getting 1.46 min, success rate of 56.66 %, time between events 19.89 sec. and the average amount of clicks 9.23 per task.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Introducción

Internet es una herramienta cada día más necesaria en la vida cotidiana de las personas, ha llegado a estar inmerso en las herramientas más simples desde una pulsera cuenta pasos, hasta un Smartphone donde puedes llevar un control preciso de todas tus actividades, ha llegado a ser indispensable en tareas laborales, en el hogar, sociales y de negocio, permitiendo que las personas estén conectadas independientemente de la parte del mundo donde se encuentren.

La diversidad de usos de internet ha provocado un rápido y amplio crecimiento de esta tecnología, debido a que varios mercados o nuevos negocios, ven en internet grandes ventajas de comunicación que les permiten crear o hacer más simples algunos de sus procesos.

El crecimiento acelerado de internet está provocando que una gran cantidad de información de diversa índole sea almacenada en distintos lugares, formatos e idiomas; Generando en cierta medida un descontrol de la información existente, teniendo como resultado, un internet comprensible únicamente para los humanos pero no para las maquinas, en este sentido, cuando se desea encontrar alguna información en específico en internet, se recurre a técnicas de búsqueda y recuperación de información que permitan cubrir las necesidades.

La técnica de búsqueda depende de la necesidad y tipo de dato requerido, Para buscar y recuperar información en internet se utilizan generalmente los denominados motores de búsqueda.

Estos motores de búsqueda utilizan varias técnicas para catalogar e indexar información en internet, una de las más recurridas es el uso de palabras clave relacionadas con cada recurso en internet, este tipo de técnicas aunado al crecimiento desordenado de internet, están provocando una merma en la usabilidad de estas herramientas, tomando en cuenta el tiempo que se requiere, la calidad de resultados y el número de intentos para que el usuario encuentre la información de su interés.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Desde hace ya más de una década, grupos de colaboradores y visionarios han hecho propuestas sobre cómo debe evolucionar internet, el primero en proponer una evolución fue Tim Berners-Lee en el año 2000 (Passin, 2004), la evolución denominada Web 3.0 o Web Semántica ha sido la más aceptada, la cual básicamente consiste en dotar no solo de sentido explícito a la información, si no de atribuirle características que permitan una interpretación implícita de la información contenida en internet.

Estas características semánticas pueden ser atribuidas a la información en internet usando una estructura bien definida de tecnologías, entre las cuales destacan XML, RDF, OWL, SPARQL, etc. Para que la información en internet pueda tener una interpretación semántica es necesario un arduo proceso de migración de la información actualmente existente y nuevos mecanismos para expresarla en internet.

Una vez creada esta estructura en la web, basada en tecnologías semánticas, no solo la información será interpretada por humanos sino también por las maquinas, haciendo posible el desarrollo de herramientas con la habilidad de explotar la información con base en consultas cercanas al lenguaje natural humano.

En este contexto de búsquedas semánticas, una herramienta que tendría un cambio fundamental son los motores de búsqueda, implicando entre otras cosas, la modificación de la estructura interna de las bases de datos, construidas principalmente con modelos relacionales, hacia un paradigma semántico que permita una consulta más natural de la información.

Es así que la presente tesis describe un camino por el cual se puede transformar la información almacenada en bases de datos relacionales hacia una representación semántica. Se describe una primera etapa en donde se mide el esfuerzo de transformación de una base de datos relacional hacia su representación semántica basada en RDF, mientras que una segunda etapa es enfocada en la medición de usabilidad y construcción de un motor de búsquedas semánticas que ayude a explotar el repositorio RDF generado en el paso anterior, con base en consultas expresadas en lenguaje natural por parte del usuario.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

1. Antecedentes

1.1. La evolución de internet

El inicio de internet se remonta a los años 60's durante el periodo de la guerra fría, cuando Estados Unidos y la Unión Soviética estaban compitiendo por expandir sus relaciones en todo el mundo. Estados Unidos motivado por la necesidad de mantenerse como líder tecnológico de la época, asignó al departamento de defensa establecer la agencia de proyectos de investigación avanzados (ARPA), diseñada para producir ideas innovadoras de investigación, dentro de esta agencia existía la oficina de información y técnicas de procesamiento la cual financió la investigación enfocada a movilizar a las universidades y laboratorios de investigación para construir una red de comunicación estratégica que mejoraría las capacidades de mensajería del gobierno, durante las dos siguientes décadas esta red fue creciendo y mejorándose, pero aún seguía bajo uso exclusivo del gobierno de los Estados Unidos.

Durante la mitad de la década de los 80's el público en general comenzó a tener acceso a esta red ahora con el nombre de ARPANET, rápidamente el número de host llego a los 10,000, provocando congestiones en la red debido a la limitada capacidad de las líneas telefónicas. Para distribuir la carga de trabajo surgieron varias nuevas redes, entre ellas NSFnet, promovida por la Fundación Nacional de Ciencias, en 1988 NSFnet era capaz de manejar más de 75millones de paquetes al día, esta característica de inmediato provoco una expansión aun mayor de internet llegando a los 100,000 host en poco tiempo.

Un año más tarde, Tim Berners-Lee un investigador de la Organización Europea para la Investigación Nuclear (CERN), propuso la idea de un sistema internacional de protocolos, incluía la construcción de un servidor de hipermedia distribuida que permitiría a los usuarios preparar documentos electrónicos compuestos de enlaces a recursos de diferentes tipos repartidos por todo el mundo, creando un espacio global de hipertexto en el que la información puede ser referenciada por un simple UDI (Identificador Universal de Documento), Tim Berners-Lee lo llamó World Wide Web (WWW), esta idea es la base de lo que hasta el momento conocemos como internet (Cohen-Almagor, 2011).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

En las últimas dos décadas, el internet ha evolucionado gracias al nacimiento de nuevas tecnologías que permiten la conexión entre la información y personas, estas tecnologías permiten conexiones cada vez más sofisticadas de información, permitiendo generar relaciones más fuertes entre las personas, un ejemplo de esta idea es propuesta por Nova Spivack, un destacado visionario de internet, que propone un enfoque particular de como el internet ha ido evolucionando y probablemente a donde seguirá creciendo en los próximos años, este enfoque se muestra en la Figura 1 Evolución de Internet (Networks & Nova, 2007)..

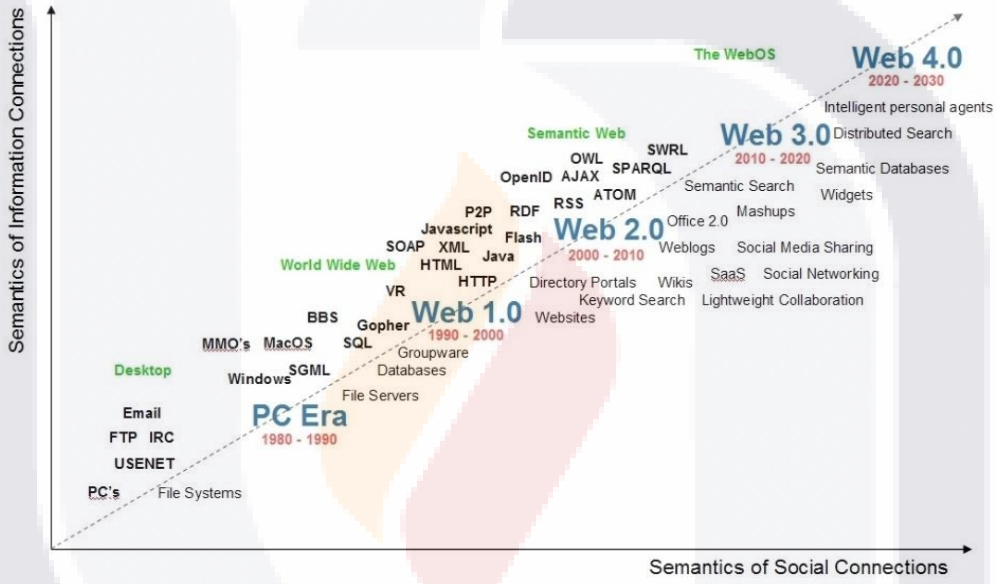

Figura 1 Evolución de Internet (Networks & Nova, 2007).

La mayoría de los autores identifican tres etapas principales en la evolución de internet, web 1.0, 2.0 y 3.0 (web semántica). La Web 1.0 es principalmente de solo lectura, una web estática y mono-direccional, las empresas pueden ofrecer catálogos para presentar sus productos y los usuarios pueden leerlos, estos sitios incluyen paginas HTML (HyperText Markup Language) estáticas que se actualizan con frecuencia, el objetivo de esta web es publicar información para cualquier persona en cualquier momento y establecer una presencia en línea. Los usuarios pueden solo visitar el sitio sin ningún impacto o contribución, las tecnologías más utilizadas en este web es HTTP (Hypertext Transfer Protocol), HTML y URI (Uniform Resource Identifier) (Sareh & Mohammad, 2012).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Web 2.0 es una red bidireccional que sirve como plataforma donde los usuarios pueden interactuar y disponen de un ambiente más flexible, diseños más sofisticados, reutilización de componentes, actualizaciones, creación y modificación de contenido colaborativo, donde esta última característica se vuelve la más representativa de esta etapa de la web y que sirve como repositorio de la denominada “inteligencia colectiva”, la inteligencia colectiva puede ser vista como una red de individuos, donde cada uno está especializado en una tarea en particular, por lo que no hay dos personas que compartan el mismo estado cognitivo (Halpin, 2008).

Esta inteligencia colectiva dio nacimiento a varias clases de plataformas de colaboración, como por ejemplo los blogs, wikis, video-sharing, redes sociales, etc. Creando así una amplia gama de sitos en internet de varios temas y varios recursos de distintos tipos.

La Web 3.0 o Web Semántica, inicialmente propuesta por Tim Berners-Lee, busca dotar de semántica a la información en internet, dando un significado comprensible por los humanos y las maquinas por igual, proponiendo una plataforma basada en varias tecnologías web que permiten el surgimiento de software especializado que mejore el rendimiento del usuario de la web.

1.2. Aprendizaje de la programación

El aprendizaje de la programación (Jenkins, 2002) es un tema ampliamente documentado desde diversas perspectivas. La línea de investigación de mayor relevancia se da en el contexto de la psicología cognitiva, la cual está directamente relacionada con la capacidad del cerebro humano para resolver problemas. Desde esta perspectiva se han comparado los modelos mentales (Bornat, Dehnadi, & Simon, 2008; Ma, Ferguson, Roper, & Wood, 2007) de programadores expertos y novatos, se han hecho interesantes aportaciones acerca de la creación y uso de estrategias estereotipadas de solución de problemas de programación (Rist, 1996, 2004), y en fechas recientes, se ha puesto especial atención al tema de la metacognición (Aleven & Azevedo, 2013; Azevedo, 2007; Flavell, 1979), tanto como habilidad que debe ser desarrollada, como por su faceta de estrategia pedagógica, que debe ser fomentada.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

El aprendizaje de la programación también ha sido estudiado en el ámbito conductual, en el que se han realizado esfuerzos por proponer y refinar constructos que puedan tener valor predictivo sobre el desempeño de los estudiantes, tales como la auto-eficacia, la motivación intrínseca y la percepción inicial de éxito (Downs & McAllen, 2012; Ramalingam, 2004; Schiefele, 1991)

Finalmente, pero no de menor importancia, existe una rica tradición de herramientas para apoyar el aprendizaje de la programación (Kelleher & Pausch, 2005) que proveen elementos interactivos y retroalimentación visual y que buscan de una u otra manera disminuir la carga cognitiva del estudiante. En esta línea de trabajo se pueden encontrar los primeros programas orientados a fomentar el aprendizaje de la programación, tales como LOGO (Begel, 1996), hasta los artefactos más recientes, como lo son los tutores cognitivos (Koedinger & Aleven, 2004).

El presente trabajo se enmarca en la última categoría, ya que un subproducto importante es el buscador semántico por sí mismo, pero de forma tangencial también contiene ciertos elementos de la psicología cognitiva, ya que la base de datos transformada a su equivalente semántico contiene información metacognitiva en forma de protocolos verbales, sobre la solución de problemas de programación por parte de expertos.

1.2.1. Uso de protocolos verbales para el apoyo al aprendizaje de la programación

Un protocolo verbal (PV) es un mecanismo para estudiar el contenido de la memoria de corto plazo de las personas en el proceso de resolver problemas o llevar a cabo una tarea predefinida (Ericsson & Simon, 1993). Un protocolo verbal provee elementos de análisis del proceso cognitivo de un programador, mostrando como éste selecciona una estrategia específica y como la adapta o descarta según sea el avance hacia los objetivos del programa.

El proceso de grabación de los PVs consiste en presentar un problema no ensayado previamente a una persona, solicitándole que “piense en voz alta” al momento de resolverlo. En tiempos recientes se usa software de captura para registrar videos y audio de verbalizaciones que se editan y posteriormente se analizan.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

En (Arévalo, Muñoz, & Gómez, 2011; Arevalo & Solano, 2012) se parte del principio de que los PVs pueden también utilizarse como material didáctico en el contexto del aprendizaje de la programación y de que una aplicación diseñada para visualizar protocolos de programadores en el proceso de resolver problemas de programación, puede ser una herramienta útil para los aprendices de esta disciplina, en el sentido de mostrar de forma explícita el comportamiento metacognitivo de un programador con experiencia.

Como prueba de concepto, se desarrolló un sistema para la visualización de PVs de programadores expertos[1], resolviendo problemas básicos de programación.

1.3. La búsqueda de información en internet

La búsqueda de información o "*information seeking*" son estrategias para encontrar información en internet, estas estrategias dependen de la clase, el propósito y el nivel de detalle necesario en la información, las principales estrategias son los siguientes según Levene Mark (Levene, 2010).

Navegación directa: esta estrategia consiste en ingresar directamente la URL del sitio en el navegador web, es la más simple y exitosa estrategia que podemos utilizar, si estamos buscando páginas principales de compañías, escuelas, sitios de e-commerce, etc. En contraste, puede ser la menos exitosa si estamos buscando productos en particular, ya que no necesariamente estos se asignan directamente en la URL, el usuario puede complementar esta estrategia con una estrategia de navegación después de ingresar a la página principal donde desea encontrar el producto.

Navegación dentro de un directorio: Un portal web es un sitio web que provee una puerta o punto de entrada a otros recursos en la web, estos portales proporcionan una gran cantidad de características y contenidos que regularmente se organizan en forma de directorios, por ejemplo Yahoo Directory (http://dir.yahoo.com) u Open Directory (http://www.dmoz.org/), estos portales constan de una categorización de temas, que incluyen por ejemplo, Arte, Negocios, Computación, Juegos, Salud, etc., y donde cada

[1] Ver: http://www.igualproject.org, http://aprende.igualproject.org/

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

uno de estos temas tiene subtemas formando una estructura de directorios en forma de árbol. Para encontrar un tema en específico se tiene que navegar en la jerarquía de directorios hasta encontrar una serie de sitios que hacen referencia a la categoría seleccionada y donde se tiene que investigar para encontrar información del tema deseado.

Navegación usando motores de búsqueda: Esta estrategia de búsqueda en los últimos años se ha vuelto la más utilizada, debido al creciente número de motores de búsqueda en la web. Los usuarios generalmente en esta estrategia iteran a través de los siguientes pasos:

1. *Formulación de consulta*: El usuario envía al motor de búsqueda una consulta que normalmente consta de una o varias palabras clave de entrada.
2. *Selección:* El motor de base de datos regresa una lista de resultados donde el usuario selecciona una página dando un clic sobre un enlace, la cual se mostrara al ser cargada en el navegador.
3. *Navegación ("Surfing"):* El usuario inicia una sesión de navegación, que es el proceso de hacer clic en los enlaces y navegar por las páginas mostradas utilizando diversas señales y herramientas para para aumentar su actividad de navegación.
4. *Modificación de consulta:* Esto sucede cuando el usuario decide reformular la consulta original y enviarla al motor de búsqueda, interrumpiendo de esta manera la sesión de navegación, en este caso el usuario regresa al paso 1.

1.3.1. Motor de búsqueda

Los motores de búsqueda en internet no podrían existir sin la ciencia de la búsqueda y recuperación de información (IRS), básicamente un IRS es un programa computacional para la consulta y almacenamiento de documentos. Generalmente, tales sistemas solo ayudan a sus usuarios a localizar y recuperar la información que necesitan (Alfredo, 2013).

Un enfoque general de como un IRS es construido para trabajar con una específica colección de documentos, que podría ser una librería digital, un conjunto de

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

documentos especializados en un área o en toda la Word Wide Web se muestra en la Figura 2.

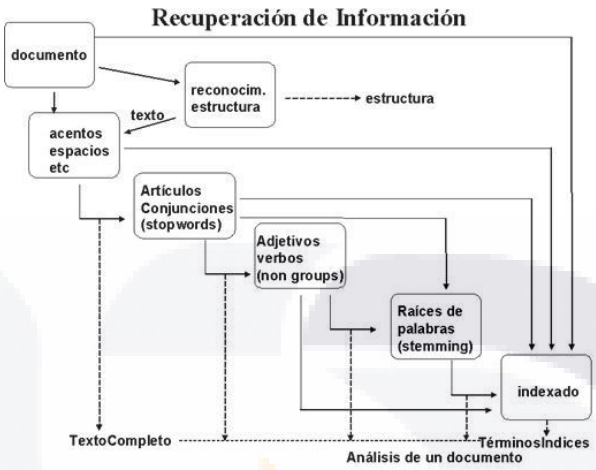

Figura 2 Procesamiento de un documento en un IRS (Proal, 2013).

Un usuario quien desea encontrar alguna información en la colección de documentos puede describir su requerimiento de información en una consulta, una consulta puede ser una larga sentencia o incluso en algunos casos un ejemplo de documento, pero usualmente es un contenido muy corto, en cualquier caso se utilizan procesos similares de indexación con el fin de "comprender", en cierta medida, lo que el usuario desea encontrar o de lo que trata el documento, estos procesos implican principalmente la extracción de palabras clave importantes que representan los principales contenidos (Nie, 2010) y debido a la gran cantidad de información en la web, algunos de los motores de búsqueda actuales utilizan esta técnica.

Un motor de búsqueda es una página web que ayuda al usuario a encontrar información relevante, una vez que este encuentra una o varias páginas presenta un resumen, donde el usuario puede seleccionar que paginas desea visitar y navegar entre ellas siguiendo los links.

La interfaz principal de Google se muestra en la Figura 3, este motor de búsqueda es uno de los líderes en el mercado, donde se presenta una caja de texto alargada, el usuario ingresa su consulta de información, en este caso queremos encontrar información acerca de "problemas de programación" y es lo que se escribe en la caja de texto.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Figura 3 Motor de búsqueda de Google (Google, 2014a)

Google responde instantáneamente, en promedio en 0.28 segundos, mostrando en toda la pantalla un listado de resultados, como se muestra en la Figura 4 Resultados de búsqueda en Google, en este caso Google muestra alrededor de 2.68 millones de resultados relevantes acerca de problemas de programación ordenados de acuerdo a la relevancia de la página.

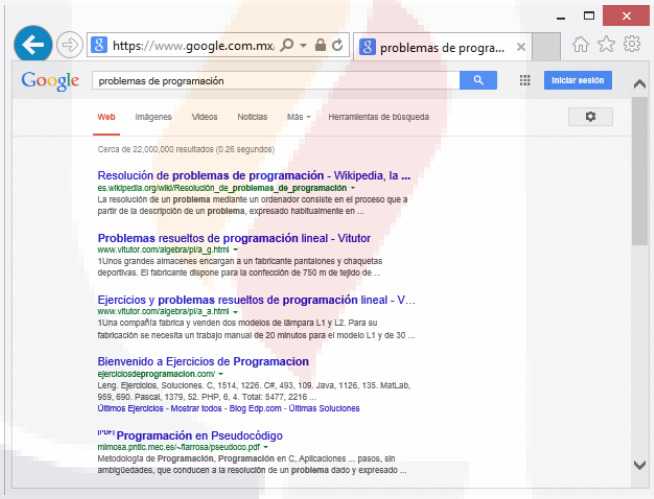

Figura 4 Resultados de búsqueda en Google (Google, 2014b)

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

1.4. La web semántica

La web semántica permite buscar no solo información, si no también conocimiento. El objetivo principal es introducir estructuras y contenido semántico en la enorme cantidad de conocimiento estructurado y no estructurado disponible en la web, enfocándose en la creación de ontologías que describen los conceptos y sus relaciones en un determinado campo de conocimiento (Badr, 2010).

La web semántica fue concebida en los inicios de la web en 1989 por Tim Berners-Lee, pero su desarrollo hasta hace unos años había quedado en un segundo plano, a pesar que en la primera conferencia de la W3C en 1994 ya se hablaba de diferentes estándares web y pautas como URI, XML y RDF.

A partir del 2001 cuando la W3C definió formalmente la web semántica, las tecnologías inmersas han evolucionado progresivamente, permitiendo que el concepto de web semántica sea una realidad cada vez más tangible. La web semántica se compone de varias capas de estándares, iniciando con “unicode” cuyo objetivo es proporcionar el medio por el cual un texto en cualquier idioma puede ser codificado para el uso informático, en el mismo nivel se encuentran las cadenas de caracteres conocidas como “URIs” que permiten identificar y acceder a cualquier recuerdo de la web, en la capa siguiente “XML+NS+xmlschema” son tecnologías que posibilitan la comunicación entre agentes, XML (Extensible Markup Language) para el intercambio de documentos, NS (Namespaces) proporciona un método para cualificar elementos y atributos de nombres, XML Schema como lenguaje para describir la estructura y registrar el contendió de documentos XML.

La siguiente capa es denominada “RDF+rdfschema”, donde RDF sirve como lenguaje que define un modelo de datos para describir recursos mediante triples sujeto - predicado – objeto y RDF Schema es un vocabulario RDF que permite una descripción de recursos con un enfoque orientado a objetos. La capa media de la estructura es la de las “ontologías” que ayuda en la clasificación de la información, extendiendo la funcionalidad de la Web Semántica agregando nuevas clases y propiedades para describir recursos.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

La siguiente capa es la “lógica”, que ayuda a precisar reglas de inferencia y posterior a esta capa es la de “pruebas”, donde se intercambian pruebas escritas en el lenguaje unificador de la Web Semántica, posibilitando las inferencias lógicas, la capa superior de la estructura es la de “confianza” donde los agentes deberían ser muy escépticos de la información que leen en la web. Otra capa que tendrá que ser contemplada es la capa de “firmas digitales” que será utilizada para verificar si la información ha sido ofrecida por una fuente de confianza, la estructura de estas tecnologías se muestra en la Figura 5.

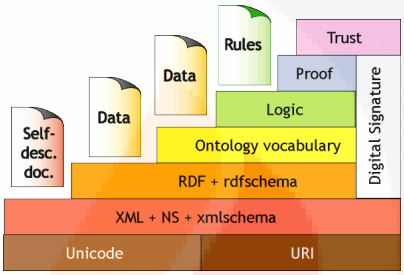

Figura 5 Tecnologías inmersas en la web semántica (Berners-Lee, 2000)

La web semántica es una red de información comprensible por maquinas, cuyo significado es bien definido por medio de estándares, es absolutamente necesaria la infraestructura interoperable que solo los protocolos estándar globales pueden proporcionar (Fensel, 2005).

Un concepto fundamental de la web semántica es la “búsqueda semántica”, la cual puede ser descrita como el esfuerzo para mejorar la exactitud del proceso de búsqueda basado en el entendimiento del contexto y la limitación de la ambigüedad (De Virgilio, Guerra, & Velegrakis, 2012), este concepto es ampliamente usado para referirse a diferentes enfoques de sistemas como (Elbedweihy & Wrigley, 2012):

- “gateways” usado para localización de ontologías y documentos.
- Razonamiento sobre la información que se encuentra dentro de los documentos y ontologías.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

- Interfaces que permiten a los usuarios explorar el espacio de búsqueda mientras formulan sus consultas.
- Mashups que integran datos de diferentes fuentes para proveer una descripción precisa acerca de objetos de la web semántica.

1.4.1. Buscador semántico

La base de un buscador semántico es la interpretación del conocimiento subyacente de la información en internet, este conocimiento es expresando dentro de una o varias ontologías. Los conceptos definidos en las ontologías se identifican en el texto de cada documento regularmente expresadas en algún lenguaje de etiquetado, por lo que la fase de mapeo es claramente crucial para la calidad de los resultados (Bry & Małuszyński, 2009), además otra herramienta importante para la búsqueda en la web semántica es el concepto de contexto y su correspondencia con las ontologías.

Las diferencias fundamentales de los motores de búsqueda semántica con respecto a los motores de búsqueda tradicionales son principalmente (Sheykh Esmaili & Abolhassani, 2006):

- Uso de un marco lógico que permite una recuperación más inteligente.
- Existen relaciones más complejas en los documentos resultantes, según la importancia del tema, el mantenimiento de los metadatos y la clasificación.
- La especificación de las relaciones entre los objetos destaca la necesidad de mejores técnicas de visualización de resultados.

1.4.2. Estado actual de los buscadores semánticos.

Al momento existen varios motores de búsqueda semántica en la web entre los cuales los más destacados son los siguientes:

Swoogle: Es un proyecto de investigación sin fines de lucro, llevado a cabo por el grupo de investigación Ebiquity en el Departamento de Ciencias de la Computación e Ingeniería Eléctrica en la Universidad de Maryland, condado de Baltimore (UMBC). La interfaz del buscador semántico se muestra en la Figura 6 Interfaz de Swoogle y su última

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

actualización se realizó en el 2007, Este buscador rastrea en los documentos web RDF y ofrece los siguientes servicios:

- Búsqueda de ontologías en la web semántica.
- Buscar datos en instancias de la web semántica.
- Búsqueda de terminologías de la web semántica como URIs, clases y propiedades.
- Archivar diferentes versiones de documentos de la web semántica.

Figura 6 Interfaz de Swoogle (Swoogle, 2014a)

Actualmente solo están indexados algunos metadatos de los documentos de la web semántica y para realizar una búsqueda es necesario conocer la sintaxis de las ontologías inmersas en el contexto de la información, por lo que se requiere un conocimiento previo del lenguaje de ontologías, por ejemplo en la Figura 7 Ejemplos de búsquedas en Swoogle, se muestran algunos ejemplos de búsquedas que se pueden hacer en el motor, demostrando que en cierta medida carecer de características propias del uso de lenguaje natural (“Swoogle Semantic Web Search,” 2007).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

url:foaf	search documents having "foaf" as part of their URLs
url:"http://www.w3.org/2000/01/rdf-schema"	search a particular SWD with the URL "http://www.w3.org/2000/01/rdf-schema"
desc:timbl	search documents having "timbl" in their document annotations
def:food	search documents explicitly defining the term(classes/properties) that include a token "food". Note that the term food is case-insensitive.
ref:food	search documents implicitly defining the term(classes/properties) that include a token "food". That is, the term being a class/property is deduced by the domain and range definition of RDF/RDFS and OWL predicates. Note that in the results returned by Swoogle search, only the fields 'desc' and 'def' are highlighted, the field 'ref' and 'pop' are neglected for a neat output.
pop:person	search documents that populate the class with a token of 'person' with instances OR use the property with a token of 'person' as predicates
pop:knows	search documents that populate the class with a token of 'knows' with instances OR use the property with a token of 'knows' as predicates

Figura 7 Ejemplos de búsquedas en Swoogle (Swoogle, 2014b)

Lexxe Beta: Fundado en 2005, Lexxe ha sido el desarrollo de un motor de tercera generación con tecnologías de procesamiento del lenguaje natural, la interfaz del motor se muestra en la Figura 8 Interfaz Lexxe Beta. Lexxe ha estado explorando maneras más inteligentes para encontrar información para los usuarios, el 2011 lanzo su versión beta y siguen trabajando en la mejora de sus técnicas de búsqueda.

Figura 8 Interfaz Lexxe Beta (Lexxebeta, 2014b)

Este buscador semántico utiliza una palabra clave "Semantic Key Word" para determinar el área de búsqueda o contexto de búsqueda, seleccionando primero de una lista la palabra clave y en seguida el complemento de la oración como se muestra en Figura 9 Ejemplo búsqueda en Lexxe, un ejemplo de listado de llaves semánticas se muestra en la Figura 10 Llaves semánticas Lexxe.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Figura 9 Ejemplo búsqueda en Lexxe (Lexxebeta, 2014a)

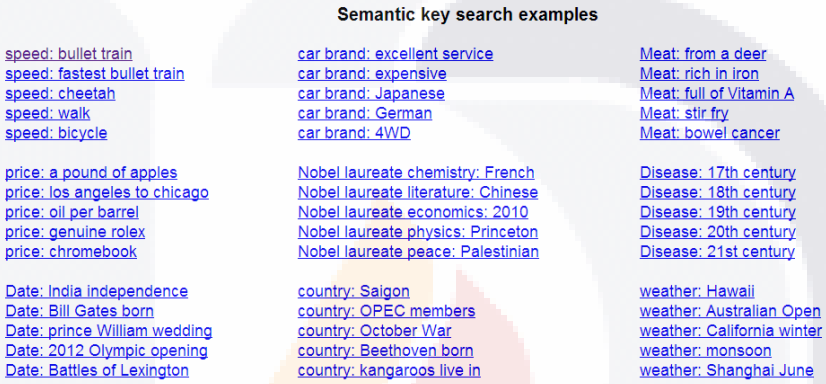

Figura 10 Llaves semánticas Lexxe (Lexxebeta, 2014c)

WolframAlpha: Es un motor de búsqueda semántico que tiene la capacidad de responder a preguntas expresadas en lenguaje natural, dando una respuesta puntual y en caso de posibles ambigüedades presenta opciones para ser más específico, la interfaz del buscador se muestra en la Figura 11 Interfaz WolframAlpha, donde se puede ver una respuesta puntual a la búsqueda sobre la longitud total de todos los caminos en España. Y en la Figura 12 Opción de contexto WolframAlpha se muestra un contexto específico para resolver las ambigüedades de la expresión en lenguaje natural.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

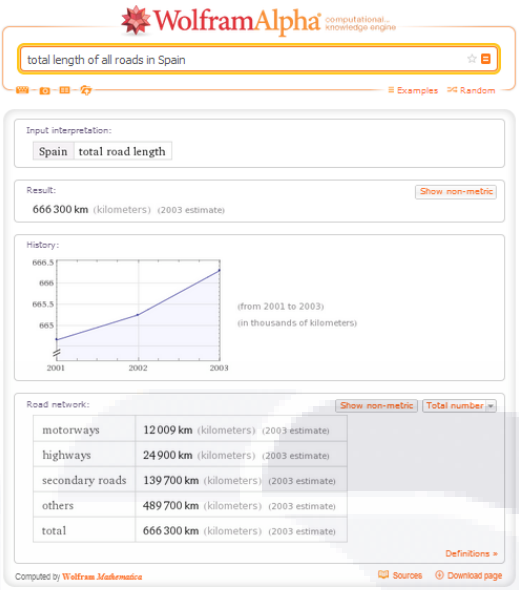

Figura 11 Interfaz WolframAlpha (Wolframalpha, 2014a)

Figura 12 Opción de contexto WolframAlpha (Wolframalpha, 2014b)

La meta de este buscador es hacer todo el conocimiento sistémico inmediatamente computable y accesible a todos, reuniendo y conservando todos los datos clave aplicando cada modelo, método y algoritmo conocido, permitiendo calcular lo que se pueda calcular de cualquier cosa, construyendo sobre los logros de la ciencia y otras sistematizaciones del conocimiento para ofrecer una única fuente que pueda ser invocada por todos (“WolframAlpha,” 2014).

2. Problemática

Actualmente para los usuarios de internet existen servicios de varias clases totalmente gratis, desde tener su propio sitio web, blog, wiki y redes sociales, hasta tener sus aplicaciones en la nube con un almacenamiento ilimitado. Todos estos servicios disponibles están provocando un crecimiento exponencial de internet como se muestra en

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

la Figura 13 Número de sitios web . Según Netcraft existe un rápido crecimiento en los últimos 10 años, desde tener 21,3 millones de sitios web en 2004 hasta tener 182,1 millones en abril del 2014, representando un crecimiento del 850%.

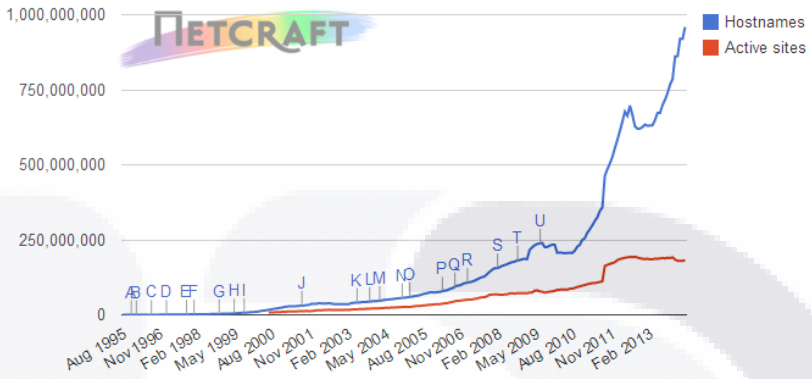

Figura 13 Número de sitios web (Netcraft, 2014)

Este crecimiento no toma en cuenta el concepto de “deep web” que hace referencia a los datos contenidos en las base de datos y donde estos no son directamente accesibles por los motores de búsqueda, ya que se requieren de interfaces web que hagan consultas en la base de datos y muestren los resultados, existen estadísticas sobre el número de base datos conectadas a la web, donde aproximadamente eran 0.45 millones en el 2004 (Levene, 2010).

Internet se ha convertido en el repositorio de información más grande de todo el mundo y encontrar información en este repositorio se ha vuelto en una tarea cada vez más complicada, debido a que la mayoría de la información se encuentra poco estructurada u organizada. Como se muestra en la Figura 14 Factores de dificultad en la Búsqueda de Información en la web, existen numerosos obstáculos para buscar información de manera eficaz, estos obstáculos van desde problemas de experiencia hasta temas de diseño de interfaces. A un porcentaje significativo de usuarios web les cuesta mucho tiempo buscar información como documentos específicos o páginas web determinadas. Otros temas se relacionan directamente con la dificultad de recuperar información útil y comprensible.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

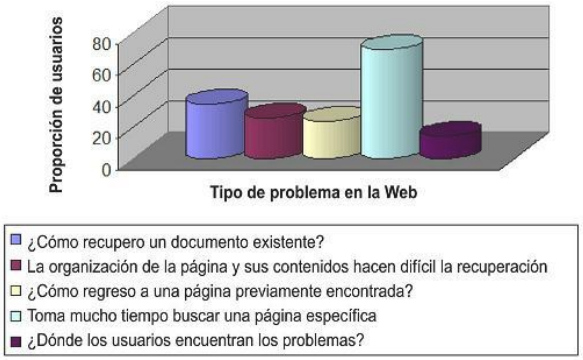

Figura 14 Factores de dificultad en la Búsqueda de Información en la web (Anita Ferreira & Atkinson, 2013)

Para poder medir realmente la necesidad de nuevos mecanismos de búsqueda de información, debemos tomar en cuenta la cantidad de datos aunado a los problemas presentes en la recuperación en la web, además es preciso conocer el número de búsquedas puntuales que se realizan en los motores de búsqueda actuales, en Google Sites se realizan 13,067 millones de búsquedas al mes, representando el 67.5% de las búsquedas en internet en Marzo del 2014, esto nos muestra una perspectiva más amplia de la cantidad de veces que los usuarios intentan encontrar información en la web, como se muestra en la Figura 15 Búsquedas explicitas relativas y Figura 16 Búsquedas explicitas absolutas .

comScore Explicit Core Search Share Report*
March 2014 vs. February 2014
Total U.S. – Home & Work Locations
Source: comScore qSearch

Core Search Entity	Explicit Core Search Share (%)		
	Feb-14	Mar-14	Point Change
Total Explicit Core Search	*100.0%*	*100.0%*	*N/A*
Google Sites	67.5%	67.5%	0.0
Microsoft Sites	18.4%	18.6%	0.2
Yahoo Sites	10.3%	10.1%	-0.2
Ask Network	2.4%	2.5%	0.1
AOL, Inc.	1.3%	1.3%	0.0

Figura 15 Búsquedas explicitas relativas (comScore, 2014)

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

comScore Explicit Core Search Query Report
March 2014 vs. February 2014
Total U.S. – Home & Work Locations
Source: comScore qSearch

Core Search Entity	Explicit Core Search Queries (MM)		
	Feb-14	Mar-14	Percent Change
Total Explicit Core Search	*17,687*	*19,358*	*9%*
Google Sites	11,941	13,067	9%
Microsoft Sites	3,257	3,594	10%
Yahoo Sites	1,822	1,960	8%
Ask Network	431	478	11%
AOL, Inc.	235	259	10%

Figura 16 Búsquedas explicitas absolutas relativas (comScore, 2014)

Todos estos datos nos dejan ver la necesidad de una evolución en la web, que haga uso de nuevos métodos de búsqueda y organización de información, creando así una mejor experiencia del usuario y permitiendo generar nuevas herramientas que ayuden aún más en la difusión del uso de internet.

2.1. Limitantes de la web semántica

La web semántica es una web basada en ontologías que consta de almacenar datos con significados formales, esta web contrasta con la web actual que contiene básicamente información limitada en forma de documentos, por lo que hacer una transformación de la información ya existente en la web a una representación semántica podría resultar una tarea bastante compleja.

Todas las nuevas tecnologías al ser introducidas tienen que hacer frente a ciertas dificultades de distintas índole, la web semántica y los sistemas automatizados de razonamiento tienen que hacer frente a estas dificultades para cumplir la promesa de la web semántica, algunas de las principales dificultades para esta web son (Pandey, 2012):

- Desarrollo de ontologías.
- Semánticas formales en lenguajes apropiados.
- Prueba y confianza.
- Disponibilidad de contenidos.
- Escalabilidad.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

- Multilingüismo.
- Visualización.
- Estabilidad de los lenguajes de la web semántica.

El desarrollo de la web semántica depende de las ontologías y hay varios aspectos que se deben tomar en cuenta como los lenguajes para su representación, los enfoques de aprendizaje y los sistemas de bibliotecas de ontologías que gestionan, adaptan y estandarizan.

Referente a la confianza en la información de la web semántica, existen diferentes enfoques y la mayor preocupación es la falta de una regulación o especificación formal para proporcionar esta característica, ya que actualmente cualquiera puede decir lo que sea de quien sea y puede provocar conflictos y contradicciones a través de los puntos de vista de las personas alrededor del mundo, un ejemplo de esto es la información médica en internet donde el engaño y el peligro de una incorrecta interpretación o el mal uso están muy presentes.

Por otro lado, la mayoría de las páginas web están escritas en HTML, y esta tecnología no soporta anotaciones semánticas, por lo que se está dando un cambio importante en este sentido, al utilizar tecnologías como RDF o HTML5 para hacer estas anotaciones, pero esta tarea resulta un tanto tediosa para los programadores, ya que de momento no ven una ventaja significativa en dedicar tiempo a crear estas anotaciones semánticas, cuando pueden únicamente hacer uso del fácil y probado HTML.

2.2. Bases de datos relacionales.

Una base de datos es una colección de datos organizados y estructurados según un determinado modelo de información que refleja no sólo los datos en sí mismos, sino también las relaciones que existen entre ellos. Una base de datos se diseña con un propósito específico y debe ser organizada con una lógica coherente. Los datos podrán ser compartidos por distintos usuarios y aplicaciones, pero deben conservar su integridad y seguridad al margen de las interacciones de ambos. La definición y descripción de los datos han de ser únicas para minimizar la redundancia y maximizar la independencia en su utilización.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Los modelos clásicos de tratamiento de los datos son:

- Jerárquico.
- En red.
- Relacional

Donde este último es el modelo más utilizado, ya que permite una mayor eficacia, flexibilidad y confianza en el tratamiento de los datos. La mayor parte de las bases de datos y sistemas de información actuales se basan en el modelo relacional ya que ofrece numerosas ventajas sobre los dos modelos anteriores, como es el rápido aprendizaje por parte de usuarios que no tienen conocimientos profundos sobre sistemas de bases de datos (Lamarca, 2013).

Muchos contenidos en la web tienen información no rastreable, oculta atrás de interfaces de búsqueda, donde la información solo es descubierta por la ejecución de una consulta específica a una base de datos, este tipo de información es conocida como Web Profunda o “Deep-Web” y se debe a la falta de capacidad por parte de los motores de búsqueda convencionales de indexar información almacenada en bases de datos. La disponibilidad de información pública en la llamada “Deep-Web” es alrededor de 550 veces mayor que en el Web Superficial, lo que comúnmente llamamos World Wide Web (Lourdes & Carro, 2004).

2.3. Limitantes de los motores de búsqueda actuales.

Hay varios problemas con la recuperación de información en internet y los motores de búsqueda se enfrentan a ellos con mayor frecuencia, el primero de ellos es lo abierto que es internet, permitiendo un constante cambio, nuevos sitios aparecen, los sitios viejos cambian o desaparecen y en general el contenido es emergente y no previsto, esto implica que los resultados no sean estables y que el usuario adapte varias estrategias para mejorar sus resultados. Otra limitante es la calidad variante de la información en la web, donde el usuario tiene que hacer un juicio para discriminar la calidad de los resultados ya que por lo regular los motores de búsqueda no toman en cuenta este parámetro. Por último el alcance de la web actual no es fijo y en muchos casos no sabemos de antemano si la información está disponible y se crea un ambiente de

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

incertidumbre al no encontrar lo que estamos buscando en el primer intento, hasta llegar en algunos casos a tener varias sesiones de búsqueda o simplemente darnos por vencidos (Levene, 2010) .

2.4. Proceso actual de búsqueda en el sitio de protocolos verbales.

Actualmente el método de búsqueda implementado en el sitio web de protocolos verbales está basado en el uso de palabras clave las cuales son anexadas de manera manual al dar de alta un protocolo en el sistema, este proceso puede estar sujeto a subjetividad por parte del editor del protocolo y puede ser ineficiente desde el punto de vista de la calidad de los resultados de la búsqueda.

La interfaz para crear un problema dentro del sitio visor de protocolos se muestra en la Figura 17 Interfaz para alta de problema, donde en el campo palabras clave, se listan todas las palabras representativas del problema.

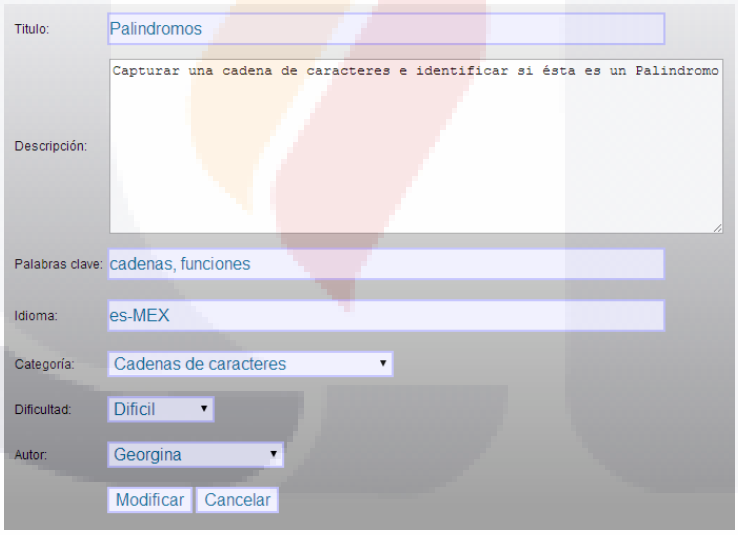

Figura 17 Interfaz para alta de problema (Sistema Visor de Protocolos Verbales, 2014)

Posteriormente el usuario puede hacer búsquedas en base a estas palabras clave, donde la interfaz del buscador implementado actualmente en el sitio web de protocolos verbales se muestra en la Figura 18 Interfaz buscador por palabra clave (Sistema Visor de Protocolos Verbales, 2014).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

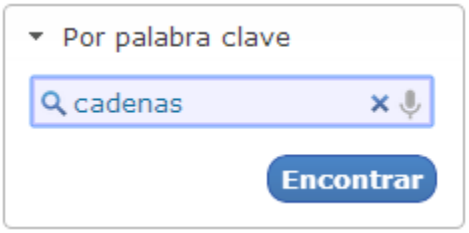

Figura 18 Interfaz buscador por palabra clave (Sistema Visor de Protocolos Verbales, 2014)

3. Formulación del problema de investigación

La presente investigación se enfoca en el área de la búsqueda de información en internet, tomando en cuenta la evolución de la web con ayuda de nuevas tecnologías para el manejo de notaciones semánticas en la información

Teniendo como fin medir y describir el proceso de transformación de una base de datos relacional a una representación semántica, en base a la construcción de un prototipo de buscador semántico, delimitado a una particular área de conocimiento expresada en una base de datos relacional.

Para la construcción de este prototipo se propone una metodología que describe los pasos técnicos necesarios para su construcción, finalizando con la aplicación de un test, que proporcione una aproximación de usabilidad del buscador semántico.

3.1. Tipo de investigación

Para esta tesis se emplearon dos clases de estudios, estudio exploratorio y descriptivo, comenzando como exploratorio al examinar las tecnologías inmersas en la Web Semántica, en particular las relacionadas al proceso de conversión relacional-semántico y a la construcción de un motor de búsqueda semántica, identificando conceptos y herramientas esenciales en proceso de representación de información semántica en la web.

Posteriormente se realizó un estudio descriptivo que especifica las propiedades de cada tecnología, proponiendo una metodología para construcción de un buscador semántico basado en datos relacionales y aplicándola a un caso específico, describiendo

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

el proceso de construcción y tomando mediciones tanto del esfuerzo en horas requerido para la transformación y la usabilidad del prototipo de buscador semántico.

3.2. Objetivos de investigación

Este estudio de tesis está enfocado en brindar una descripción detallada del proceso de construcción de un buscador semántico, delimitado a un sitio web de apoyo para el aprendizaje de la programación basado en protocolos verbales, donde se tomó una base de datos relacional, se hizo la transformación a una versión semántica y se construyó un prototipo de búsquedas semánticas.

3.2.1. Objetivo general

El objetivo general de este estudio de tesis es el siguiente:

"Describir y evaluar las implicaciones técnicas y medir las ventajas en cuanto a usabilidad y esfuerzo, de la migración de una aplicación bajo RDBMS, hacia un modelo semántico basado en RDF, mediante la creación de un prototipo de buscador semántico."

3.2.2. Objetivos específicos

En el desarrollo de esta tesis se busca cumplir con los siguientes objetivos específicos que son planteados de acuerdo a ciertos puntos clave dentro de la metodología.

- Identificar herramientas para migrar de RDBMS a RDF.
- Aplicar la herramienta más adecuada al caso específico de protocolos verbales creando un repositorio RDF.
- Medir el esfuerzo del proceso de migración RDMBS a RDF.
- Identificar herramientas para explotar el repositorio RDF mediante SPARQL.
- Adaptar la herramienta más adecuada al repositorio RDF.
- Crear una interfaz que interprete expresiones en lenguaje natural y construya la correspondiente consulta SPARQL.
- Medir la usabilidad del prototipo de búsquedas semánticas.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

3.3. Preguntas de investigación

Basado en los objetivos específicos planteados en sección anterior se proponen las siguientes preguntas de investigación que serán contestadas en el desarrollo de esta tesis.

- ¿Qué herramientas existen para migrar RDBMS a RDF?
- ¿Cuál herramienta es más compatible con el caso específico de protocolos verbales?
- ¿Qué pasos se requieren para transformar de un paradigma relacional a uno semántico, construyendo un repositorio RDF?
- ¿Qué herramientas existen para hacer consultas SPARQL en repositorios RDF?
- ¿Qué características debe tener el prototipo de búsquedas semánticas?
- ¿Cuál es el esfuerzo del proceso de migración de RDBMS a RDF?
- ¿Qué grado de usabilidad tiene el prototipo de búsquedas semánticas?

3.4. Justificación

Hoy en día el internet ha crecido de una manera desordenada, provocando que las estrategias de búsqueda cada vez sean menos eficientes debido a la variedad de contenidos. Actualmente existe un nuevo paradigma para representar recursos web que permite explotar de una manera más sistemática la información contenida en internet.

Este nuevo paradigma es denominado Web 3.0 o Web Semántica y consiste en dar un significado a la información por medio de anotaciones basadas en ontologías, de esta manera la información puede ser comprendida por las maquinas, dando paso a una nueva era de herramientas web.

Una de las herramientas web que está evolucionando gracias a este nuevo paradigma son los motores de búsqueda, creando versiones semánticas que se valen de estos recursos para ofrecer servicios más robustos y eficientes. Actualmente existen varios motores de búsqueda semántica, pero la mayoría sigue en fases de desarrollo o primeras versiones.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Debido a esta evolución de la web se cree importante proponer una metodología para la construcción de un prototipo de buscador semántico, en este contexto el proceso de dotar de sentido semántico a la información en internet en particular la información inmersa en bases de datos, es un proceso sumamente complejo y el practicante que opte por seguir esta ruta, se enfrenta a muchos obstáculos y términos nuevos. Al proporcionar un proceso descriptivo, se busca tanto reducir el esfuerzo implicado, como sugerir alternativas de mejora que perfeccionen el proceso de construcción de un buscador semántico.

Este prototipo de buscador semántico será delimitado a un sitio web de apoyo al aprendizaje de la programación por medio de protocolos verbales. La metodología en primera instancia se enfoca a la transformación de una base de datos relacional a su representación semántica, posteriormente en la búsqueda he implementación de una herramienta que ayude en la manejo de esta representación semántica y finalmente en la construcción de una interfaz que interprete oraciones en lenguaje natural y presente los resultados más adecuados. Proporcionando una descripción detallada de todo el proceso permitiendo replicar este estudio en investigaciones futuras, aportando de esta manera un primer avance en el proceso de adaptación a esta nueva web semántica.

4. Marco Teórico

La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante (World Wide Web Consortium, 2014).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.1. Web semántica e internet

La Web Semántica es la siguiente evolución de la web 2.0. Hablar de semántica en la web implica hablar del significado de los datos y como puede ser descubierto no solo por las personas sino también por las computadoras. Este advenimiento tecnológico está provocando un replanteamiento de enfoques sobre las herramientas web disponibles hoy en día, algunos de estos enfoques son descritos a continuación (Passin, 2004).

- **Datos de lectura mecánica:** La idea es tener datos definidos y enlazados de una manera que puedan ser utilizados por maquinas no solo para mostrar, si no para automatizar, integrar y reutilizar atreves de varias aplicaciones.
- **Agentes inteligentes:** la idea es hacer de la web más comprensible para las maquinas, para permitir a los agentes inteligentes recuperar y manipular información.
- **Base de datos distribuida:** Referente a proveer una flexibilidad que permita representar todas las bases de datos y reglas lógicas vinculadas entre ellas para generar un valor.
- **Infraestructura automatizada:** La web semántica es vista más como una estructura y no como una aplicación, donde el verdadero problema de la web actual es la falta de un sencillo marco de trabajo para describir recursos y sus relaciones.

La W3C ha sido una organización líder en el desarrollo de tecnologías web, y su primer enfoque sobre la evolución a la web semántica propuesta por Tim Berners-Lee se muestra en la Figura 19 Estructura original en capas de la Web Semántica (Passin, 2004), donde se inicia con la capa de autentificación y fiabilidad de las declaraciones, capa referente a la verdad de las declaraciones y la inferencia de hechos no declarados, capa de vocabularios y significados compartidos, capa de tipos de recursos RDF, Capa de metadatos, capa de estructura y tipos de datos y por ultimo capa de sintaxis común.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

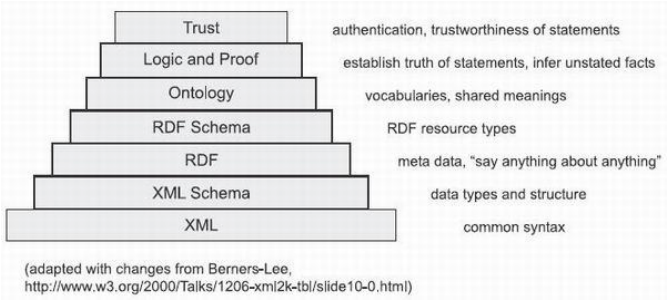

Figura 19 Estructura original en capas de la Web Semántica (Passin, 2004)

Posteriormente con la evolución y aparición de nuevas tecnologías web se ha ido modernizando la estructura y en la Figura 20 Estructura en capas reciente de la Web Semántica, muestra la estructura actual de la web semántica donde se conserva la idea original pero se refuerza con nuevas tecnologías y conceptos.

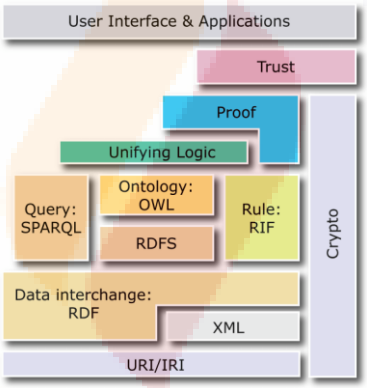

Figura 20 Estructura en capas reciente de la Web Semántica(Passin, 2004)

En general la estructura de la web semántica está compuesta de varias tecnologías como lo son:

- **XML (Estensible Markup Language):** Lenguaje de marco de trabajo usado para definir lenguajes que son usados para intercambio de datos en la web.
- **XML Schema:** Lenguaje usado para definir una estructura especifica en XML.
- **RDF (Resource Description Framework):** Un lenguaje flexible capaz de describir el orden de la información y metadatos. Aporta una semántica básica para el modelo de datos.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

- **RDF Schema:** vocabulario para describir propiedades y clases de los recursos RDF.
- **Ontology:** Lenguaje usado para definir vocabularios y establecer el uso de palabras y términos en el contexto de un vocabulario específico, OWL es un lenguaje para definir ontologías mediante la descripción detallada de propiedades y clases.
- **Lógica y pruebas:** El razonamiento lógico es utilizado para establecer la consistencia del conjunto de información e inferir conclusiones implícitas que son consistentes con el área de conocimiento de la información.
- **Confianza:** un medio para proporcionar autentificación de identidad y confiabilidad de los datos.

4.1.1. Representación semántica de una base de datos relacional

En la actualidad, existe un sinfín de bases de datos ligadas a la web, donde la información contenida en ellas pocas veces es utilizada de manera compartida y es de difícil acceso para motores de búsqueda debido a que no es información indexada. Esta información se le conoce como internet profunda o “Deep Web” y se calcula que es 500 veces más información que el contenido indexado en la web (Bergman, 2001).

Para poder exponer estos datos estructurados en la web almacenados en bases de datos relacionales se utiliza la descripción de recursos RDF. Las ventajas de crear estas vistas RDF se pueden resumir enlistando las tareas que facilitan (W3C, 2010):

- **Integración:** Se podrían enlazar información de distintos tipos, información financiera, geográfica, estadística, etc. Información con semántica incorporada y de esta manera se puede ver a internet como una gran base de datos.
- **Recuperación:** Una vez que los datos se publican en la web, las consultas pueden abarcar diferentes fuentes de datos y métodos de recuperación más efectivos pueden ser construidos.

Existen 3 enfoques para transformar una base de datos relacional a RDF, mapeo directo, mapeo directo más mapeo de ontología, base de datos a mapeo de ontología. Para este estudio de tesis se tomará el mapeo directo, básicamente consiste en el

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

subministro de una base de datos y una estructura de URIs para definir un grafo RDF que emule un esquema relacional, como se muestra en la Figura 21 Mapeo directo propuesto por el RDB2RDF Working Group.

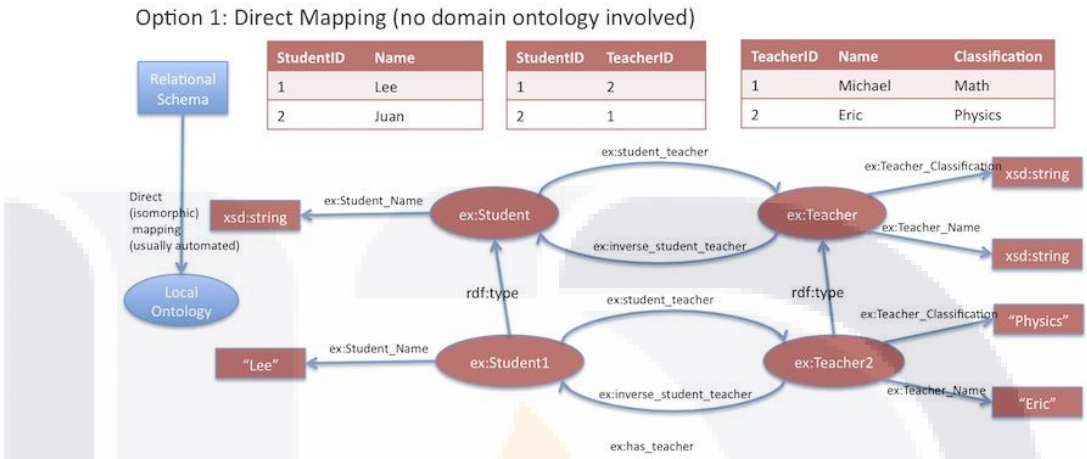

Figura 21 Mapeo directo propuesto por el RDB2RDF Working Group (W3C, 2010)

4.1.2. Avances de la web semántica

La Web Semántica es una estructura en capas que está en proceso de construcción, donde las primeras capas o capas base existen tecnologías bien probadas y con amplio desarrollo, por otra parte existen investigaciones y trabajos sobre las capas superiores de lógica y confianza, donde se intenta buscar mecanismos como el cifrado de información o el uso de firmas digitales para brindar estas características a la información en internet.

Actualmente la Web Semántica se está abriendo paso en varios campos de aplicación como las redes sociales, blogs, plataformas colaborativas, búsqueda de información, clasificación bibliográfica, aplicaciones industriales, investigación y desarrollo. Grandes empresas como Facebook, Google y BestBuy han hecho esfuerzos para incorporar la estructura de la Web Semántica en nuevos proyectos que ayuden a mejorar la colaboración entre sitios web (Kioskea.net, 2014).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.2. Extensible Markup Language (XML)

El lenguaje de marcas extensible (XML) es un sencillo formato de texto para representar información estructurada, como documentos, datos, configuración, libros, transacciones, facturas y mucho más. Fue derivado de un formato estándar antiguo llamado SGML (ISO 8879), con el fin de ser más adecuado para el uso en la web (Quin, 2014).

XML es un meta-lenguaje el cual puede ser usado como mecanismo para la representación de otros lenguajes de una forma estandarizada, XML describe el diseño de los datos de un documento y su estructura como un árbol de etiquetas anidadas. Los motores de búsqueda aprovechan esta estructura permitiendo buscar documentos donde las palabras clave y frases aparezcan dentro de los elementos del XML (Davies, 2006), por ejemplo buscar el nombre “Fletcher” dentro de todos los elementos “name” de un conjunto de documentos XML como se muestra en la Figura 22 Ejemplo de un documento XML.

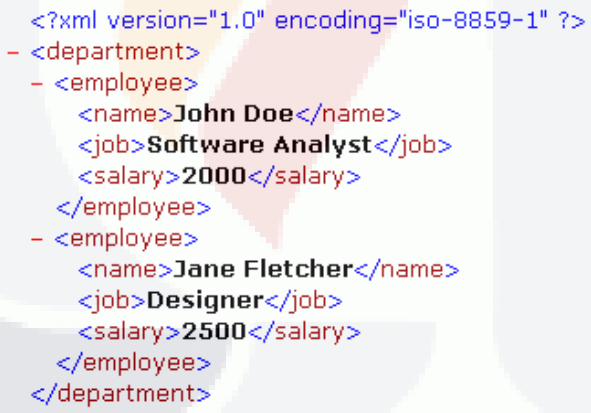

Figura 22 Ejemplo de un documento XML

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.3. Correlación de base de datos relacionales a RDF

Una de las maneras más comunes para publicar recursos semánticos en la web basado en una base de datos relacional, es haciendo uso de tecnologías como RDF y buenas prácticas como el correcto manejo de URIs, de esta manera se logra una descripción completa y flexible que puede ser insumo para diferentes herramientas informáticas como los agentes de software, esta forma de trabajo se muestra en la Figura 23 Representación general de un sistema .

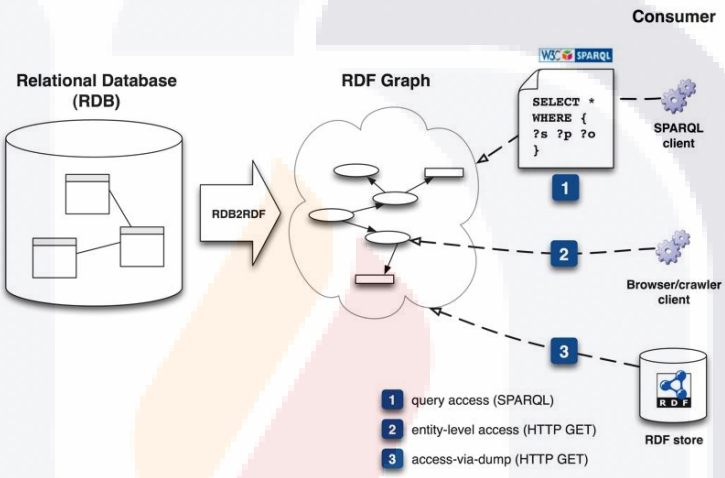

Figura 23 Representación general de un sistema RDB2RDF (W3C, 2010)

En este tipo de sistemas, el usuario puede consultar los grafos RDF de distintas maneras:

1. *Acceso por consulta:* Donde existe un agente de software que realiza una consulta SPAQRL en un punto de acceso al sistema y procesa los resultados, regularmente expresados en XML o JSON.
2. *Acceso a nivel de entidad:* Donde el agente de software realiza una petición HTTP GET a un URI expuesto por el sistema.
3. *Volcado de acceso:* Donde el agente de software realiza una petición HTTP GET y muestra todo el grafo RDF.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.4. Resource Description Framework (RDF)

En la web hay una gran cantidad de contenidos denominados “Web Sintáctica” basados en tecnologías como HTML, legible por los humanos pero no por las maquinas. Por otro lado, existe una gran variación de la calidad, oportunidad y pertinencia de los recursos web, haciendo difícil a los programas evaluar cada recurso.

La visión de la Web Semántica es aumentar la Web Sintáctica para que los sistemas informáticos puedan interpretar fácilmente los recursos, las mejoras se lograran a través de marcas semánticas que son anotaciones compresibles asociadas a los recursos web (Staab & Studer, 2009).

Para codificar marcas semánticas en la web, se adopta una notación de lenguaje desarrollado por el W3C llamado RDF (Marco de Descripción de Recursos), RDF se basa en los estándares de URIs y Unicode, teniendo un sintaxis similar a XML permite escribir meta-información para cualquier tipo de datos y trabajar con ellos, facilitando la interoperabilidad entre distintas aplicaciones sin pérdida de significado de los datos (“RDF y RDF schema,” 2013).

Una función de los archivos RDF es describir ontologías previamente definidas en un vocabulario ya sea RDF Schema u OWL, una ontología básica se compone de tres características principales como se muestra en la Figura 24 Ontología básica, en donde el sujeto es el recurso, el predicado es un arco que representa las propiedades o la relación. El objeto es el valor de la propiedad o el otro recurso con el que se establece la relación.

Figura 24 Ontología básica (W3C, 2004)

Tomando como ejemplo la Figura 25 Grafo RDF, Se presenta una visión general de cómo se estructura un archivo RDF, en este caso podemos ver que el URL http://example.org/Ganesh.html es la página web del recurso Ganesh, el cual es un

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

elefante y come pasto, en si este ejemplo no explica el significado de un elefante, por ello, existen otros lenguajes como RFD Schema y OWL que ayudan dotando de semántica a los recursos.

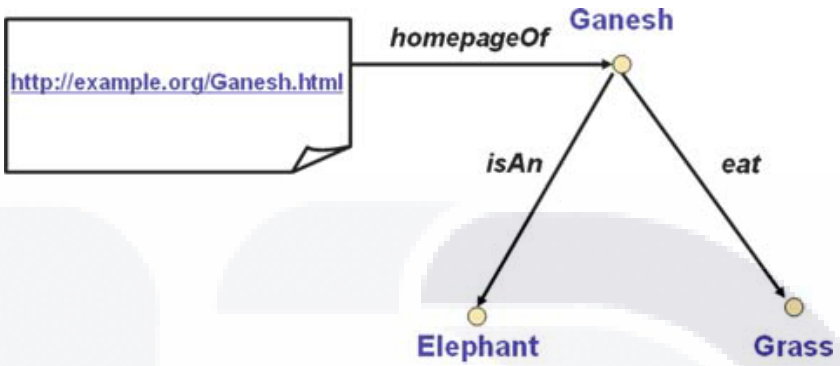

Figura 25 Grafo RDF (Pan, 2004)

RDF describe recursos en términos de propiedades con nombre, los valores de las propiedades con nombre son llamados objetos y pueden ser URIs de un recurso web o literal, un ejemplo de este tipo de representación es presentado en la Figura 26 Ejemplo de representación de un enunciado en RDF, donde podemos observar que la flecha apunta hacia el objeto y se representa la el enunciado “la página web http://www.example.org/index.html fue creada por Carlos y el día 30 de febrero”.

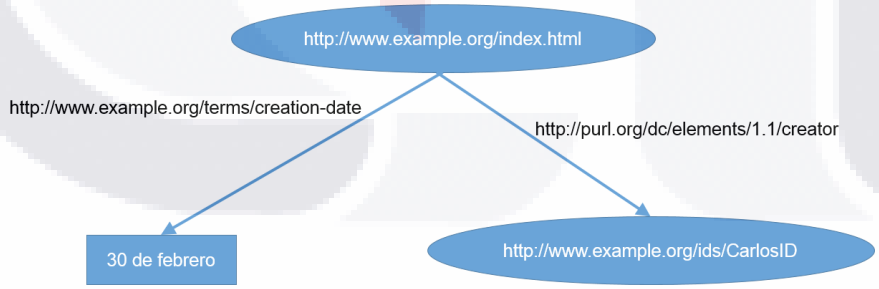

Figura 26 Ejemplo de representación de un enunciado en RDF

En el ejemplo, Carlos resulta ser un recurso porque podemos decir cosas sobre él, como su dirección, la que se convertiría en información de la información que estamos dando sobre la página. Cuando existe más de una sentencia RDF se le conoce como

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Grafo y en su representación cabe destacar que los recursos son representados como una elipse y las literales con un rectángulo.

4.5. Web Ontology Language (OWL)

OWL (Lenguaje de ontologías web) es un lenguaje de la web semántica diseñado para representar de una manera enriquecida y compleja el conocimiento acerca de las cosas, grupos de cosas y sus relaciones. OWL es un lenguaje basado en lógica computacional, de tal manera, que el conocimiento expresado en OWL puede ser explotado por los programas computacionales, por ejemplo, para verificar la consistencia del conocimiento o para hacer explícito el conocimiento implícito (OWL Working Group, 2012).

Los documentos OWL, conocidos como ontologías, pueden ser publicados en la web y hacer referencia o ser derivados desde otras ontologías OWL. OWL es parte de la tecnología de la web semántica del W3C. La primera versión de OWL fue desarrollada en el 2009 y la segunda edición publicada en el 2012.

4.6. EasyRDF

EasyRDF es una librería PHP diseñada para facilitar la producción y consumo de archivos RDF. Librería enfocada a desarrolladores con o sin experiencia, escrita con un enfoque orientado a objetos y ampliamente probada en varios proyectos. Después del análisis con EasyRDF se acumula un gráfico de objetos PHP que puede ser recorrido con el fin de obtener datos para mostrar en el sitio web. La carga de datos a un almacén de RDF se realiza con la clase EasyRDF_GrapStore, que implementa funcionalidades para gestionar una colección de archivos RDF vía HTTP (Humfrey, 2014).

Esta librería PHP es un proyecto OpenSource disponible en el sitio http://www.easyrdf.org/, se encuentra en constante mantenimiento y desarrollando de nuevas mejoras.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.7. Linked Data Conectando datos distribuidos en la web

Linked Data es la forma que tiene la Web Semántica para conectar datos relacionados que no estaban vinculados con anterioridad, ayudando a disminuir las barreras en la conexión de datos actualmente conectados con otros métodos.

En otras palabras, la web nos permite vincular documentos relacionados, del mismo modo que nos permite vincular los datos relacionados. El termino Linked Data se refiere a un conjunto de mejores prácticas para la publicación y la conexión de datos estructurados en la web. Las tecnologías clave que soportan datos vinculados son los URI (un medio genérico para identificar entidades o conceptos en el mundo), HTTP (un simple y universal mecanismo para recuperar recursos o las descripciones de los recursos) y RDF (un modelo de datos basado en grafos genéricos con que se enlazan y estructuran datos que describen cosas en el mundo) (Bizer & Heath, 2009).

El ejemplo más tangible de la adopción y aplicación de los principios de Linked Data ha sido “Linking Open Data Project”, un esfuerzo de la comunidad web fundada en Enero del 2007 y con el apoyo de la W3C. El objetivo del proyecto es arrancar la web de datos mediante la identificación de conjuntos de datos existentes que están disponibles bajo licencia abierta, la conversión de estos a RDF de acuerdo con los principios Linked Data y su publicación en la web.

Investigadores, desarrolladores, compañías de varios tamaños han sido participantes en el proyecto, permitiendo un rápido crecimiento gracias a la naturaleza abierta del mismo, donde cualquiera puede participar publicando un vocabulario de acuerdo a los principios Linked Data y vinculándolo con los ya existentes, un ejemplo de los datos vinculados en el proyecto se muestra en la Figura 27 Diagrama de la nube Linking Open Data, donde cada nodo en el diagrama representa un vocabulario distinto publicado con los principios Linked Data, este proyecto comenzó en Marzo del 2009 a registrar distintos vocabularios de diferentes índoles.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

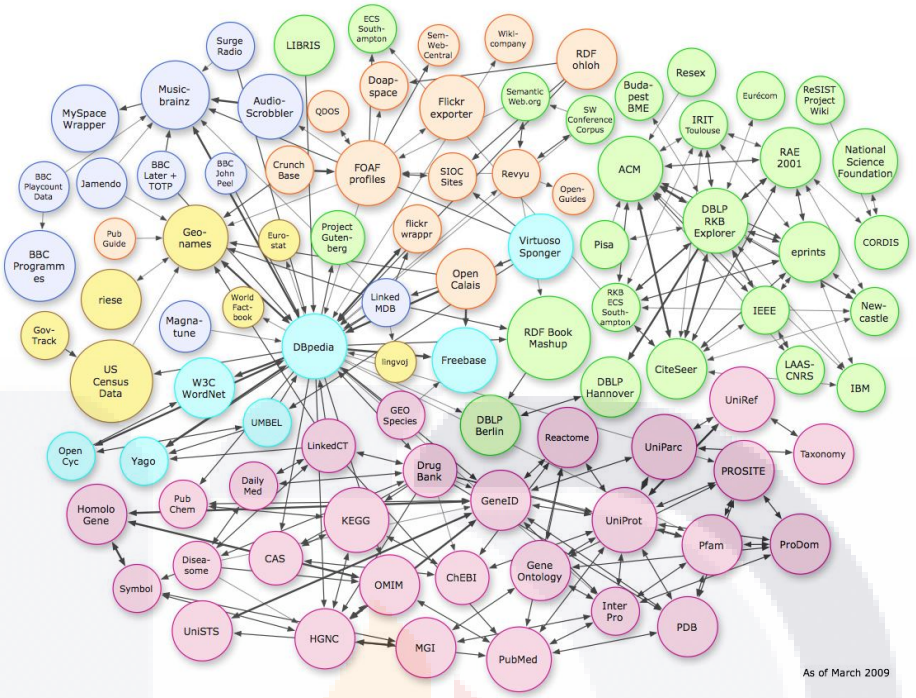

Figura 27 Diagrama de la nube Linking Open Data (Universidad Berlin, 2013)

4.8. SPARQL Protocol and RDF Query Language

La Web Semántica almacena un vasto conjunto de datos estructurados, haciendo necesario nuevos mecanismos para explotar esta información, provocando el surgimiento de nuevas herramientas informáticas que utilicen un lenguaje estandarizado para la consulta de los datos presentados en grafos RDF, este lenguaje es llamado SPARQL acrónimo de “Protocol and RDF query language”, llegando a ser un estándar de la W3C en el 2008 (DuCharme, 2013).

El lenguaje de consultas SPARQL está basado en la coincidencia de patrones, el patrón más simple es el patrón de triples, similar al utilizado en los grafos RDF, pero con la posibilidad de tener una variable en lugar de un término RDF en la posición del sujeto, predicado u objeto. Una combinación de patrones tiple da un patrón de grafico básico, donde se necesita una coincidencia exacta con un gráfico para cumplir un patrón (Staab & Studer, 2009).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Un ejemplo simple donde la consulta recupera todos los patrones de triple donde la propiedad es rdf:type y el objeto es rdfs:Class. En otras palabras, la consulta recupera todas las clases, se muestra en la Figura 28 Ejemplo de consulta SPARQL.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?c
WHERE
{
    ?c rdf:type rdfs:Class .
}
```

Figura 28 Ejemplo de consulta SPARQL

Otro ejemplo de consulta donde utilizando el vocabulario FOAF recupera la clase particular persona foaf:Person, se muestra en la Figura 29 Consulta para recuperar la clase persona.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?i
WHERE
{
  ?i rdf:type foaf:Person .
}
```

Figura 29 Consulta para recuperar la clase persona

Como SQL, las consultas SPARQL tienen la estructura SELECT-FROM-WHERE, donde SELECT especifica la protección, el número y orden de los datos recuperados. FROM es usado para especificar la fuente de consulta, esta cláusula es opcional cuando queremos especificar una particular fuente de conocimiento. WHERE ayuda para construir restricciones a posibles soluciones en forma de plantillas de patrón gráfico y restricciones booleanas (Staab & Studer, 2009).

Por ejemplo para recupera todos los correos electrónicos de las personas, donde ?x y ?y son variables, y ?x foaf:mbox ?y representa un patrón de triple recurso-propiedad-valor, se muestra en la Figura 30 Consulta para recuperar correos electrónicos. De esta

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

manera se puede ir creando patrones de grafico más elaborados y obtener información más compleja de nuestras consultas.

```
SELECT ?x ?y
WHERE
{
  ?x foaf:mbox ?y .
}
```

Figura 30 Consulta para recuperar correos electrónicos

4.9. ARC2

ARC2 es una libraría PHP para el trabajo con RDF, provee un almacén de triples basado en MySQL para el soporte a SPARQL. ARQ2 es un sistema flexible y fácil de usar para la Web Semántica y profesionales en ontologías RDF, está bajo licencia Open Source y puede correr bajo el ambiente de servidor más utilizado LAMP.

ARC2 tiene las siguientes características:

- Soporte para servidores proxy, redirecciones y negociación de contenido.
- Varios parsers como RDF/XML, N-Triples, Turtle, etc.
- Serializadores.
- Dos estructuras internas, procesamiento céntrico de recursos y sentencias.
- Almacen RDF:
- Endpoint SPARQL.

ARC comenzó en el 2004 como un sistema RDF ligero para el análisis y la serialización de archivos RDF/XML, más adelante se convirtió en un marco de trabajo más completo con el almacenamiento y consulta. Para el año 2011, ARC2 se había convertido en una de las bibliotecas RDF mas instaladas. Sin embargo, el desarrollo de códigos activo tuvo que ser interrumpido debido a la falta de fondos y a la imposibilidad de poner en práctica de manera eficiente a la creciente pila de especiaciones RDF, el código fuente continua disponible en la comunidad a través de GitHub (Corlosquet, 2014, p. 2).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.10. Procesamiento del lenguaje natural (PLN)

El procesamiento del lenguaje natural (PLN) hace referencia a las técnicas de tratamiento del lenguaje y es utilizado en diferentes aplicaciones como la traducción automática, sistemas de recuperación de información, elaboración automática de resúmenes, interfaces en lenguaje natural, y muchas más.

El PLN se concibe como el reconocimiento y utilización de la información expresada en lenguaje humano a través del uso de sistemas informáticos. En su estudio intervienen diferentes disciplinas tales como lingüística, ingeniería informática, filosofía, matemáticas y psicología. Debido a las diferentes áreas del conocimiento que participan, la aproximación al lenguaje en esta perspectiva es también estudiada desde la llamada ciencia cognitiva (Sosa, 1997).

El estudio del lenguaje natural se estructura normalmente en 4 niveles de análisis:

- Morfológico.
- Sintáctico.
- Semántico.
- Pragmático.

La relación de los pasos en el procesamiento del lenguaje natural se muestra en la Figura 31 Pasos en el procesamiento del lenguaje natural.

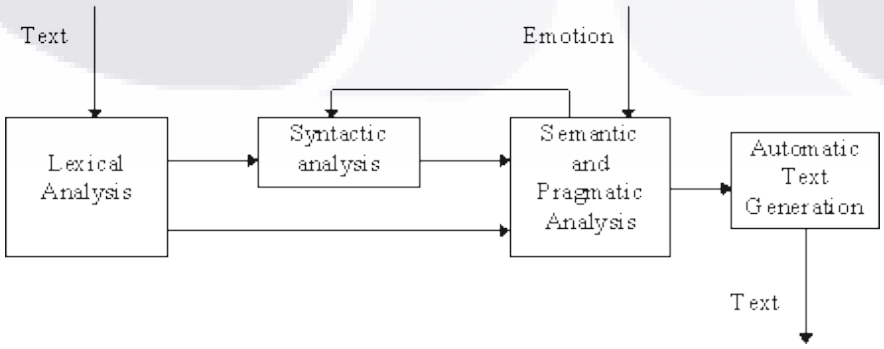

Figura 31 Pasos en el procesamiento del lenguaje natural (Nadia & Prem, 1998)

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Análisis morfológico. Su función consiste en detectar la relación que se establece entre las unidades mínimas que forman una palabra, como puede ser el reconocimiento de sufijos o prefijos. Este nivel de análisis mantiene una estrecha relación con el léxico.

Análisis sintáctico. Tiene como función etiquetar cada uno de los componentes sintácticos que aparecen en la oración y analizar cómo las palabras se combinan para formar construcciones gramaticalmente correctas. El resultado de este proceso consiste en generar la estructura correspondiente a las categorías sintácticas formadas por cada una de las unidades léxicas que aparecen en la oración.

Análisis semántico. En muchas aplicaciones del PLN los objetivos del análisis apuntan hacia el procesamiento del significado. En los últimos años las técnicas de procesamiento sintáctico han experimentado avances significativos, resolviendo los problemas fundamentales.

Análisis pragmático. Añade información adicional al análisis del significado de la frase en función del contexto donde aparece. Se trata de uno de los niveles de análisis más complejos, la finalidad del cual es incorporar al análisis semántico la aportación significativa que pueden hacer los participantes, la evolución del discurso o información presupuesta (Sosa, 1997).

4.11. NlpTools

NlpTools (Natural language processing tools) es una librería para el procesamiento de lenguaje natural escrita en PHP, su desarrollo es derivado por necesidades de clasificación de texto, clusteting, tokenizing, etc. (“NlpTools,” 2014).

Esta librería está bajo la licencia Open Source y sigue en desarrollo para colaboradores y expertos en el área lingüística y de programación, esta librería en el módulo de clasificación aún no tiene modelos predefinidos, por lo que es necesario realizar un proceso previo de entrenamiento de clasificación a modelos definidos con anterioridad.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

4.12. Usabilidad

La usabilidad se refiere a la calidad de la experiencia del usuario al interactuar con los productos o sistemas, incluyendo sitios web, software, dispositivos o aplicaciones. La usabilidad es acerca de la efectividad, eficiencia y satisfacción general del usuario (U.S. Department of Health & Human Services, 2014).

Es importante darse cuenta que la usabilidad no es una sola propiedad, de una disensión de un producto, sistema o interfaz de usuario. La usabilidad es una combinación de factores, como lo son, *diseño intuitivo*, una comprensión casi sin esfuerzo de la arquitectura y la navegación del sitio. *Facilidad de aprendizaje*, que tan rápido un usuario que nunca ha visto la interfaz del sitio antes puede realizar una tarea básica. *Eficiencia de uso*, que tan rápido un usuario experimentado puede llevar a cabo tareas. *Memorización*, después de visitar el sitio, el usuario puede recordar lo suficiente como para utilizarla de manera eficaz en futuras visitas. Frecuencia y gravedad de errores, con qué frecuencia los usuarios comenten errores al utilizar el sistema, la gravedad de los errores y como los usuarios se recuperan de los errores. *Satisfacción subjetiva*, si al usuario le gusta usar el sistema (U.S. Department of Health & Human Services, 2014).

4.13. Métodos de evaluación de usabilidad

La evaluación de usabilidad se centra en que tan bien los usuarios pueden aprender y utilizar un producto para alcanzar sus metas. También se refiere a que tan satisfechos los usuarios están con el proceso. Para obtener esta información los investigadores utilizan una variedad de métodos que recogen la opinión de los usuario acerca de un sitio o de los planes existentes en relación con el nuevo sitio (U.S. Department of Health & Human Services, 2014).

La clave para el desarrollo de sitios altamente utilizables está en emplear el diseño centrado en el usuario (UCD). La expresión, "prueba de principio y con frecuencia", es particularmente apropiado cuando se trata de pruebas de usabilidad. Como parte de la UCD se deben realizar pruebas de usabilidad a menudo, existe una amplia variedad de métodos disponibles que permitirán ayudar en el desarrollo de contenidos, arquitectura de información, diseño visual, diseño de interacción y satisfacción de los usuarios en general.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Existen distintos tipos de pruebas, como lo son:

- Pruebas de usabilidad sobre un sitio existente.
- Los grupos de enfoque, encuestas o entrevistas para establecer los objetivos del usuario.
- Pruebas Wireframe para evaluar navegación.
- Prueba del primer clic, para asegurarse que los usuarios van por el camino correcto.
- Las pruebas de usabilidad para medir la interacción del usuario.
- Las encuestas de satisfacción del sitio en el mundo real.

4.13.1. Test retrospectivo

Moderar eficazmente las pruebas de usabilidad es esencial para profundizar en el conocimiento y comprensión acerca de las necesidades de los usuarios. Los investigadores utilizan varias técnicas para moderar las sesiones de los participantes, algunas técnicas involucran consecuencias involuntarias, mientras que otras son más puras. Es importante pensar en las metas de cada estudio para seleccionar la técnica más adecuada (U.S. Department of Health & Human Services, 2014), Las ventajas y desventajas de cada técnica se muestran en la Tabla 1 Técnicas de usabilidad.

Técnica	Ventaja	Desventaja
Pensamiento en voz alta concurrente (CTA)	• Entender el pensamiento de los participantes a medida que ocurren, y en su intento de trabajar a través de los problemas que encuentran. • Obtener retroalimentación y emociones en tiempo real.	• Puede interferir con las métricas de usabilidad, como la precisión y el tiempo en la tarea.
Pensamiento en voz alta retrospectivo (RTA)	• No interfiere con métricas de usabilidad.	• Aumenta la duración de la sesión. • Dificultar para recordar los pensamientos de hasta una hora antes.
Sondeo concurrente (CP)	• Entender el pensamiento de los participantes en su intento de trabajar a través de una tarea.	• Interfiere con el proceso de pensamiento natural y el progreso que el participante hace por su cuenta es interrumplido.
Sondeo retrospectivo (RP)	• No interfiere con las métricas de usabilidad.	• Dificultar al recordar.

Tabla 1 Técnicas de usabilidad

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

En particular para el objetivo de esta investigación se utilizara un test retrospectivo RTA, donde el moderador pide a los participantes reflexionar sobre sus pasos cuando completa una sesión o durante la sesión, a menudo acciones de los participantes son grabadas en video, esta técnica permite a los participantes trabajar en silencio siendo una forma obvia de controlar los efectos negativos del CTA. El proceso de pensar en voz alta no interfiere con las métricas de usabilidad (tales como el tiempo en la tarea), aunque aumenta bastante la duración de la sesión.

Existen una serie de verdades acerca de las pruebas de usabilidad como lo son (Krug, 2006):

- Si desea un sitio web excelente, tienen que realizarse pruebas de usabilidad.
- Pruebas por solo un usuario es mejor que no probar.
- Probar por un usuario al inicio del proyecto es mejor que pruebas por cincuenta usuarios al final del proyecto.
- La importancia de utilizar participantes representativos está sobrevalorada.
- El objetivo de la prueba no es para aprobar o refutar algo, es para compartir un punto de vista.
- Las pruebas son un proceso iterativo.
- No hay nada como una reacción de la audiencia en vivo.

5. Metodología

Retomando el objetivo de la tesis descrito en la sección 2.4. El propósito general se enfoca en la descripción del proceso de creación de un prototipo de motor de búsqueda semántica basado en un paradigma relacional de base de datos, evaluando el esfuerzo de creación y la facilidad de uso.

Tomando en cuenta el estado actual de las tecnologías inmersas en la Web Semántica, se propone la siguiente metodología para la creación de un buscador semántico:

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

- Fase I “Transformación relacional semántico”
    - Búsqueda y análisis de herramientas RDF.
    - Comparativa y selección de la herramienta RDF.
    - Creación y/o selección de los vocabularios necesarios.
    - Diseño e implementación de algoritmo para generar el repositorio RDF.
- Medir el esfuerzo en horas de la transformación.
- Fase II “Implementación de consultas semánticas”
    - Búsqueda y análisis de manejadores de consulta semántica
    - Comparativa y selección del manejador de consulta semántica
    - Implementación del manejador en el repositorio RDF.
- Fase III “Creación de interfaz para el buscador semántico”
    - Búsqueda y selección de la herramienta para el análisis semántico.
    - Creación de interfaz para el motor de búsqueda semántica
- Evaluar la facilidad de uso del buscador semántico.

5.1. Transformación relacional semántico

La primera fase de la metodología para la creación de un prototipo de búsquedas semánticas es la fase de transformación relacional semántico, básicamente consiste en darle una representación semántica a una base de datos relacional con base a las tecnologías existentes de la Web Semántica, esta fase consta de cuatro etapas básicas:

- Búsqueda y análisis de herramientas RDF.
- Comparativa y selección de la herramienta RDF.
- Creación y/o selección de los vocabularios necesarios.
- Diseño e implementación de algoritmo para generar el repositorio RDF.

5.1.1. Búsqueda y análisis de herramientas RDF

Existe gran cantidad de herramientas RDF en internet, herramientas con varios propósitos dependiendo del área de aplicación y los recursos tecnológicos. En el caso de esta metodología nos interesan las herramientas de conversión a RDF, donde los formatos más populares para convertir a RDF son BibTex, Bittorrent, CSV, Debian, Email, Excel, EXIF, SQL, etc. (W3C, 2013).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Esta pasó en la metodología hace referencia a la búsqueda minuciosa en internet de herramientas RDF en el ámbito de la conversión SQL a RDF. En otras palabras, herramientas que auxilien en la creación y manipulación de un repositorio RDF basado en una base de datos relacional, recabando información como el lenguaje de programación en el que fue desarrollada la librería, el tipo de conexiones que realiza, soporte para SPAQRL, tipo de licencia, etc.

5.1.2. Comparativa y selección de la herramienta RDF

Una vez formada una perspectiva global de las herramientas existentes para la transformación de una base de datos relacional hacia un paradigma semántico, es necesario hacer la selección de la herramienta más adecuada de acuerdo a una serie de criterios, que se definen según el ambiente donde se implementará el buscador y la experiencia del recurso humano a cargo de la construcción.

Estos criterios para la selección pueden relacionarse a rubros como el lenguaje de programación que interpreta el servidor donde se alojara el motor de búsqueda, el tipo de motor de base de datos donde se almacena la información a transformar, tipo de licencia de la herramienta, cuestiones de seguridad en el acceso al servidor, etc.

5.1.3. Creación y/o selección de los vocabularios necesarios

De acuerdo a las mejores prácticas de Linked Data, para crear un repositorio RDF sobre algún tema en especial, primero se debe investigar si ya existe algún vocabulario de ontologías que defina esta área del conocimiento, de lo contrario podríamos definir algún vocabulario basado en RDFS u OWL para vincular la semántica inmersa dentro del repositorio RDF con otras áreas del conocimiento expresadas dentro de la Web Semántica.

En la actualidad existen varios proyectos cuyo objetivo es servir de repositorio de vocabularios, proyectos como SKOS (“SKOS Simple Knowledge Organization System,” 2013) , LOV Linked Open Vocabularies (“Linked Open Vocabularies (LOV),” 2014) y LinkingOpenData (“LinkingOpenData,” 2013), son solo algunos ejemplos de repositorios de vocabularios que almacenan descripciones de ontologías sobre distintas áreas del conocimiento.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

5.1.4. Diseño e implementación de algoritmo para generar el repositorio RDF

Después de la búsqueda y selección de la herramienta para el manejo de RDFs, y la creación o recopilación de los vocabularios necesarios para expresar las ontologías del área del conocimiento descrita la base de datos relacional, es momento de diseñar, codificar e implementar un algoritmo que tome en cuenta los criterios y recursos seleccionados para realizar la transformación del paradigma relacional al paradigma semántico.

En otras palabras realizar una mapeo de los datos almacenados en forma de tupla en la base de datos relacional hacia una representación en grafos RDF y almacenarlos en un repositorio para su publicación en la Web Semántica, Un ejemplo de este tipo de transformación se muestra en la Figura 32 Representación de mapeo relacional semántico (W3C, 2010).

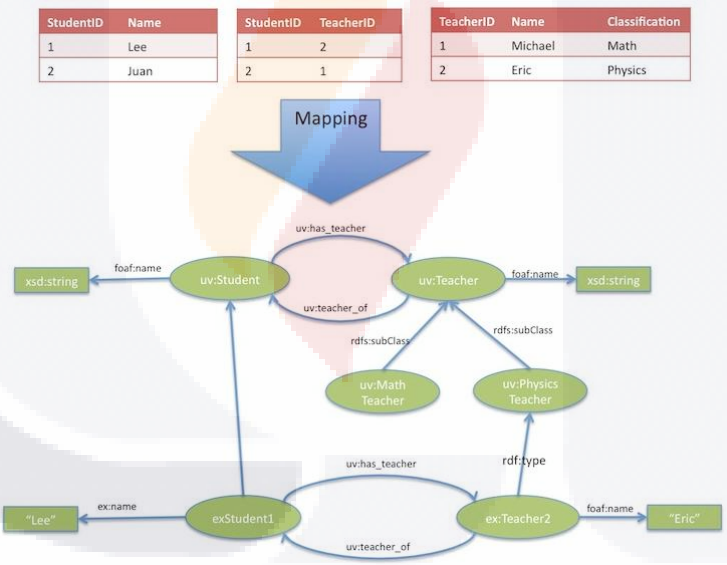

Figura 32 Representación de mapeo relacional semántico (W3C, 2010)

5.2. Medir el esfuerzo en horas de la transformación

Al terminar la fase de transformación, y con el objetivo de mostrar el esfuerzo en hora invertido en el proceso, se deberán tomar los tiempos dedicados a cada una de las actividades de transformación, de esta manera servirá como referente para el practicante

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

que opte por seguir este proceso, el registro de tiempos se basa en la las prácticas de PSP (Proceso de Software Personal).

5.3. Implementación de consultas semánticas

La segunda fase en el proceso de construcción del buscador semántico es implementar un manejador de consultas semánticas (Endpoint), que haga uso del repositorio RDF creado en una fase anterior. Del mismo modo que con las herramientas para construir repositorios RDF, se realiza una búsqueda minuciosa en internet con el fin de encontrar los manejadores de consulta semántica disponibles actualmente.

El lenguaje estándar para realizar consultas semánticas SPARQL es implementado por la mayoría de estos manejadores, por lo que es necesario complementar este criterio y otros más dependiendo del ambiente tecnológico donde interactuará el buscador semántico.

5.3.1. Búsqueda y análisis de manejadores de consulta semántica

El primer paso para implementar un manejador de consultas semánticas en un repositorio RDF es hacer una búsqueda y análisis sobre herramientas existentes, con el fin de encontrar alguna herramienta fácilmente adaptable a nuestro ambiente tecnológico, o de lo contario si no se encuentra una herramienta adecuada, replantear la construcción del repositorio RDF o investigar sobre la construcción de un manejador independiente.

En la actualidad en internet han surgido varios proyectos que pretenden desarrollar herramientas de consultas semánticas basadas en SPARQL cada vez más sofisticados, podremos encontrar una amplia lista en el sitio oficial de la W3C (W3C, 2014).

5.3.2. Comparativa y selección del manejador de consulta semántica

Una vez identificados los manejadores de consulta semántica disponibles, es necesario definir una serie de criterios que nos ayuden en la selección de la herramienta más adecuada en base al objetivo principal del buscador semántico.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Estos criterios deben ser definidos en base al ambiente tecnológico donde se implementará el buscador, por ejemplo el lenguaje de programación que interpreta el servidor, la cantidad de memoria disponibles, cuestiones de seguridad en el servidor, etc.

5.3.3. Implementación del manejador en el repositorio RDF

Después de haber hecho la selección del manejador de consultas semánticas más adecuada conforme a los criterios definidos en el paso anterior, es necesario implementar esta herramienta en el repositorio RDF, diseñando una serie de pasos con el objetivo de que las ontologías inmersas en el repositorio puedan ser utilizadas.

Existen en la actualidad manejadores de consultas semánticas disponibles como servicios web, un ejemplo de esta implementación la provee OpenLink Virtuoso en su sitio web, donde se documenta detalladamente el uso de este servicio (OpenLink Virtuoso, 2014), en la implementación del buscador semántico podría hacerse uso de este tipo de estructura, tomando en cuenta que los archivos RDF como paso previo tendrían que registrarse en el servidor donde se aloja el servicio web.

5.4. Creación de interfaz para el buscador semántico

Posterior a la implementación del manejador de consultas semánticas, es necesario ocuparse de la interfaz del buscador semántico, definir como el usuario tendrá que interactuar con el sitio web.

Lo primero que viene a la mente, gracias a que los motores de búsqueda actuales tienen este tipo de interfaz, es el uso de una sola caja de texto donde el usuario ingrese una consulta, y en consecuencia surge la incógnita, ¿Cómo dividir una consulta en lenguaje natural en sus componentes para poder ser utilizados en el manejador semántico?, la respuesta a esta pregunta nos llevaría a técnicas sofisticadas de procesamiento de lenguaje natural, a representación de ontologías y en general a temas de inteligencia artificial, donde aún existen áreas en desarrollo.

Pensando en el objetivo principal de la tesis, este paso de la metodología se enfoca únicamente a la construcción de una interfaz para el buscador semántico apoyado de herramientas que auxilien en la identificación del área del conocimiento al que se

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

refieren ciertas expresiones en lenguaje natural, permitiendo así la implementación de un prototipo de buscador semántico.

5.4.1. Búsqueda y selección de la herramienta para el procesamiento del lenguaje natural.

El penúltimo paso para la construcción del prototipo de buscador semántico, es la búsqueda y selección de una herramienta que permita procesar enunciados en lenguaje natural, clasificándolos en ciertas áreas del conocimiento. La herramienta debe cumplir ciertos criterios de acuerdo al ambiente tecnológico donde el prototipo será implementado.

La clasificación que realiza la herramienta ayuda a identificar sobre que sección del vocabulario se debe construir una sentencia SPARQL que permita explotar el repositorio RDF y recuperar información más precisa sobre el área del conocimiento que se desea encontrar.

5.4.2. Creación de interfaz para el motor de búsqueda semántica

El último paso de construcción del prototipo es crear una interfaz para el usuario final, esta construcción requiere conocimientos previos de tecnologías base en el desarrollo web, como lo es HTML, CSS, Javascript y algún lenguaje de programación de alto nivel.

Esta interfaz de preferencia debe tomar en cuenta criterios básicos de usabilidad web, que permitan una interacción amigable del usuario con el motor de búsqueda semántico.

5.5. Evaluar la facilidad de uso del buscador semántico

Como último paso de la metodología, se busca medir la facilidad de uso del buscador semántico por medio de un test de usabilidad aplicado a un grupo de control, el test deberá estar basado en usa serie de tareas diseñadas para que el usuario inicie realizando una consulta en el buscador y termine cuando encuentre lo que se pide en el test.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

De esta manera se tomaran en cuenta factores como el tiempo que toma al usuario encontrar información, la tasa de éxito de intentos de búsqueda, cantidad de clics que el usuario dio para encontrar información, etc. Con el fin de mostrar una descripción del grado de usabilidad del prototipo de buscador semántico.

6. Resultados

El alcance de la construcción del buscador semántico para esta tesis fue delimitado a un sitio web de apoyo al aprendizaje de la programación, tomando como recurso principal la base de datos de este sitio. Esta base de datos almacena información sobre diferentes maneras de solucionar problemas de programación en base a protocolos verbales.

El prototipo de buscador semántico fue pensado para su implementación dentro del mismo ambiente tecnológico donde se encuentra el sitio web, para su futura implementación dentro del mismo desarrollo. Este ambiente es denominado WAMP y se compone de la arquitectura conformada por un sistema operativo Windows y herramientas open source como un servidor HTTP Apache, motor de base de datos MySQL y lenguaje de programación PHP, Perl o Phyton.

6.1. Sitio web para el apoyo al aprendizaje de la programación utilizando protocolos verbales.

Con el fin de servir de apoyo al aprendizaje de la programación, fue creado un sitio web que por medio de protocolos verbales transfiera el conocimiento de expertos a los usuarios que tengan como objetivo aprender programación, la pantalla principal del sitio web se muestra en la Figura 33 Pantalla principal del sitio web de apoyo al aprendizaje de la programación.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Figura 33 Pantalla principal del sitio web de apoyo al aprendizaje de la programación (Sistema Visor de Protocolos Verbales, 2014)

En este sitio el usuario tiene la posibilidad de navegar entre distintos problemas de diferentes temas y sus posibles soluciones expresadas en forma de protocolos verbales, así como observar la solución paso a paso de cada problema por medio de una interfaz como se muestra en la Figura 34 Visor de protocolo verbal paso a paso.

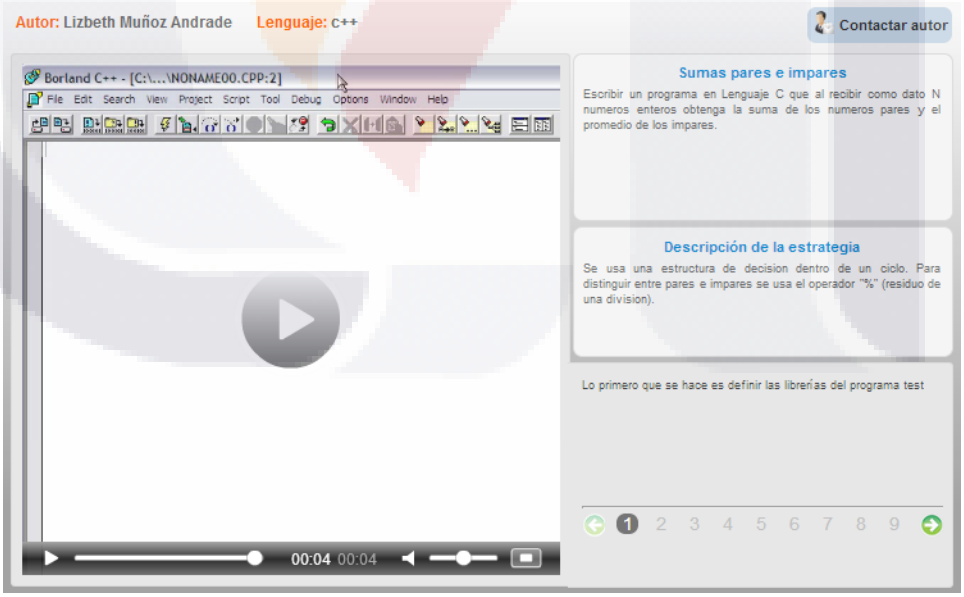

Figura 34 Visor de protocolo verbal paso a paso programación (Sistema Visor de Protocolos Verbales, 2014)

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

6.2. Transformación de base de datos relacional a su representación semántica

El sitio web de apoyo al aprendizaje de la programación almacena información en una base de datos acerca de protocolos verbales que describen soluciones a problemas de programación, el modelo entidad relación de la base de datos que utiliza el sitio web se muestra en la Figura 35 Modelo de la base de datos relacional, esta base de datos se tomó para realizar la transformación a una representación semántica, como lo son archivos RDF en un repositorio.

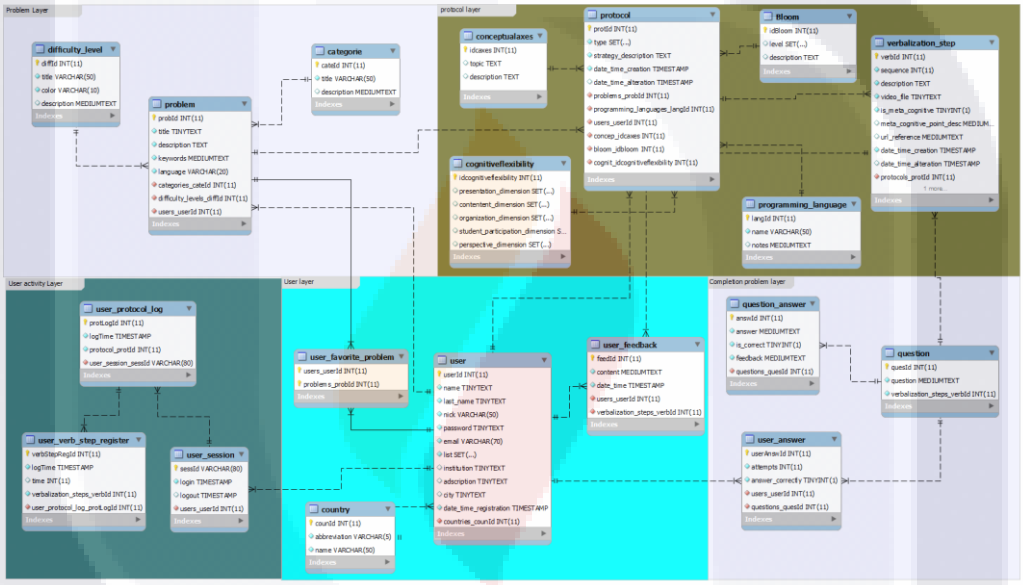

Figura 35 Modelo de la base de datos relacional

En base al modelo anterior, se delimitó la representación semántica a solo aquellas entidades y relaciones que tuvieran un mayor grado de representatividad en el contexto la verbalización de solución a problemas de programación, El segmento del modelo da la base de datos, solo con las entidades y relaciones utilizadas en el proceso de conversión a representación semántica se muestra en la Figura 36 Tablas y relaciones más representativas del modelo, donde en general se observa que un problema puede tener una o varias soluciones (protocolos) y estas soluciones pueden tener uno o varios pasos en la solución (pasos de verbalización).

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

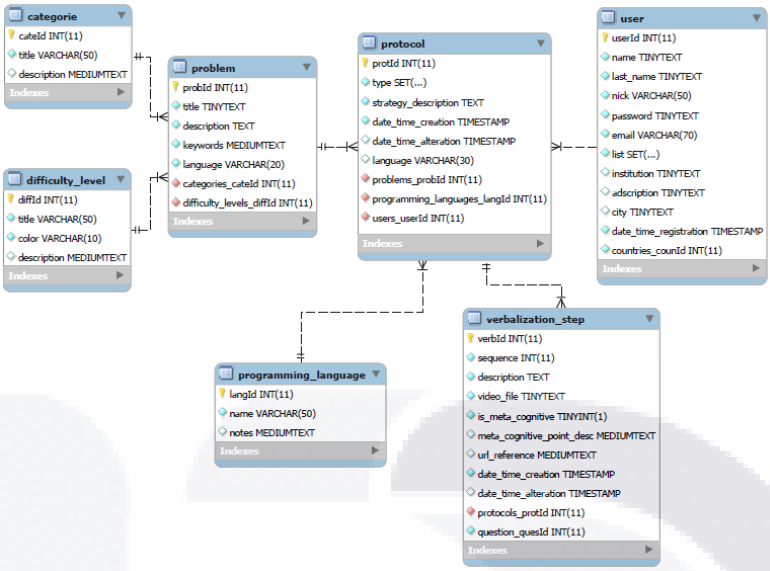

Figura 36 Tablas y relaciones más representativas del modelo

6.2.1. Herramientas RDF existentes

Una vez definidas las entidades y relaciones a trasformar, se realizó una minuciosa búsqueda en internet con el fin de encontrar las herramientas RDF existentes que puedan ser utilizadas para el proceso de conversión de la base de datos relacional a su representación semántica basada en archivos RDF, las herramientas encontradas fueron seleccionadas en base a una serie de criterios definidos de acuerdo al ambiente tecnológico donde fue implementado el prototipo de buscador semántico, estos criterios de selección son los siguientes:

- El lenguaje de programación que interpreta el servidor.
- Tipo de conexión a base de datos.
- Límite de manejo de base de datos.
- Soporte para instrucciones SPARQL.
- Tipo de licencia.
- Soporte para almacén RDF.

La comparativa de acuerdo a los criterios de selección de cada una de las herramientas encontradas en la búsqueda en internet se muestra en la Tabla 2 Comparativa de herramientas para el manejo de RDF.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Nombre	Características						
	Lenguaje	Tipo conexión	Soporta SPARQL	Licencia	Almacén RDF	Ultima Actualización (Mes/Año)	Tamaño DB
D2RQ	Lenguaje de mapeo D2RQ	MySQL , Oracle, SQL Server, Postgre SQL.	Si	Open Source	Si	jun-12	Sin limite
Triplify	PHP	MySQL	Si	Open Source	Si	may-13	< 100MB
Jena	Java	Postgre, MySQL, Oracle, SQL Server, DB2.	Si	Open Source	Si	abr-12	Sin limite
Redland	Librería en C enlazado con PHP, Perl, Python, Ruby.	Oracle, MySQL, Postgre, SQLite	Si	Open Source	Si	dic-12	Sin limite
EasyRDF	PHP	MySQL, SQL Server, Oracle.	Si	Open Source	Si	ene-12	Sin limite
Sasame	Java	MySQL, Postgre.	Si	Open Source	No	jun-13	Sin limite

Tabla 2 Comparativa de herramientas para el manejo de RDF

De acuerdo a los criterios descritos anteriormente la herramienta más acorde al ambiente tecnológico donde se implementó el prototipo es EasyRDF, ya que el servidor interpreta lenguaje PHP, el motor de base de datos es MySQL, tiene soporte para repositorio RDF y es una herramienta Open Source.

6.2.2. EasyRDF como herramienta para la transformación relacional semántico.

EasyRDF contiene una serie de características que lo hacen una buena herramienta para el proceso de conversión de una base de datos relacional a su representación semántica en base a un repositorio RDF, estas características se listan a continuación (Humfrey, 2014):

- Extensas pruebas unitarias.
    - Pruebas automatizadas con diferentes versiones de PHP.
- Analizadores y serializadores incorporados: RDF / JSON, N-Triples, RDF / XML, Turtle.
- Soporte opcional para el análisis: ARC2, enlaces Redland, rapper.
- Soporte opcional para Zend_Http_Client.
- No hay dependencias externas requeridas sobre otras bibliotecas (PEAR, Zend, etc.).
- Cumple con el estilo de codificación PSR-2.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

- Type mapper. Recursos de tipo foaf:Persona se pueden mapear en PHP como objeto de la clase Foaf_Person.
- Soporte para la visualización de los gráficos usando GraphViz.
- Compatible con Composer.
- Viene con una serie de ejemplos.

EasyRDF según su página oficial, tiene como método preferido para la descarga e instalación el uso de Composer. Composer es una herramienta para el manejo de dependencias en PHP, permite declarar las dependencias a librerías que necesita un proyecto y composer las instalara en el proyecto automáticamente (Adermann & Boggiano, 2014).

Este estudio de tesis se enfoca únicamente al uso de herramientas que estén directamente involucradas en el la construcción del prototipo, por lo que se utilizó como lenguaje de programación PHP puro, con el fin de hacer más comprensible el desarrollo del prototipo.

En este orden de ideas, para el uso de la librería EasyRDF es necesario agregar la librería en nuestro proyecto PHP como lo muestra la Figura 37 Instrucción PHP para incluir librería EasyRDF.

```php
require_once "easyrdf/lib/EasyRdf.php";
```

Figura 37 Instrucción PHP para incluir librería EasyRDF

De esta manera podremos utilizar después de la declaración, objetos de la clase EasyRdf_Graph que nos auxilia en la construcción de los archivos RDF, como se muestra en la Figura 38 Construcción básica de archivo RDF con PHP.

```php
$rdf = new EasyRdf_Graph();
$problema = $rdf->resource($uri_problema, 'pps:Problem');
$problema->set('dce:title', $titulo);
$problema->add('pps:hasSolution', $rdf->resource($uri_solucion));
$archivo = $rdf->serialise('rdfxml');
```

Figura 38 Construcción básica de archivo RDF con PHP

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Donde la variable $rdf es un objeto de tipo *EasyRdf_Graph,* permitiendo comenzar la construcción de nuestro archivo RDF, posteriormente creamos la raíz de nuestro archivo declarando un recurso de tipo problema (pps:Problem), agregando una propiedad de título (dce:title) al problema, después vinculando un recurso de tipo solución (pps:Solution) y finalizando la construcción asignándole a la variable $archivo, todo el documento RDF creado.

6.2.3. Programming Problem Solvig (PPS)

Un paso anterior a la construcción del repositorio RDF y siguiendo las buenas prácticas de Linked Data, se realizó una búsqueda minuciosa en distintos proyectos y repositorios, de un vocabulario que definiera el área del conocimiento sobre soluciones a problemas de programación y de esta manera conservar la integridad de significados expresados en vocabularios OWL o RDFS de la Web Semántica.

En esta búsqueda no se encontró algún lenguaje que expresara precisamente soluciones a problemas de programación, pero en base al modelo ER se encontraron vocabularios complementarios que pueden auxiliar en la construcción de ciertas propiedades inmersas en el área del conocimiento que compete la solución de problemas de programación.

Debido a la carencia de un vocabulario para expresar soluciones a problemas de programación por medio de protocolos verbales, se construyó una propuesta de vocabulario basado en OWL y RDFS, este vocabulario denominado “Programming Problem Solving” o PPS, abarca las áreas de conocimiento que se muestran en la Figura 39 (PPS) Programming Problem Solving, donde en general se observa la relación que existe entre distintos elementos del vocabulario y que tipo de relación o propiedad une a estos elementos. El vocabulario es capaz de otorgar los elementos semánticos necesarios para expresar archivos RDF que describan soluciones a problemas de programación, comenzando con la existencia de un problema, el cual se le asigna un título, una descripción un nivel de dificultar y una categoría, estos problemas pueden tener o no tener solución o soluciones, donde cada solución tiene asignado un lenguaje de programación en el cual se construyó la solución, una estrategia de solución y un estatus de solución, esta solución se compone por distintos pasos, y a cada paso se le asigna una secuencia en la solución, una descripción, un video representativo del paso y una

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

propiedad que define si es un paso significativo en la solución, estas soluciones son propuestas por alguna persona, la cual se define como solucionador que contiene propiedades como nombre, correo electrónico, un sobrenombre y la institución a la que pertenece.

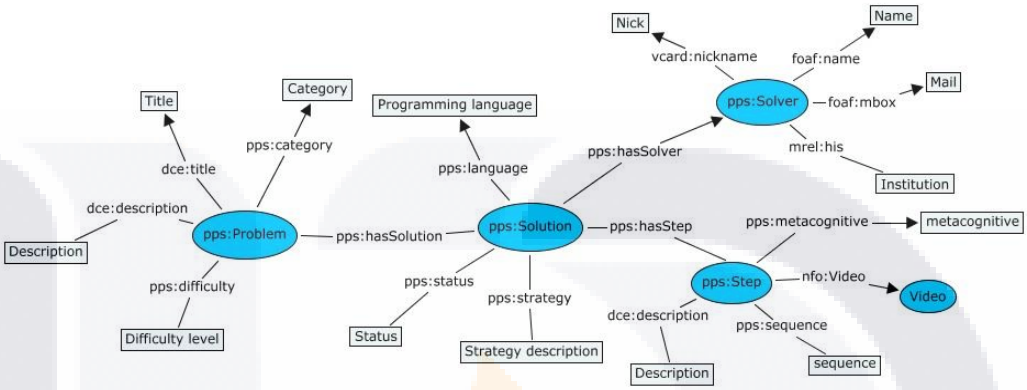

Figura 39 (PPS) Programming Problem Solving

Una representación general de la construcción del vocabulario se presenta en la Figura 40 Construcción del vocabulario PPS, donde en la primera sección se observan todas los vocabularios involucrados en la construcción, debido a que se utilizan una o varias etiquetas, por ejemplo para la construcción con etiquetas RDFS se utiliza *xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"*, permitiéndonos de esta manera utilizar elementos de este lenguaje en nuestra definición, enseguida se construye la definición del vocabulario y algunas propiedades en OWL, como el título del vocabulario con la propiedad “dc:title”. Finalmente se construye la definición de las clases, propiedades y relaciones del vocabulario con ayuda de RDFS, en el ejemplo podemos observar las definiciones de las clases “#Problem”, “#Solution” y la definición de las propiedades “#category” y “#strategy”.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xml:base="http://jrenelg.com/files/mitc/tesis/vocabulary/pps"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
    <owl:Ontology rdf:about="http://jrenelg.com/files/mitc/tesis/vocabulary/pps">
        <dc:identifier>http://jrenelg.com/files/mitc/tesis/vocabulary/pps</dc:identifier>
        <dc:date>2013-11-20</dc:date>
        <dc:title xml:lang="en">Programming Problem Solving</dc:title>
        <cc:license rdf:resource="http://creativecommons.org/licenses/by-nc-sa/3.0/es"/>
    </owl:Ontology>
    <rdfs:Class rdf:about="#Problem">
        <rdfs:isDefinedBy rdf:resource="http://jrenelg.com/files/mitc/tesis/vocabulary/pps"/>
        <rdfs:label xml:lang="en">Programming problem</rdfs:label>
    </rdfs:Class>
    <rdfs:Class rdf:about="#Solution">
        <rdfs:isDefinedBy rdf:resource="http://jrenelg.com/files/mitc/tesis/vocabulary/pps"/>
        <rdfs:label xml:lang="en">a programming problem solution</rdfs:label>
        <rdfs:subClassOf rdf:resource="#Problem"/>
    </rdfs:Class>
    <rdf:Property rdf:about="http://jrenelg.com/files/mitc/tesis/vocabulary/pps#category">
        <rdfs:isDefinedBy rdf:resource="http://jrenelg.com/files/mitc/tesis/vocabulary/pps"/>
        <rdfs:label xml:lang="en">Programming category</rdfs:label>
        <rdfs:domain rdf:resource="#Problem"/>
    </rdf:Property>
    <rdf:Property rdf:about="http://jrenelg.com/files/mitc/tesis/vocabulary/pps#strategy">
        <rdfs:isDefinedBy rdf:resource="http://jrenelg.com/files/mitc/tesis/vocabulary/pps"/>
        <rdfs:label xml:lang="en">Strategy description</rdfs:label>
        <rdfs:domain rdf:resource="#Solution"/>
    </rdf:Property>
</rdf:RDF>
```

Figura 40 Construcción del vocabulario PPS

6.2.4. Algoritmo para generar el repositorio RDF

Una vez definida la base de datos que transformar, la herramienta RDF a utilizar y comprobar la existencia y/o crear los vocabularios necesarios para expresar la información relacional en una representación semántica, es necesario diseñar e implementar un algoritmo que auxilie en esta tarea. En el caso de esta tesis se construyó un algoritmo en el lenguaje de programación PHP utilizando conexiones a bases de datos en MySQL, Una representación del flujo diseñado e implementado para la construcción del repositorio de archivos RDF representando protocolos verbales que dan solución a problemas de programación se muestra en la Figura 41 Algoritmo para la creación del repositorio RDF.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

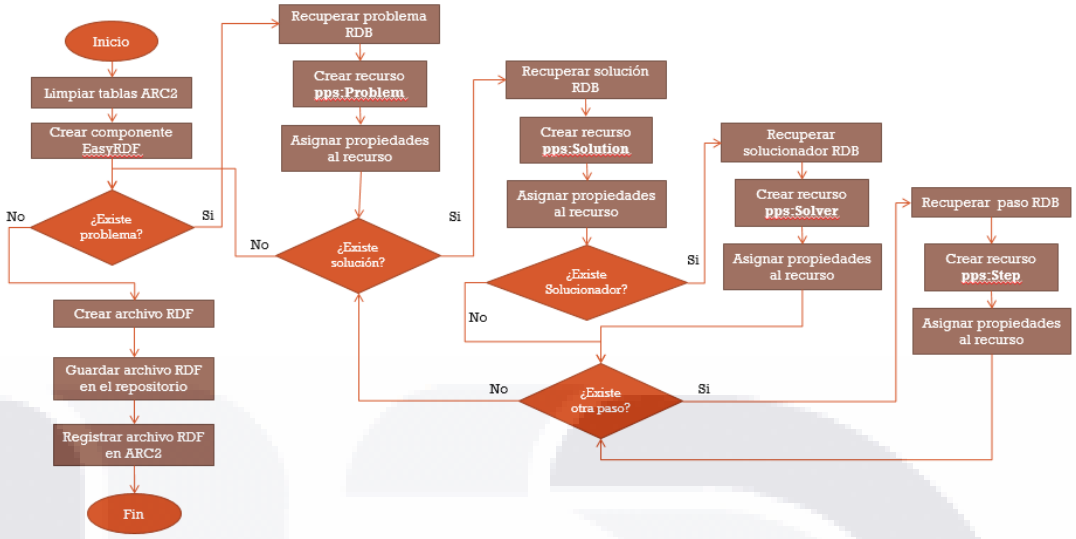

Figura 41 Algoritmo para la creación del repositorio RDF

Después de ser implementado el algoritmo y posterior a las pruebas, se corrió este proceso de transformación en la base de datos completa del sitio web de protocolos verbales, obteniéndose un conjunto de archivos RDF con una estructura similar a la que se muestra en la Figura 42 Ejemplo de archivo RDF, donde podemos ver que la primera sección son los vocabularios necesarios para la definición del archivo, posteriormente la definición de las distintas clases y propiedades según lo existente en la base de datos, por ejemplo en la imagen se observa la definición del problema “Sumas pares e impares”, que contiene una solución que comienza con la estrategia “Se usa una estructura…”, esta solución fue propuesta por el solucionador Lizbeth Muñoz, y la solución contiene solo un paso de solución.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:pps="http://jrenelg.com/files/mitc/tesis/vocabulary/pps#"
         ... >
  <pps:Problem rdf:about="URI#Problem">
    <dce:title>Sumas pares e impares</dce:title>
    <dce:description>Escribir un programa ....</dce:description>
    <pps:has_solution>
      <pps:Solution rdf:about="URI#Solution">
        <pps:strategy>Se usa una estructura ...</pps:strategy>
        <pps:has_solver>
          <pps:Solver rdf:about="URI#Solver">
            <foaf:name>Lizbeth Muñoz Andrade</foaf:name>
            <aiiso:Institution rdf:resource="http://www.uaa.mx/"/>
          </pps:Solver>
        </pps:has_solver>
        <pps:has_step>
          <pps:Step rdf:about="URI#Step">
            <dce:description>Lo primero ...</dce:description>
            <pps:sequence>1</pps:sequence>
          </pps:Step>
        </pps:has_step>
      </pps:Solution>
    </pps:has_solution>
  </pps:Problem>
</rdf:RDF>
```

Figura 42 Ejemplo de archivo RDF

Al terminar este proceso, se obtiene un repositorio de hasta el momento 132 archivos RDF, cada uno con la descripción de algún problema de programación de distintos niveles y en distintos lenguajes de programación.

6.3. Esfuerzo de la transformación

El proceso de transformar la base de datos relacional a un repositorio RDF interpretable semánticamente, llevó consigo esfuerzo en las distintas etapas del proceso, El cálculo del esfuerzo de transformación medido en horas efectivas dedicadas a cada tarea se muestra en la Tabla 3 Medición de esfuerzo (horas) de la transformación, de esta manera podemos observar que hacer una búsqueda minuciosa en internet sobre herramientas que ayuden en la creación de archivos RDF y la evaluación de las mismas, son las tareas que consumen más tiempo en el proceso.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Actividad	Esfuerzo absoluto (hr)	Esfuerzo relativo
Búsqueda y análisis de herramientas RDF	36	39.56 %
Evaluación de herramientas RDF (8)	28	30.76 %
Comparativa y selección de la herramienta RDF.	1	01.09 %
Definición del vocabulario	7	07.69 %
Diseño de algoritmo de transformación.	3	03.29 %
Implementación del algoritmo	11	12.08 %
Prueba del algoritmo	5	05.49 %
Total	**91**	**100.00 %**

Tabla 3 Medición de esfuerzo (horas) de la transformación

6.4. Implementación de consultas semánticas

Un paso importante en la creación de un prototipo de búsquedas semánticas es definir la forma en que vamos a realizar las consultas a nuestra representación semántica de la información, para este propósito podemos hacer uso de distintas herramientas disponibles en internet, a estas herramientas utilizadas para realizar consultas semánticas se les conoce como “endpoint”, las cuales por lo regular interpretan SPARQL, el lenguaje de consultas semánticas adaptado como estándar en los últimos años.

Podemos adaptar dos esquemas para implementar consultas semánticas en nuestro repositorio, en un esquema podemos hacer uso de endpoints públicos en internet, típicamente utilizados como servicios web, donde nuestro repositorio tendríamos que hacer un registro previo de nuestros archivos en el catálogo de información del servicio web. Otro esquema es el uso de una implementación propia de un endpoint dentro de nuestro prototipo. Para esta tesis se decidió utilizar el segundo esquema, implementar una herramienta para la creación de un endpoint, el cual hará uso del repositorio de archivos RDF creados en un paso anterior.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

6.4.1. Herramientas de consulta SPARQL existentes

Actualmente hay varios endpoint disponibles en internet, de los cuales destacan dos herramientas, por su amplio uso por distintos proyectos y basta documentación en internet. Estas herramientas son:

Open Link Virtuoso Unviersal Server. Es un servidor de datos multi-modelo para empresas e individuos. Ofrece una plataforma para la gestión de datos, el acceso y la integración. La arquitectura de servidor híbrido de Virtuoso le permite ofrecer la funcionalidad de servidor en un solo producto que cubre las siguientes áreas:

- Gestión de Datos Relacional.
- Gestión de datos RDF.
- XML Data Management.
- Gestión de contenido e indización de texto completo.
- Servidor web de documentos.
- Linked Data Server.
- Servidor de aplicaciones Web.
- Servicios de implementación web (SOAP o REST).

Semsol ARC2. Es un sistema RDF flexible para la web semántica y desarrolladores PHP, es gratis, Open Soure y de fácil uso y funciona en la mayoría de los entornos de servidores web.

- Soporte para proxies, redirección y negociación de contenidos.
- Varios parsers, como RDF/XML, N-Triples, Turtle, SPARQL + SPOG, Legacy XML, HTML tag soup, RSS 2.0, Google Social Graph API JSON, etc..
- Serializadores, como N-Triples, RDF/JSON, RDF/XML, Turtle, SPOG dumps, etc..
- Dos estructuras internas.
- Almacén RDF (usando MySQL).
- Instrucciones SPARQL como, SELECT, ASK, DESCRIBE, CONSTRUCT, LOAD, INSERT y DELETE.
- SPARQL Endpoint.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

6.4.2. ARC2 como herramienta para consultas SPARQL

De las herramientas mencionadas en el paso anterior, se decidió utilizar ARC2 para la creación del endpoint SPARQL utilizado en el prototipo de búsquedas semánticas, ya que puede ser implementado bajo el mismo lenguaje de programación utilizado en la construcción y no requiere levantar un servicio que requiera privilegios especiales de usuario en el servidor web como se requiere con Open Link Virtuso.

6.4.3. Implementación de ARC2 en el repositorio RDF de protocolos verbales

La implementación de ARC2 dentro del prototipo de búsquedas semánticas comienza integrando la librería dentro del desarrollo, esto se hace con la instrucción que se muestra en la Figura 43 Instrucción para incluir la librería ARC2.

```
require_once 'arc2/ARC2.php';
```

Figura 43 Instrucción para incluir la librería ARC2

Después de esta declaración es necesario configurar el endpoint ARC2, esto se hace declarando una estructura que contenga los parámetros de configuración, como se muestra en la Figura 44 Arreglo de configuración ARC2.

```
$config = array(
    'db_host' => HOSTNAME_DB_CONN,
    'db_name' => DATABASE_DB_CONN,
    'db_user' => USERNAME_DB_CONN,
    'db_pwd' => PASSWORD_DB_CONN,
    'store_name' => 'acr2',
    'max_errors' => 1000,
    'endpoint_features' => array(
        'select', 'construct', 'ask',
        'describe',
        'load', 'insert', 'delete', 'dump'
    )
);
```

Figura 44 Arreglo de configuración ARC2

Una vez configurada la herramienta, es necesario registrar cada archivo RDF de nuestro repositorio en nuestro endpoint, este paso se puede implementar junto al

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

algoritmo de trasformación relacional semántico, de esta manera, recién construido el archivo RDF, puede ser registrado en la herramienta ARC2, como la instrucción que se muestra en la Figura 45 Instrucción para el registro de archivo RDF en ARC2.

```
$endpoint = ARC2::getStore($config);
$endpoint->setUp();
$endpoint->query('LOAD <' . ARCHIVORDF . '>');
```

Figura 45 Instrucción para el registro de archivo RDF en ARC2

Este registro de archivos RDF en ARC2 queda expresado como el elemento "Registrar archivo RDF en ARC2" que se muestra en la Figura 41 Algoritmo para la creación del repositorio RDF. Posterior al registro de los archivos RDF, es posible ejecutar instrucciones SPARQL que hagan uso de nuestro repositorio RDF, la ejecución de instrucciones para la recuperación de información semántica se muestra en la Figura 46 Ejecución de consulta SPARQL en ARC2, donde en primera instancia se obtiene una representación del endpoint ARC2 y posteriormente se ejecuta una instrucción SPARQL almacenada en la variable $sparql, el resultado de esta consulta es almacenado en forma de arreglo en la variable $rows, que posteriormente podremos utilizar esta variable para interactuar con la información resultado de la consulta.

```
$endpoint = ARC2::getStore($config);
$rows = $endpoint->query($sparql, 'rows'));
```

Figura 46 Ejecución de consulta SPARQL en ARC2

6.5. Fase III "Creación de interfaz para el buscador semántico"

La creación de la interfaz de usuario para el buscador semántico es considerada según la metodología la última fase en la construcción del prototipo. Gracias al uso constante de otros buscadores utilizados con regularidad en internet, la interfaz más comúnmente implementada consta de una sola caja de texto, donde el usuario ingresa algún criterio de búsqueda, por ejemplo el uso de palabras clave en el buscador Google, tomando como tarea que se desea encontrar ejemplos de programabas para validar si

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

una palabra es un palíndromo se muestra en la Figura 47 Búsqueda "palíndromos java" en Google.

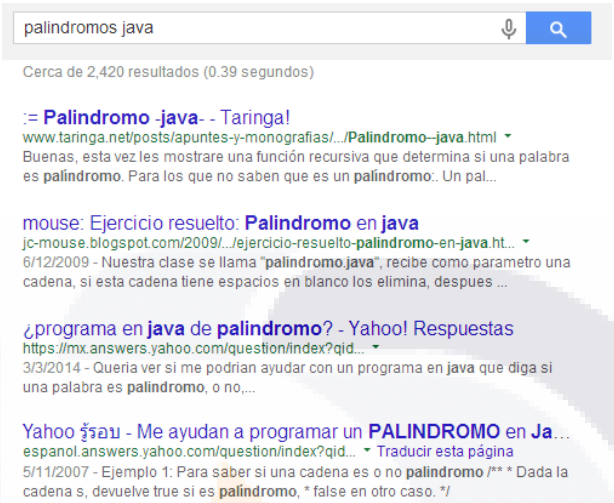

Figura 47 Búsqueda "palíndromos java" en Google

Tomando como base este tipo de interfaz para el prototipo de buscador semántico, se debe solucionar la pregunta ¿Cómo dividir una consulta en lenguaje natural en sus componentes para poder ser utilizados en el manejador semántico?, para hacer frente a esta pregunta esta tesis propone el uso de un analizador semántico o “parser” para poder distinguir a que elementos del vocabulario semántico hace referencia la expresión en lenguaje natural.

Antes del uso de la herramienta para el análisis semántico y con el fin de mostrar el proceso de construir consultas SPARQL a partir de expresiones en lenguaje natural, se creó una interfaz basada en combos anidados de sujeto, predicado y objeto, los cuales se pueden ir combinando para armar oraciones en lenguaje natural predefinidas y poder analizar las consultas SPARQL resultantes.

6.5.1. Interfaz de combos sujeto, predicado y objeto

Una primera propuesta de interfaz para el buscador semántico es el uso de tres combos anidados, combos con las diferentes combinaciones de ontologías simples en el vocabulario, estos combos se dividen en sujeto, predicado y objeto en los cuales se

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

pueden formar oraciones con distintas combinaciones de estos tres elementos, la interfaz que simulan oraciones expresadas en lenguaje natural utilizando combos se muestra en la Figura 48 Interfaz utilizando combos sujeto, predicado y objeto, en esta caso la oración “Me gustaría aprender problemas con dificultad difícil”.

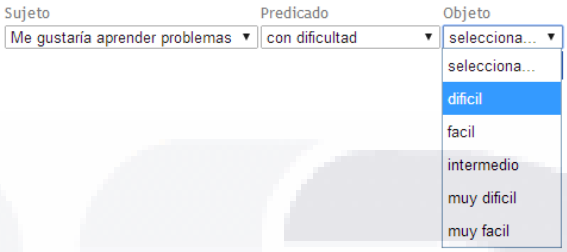

Figura 48 Interfaz utilizando combos sujeto, predicado y objeto

Esta interfaz genera los parámetros necesarios para construir sentencias SPARQL que posteriormente pueden ser implementadas en ARC2 y mostrar los resultados de los protocolos verbales que cumplen esta combinación de elementos, la consulta SPARQL generada con el ejemplo anterior se muestra en la Figura 49 Consulta SPARQL construida por interfaz de combos.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
PREFIX pps: <http://jrenelg.com/files/mitc/tesis/vocabulary/pps#> .
PREFIX dce: <http://purl.org/dc/elements/1.1#> .
SELECT *
WHERE {
    ?uri rdf:type ?tip
    FILTER regex(?tip, "pps#Problem", "i").
    ?uri dce:title ?c1 .
    ?uri pps:category ?c2 .
    ?uri pps:difficulty ?c3 .
    ?uri pps:difficulty "dificil".
    }
```

Figura 49 Consulta SPARQL construida por interfaz de combos

Después de la construcción y ejecución de la sentencia SPARQL, el resultado es mostrado en una lista de enlaces que vinculan los protocolos verbales que cumplen con el criterio de búsqueda definido en la oración, al seleccionar cualquier resultado nos abre el protocolo verbal dentro de la plataforma del sitio web para el apoyo del aprendizaje de la

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

programación básica, como se muestra en la Figura 50 Resultados en la consulta semántica.

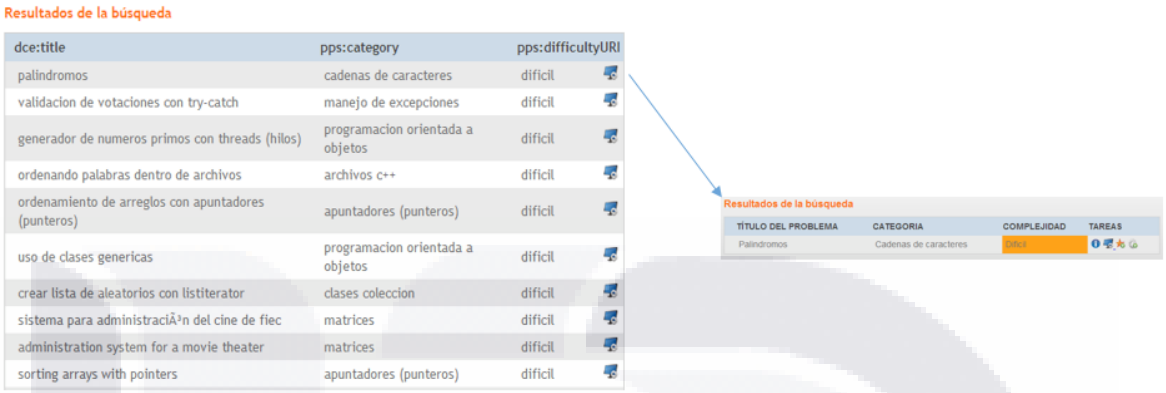

Figura 50 Resultados en la consulta semántica

6.5.2. Herramientas para el análisis semántico

Para la creación de una interfaz más amigable del prototipo, es necesario la búsqueda de una herramienta que nos auxilie en el análisis semántico de una oración en lenguaje natural, actualmente en internet existen varias herramientas con este propósito, pero basados en los criterios de selección, como el lenguaje de programación que interpreta el servidor web, el tipo de licencia del proyecto y los privilegios de usuario en el servidor, se encontraron al momento de la investigación dos herramientas en esta área.

NlpTools. Es una librería para el procesamiento del lenguaje natural escrita en PHP, el desarrollo de la librería se deriva de las necesidades de clasificación de texto, clustering, tokenizing, stemming, etc. Cabe destacar que esta librería sigue siendo un proyecto en desarrollo (“NlpTools,” 2014).

Open Calais Tags. Es una clase de PHP para la extracción de entidades y etiquetas de algún texto usando Open Calais (Reuters, 2013), Open Calais realiza un análisis semántico del texto, utilizando procesamiento del lenguaje natural para identificar conceptos como personas, empresas y tecnologías que se tratan en el texto, estos conceptos son especialmente útiles para sugerir etiquetas para el contenido de la web.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

6.5.3. NlpTools como herramienta para el procesamiento del lenguaje natural

Para el procesamiento de las búsquedas expresadas en lenguaje natural dentro del prototipo, se decidió utilizar la herramienta NlpTools ya que permite definir los conceptos clave a identificar sobre ciertas áreas del conocimiento, al contrario de Open Calais Tags que al ser un servicio disponible en internet no permite definir conceptos personalizados por el usuario del servicio.

Para el uso de NlpTools es necesario invocar en el proyecto la instrucción que se muestra en la Figura 51 Instrucciones para incluir librería NlpTools, donde a partir de esta línea es posible utilizar elementos de la clase FeatureBasedNB, con la cual podremos analizar semánticamente algún texto en lenguaje natural.

```
@require_once 'nlp-tools/autoloader.php';
use NlpTools\Tokenizers\WhitespaceTokenizer;
use NlpTools\Models\FeatureBasedNB;
use NlpTools\Documents\TrainingSet;
use NlpTools\Documents\TokensDocument;
use NlpTools\FeatureFactories\DataAsFeatures;
use NlpTools\Classifiers\MultinomialNBClassifier;
```

Figura 51 Instrucciones para incluir librería NlpTools

Para implementar una funcionalidad de analizador semántico y poder definir a que conceptos del vocabulario hace referencia una expresión en lenguaje natural, NlpTools requiere un entrenamiento previo de posibles enunciados relacionados con un concepto en especial, para realizar este entrenamiento se utilizan enunciados encontrados en la descripción de los protocolos verbales y de sus respectivos pasos de verbalización, como se muestra en Figura 52 Entrenamiento de NlpTools, donde la variable $entrenamiento es declarado como objeto de tipo TrainingSet, al cual se agregan los enunciados por medio de la variable $enunciado con el método addDocument, al finalizar de ingresar el conjunto de enunciados y para realizar el entrenamiento se declara un objeto de la clase FeatureBasedNB y se invoca el método train, al cual se le envía como parámetro el objeto instancia de la clase TrainingSet.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

```
$entrenamiento = new TrainingSet();
$tok = new WhitespaceTokenizer();
$data = new DataAsFeatures();
$entrenamiento->addDocument($concepto,
                    new TokensDocument(
                        $tok->tokenize($enunciado)
                        )
                    );
$model = new FeatureBasedNB();
$model->train($data, $entrenamiento);
```

Figura 52 Entrenamiento de NlpTools

Posterior al entrenamiento de NlpTools, es posible utilizar el modelo para analizar semánticamente las búsquedas expresadas en lenguaje natural como se muestra en la Figura 53 Análisis semántico en NlpTools, es necesario crear un objeto de la clase MultinomialNBClassifier el cual recibe el modelo del entrenamiento, enseguida se crea un objeto de la clase TokensDocument al cual se le asigna la expresión en lenguaje natural escrita por el usuario, previamente almacenada en la variable $expre, y al finalizar se invoca el método classify, que realiza el análisis semántico y regresa el posible concepto del vocabulario al cual hace referencia la expresión, esta concepto se almacena en la variable $prediction.

```
$cls = new MultinomialNBClassifier($data, $model);
$tkd = new TokensDocument($tok->tokenize($expre));
$prediction = $cls->classify($myArr, $tkd);
```

Figura 53 Análisis semántico en NlpTools

6.5.4. Creación de interfaz para el motor de búsqueda semántica

Como último paso en la construcción del prototipo y después de realizar el análisis semántico, el único paso restante es crear la interfaz de usuario. En esta tesis se propone la interfaz que se muestra en la Figura 54 Interfaz del buscador semántico, donde en la única caja de texto el usuario escribe su expresión en lenguaje natural, y posteriormente se realiza en análisis semántico de la expresión con NlpTools para identificar el concepto del vocabulario PPS al que se hace referencia, y con este concepto poder construir la instrucción SPARQL para ser ejecutada en el Endpoing ARC2, el cual recupera la información semántica almacenada en el repositorio RDF creado con EasyRDF.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

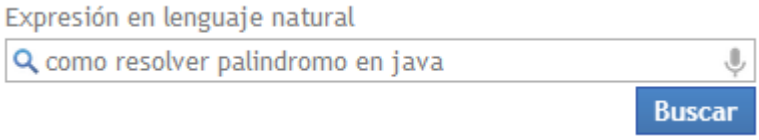

Figura 54 Interfaz del buscador semántico

El resultado de esta búsqueda semántica se divide en tres secciones, la primera sección se muestra en la Figura 55 Resultado del análisis semántico, donde se observa el resultado del análisis semántico el cual identifica el sujeto y predicado al que hace referencia la expresión en lenguaje natural.

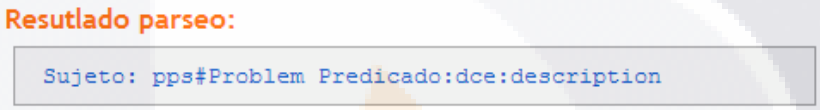

Figura 55 Resultado del análisis semántico

Enseguida se muestra la consulta SPARQL construida en base al resultado del análisis semántico como se muestra en la Figura 56 Consulta SPARQL generada por el prototipo.

Figura 56 Consulta SPARQL generada por el prototipo

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Como última sección del resultado de la búsqueda semántica es el listado de los posibles protocolos verbales que dan solución a la expresión en lenguaje natural expresada por el usuario, como se muestra en la Figura 57 Listado de protocolos verbales resultantes de la búsqueda semántica, al dar un clic en la columna de URI correspondiente al protocolo que queramos utilizar, nos envía al protocolo verbal dentro del sitio web para el apoyo al aprendizaje de la programación.

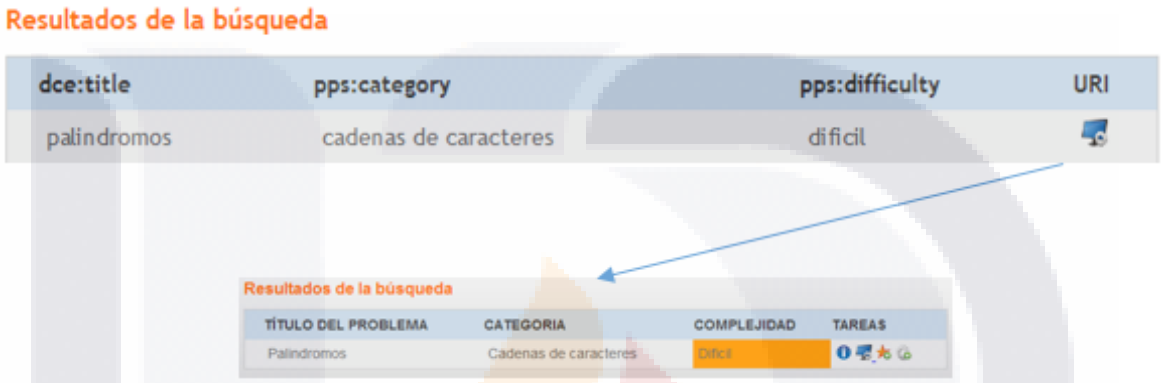

Figura 57 Listado de protocolos verbales resultantes de la búsqueda semántica

6.6. Diseño del test retrospectivo

Con base a los objetivos de la tesis, donde se busca medir la usabilidad del prototipo de búsquedas semánticas, se diseñó un test de usabilidad de tipo RTA o pensamiento en voz alta retrospectivo, donde se graba al participante realizando una serie de tareas y verbalizando sus pensamientos, este test sigue un guión y unas tareas previamente definidas.

El test se diseñó basado en lo propuesto por el autor Steve Krug y un ejemplo de este guión se muestra en el Anexo B, el listado de tareas que realizó cada participante se muestra en el anexo C, mientras que un ejemplo del formato de consentimiento de la grabación se muestra en el anexo A.

El software utilizado para la prueba de usabilidad fue TechSmith Morae, el cual fue pensado en estudios de usabilidad a grupos de enfoque, Morae ayuda a entender mejor las experiencias de los usuarios al proporcionar datos de gran alcance, grabar y

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

remotamente observar las interacciones del usuario, analizar de manera eficiente los resultados y compartir los hallazgos fácilmente.

El test consistió en la ejecución de cinco búsquedas básicas en el prototipo, de las cuales fueron recabados datos cuantitativos y cualitativos por medio de la grabación, las actividades realizadas por cada participante se muestran en el Anexo C. El test fue aplicado a seis participantes de los cuales se describen características como edad, profesión, nivel de estudios y horas a la semana en internet la, las características de los participantes a los que se les aplicó el test se muestran en la Tabla 4 Listado de participantes del test. Los participantes fueron seleccionados según un perfil informático y tiempo en el uso de internet similar, buscando reducir el sesgo que puede generar la falta de conocimiento de algunos conceptos informáticos necesarios para la comprensión de las tareas y la poca experiencia en el uso de internet.

Participante	Edad	Nivel de estudios	Profesión	Horas semanales en internet
Alejandro	29	Licenciatura	Analista programador	72
Carolina	28	Licenciatura	Programador Mainframe	60
Roberto	27	Licenciatura	Ingeniero en electrónica	53
Mario	33	Maestría	Ingeniero en sistemas	40
José Luis	28	Licenciatura	Desarrollador de SW	100
Ramón	26	Especialidad	Ingeniero de software	62

Tabla 4 Listado de participantes del test

6.7. Usabilidad del prototipo

Una vez aplicados los test a los participantes, se reunieron los resultados con el fin de presentar distintos indicadores de usabilidad del prototipo de buscador semántico, como se muestra en la Tabla 5 Indicadores de usabilidad.

Indicador	Objetivo
Tiempo de ejecución	Mostrar el tiempo dedicado a resolver la tarea
Nivel de cumplimiento	Mostar el grado de satisfacción con el que se cumplió la tarea
Tiempo entre eventos	Mostar el tiempo promedio en que tarda el usuario en encontrar el resultado en el listado
Cantidad de clics	Mostar el total de clics que utiliza el usuario en completar la tarea

Tabla 5 Indicadores de usabilidad

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Tiempo de ejecución. Es el tiempo promedio dedicado a cada una de las tareas, como se muestra en la Figura 58 Tiempo promedio por tarea, la tarea con mayor consumo de tiempo es la tarea dos con un promedio de 1.87 minutos, mientras que la de menor tiempo fue la tarea tres con 1.14 minutos y en promedio el tiempo requerido en el buscador semántico para realizar una búsqueda y llegar al resultado es de 1.46 minutos, una descripción más detallada de los tiempos por tarea se presenta en la Tabla 6.

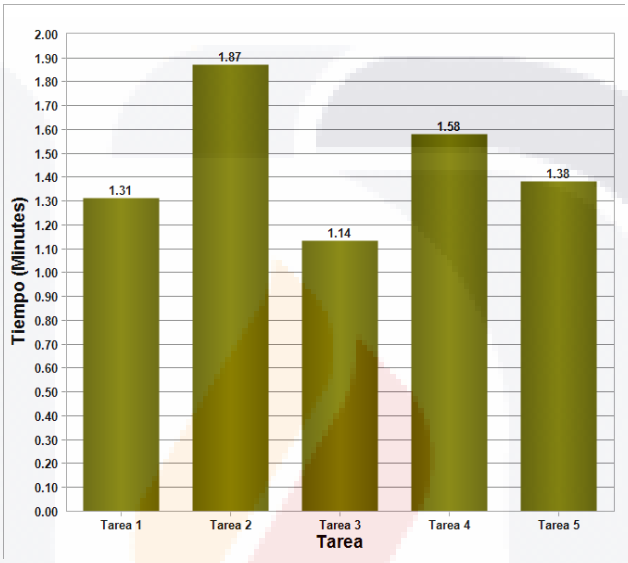

Figura 58 Tiempo promedio por tarea

Participante	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Promedio
Alejandro	1.43	1.88	1.25	1.46	1.73	**1.55**
Carolina	0.81	1.63	1.54	2.12	1.42	**1.50**
José Luis	0.77	2.61	0.58	1.17	0.47	**1.12**
Mario	2.48	1.65	1.12	1.95	2.43	**1.93**
Ramón	0.9	1.76	1.13	1.39	1.43	**1.32**
Roberto	1.48	1.71	1.2	1.39	0.82	**1.32**
Promedio	**1.31**	**1.87**	**1.14**	**1.58**	**1.38**	**1.46**

Tabla 6 Tabla de tiempos por tarea

Nivel de cumplimiento. Se refiere a la manera de cómo se completó la tarea, existen tres valores posibles, los cuales fueron asignados de la siguiente manera, “completada fácilmente” cunado el participante cumplió el objetivo de la actividad directamente sin necesidad de rescribir la expresión en lenguaje natural, “completada con dificultar” cuando

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

el participante se vio a la necesidad de reescribir la expresión ya se por falta de resultados o por que los resultados no cumplieron con lo que el participante deseaba encontrar y “No completada” cuando el usuario no encontró lo que se pretendía buscar y termino la tarea sin llegar a un resultado. La proporción de cumplimiento por tarea se muestra en la Figura 59 Nivel de cumplimiento por tarea, donde se observa que la mayoría de las tareas fueron completadas fácilmente, excepto la actividad cuatro donde un 66.67% de los casos fueron tareas completadas con dificultar y existiendo un 16.67% de tareas no completadas y en la actividad cinco donde el 66.67% de los casos fueron tareas completadas con dificultad.

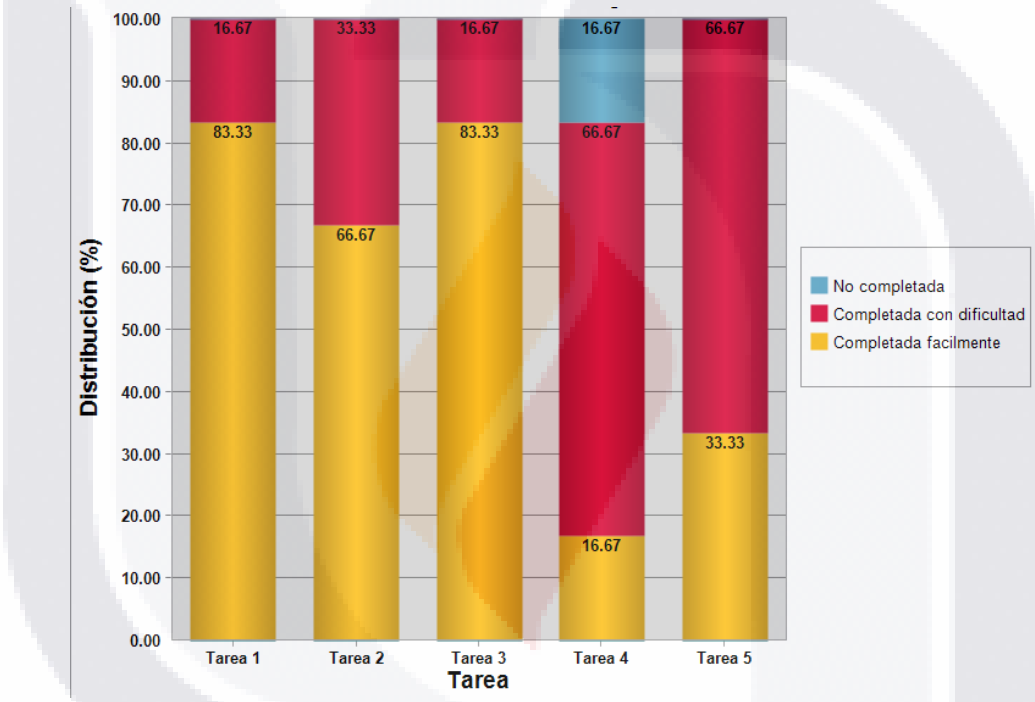

Figura 59 Nivel de cumplimiento por tarea

Tiempo entre eventos. Este indicador muestra el tiempo máximo entre eventos en cada tarea, el cual puede indicar el tiempo que transcurre desde que el participante lanza la búsqueda y navega entre los resultados hasta encontrar alguno que satisfaga la tarea, de esta manera podremos captar que tan precisa fue la expresión en lenguaje natural ya que entre más ambigua sea más resultados lanzara la buscador, tomando más tiempo en encontrar lo que se desea. La terea dos es la que se consume más tiempo en buscar un resultado que resulte satisfactorio tomando un promedio de 31.30 segundos como se muestra en la Figura 60. El desglose por participante del máximo de tiempo entre eventos

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

por tarea se muestra en la Tabla 7, en general toma 19.89 segundos encontrar un resultado después de lanzar la búsqueda en el motor.

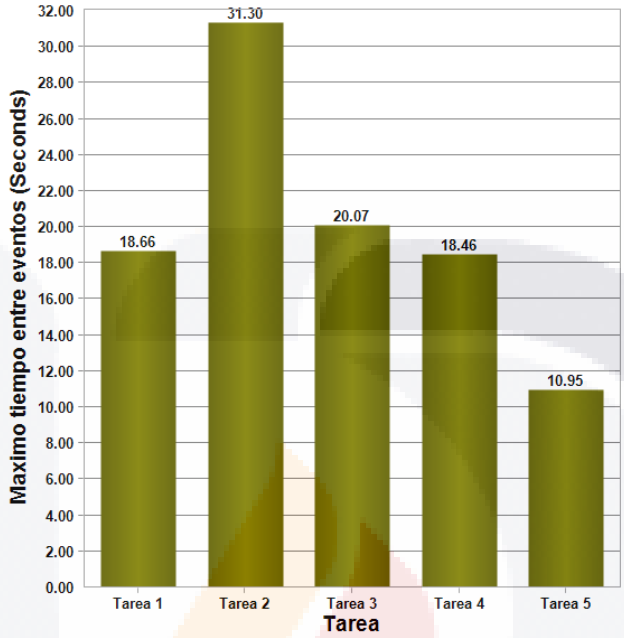

Figura 60 Máximo de tiempo entre eventos

Participante	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Promedio
Alejandro	31.16	30.28	30.36	16.53	12.67	**24.20**
Carolina	17.92	35.69	29.81	17.39	11.22	**22.41**
José Luis	11.65	21.13	10.98	21.13	7.94	**14.57**
Mario	21	23.63	17.97	23.78	11.92	**19.66**
Ramón	15.99	26	16.09	16.98	14.03	**17.82**
Roberto	14.22	51.1	15.17	14.92	7.89	**20.66**
Promedio	**18.66**	**31.3**	**20.07**	**18.46**	**10.95**	**19.89**

Tabla 7 Desglose por participante del tiempo máximo por tarea (segundos)

Numero de clics. Es el total de clics realizados para completar cada tarea, entre mayor número de clics, mayor el esfuerzo requerido para culminarla, como se muestra en la Figura 61 Promedio de clics por tarea, la tarea que requiere mayor número de clics es la tarea cinco con 12.67 clics, mientras que la tarea con menor número de clics requeridos fue la tarea uno con únicamente 7 clics, en promedio se requieren 9.23 clics en el buscador semántico para llegar a una solución, El desglose de los clics necesarios por

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

cada tarea y participante se muestra en la Tabla 8 Desglose de clics por participante y tarea.

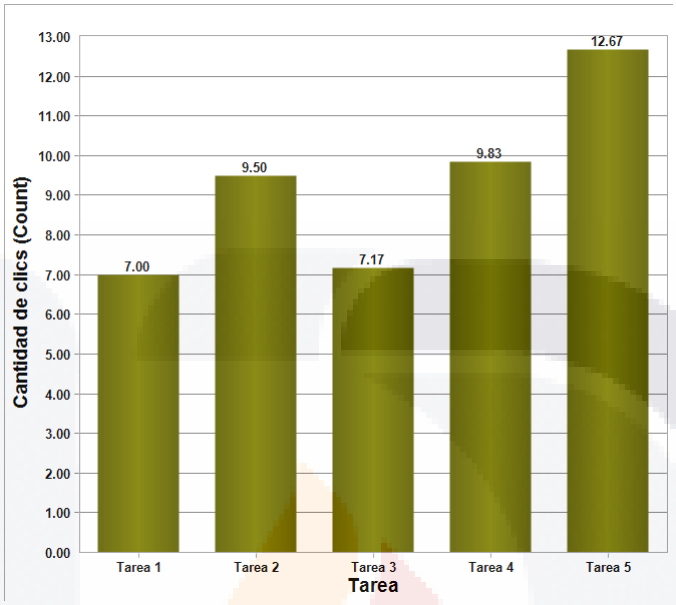

Figura 61 Promedio de clics por tarea

Participante	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Promedio
Alejandro	3	3	5	6	8	**5.00**
Carolina	4	10	8	17	18	**11.40**
José Luis	6	28	7	8	5	**10.80**
Mario	21	7	13	16	30	**17.40**
Ramón	2	5	5	4	8	**4.80**
Roberto	6	4	5	8	7	**6.00**
Promedio	**7**	**9.5**	**7.17**	**9.83**	**12.67**	**9.23**

Tabla 8 Desglose de clics por participante y tarea

7. Discusión de resultados

Esta tesis busca describir detalladamente el proceso de la construcción de un prototipo de búsquedas semánticas basado en una fuente de información relacional y durante este proceso evaluar el esfuerzo y las implicaciones técnicas de la construcción, así como medir las ventajas en cuanto a usabilidad del prototipo.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Pala lograr esta investigación se realizó una búsqueda minuciosa de posibles métodos existentes para la creación de un buscador semántico basado en información almacenada en una base de datos relacional, debido a que esta tecnología se encuentra en proceso de desarrollo, aun no hay un proceso automatizado para alcanzar este propósito, por lo cual se decidió dividir la construcción en fases importantes como la transformación relacional a semántico, la implementación de un motor de búsqueda semántica y el uso de una analizador semántico de expresiones en lenguaje natural.

Para el logro de estas tres etapas se buscaron herramientas que pueden ayudar en cada una de las tareas, se evaluaron de acuerdo a criterios basados en el ambiente tecnológico donde el prototipo de búsquedas semánticas fue implementado y posteriormente se realizó una combinación de las herramientas para conseguir un prototipo de buscador semántico.

Durante la construcción del prototipo, se tomaron datos sobre el esfuerzo necesario para del proceso de transformación relacional semántico, dividiendo este proceso en distintas tareas y tomando el esfuerzo en horas dedicada a cada una, como se muestra en la Tabla 3 Medición de esfuerzo (horas) de la transformación. Estos resultados en próximas investigaciones que decidan utilizar esta metodología podrían ayudar a estimar mejor los tiempos de desarrollo de la investigación.

En cuanto a la usabilidad del prototipo de búsquedas semánticas se realizó un test en voz alta retrospectivo o RTA, el cual fue basado en un guion presentado en el Anexo B construido siguiendo lo propuesto por Steve Krug un buen referente en estudios de usabilidad, el test fue aplicado a un conjunto de seis participantes los cuales realizaron cinco tareas de búsquedas en el prototipo, la característica de los participantes se presenta en la Tabla 4 Listado de participantes del test. Los participantes fueron grabados con el fin de recabar datos cuantitativos y cualitativos de la ejecución del test, posteriormente los datos fueron analizados según una serie de indicadores como el tiempo de ejecución por tarea como se muestra en la Figura 58 Tiempo promedio por tarea, nivel de cumplimiento como se muestra en la Figura 59 Nivel de cumplimiento por tarea, tiempo entre eventos como se muestra en la Figura 60 Máximo de tiempo entre eventos y numero de clics como se muestra en la Figura 61 Promedio de clics por tarea. Este estudio se limita a mostrar los resultados del test de usabilidad, pero se cree

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

conveniente en próximas investigaciones hacer una comparación con algún buscador por palabra clave que abarque la misma área del conocimiento donde se construya el prototipo.

Este estudio de tesis puede generalizarse y aplicarse a otros contextos de búsquedas en internet, las tareas de buscar y/o crear los vocabularios y la construcción de la instrucción SPARQL son las únicas tareas que tendrían que especializarse de acuerdo al ámbito del conocimiento donde se desee utilizar esta metodología. Algunos de los aspectos que podrían mejorarse en próximas investigaciones es el uso del analizador semántico para identificar el concepto del vocabulario al que se hace referencia, el analizador semántico necesita ser entrenado con ejemplos de expresiones para cada uno de los conceptos del lenguaje y de esta manera mejorar las predicciones del analizador, en este sentido próximas investigaciones pueden dedicar esfuerzos a mejorar el uso del analizador semántico.

Conclusiones

Conclusión general

Este estudio de tesis hace una revisión del estado actual de la web semántica, por medio de la propuesta de una metodología de construcción y la implementación de un prototipo de buscador semántico, el cual tiene como fin medir el esfuerzo de transformación relacional-semántico y la usabilidad del prototipo, contribuyendo para próximas investigaciones que decidan utilizar esta metodología en la implementación de un buscador semántico en algún otra área del conocimiento.

Con base a los resultados obtenidos en el proceso de construcción del prototipo y tomando en cuenta el objetivo general de este estudio, podemos concluir que la adopción de algunas tecnologías propias de la web semántica continúan siendo un trabajo que requiere alto conocimiento técnico y es crítico contar con las herramientas más adecuadas para cada caso en particular, contrastando con la usabilidad del prototipo cuyos indicadores de satisfacción muestran tiempos cortos y un porcentaje de efectividad alto.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

El estudio representa una buena implementación de búsquedas semánticas en el caso específico del uso de protocolos verbales para el apoyo en el aprendizaje de la programación, esta implementación facilita a los usuarios encontrar con poco esfuerzo aquellas soluciones que sean de su interés.

La web semántica es una tecnología aún en proceso de desarrollo, la cual actualmente se encuentra en un estado avanzado que permite la construcción de herramientas que comiencen a mostrar los beneficios que conlleva la adopción de esta tecnología.

Identificar herramientas para migrar de RDBMS a RDF

Uno de los aspectos a considerar en la adopción de esta tecnología es la transformación de la información existente en internet a una representación semántica, para esta transformación existen actualmente una gran variedad de herramientas, como lo es D2RQ, Triplify, Jena, Redland, EasyRDF, Sasame, etc.

Aplicar la herramienta más adecuada al caso específico de protocolos verbales creando un repositorio RDF

Para esta tesis se definieron una serie de criterios en base al ambiente tecnológico donde fue implementado el prototipo de buscador semántico, estos criterios para la selección de la herramienta a utilizar en la transformación fueron:

- Programada en el lenguaje de programación PHP.
- Compatibilidad con bases de datos en MySQL.
- Implementación directa en algún proyecto, sin necesidad de algún servicio adicional.

De acuerdo a estos criterios, para esta tesis se decidió utilizar la herramienta EasyRDF ya que cumple con todos los criterios y además es una herramienta que permite la construcción de un repositorio semántico de una manera sencilla.

Medir el esfuerzo del proceso de migración RDMBS a RDF.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Para lograr la transformación relacional semántica se divido el proceso en una serie de tareas, las cuales fueron:

1. Búsqueda y análisis de herramientas RDF.
2. Evaluación de las herramientas RDF.
3. Comparativa y selección de la herramienta RDF.
4. Definición del vocabulario.
5. Diseño del algoritmo de transformación.
6. Implementación del algoritmo.
7. Pruebas del algoritmo.

Para cada tarea de este proceso de transformación fue medido el esfuerzo en horas dedicadas, teniendo como resultado 91 horas en total.

Identificar herramientas para explotar el repositorio RDF mediante SPARQL.

Una vez transformada la información de la base de datos a una representación semántica basada en RDF, fue implementado un motor de búsquedas semánticas, actualmente existen herramientas que permiten realizar consultas SPARQL en repositorios RDF, ente las que destacan Open Link Virtuoso Unviersal Server y Semsol ARC2.

Adaptar la herramienta más adecuada al repositorio RDF.

Para la implementación de este motor de búsquedas semánticas en esta tesis se utilizó ARC2, debido a que es una herramienta basada en PHP y no requiere un servicio adicional disponible en el servidor.

Crear una interfaz que interprete expresiones en lenguaje natural y construya la correspondiente consulta SPARQL.

En seguida de la implementación del motor de búsquedas semánticas y como característica principal del prototipo de buscador semántico, es necesario construir una interfaz que permita al usuario generar consultas SPARQL de una manera transparente,

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

para ello se utilizó un analizador semántico llamado NlpTools que en base al entrenamiento con expresiones en lenguaje natural ligadas a los conceptos del vocabulario es posible inferir a que área del conocimiento hace referencia el usuario con su expresión ingresada en el buscador.

Medir la usabilidad del prototipo de búsquedas semánticas.

Una vez terminada la construcción del prototipo se evaluó la usabilidad del mismo en base a un test retrospectivo RTA, que recabó datos cuantitativos y cualitativos de usabilidad en base a una serie de indicadores como lo son el tiempo promedio por tarea, el nivel de cumplimiento, tiempo entre eventos y cantidad de clics. Teniendo un promedio que por búsqueda en el prototipo se requiere 1.46 minutos de los cuales se emplean 19.89 segundos en promedio encontrar el protocolo satisfactorio en la lista de resultados, utilizando 9.23 clics en llegar a un resultado, obteniendo un promedio de 56.66% de tareas completadas sin dificultad, un 40% de tareas completadas con algún tipo de dificultar y un 3.34% de tareas no completadas.

Acerca de las tecnologías inmersas en la web semántica

Como se muestra en el desarrollo de los resultados de esta tesis, actualmente no existe una tecnología que automáticamente construya una representación semántica de un área del conocimiento específica y que posteriormente esta representación pueda ser empleada por un buscador semántico para el mejoramiento de sus resultados.

La falta de una herramienta que cumpla con las características mencionadas anteriormente, se debe a que algunas de las tecnologías inmersas en la web semántica continúan en proceso de maduración, lo cual provoca que la construcción de herramientas semánticas actualmente requiera de un alto grado de conocimientos sobre el tema y buenas habilidades técnicas que permitan una correcta aplicación de las tecnologías requeridas para que esta web evolucione correctamente.

Sobre el estudio y visión futura

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Este estudio de tesis, fue delimitado al caso específico de la representación semántica en base a un modelo relacional sobre soluciones a problemas de programación por medio de protocolos verbales, pero se considera oportuno plantear nuevas líneas de investigación en dos vertientes, la primera referente a proponer mecanismos no solo para la transformación en base a modelos relacionales, si no también transformar la información existente en otros formatos, como lo son paginas HTML, RSS, XML, etc. La segunda sobre nuevos mecanismos que permitan explotar con mayor eficacia la totalidad de los vocabularios inmersos en la representación semántica, por medio de mejoras y/o nuevos mecanismos de análisis semántico.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Glosario

Concepto	Significado
Web Semántica	La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida.
Protocolo verbal	Un protocolo verbal (PV) es un mecanismo para estudiar el contenido de la memoria de corto plazo de las personas en el proceso de resolver problemas o llevar a cabo una tarea predefinida (Ericsson & Simon, 1993).
Motor de búsqueda	Son sistemas de búsqueda por palabras clave. Son bases de datos que incorporan automáticamente páginas web mediante "robots" de búsqueda por la red.
Ontología	No existe una división clara "vocabulario" y "ontología" en informática. La tendencia es a utilizar la palabra "ontología", siendo una colección bastante formal de términos, mientras que el término "vocabulario" se utiliza cuando tal formalismo estricto no se utiliza necesariamente o sólo en un sentido menos estricto. Los vocabularios son los bloques de construcción básicos de las técnicas de inferencia sobre la Web Semántica.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Bibliografía.

Adermann, N., & Boggiano, J. (2014). Composer. Retrieved from https://getcomposer.org/

Aleven, V., & Azevedo, R. (2013). Overview of Current Interdisciplinary Research. In *Metacognition and Learning Technologies*. New York: Springer International Handbooks of Education.

Alfredo, L. M. (2013, Enero). *Recuperación de información para respuesta a preguntas en documentos legales*. INSTITUTO POLITÉCNICO NACIONAL.

Anita Ferreira, & Atkinson, J. (2013, June 21). Disminución de la sobrecarga de información en la World Wide Web a partir de interacciones dialógicas hombre-computador. Retrieved June 21, 2013, from http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-09342009000100001&lng=en&nrm=iso&ignore=.html

Arévalo, C., Muñoz, L., & Gómez, J. (2011). A Software Tool to Visualize Verbal Protocols to Enhance Strategic and Metacognitive Abilities in Basic Programming. *International Journal of Interactive Mobile Technologies (iJIM)*, *5*(3), 6–11.

Arevalo, C., & Solano, L. (2012). The Use of Verbal Protocols as Learning Materials for Introductory Programming. In *Proceedings of the 1st International Workshop on Technology Transfer and Learning Solutions for Programming Education* (Vol. 1, pp. 18–29). Bucharest, Rumania: Conspress, Bucuresti.

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: an introduction. *Metacognition Learning*, *2*, 57–65.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Badr, Y. (Ed.). (2010). *Emergent Web intelligence: advanced semantic technologies*. London ; New York: Springer.

Begel, A. (1996). *LogoBlocks: A Graphical Programming Language for Interacting with the World*. MITEditor.

Bergman, M. K. (2001). White Paper: The Deep Web: Surfacing Hidden Value. Retrieved from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104

Berners-Lee, T. (2000). Architecture. Retrieved from http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

Bizer, C., & Heath, T. (2009). Linked Data - The Story So Far.

Bornat, R., Dehnadi, S., & Simon. (2008). Mental models, Consistency and Programming Aptitude. Presented at the Psychology of Programming interested Group (PPIG).

Bry, F., & Małuszyński, J. (Eds.). (2009). *Semantic techniques for the web: the REWERSE perspective*. Berlin ; New York: Springer.

Cohen-Almagor, R. (2011). Internet History. University of Hull, UK.

comScore. (2014). comScore Releases March 2014 U.S. Search Engine Rankings. Retrieved from http://www.comscore.com/Insights/Press_Releases/2014/4/comScore_Releases_March_2014_U.S._Search_Engine_Rankings

Corlosquet, S. (2014). Semsol ARC2. Retrieved from https://github.com/semsol/arc2/wiki

Davies, J. (2006). *Semantic Web technologies: trends and research in ontology-based systems*. Chichester, England ; Hoboken, NJ: John Wiley & Sons.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

De Virgilio, R., Guerra, F., & Velegrakis, Y. (2012). *Semantic search over the web*. Berlin; New York: Springer. Retrieved from http://dx.doi.org/10.1007/978-3-642-25008-8

Downs, G. H., & McAllen, D. K. (2012). The Effect of Intrinsic Motivation on Success in a Technology Management Undergraduate Program. In *Proceedings of PICMET '12: Technology Management for Emerging Technologies.* Vancouver, Canada.

DuCharme, B. (2013). *Learning SPARQL*.

Elbedweihy, K., & Wrigley, S. N. (2012). Evaluating Semantic Search Systems to Identify Future Directions of Research.

Ericsson, K., & Simon, H. A. (1993). *Protocol Analysis. Verbal reports as data* ((rev. ed).). Cambridge Massachusets.: MIT Press.

Fensel,, D. (2005). *Spinning the Semantic Web*.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, *34*(10), 906–911.

Google. (2014a). Motor de búsqueda de Google. Retrieved from https://www.google.com.mx/?gws_rd=ssl

Google. (2014b). Resultados de búsqueda en Google. Retrieved from https://www.google.com.mx/?gws_rd=ssl#q=problemas+de+programacion

Halpin, H. (2008). Foundations of a Philosophy of Collective Intelligence. University of Edinburgh.

Humfrey, N. (2014). EasyRDF. Retrieved from http://www.easyrdf.org/

Jenkins, T. (2002). On the difficulty of learning to program. Loughborough University: LTSN Centre of information and computer sciences.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Kelleher, C. P., & Pausch, R. (2005). Lowering the Barriers to Programming: a survey of programming environments and languages for novice programmers. *ACM Computing Surveys (CSUR)*, *37*(2), 83 – 137.

Kioskea.net. (2014). Web semántica: las aplicaciones actuales. Retrieved from http://es.kioskea.net/faq/7082-web-semantica-las-aplicaciones-actuales

Koedinger, K., & Aleven, V. (2004). Toward a Rapid Development Environment for Cognitive Tutors (pp. 167–179). Presented at the Engineering Advanced Web Applications: Proceedings of Workshops in Connection with the 4th International Conference on Web Engineering, Princeton: Rinton Press.

Krug, S. (2006). *Don't make me think!: a common sense approach to Web usability* (2nd ed.). Berkeley, Calif: New Riders Pub.

Lamarca, M. (2013). Hipertexto, el nuevo concepto de documento en la cultura de la imagen. Retrieved from http://www.hipertexto.info/documentos/b_datos.htm

Levene, M. (2010). *An introduction to search engines and web navigation* (2nd ed.). Hoboken, N.J: John Wiley.

Lexxebeta. (2014a). Ejemplo búsqueda en Lexxe. Retrieved from http://www.lexxe.com/cc?sstring=programming+language%3A+cycles+java&src=hp

Lexxebeta. (2014b). Lexxebeta. Retrieved from http://www.lexxe.com/

Lexxebeta. (2014c). Llaves semánticas Lexxe. Retrieved from http://www.lexxe.com/semkeylist.html

Linked Open Vocabularies (LOV). (2014). Retrieved from lov.okfn.org/dataset/lov/index.html

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

LinkingOpenData. (2013). Retrieved from http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

Lourdes, V., & Carro, J. (2004). Para acceder al web profundo: conceptos y herramientas.

Ma, L., Ferguson, J., Roper, M., & Wood, M. (2007). Investigating the viability of mental models held by novice programmers. *ACM SIGCSE Bulletin*, *39*(1), 499–503.

Nadia, M., & Prem, K. (1998). Face to Virtual Face. Retrieved from http://ivizlab.sfu.ca/arya/Papers/Others/Face%20to%20Virtual%20Face.htm

Netcraft. (2014). Número de sitios web. Retrieved from http://news.netcraft.com/archives/2014/06/06/june-2014-web-server-survey.html

Networks, R., & Nova, S. (2007). The past, present and future of the Web. Retrieved from http://novaspivack.typepad.com/RadarNetworksTowardsAWebOS.jpg

Nie, J.-Y. (2010). *Cross-Language Information Retrieval*.

NlpTools. (2014). Retrieved from http://php-nlp-tools.com/

OpenLink Virtuoso. (2014). SPARQL Implementation Details. Retrieved from http://docs.openlinksw.com/virtuoso/rdfsparql.html#rdfsupportedprotocolendpoint

OWL Working Group. (2012). OWL. Retrieved from http://www.w3.org/2001/sw/wiki/OWL

Pan, J. rey. (2004). *Description logics: reasoning support for the semantic web*. Retrieved from http://homepages.abdn.ac.uk/jeff.z.pan/pages/pub/thesis.pdf

Pandey, G. (2012). The Semantic Web: An Introduction and Issues.

Passin, T. B. (2004). *Explorer's guide to the Semantic Web*. Greenwich: Manning.

Proal, C. (2013). Almacenamiento y Recuperación de Información. Retrieved from http://www.carlosproal.com/bda/capitulo4.html

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Quin, L. R. E. (2014). XML ESSENTIALS. Retrieved from http://www.w3.org/standards/xml/core

Ramalingam, V. L. (2004). Self-Efficacy and mental models in learning to program (pp. 171 – 175). Leeds, United Kingdom: ACM Press New York, NY, USA.

RDF y RDF schema. (2013, June 21). Retrieved June 21, 2013, from http://www.matem.unam.mx/~grecia/semantic_web/rdf.html

Reuters, T. (2013). CALAIS. Retrieved from http://www.opencalais.com/

Rist, R. S. (1996). System Structure and Design. In *Empirical studies of programmers: sicth workshop*. Norwood, New Jersey: Ablex Publishing Corporation.

Rist, R. S. (2004). Learning to program: schema creation, application and evaluation. In *Computer Science Education and Research* (pp. 175–197). Netherlands: Taylor & Francis Group.

Sareh, A., & Mohammad, N. (2012). EVOLUTION OF THE WORLD WIDE WEB: FROM WEB 1.0 TO WEB 4.0.

Schiefele, U. (1991). Interest, Learning and Motivation. *Educational Psychologist*, *26*(3&4), 299–323.

Sheykh Esmaili, K., & Abolhassani, H. (2006). A Categorization Scheme for Semantic Web Search Engines. Sharif University of Technology, Tehran, Iran.

Sistema Visor de Protocolos Verbales. (2014). Sistema Visor de Protocolos Verbales. Retrieved from http://dsi.ccbas.uaa.mx/Protocolos3/

SKOS Simple Knowledge Organization System. (2013). Retrieved from http://skos.um.es/

Sosa, E. (1997). Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones (Parte I). Retrieved from

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

http://www.elprofesionaldelainformacion.com/contenidos/1997/enero/procesamiento_del_lenguaje_natural_revisin_del_estado_actual_bases_tericas_y_aplicaciones_parte_i.html

Staab, S., & Studer, R. (2009). *Handbook on Ontologies*. Springer.

Swoogle. (2014a). Swoogle. Retrieved from http://swoogle.umbc.edu/

Swoogle. (2014b). Swoogle Manual. Retrieved from http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=search_swd

Swoogle Semantic Web Search. (2007). Retrieved from http://swoogle.umbc.edu/

U.S. Department of Health & Human Services. (2014). usability.gov. Retrieved from http://www.usability.gov/

Universidad Berlin. (2013, June 21). The Linking Open Data cloud diagram. Retrieved June 21, 2013, from http://lod-cloud.net/

W3C. (2004). Resource Description Framework (RDF). Retrieved from http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/Graph-ex.gif

W3C. (2010). Use Cases and Requirements for Mapping Relational Databases to RDF.

W3C. (2013, June 21). ConverterToRdf - W3C Wiki. Retrieved June 21, 2013, from http://www.w3.org/wiki/ConverterToRdf

W3C. (2014). SparqlEndpoints. Retrieved from http://www.w3.org/wiki/SparqlEndpoints

Wolframalpha. (2014a). Interfaz WolframAlpha. Retrieved from http://www.wolframalpha.com/

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Wolframalpha. (2014b). Opción de contexto WolframAlpha. Retrieved from http://www.wolframalpha.com/input/?i=how+to+program+java+cycles

WolframAlpha. (2014). *Computational knowledge engine*. Retrieved from http://www.wolframalpha.com/

World Wide Web Consortium. (2014). Guía Breve de Web Semántica. Retrieved from http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica

Adermann, N., & Boggiano, J. (2014). Composer. Retrieved from https://getcomposer.org/

Aleven, V., & Azevedo, R. (2013). Overview of Current Interdisciplinary Research. In *Metacognition and Learning Technologies*. New York: Springer International Handbooks of Education.

Alfredo, L. M. (2013, Enero). *Recuperación de información para respuesta a preguntas en documentos legales*. INSTITUTO POLITÉCNICO NACIONAL.

Anita Ferreira, & Atkinson, J. (2013, June 21). Disminución de la sobrecarga de información en la World Wide Web a partir de interacciones dialógicas hombre-computador. Retrieved June 21, 2013, from http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-09342009000100001&lng=en&nrm=iso&ignore=.html

Arévalo, C., Muñoz, L., & Gómez, J. (2011). A Software Tool to Visualize Verbal Protocols to Enhance Strategic and Metacognitive Abilities in Basic Programming. *International Journal of Interactive Mobile Technologies (iJIM)*, *5*(3), 6–11.

Arevalo, C., & Solano, L. (2012). The Use of Verbal Protocols as Learning Materials for Introductory Programming. In *Proceedings of the 1st International Workshop on*

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Technology Transfer and Learning Solutions for Programming Education (Vol. 1, pp. 18–29). Bucharest, Rumania: Conspress, Bucuresti.

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: an introduction. *Metacognition Learning*, *2*, 57–65.

Badr, Y. (Ed.). (2010). *Emergent Web intelligence: advanced semantic technologies*. London ; New York: Springer.

Begel, A. (1996). *LogoBlocks: A Graphical Programming Language for Interacting with the World*. MITEditor.

Bergman, M. K. (2001). White Paper: The Deep Web: Surfacing Hidden Value. Retrieved from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104

Berners-Lee, T. (2000). Architecture. Retrieved from http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

Bizer, C., & Heath, T. (2009). Linked Data - The Story So Far.

Bornat, R., Dehnadi, S., & Simon. (2008). Mental models, Consistency and Programming Aptitude. Presented at the Psychology of Programming interested Group (PPIG).

Bry, F., & Małuszyński, J. (Eds.). (2009). *Semantic techniques for the web: the REWERSE perspective*. Berlin ; New York: Springer.

Cohen-Almagor, R. (2011). Internet History. University of Hull, UK.

comScore. (2014). comScore Releases March 2014 U.S. Search Engine Rankings. Retrieved from

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

http://www.comscore.com/Insights/Press_Releases/2014/4/comScore_Releases_March_2014_U.S._Search_Engine_Rankings

Corlosquet, S. (2014). Semsol ARC2. Retrieved from https://github.com/semsol/arc2/wiki

Davies, J. (2006). *Semantic Web technologies: trends and research in ontology-based systems*. Chichester, England ; Hoboken, NJ: John Wiley & Sons.

De Virgilio, R., Guerra, F., & Velegrakis, Y. (2012). *Semantic search over the web*. Berlin; New York: Springer. Retrieved from http://dx.doi.org/10.1007/978-3-642-25008-8

Downs, G. H., & McAllen, D. K. (2012). The Effect of Intrinsic Motivation on Success in a Technology Management Undergraduate Program. In *Proceedings of PICMET '12: Technology Management for Emerging Technologies.* Vancouver, Canada.

DuCharme, B. (2013). *Learning SPARQL*.

Elbedweihy, K., & Wrigley, S. N. (2012). Evaluating Semantic Search Systems to Identify Future Directions of Research.

Ericsson, K., & Simon, H. A. (1993). *Protocol Analysis. Verbal reports as data* ((rev. ed).). Cambridge Massachusets.: MIT Press.

Fensel,, D. (2005). *Spinning the Semantic Web*.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, *34*(10), 906–911.

Google. (2014a). Motor de búsqueda de Google. Retrieved from https://www.google.com.mx/?gws_rd=ssl

Google. (2014b). Resultados de búsqueda en Google. Retrieved from https://www.google.com.mx/?gws_rd=ssl#q=problemas+de+programacion

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Halpin, H. (2008). Foundations of a Philosophy of Collective Intelligence. University of Edinburgh.

Humfrey, N. (2014). EasyRDF. Retrieved from http://www.easyrdf.org/

Jenkins, T. (2002). On the difficulty of learning to program. Loughborough University: LTSN Centre of information and computer sciences.

Kelleher, C. P., & Pausch, R. (2005). Lowering the Barriers to Programming: a survey of programming environments and languages for novice programmers. *ACM Computing Surveys (CSUR)*, *37*(2), 83 – 137.

Kioskea.net. (2014). Web semántica: las aplicaciones actuales. Retrieved from http://es.kioskea.net/faq/7082-web-semantica-las-aplicaciones-actuales

Koedinger, K., & Aleven, V. (2004). Toward a Rapid Development Environment for Cognitive Tutors (pp. 167–179). Presented at the Engineering Advanced Web Applications: Proceedings of Workshops in Connection with the 4th International Conference on Web Engineering, Princeton: Rinton Press.

Krug, S. (2006). *Don't make me think!: a common sense approach to Web usability* (2nd ed.). Berkeley, Calif: New Riders Pub.

Lamarca, M. (2013). Hipertexto, el nuevo concepto de documento en la cultura de la imagen. Retrieved from http://www.hipertexto.info/documentos/b_datos.htm

Levene, M. (2010). *An introduction to search engines and web navigation* (2nd ed.). Hoboken, N.J: John Wiley.

Lexxebeta. (2014a). Ejemplo búsqueda en Lexxe. Retrieved from http://www.lexxe.com/cc?sstring=programming+language%3A+cycles+java&src=hp

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Lexxebeta. (2014b). Lexxebeta. Retrieved from http://www.lexxe.com/

Lexxebeta. (2014c). Llaves semánticas Lexxe. Retrieved from http://www.lexxe.com/semkeylist.html

Linked Open Vocabularies (LOV). (2014). Retrieved from lov.okfn.org/dataset/lov/index.html

LinkingOpenData. (2013). Retrieved from http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

Lourdes, V., & Carro, J. (2004). Para acceder al web profundo: conceptos y herramientas.

Ma, L., Ferguson, J., Roper, M., & Wood, M. (2007). Investigating the viability of mental models held by novice programmers. *ACM SIGCSE Bulletin*, *39*(1), 499–503.

Nadia, M., & Prem, K. (1998). Face to Virtual Face. Retrieved from http://ivizlab.sfu.ca/arya/Papers/Others/Face%20to%20Virtual%20Face.htm

Netcraft. (2014). Número de sitios web. Retrieved from http://news.netcraft.com/archives/2014/06/06/june-2014-web-server-survey.html

Networks, R., & Nova, S. (2007). The past, present and future of the Web. Retrieved from http://novaspivack.typepad.com/RadarNetworksTowardsAWebOS.jpg

Nie, J.-Y. (2010). *Cross-Language Information Retrieval*.

NlpTools. (2014). Retrieved from http://php-nlp-tools.com/

OpenLink Virtuoso. (2014). SPARQL Implementation Details. Retrieved from http://docs.openlinksw.com/virtuoso/rdfsparql.html#rdfsupportedprotocolendpoint

OWL Working Group. (2012). OWL. Retrieved from http://www.w3.org/2001/sw/wiki/OWL

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Pan, J. rey. (2004). *Description logics: reasoning support for the semantic web*. Retrieved from http://homepages.abdn.ac.uk/jeff.z.pan/pages/pub/thesis.pdf

Pandey, G. (2012). The Semantic Web: An Introduction and Issues.

Passin, T. B. (2004). *Explorer's guide to the Semantic Web*. Greenwich: Manning.

Proal, C. (2013). Almacenamiento y Recuperación de Información. Retrieved from http://www.carlosproal.com/bda/capitulo4.html

Quin, L. R. E. (2014). XML ESSENTIALS. Retrieved from http://www.w3.org/standards/xml/core

Ramalingam, V. L. (2004). Self-Efficacy and mental models in learning to program (pp. 171 – 175). Leeds, United Kingdom: ACM Press New York, NY, USA.

RDF y RDF schema. (2013, June 21). Retrieved June 21, 2013, from http://www.matem.unam.mx/~grecia/semantic_web/rdf.html

Reuters, T. (2013). CALAIS. Retrieved from http://www.opencalais.com/

Rist, R. S. (1996). System Structure and Design. In *Empirical studies of programmers: sicth workshop*. Norwood, New Jersey: Ablex Publishing Corporation.

Rist, R. S. (2004). Learning to program: schema creation, application and evaluation. In *Computer Science Education and Research* (pp. 175–197). Netherlands: Taylor & Francis Group.

Sareh, A., & Mohammad, N. (2012). EVOLUTION OF THE WORLD WIDE WEB: FROM WEB 1.0 TO WEB 4.0.

Schiefele, U. (1991). Interest, Learning and Motivation. *Educational Psychologist*, *26*(3&4), 299–323.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Sheykh Esmaili, K., & Abolhassani, H. (2006). A Categorization Scheme for Semantic Web Search Engines. Sharif University of Technology, Tehran, Iran.

Sistema Visor de Protocolos Verbales. (2014). Sistema Visor de Protocolos Verbales. Retrieved from http://dsi.ccbas.uaa.mx/Protocolos3/

SKOS Simple Knowledge Organization System. (2013). Retrieved from http://skos.um.es/

Sosa, E. (1997). Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones (Parte I). Retrieved from http://www.elprofesionaldelainformacion.com/contenidos/1997/enero/procesamiento_del_lenguaje_natural_revisin_del_estado_actual_bases_tericas_y_aplicaciones_parte_i.html

Staab, S., & Studer, R. (2009). *Handbook on Ontologies*. Springer.

Swoogle. (2014a). Swoogle. Retrieved from http://swoogle.umbc.edu/

Swoogle. (2014b). Swoogle Manual. Retrieved from http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=search_swd

Swoogle Semantic Web Search. (2007). Retrieved from http://swoogle.umbc.edu/

U.S. Department of Health & Human Services. (2014). usability.gov. Retrieved from http://www.usability.gov/

Universidad Berlin. (2013, June 21). The Linking Open Data cloud diagram. Retrieved June 21, 2013, from http://lod-cloud.net/

W3C. (2004). Resource Description Framework (RDF). Retrieved from http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/Graph-ex.gif

W3C. (2010). Use Cases and Requirements for Mapping Relational Databases to RDF.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

W3C. (2013, June 21). ConverterToRdf - W3C Wiki. Retrieved June 21, 2013, from http://www.w3.org/wiki/ConverterToRdf

W3C. (2014). SparqlEndpoints. Retrieved from http://www.w3.org/wiki/SparqlEndpoints

Wolframalpha. (2014a). Interfaz WolframAlpha. Retrieved from http://www.wolframalpha.com/

Wolframalpha. (2014b). Opción de contexto WolframAlpha. Retrieved from http://www.wolframalpha.com/input/?i=how+to+program+java+cycles

WolframAlpha. (2014). *Computational knowledge engine*. Retrieved from http://www.wolframalpha.com/

World Wide Web Consortium. (2014). Guía Breve de Web Semántica. Retrieved from http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Anexos

Anexo A. Formato de consentimiento de grabación de prueba.

Anexo B. Guion para aplicar prueba de usabilidad a prototipo.

Anexo C. Descripción de las tareas.

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Anexo A. Formato de consentimiento de grabación de prueba.

Gracias por participar en la investigación de usabilidad para un prototipo de búsquedas semánticas en internet.

Su sesión será grabada con el objetivo de obtener un indicador para medir la facilidad de uso del prototipo de búsquedas semánticas desarrollado.

Por favor lea las siguientes declaraciones y firme si está de acuerdo.

- Estoy de acuerdo de que mi sesión de pruebas de usabilidad se grabe.
- Permito que el C. Julio René López Guerrero utilice la grabación para su investigación de tesis.

Nombre: _____

Firma: _____

Fecha: _____

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Anexo B. Guion para aplicar prueba de usabilidad a prototipo.

Guion basado en (Krug, 2006).

Entrevistado: _____

Entrevistador: _____

(El navegador de internet debe estar abierto en Google o en alguna otra página)

Hola, {Entrevistado} mi nombre es {Entrevistador} y voy a acompañarlo a través de esta sesión.

Probablemente ya lo sabe, pero permítame explicar porque le hemos pedido que venga aquí hoy. Estamos probando la facilidad de uso de un buscador web en el que estamos trabajando y queremos ver qué le parece a las personas que la usan.

Quiero que quede claro que estamos probando el buscador, no a usted. No puede equivocarse aquí, de hecho, es probable que sea este el único lugar donde hoy no tendrá que preocuparse por cometer algún error.

Queremos escuchar exactamente qué es lo que piensa, por favor no se preocupe si hiere nuestros sentimientos, nosotros queremos mejorarlo, de ahí nuestra necesidad de conocer exactamente lo que piensa.

A medida que avancemos le iré pidiendo pensar en voz alta, para que me diga lo que pasa exactamente por su mente. Todo esto nos ayudará. Si tiene alguna pregunta, hágala. Puede que no tenga respuesta inmediata porque de lo que se trata es de ver cómo reacciona usted sin alguien a su lado, no obstante, tratare de responder a cualquier pregunta que tenga cuando hayamos terminado. Y si usted necesita tomar un descanso en cualquier momento, solo hágamelo saber.

Con su permiso, vamos a grabar la pantalla del ordenador y lo que usted tenga que decir. La grabación solo la utilizaremos para ver de qué forma podemos mejorar el

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

buscador, y no será vista por nadie que no esté trabajando en el proyecto, también me ayuda a mí, porque así no tomo notas.

Si nos lo permite, le voy a pedir firmar algo que simplemente da su consentimiento para grabar, pero esta grabación solo será vista por personas que trabajen en el proyecto.

(Dele el formulario de consentimiento de la grabación y un bolígrafo, mientras ellos firman, inicie la grabación)

¿Tiene alguna pregunta?

Antes de ver el sitio, me gustaría hacerle unas preguntas rápidas.

¿Cuántos años tiene?

¿Cuál es su nivel de estudios?

¿A qué se dedica?

¿Cuántas horas a la semana podría decir que usa internet incluyendo el email?

¿Qué motor de búsqueda utiliza normalmente?

¿Por qué utiliza el motor de búsqueda mencionado?

Muy bien, ya hemos terminado con las preguntas, ahora empezaremos a ver cosas.

(Haga clic en la página principal del sitio http://jrenelg.com/files/mitc/tesis/buscador/)

Este es el sitio donde se implementó el motor de búsqueda semántico, no estamos evaluando el diseño del sitio, únicamente queremos evaluar la usabilidad del buscador, le voy a pedir que de un clic sobre el link “Buscador S.” (Esperar a que de un clic en el link), este es el buscador semántico, como puede observar la interfaz es muy similar a la de los

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

motores de búsqueda convencionales ya que el objetivo de la prueba es evaluar la forma de búsqueda y la usabilidad del buscador.

Entrando en contexto, Ahora le voy a pedir que haga unas tareas específicas en base a búsquedas en el motor, voy a leer cada una de ellas y le proporcionaré una copia impresa.

Déjeme repetir que cuanto más piense en alto más nos ayudara a saber lo que realmente pasas por su cabeza.

(Repita este procedimiento para cada tarea o hasta que el tiempo termine)

Gracias, esto fue de mucha ayuda.

¿Tiene alguna pregunta para mí, ahora que hemos terminado?

(Dele las gracias por su participación, detenga la grabación y guarde el archivo)

TESIS TESIS TESIS TESIS TESIS

TESIS TESIS TESIS TESIS TESIS

Anexo C. Descripción de las tareas.

Tarea 1. Ordenar arreglo con punteros.

Tarea 2. Validar si una palabra es un palíndromo utilizando enfoque orientado a objetos.

Tarea 3. Ordenar palabras dentro de un archivo.

Tarea 4. Generar números primos usando hilos.

Tarea 5. Uso de colecciones genéricas.

TESIS TESIS TESIS TESIS TESIS