



Universidad Austral de Chile

Facultad de Ciencias de la Ingeniería
Escuela de Ingeniería Civil en Informática

Pautas Para Implementar Servicios Basados En Web Semántica En Una Empresa de Telecomunicaciones Móviles.

Proyecto para optar al título de
Ingeniero Civil en Informática

PROFESOR PATROCINANTE:

Luis Vidal V.

Ingeniero Civil Informático.

PROFESOR CO-PATROCINANTE

Claudio Cerda S.

Ingeniero Civil Informático.

PROFESOR INFORMANTE

Eliana Scheihing

TÍTULOS Y GRADOS DEL INFORMANTE

PAOLA CECILIA MIRANDA MUTIZÁBAL.

VALDIVIA – CHILE

2009

DEDICATORIA

El presente proyecto de Tesis se lo dedico con todo mi amor a mis padres Mariela y Gastón quienes, en todo momento, se esmeraron por darme la educación que logro alcanzar el día de hoy. Y a mis hermanos Javier y Carola a quienes amo mucho.

AGRADECIMIENTOS

Mis sinceros agradecimientos a ENTELPCS, en especial, a Claudio Cerda, por la oportunidad que me dio para desarrollar mi proyecto en la empresa.

Igualmente, quiero expresar mi especial gratitud a Ignacio Silva e Ivan Kruze quienes me guiaron y apoyaron incondicionalmente durante todo mi proceso de tesis. Gracias por su gran compañerismo y amistad.

ÍNDICE

I. ÍNDICE DE FIGURAS	4
II. RESUMEN.....	5
III.ABSTRACT	6
1. Introducción.....	7
1.1 Objetivo General	10
1.2 Objetivos Específicos.....	10
2. Marco Teórico.....	11
2.1 La Web de Hoy y la Web Semántica	11
2.2 De la Web Actual a la Web Semántica	17
2.3 Tecnologías que Involucra la Web Semántica. 	19
2.3.1 Unicode	20
2.3.2 URI.....	20
2.3.3 XML + NS + XML Schema	21
2.3.4 RDF + RDF Schema.....	21
2.3.5 Ontology Vocabulary	24
2.3.6 Logic.....	26
2.3.7 Proof	27
2.3.8 Trust.....	28
2.3.9 Digital Signature.....	28
2.3.10 Lenguajes de Consulta.....	29
2.3.10.1 SPARQL	29
2.3.10.2 Otros Lenguajes	31
2.4 Proyectos realizados con las tecnologías de la Web Semántica.....	33
2.4.1 Vocabularios RDF	33
2.4.1.1 FOAF.....	33
2.4.1.2 RSS	34
2.4.1.3 Dublin Core	36
2.4.2 Aplicaciones Semánticas	39
2.5 Entorno de la Telefonía.....	41
2.6 La Telefonía móvil en Chile	47
3. Propuesta del Proyecto.....	52
3.1 Propuesta de prototipo de servicio a implementar	52
3.2 Tecnologías a Utilizar	53
3.2.1 JAVA.....	53
3.2.2 JENA	54
3.2.3 Tomcat.....	54
3.2.4 RDF	55
3.2.5 SPARQL.....	58
4. Desarrollo del Buscador de Equipos Móviles.....	62
4.1 Especificación de requisitos	62
4.1.1 Requisitos Funcionales	62
4.1.2 Requisitos No Funcionales	62
4.2 Análisis de Requisitos	63
4.2.1 Casos de Uso Buscar Equipos Móviles	63
4.2.2 Diagrama de Secuencia	66

4.2.3	Diagrama de Estados	66
4.3	Diseño	67
4.3.1	Diagrama de Clases	67
4.4	Implementación.....	68
4.4.1	Implementación del Buscador	68
4.4.2	Implementación de la Interfaz Web.....	69
4.4.3	Interfaz de Usuario	69
4.4.4	Tiempos de Respuesta	73
4.4.4.1	Hardware.....	73
4.4.4.2	Software	73
4.4.4.3	Almacenamiento de la Información.....	73
4.4.4.4	Medición de los Tiempos.....	74
5.	Pautas.....	77
5.1	Respecto al Diseño.....	77
5.2	Respecto a las Herramientas	79
6.	Conclusiones.....	85
7.	Referencias	88

I. ÍNDICE DE FIGURAS

Figura 1: Resultados obtenidos con un buscador normal	12
Figura 2: Resultados obtenidos con un buscador semántico	13
Figura 3: La Web actual está orientada al consumo humano	14
Figura 4: La Web actual v/s la Web semántica.	15
Figura 5: Arquitectura de la Web Semántica	19
Figura 6: RDF y RDFS	24
Figura 7: Ejemplo FOAF	34
Figura 8: Ejemplo RSS.....	36
Figura 9: Ejemplo Dublin Core.....	38
Figura 10: Mercado mundial de los servicios de telecomunicaciones por segmento.....	42
Figura 11: Número de clientes móviles y penetración cada 100 habitantes	47
Figura 12: Evolución del número de clientes por empresa	51
Figura 13: Partes de una declaración RDF.....	56
Figura 14: Diagrama de nodos y arcos	57
Figura 15: Fragmento RDF	58
Figura 16: Sintaxis del operador SELECT.....	59
Figura 17: Sintaxis del operador CONSTRUCT.....	59
Figura 18: RDF generado por consulta CONSTRUCT.....	60
Figura 19: Sintaxis del operador ASK	60
Figura 20: Sintaxis del operador DESCRIBE	61
Figura 21: Requisitos Funcionales	62
Figura 22: Requisitos No Funcionales	63
Figura 23: Caso de Uso Buscar equipos móviles	63
Figura 24: Curso normal de eventos de Caso de Uso Buscar Equipos Móviles.....	65
Figura 25: Curso alternativo de eventos de Caso de Uso Buscar Equipos Móviles.	65
Figura 26: Diagrama de secuencia	66
Figura 27: Diagrama de estados	67
Figura 28: Diagrama de clases	67
Figura 29: Interfaz inicial del buscador.....	70
Figura 30: Ejemplo de búsqueda.....	70
Figura 31: Interfaz de usuario que muestra resultados de la búsqueda	71
Figura 32: Descripción del equipo móvil escogido por el usuario.	72
Figura 33: Tabla de Características del hardware utilizado	73
Figura 34: Tabla de características de software utilizado.....	73
Figura 35: Tabla para prueba de consulta.....	74
Figura 36: Gráfico Tiempo de respuesta v/s Número de equipos.	75
Figura 37: Ilustración del ejemplo de consulta de eventos de calendario	82
Figura 38: Tabla de Tecnologías/Herramientas según su función.....	84

II. RESUMEN

El siguiente proyecto de tesis propone pautas para implementar un servicio basado en las tecnologías de la Web Semántica en una empresa de telecomunicaciones móviles. Es decir, pautas que le permitan a la empresa conocer cuál es el estado actual de las tecnologías que propone esta nueva idea de Web, cuáles son estas tecnologías y cuándo es conveniente usar una u otra. Además, de dar a conocer las herramientas más utilizadas actualmente.

Para lo anterior, fue necesario el estudio del marco teórico que involucra la Web Semántica, el estudio en relación al entorno de la telefonía móvil a nivel mundial y nacional, como también los objetivos que los operadores móviles hoy persiguen.

Posteriormente, en base a los estudios hechos, se propone la implementación de un prototipo de servicio basado en las tecnologías de la Web Semántica. Este servicio es un buscador Web de equipos móviles, que permite a los usuarios ingresar en lenguaje natural lo que desea buscar. Para esto, se realizó el análisis, diseño e implementación de la aplicación propuesta.

Finalmente, el proyecto termina con la definición de pautas que permitan exponer las principales propuestas en relación al diseño y uso de herramientas para implementar un servicio basado en las tecnologías de la Web Semántica dentro de la empresa.

III.ABSTRACT

The following thesis proposes guidelines to set up a service based on Semantic Web technologies in mobile telecommunication companies. That is to say, the guidelines which permit to the company know the current state of the technologies that put forward this new idea about the Web; which are these technologies and when is proper to use each one. Besides to publish the most used tools nowadays.

To make the previous, it was necessary study the theoretical framework involving in the Semantic Web, the research in relation to the mobile phone environment worldwide; as well as the objectives pursued these days by mobile network operators.

Subsequently, as a result of the researches carried out, is proposed the implementation of a service prototype based on the Web Semantic technologies. This service is a search engine Web of mobiles phones, permitting users enter in natural language that they want looking for. For that, it was made an analysis, design and implementation of the suggested application.

Eventually, the project finishes defining guidelines that permit to set out the mean proposals in relation with the design and the use of tools to install a service based on the Web Semantic technologies in a company.

1. Introducción

A lo largo de los años, la tecnología ha ido evolucionando constantemente, lo que a su vez ha tenido un impacto en la sociedad. El desarrollo de nuevas tecnologías, se ha traducido en transformaciones de las conductas sociales que han influido de forma positiva en las industrias vinculadas a la tecnología de Internet y la telefonía móvil, las cuales serán abordadas durante el desarrollo del presente trabajo con un enfoque que fusiona diferentes aspectos de ellas.

La World Wide Web¹ (WWW), uno de los servicios que más éxito ha tenido en Internet, ha llegado a ser más global, más rápida y más fácil de usar, que hace aproximadamente 28 años atrás. Esto, gracias a los rápidos avances tecnológicos de finales del siglo pasado. Hoy, a través, de distintos trabajos de investigación y desarrollo ya se están dando los primeros pasos tendientes a transformar esta Web en su nueva generación: *“La Web Semántica”*.

La semántica, es la ciencia que estudia el significado, sentido o interpretación de las expresiones del lenguaje del hablante. Esta ciencia ha sido fundamental para que las personas se comuniquen y se entiendan, pues para que esto ocurra es necesario partir sobre un acuerdo semántico, es decir, un acuerdo sobre el significado de las palabras que se emplean. Esto es lo que ahora se busca a nivel informático, específicamente a nivel Web, que la información presente en la WWW sea comprensible no sólo por los humanos sino también por las propias máquinas. Lo anterior, se pretende lograr catalogando la información de los recursos Web, páginas HTML, videos, archivos de sonido, etc., mediante lo que se denomina ontologías², es decir, mediante el significado de las palabras y no mediante palabras claves. Tim Berners-Lee, creador de la WWW y precursor de la Web Semántica, afirma que en un futuro no lejano, se producirán desarrollos importantes que introducirán prestaciones nuevas, al lograr que las máquinas multipliquen su capacidad de procesar y “comprender” los datos dispersos en la Web y que hoy solamente se exhiben en la pantalla.

¹ Medio, a través de cual, los usuarios pueden tener acceso a diversos servicios (correo electrónico, buscadores, audio, video, etc) por medio de un programa instalado en el computador y conocido como navegador.

² Se refiere a la formulación de un exhaustivo y riguroso modelo conceptual dentro de uno o varios dominios dados; con la finalidad de facilitar la comunicación y la compartición de la información entre diferentes sistemas y entidades.

Hasta ahora, la Web se ha desarrollado con suma rapidez, pero concebida mucho más como medio de proporcionar documentos a los humanos, que para la manipulación de datos e informaciones procesables de forma automática. La Web Semántica aspira a cubrir esa deficiencia, es decir, pretende mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos y así reducir la necesaria mediación de operadores humanos y, en muchas ocasiones, la interpretación de la información por parte del usuario final.

La telefonía móvil, por su parte, está marcada por un crecimiento explosivo a nivel mundial como nacional, lo que constituye una señal positiva del desarrollo que ha alcanzado el sector de las telecomunicaciones. Hoy este sector, está volcado, entonces, a la innovación y el lanzamiento de nuevos productos. El objetivo es diferenciarse de la competencia y ofrecer a los usuarios un cada vez más amplio abanico de servicios que satisfaga, con una experiencia de usuario positiva, sus crecientes necesidades de comunicación. Lo anterior, con el fin de lograr fidelizar a los clientes, pues en este mercado, incluyendo el chileno, donde hoy todos tienen un celular o incluso más de uno, el principal objetivo de estas empresas se centra en cautivarlos. Para ello, se aplican distintas estrategias, las cuales incluyen destinar recursos para mejorar los servicios tecnológicos propios de la telefonía móvil como también la plataforma de negocio de estas empresas, con el propósito de entregar productos de calidad con un enfoque de atención al cliente cada vez más personalizado, potenciando el valor agregado del producto principal.

Dado estos dos escenarios, se puede visualizar que la incorporación de la Web Semántica en el negocio de la telefonía móvil, puede ayudar a lograr los fines de esta última. La propuesta de la Web Semántica permitiría a los operadores móviles diferenciarse de su competencia, ofreciendo un mejor servicio y fidelizando a sus clientes.

El presente trabajo de tesis, se desarrolla en torno a lo mencionado anteriormente y tiene como principal objetivo, generar pautas para implementar aplicaciones Web Semánticas en un entorno móvil. Para lograr esto, se estudiaron los distintos estándares que conforman las tecnologías de la Web Semántica y sus distintas aplicaciones, además de estudiar el crecimiento de la telefonía móvil en los últimos años y los objetivos que esta persigue. Luego de esto, se definió qué aplicación Web Semántica aportaría mayor

valor al ser implementada en un entorno móvil, ya sea una aplicación móvil propiamente tal o una que potencie la plataforma de negocio de estas empresas. Se realizó el análisis, diseño e implementación del prototipo de servicio escogido, finalizando con la descripción de las pautas requeridas por una empresa TELCO³ para la implementación de aplicaciones Web Semántica.

³ Empresa de Telecomunicaciones.

1.1 Objetivo General

Analizar, diseñar e implementar un prototipo empleando Web Semántica en un entorno móvil (empresa de telecomunicaciones móviles), con el objeto de entregar pautas que permitan desarrollos futuros para la industria.

1.2 Objetivos Específicos.

1. Describir el conjunto de estándares y recomendaciones más importantes que conforman el marco de la tecnología que permite implementar Web Semánticas.
2. Analizar y Diseñar prototipo basado en Web Semántica para un entorno móvil.
3. Implementar un prototipo basado en Web semántica en un entorno móvil.
4. Formular pautas de implementación de sistemas que empleen Web semántica en un entorno móvil.

2. Marco Teórico

2.1 La Web de Hoy y la Web Semántica

La información actualmente presente en la Web se encuentra estructurada mediante lenguajes de etiquetado que únicamente describen la forma en que dicha información debe ser presentada al usuario por el navegador (colores, interacción, entre otros), pero no expresan nada sobre su significado, es decir, nada sobre su semántica.

El último año ha sido el año de la Web social, tal y como atestigua el triunfo de sitios de comunicaciones sociales del tipo de *LinkedIn*⁴, *MySpace*⁵, *YouTube*⁶ o *Facebook*⁷, entre otros y donde la tecnología social como los “*blogs*”⁸ o los “*wikis*”⁹ se ha extendido de manera muy rápida. Junto con esto, la idea de una nueva generación de la Web, va tomando cada vez más fuerza. La Web Semántica, se define como una Web que extiende a la actual y no como una Web independiente, 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. La forma en que se procesará la información no sólo será en términos de entrada y salida de parámetros sino en términos de su Semántica. La Web Semántica como infraestructura basada en metadatos aporta un camino para “razonar” en la Web, extendiendo así sus capacidades.

Supongamos, por ejemplo, que la Web tiene la capacidad de construir una base de conocimiento sobre las preferencias de los usuarios y que, a través de una combinación entre su capacidad de conocimiento y la información disponible en

⁴ Sitio de red profesional orientada a los negocios.

⁵ Sitio web de interacción social formado por perfiles de usuarios, redes de amigos, grupos, blogs, fotos, videos, música, además de una red interna de mensajería que permite comunicarse a los usuarios con otros.

⁶ Sitio web que permite a los usuarios compartir videos digitales a través de Internet.

⁷ Sitio web de redes sociales que actualmente tiene más de 150 millones de usuarios registrados en todo el mundo. Formado por perfiles de usuario, redes de amigos, grupos y páginas para reunir personas con intereses comunes, espacios en cada perfil de usuario para escribir / recibir mensajes o subir animaciones flash, espacio para subir fotos, etc.

⁸ Sitio web periódicamente actualizado que recopila cronológicamente textos o artículos de uno o varios autores, apareciendo primero el más reciente, donde el autor conserva siempre la libertad de dejar publicado lo que crea pertinente.

⁹ Sitio web cuyas páginas pueden ser editadas por múltiples voluntarios a través del navegador web.

Internet, es capaz de atender de forma exacta las demandas de información por parte de los usuarios en relación, por ejemplo, a reserva de hoteles, vuelos, médicos, libros, etc. Si esto ocurriese así en la vida real, el usuario, en su intento, por ejemplo, por encontrar todos los vuelos a Valdivia para mañana por la mañana, obtendría unos resultados exactos sobre su búsqueda. Sin embargo la realidad es otra. La figura 1 muestra los resultados inexactos que se obtendrían con el uso de cualquier buscador actual, el cual ofrecería información variada sobre Valdivia, pero que no tiene nada que ver con lo que realmente el usuario buscaba. El paso siguiente por parte del usuario es realizar una búsqueda manual entre esas opciones que aparecen, con la consiguiente dificultad y pérdida de tiempo [W3C08].

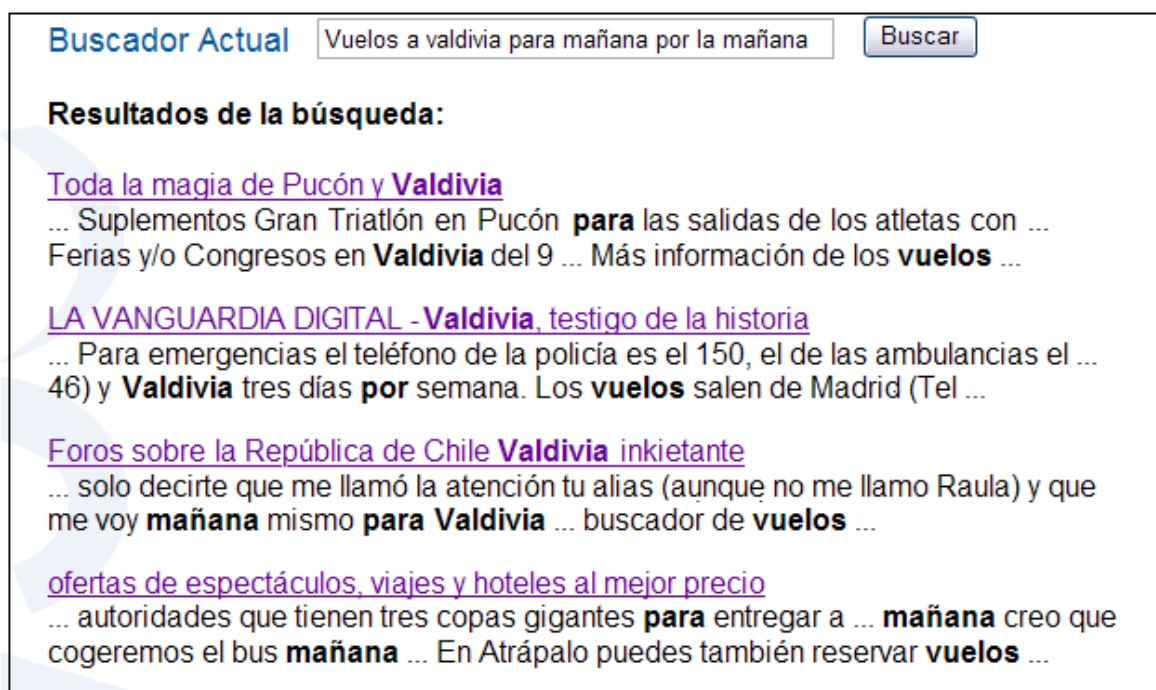


Figura 1: Resultados obtenidos con un buscador normal

Con la incorporación de semántica a la Web los resultados de la búsqueda serían exactos. La figura 2 muestra los resultados obtenidos a través de un buscador semántico. Estos resultados ofrecen al usuario la información exacta que estaba buscando. La ubicación geográfica desde la que el usuario envía su pregunta es detectada de forma automática sin necesidad de especificar el punto de partida, elementos de la oración como "mañana" adquirirían significado, convirtiéndose en un día concreto calculado en función de un "hoy". Algo semejante ocurriría con el segundo "mañana", que sería interpretado como un momento determinado del día. Todo ello a través de una Web en la que los datos pasan a ser información llena de significado. El resultado final sería la

obtención de forma rápida y sencilla de todos los vuelos a Valdivia para mañana por la mañana [W3C08].

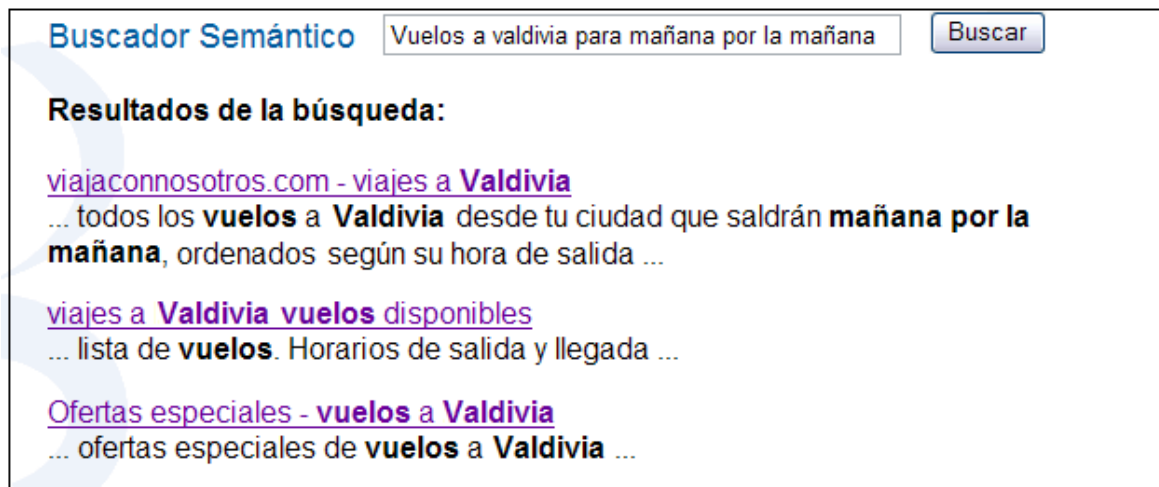


Figura 2: Resultados obtenidos con un buscador semántico

Hoy casi todo está representado de una u otra forma en la Web, y con la ayuda de un buen buscador, podemos encontrar información sobre casi cualquier cosa que necesitemos. La Web está cerca de convertirse en una enciclopedia universal del conocimiento humano. Por otra parte, la Web nos permite realizar diferentes actividades de nuestra vida diaria con una comodidad, economía y eficiencia sin precedentes: sin moverse de la casa se puede comprar todo tipo de productos y servicios, gestionar una cuenta bancaria, buscar un restaurante, consultar la cartelera, leer la prensa, localizar a una persona, matricularse en la universidad, o trabajar desde el domicilio [Cas].

No obstante, en este panorama tan favorable hay espacio para mejoras. Por ejemplo, el enorme tamaño que ha alcanzado la Web, a la vez que es una de las claves de su éxito, hace que algunas tareas (por ejemplo, encontrar la planificación óptima con transporte, alojamiento, etc., entre todas las posibles para un viaje bajo ciertas condiciones), requieran un tiempo excesivo para una persona o resulten sencillamente inabarcables. Desarrollar programas que realicen estas tareas en nuestro lugar es enormemente complicado, ya que es muy difícil reproducir, y más costoso aún mantener, en una máquina la capacidad de una persona para comprender los contenidos de la Web tal y como están codificados actualmente [Cas].

La Web de hoy no expresa significado. Los contenidos y servicios en la Web se presentan en formatos (por ejemplo, HTML¹⁰) e interfaces (por ejemplo, formularios) comprensibles por personas, pero no por máquinas. La figura 3 ejemplifica esta situación con una versión simplificada de una página de información meteorológica. Mientras que la presentación de los datos en el navegador (izquierda) es inmediatamente comprendida por una persona, es muy difícil para el computador entender cuál es la temperatura, el estado del cielo, y demás semántica del documento, al estar entremezclada con las etiquetas de formato (derecha).

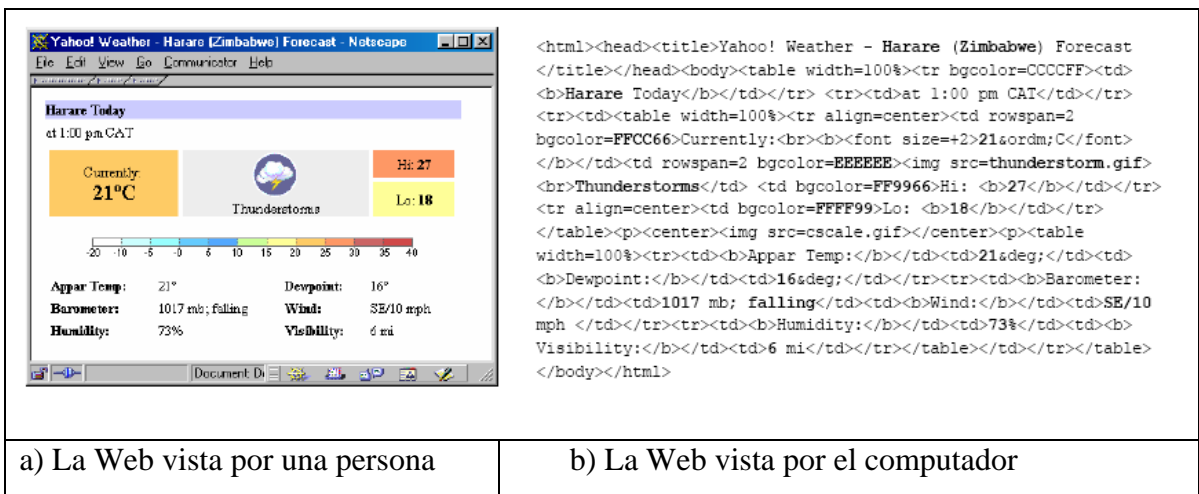


Figura 3: La Web actual está orientada al consumo humano.¹¹

En estas condiciones es poco viable automatizar tareas mediante softwares en sustitución del humano. Un programa puede llevar al usuario hasta lugares en la Web, generar, transportar y ofrecer la información a las personas, pero la máquina sencillamente no sabe lo que esta información significa, y por tanto su capacidad de actuación autónoma es muy limitada. Esta misma limitación expresiva hace que la noción de semántica que manejan los buscadores Web se limite a palabras clave con pesos, pero planas e inconexas, lo que no permite reconocer ni solicitar significados más elaborados [Cas].

La Web semántica, propone superar las limitaciones de la Web actual mediante la introducción de descripciones explícitas del significado, la estructura interna y la estructura global de los contenidos, y servicios disponibles en la WWW. Frente a la

¹⁰ HyperText Markup Language (Lenguajes de Marcas de Hipertexto), es el lenguaje de marcado predominante para la construcción de páginas Web.

¹¹ [Cas]

semántica implícita, el crecimiento caótico de recursos, y la ausencia de una organización clara de la Web actual, la Web semántica aboga por clasificar, dotar de estructura y anotar los recursos con semántica explícita procesable por máquinas. La figura 4 ilustra esta propuesta. Actualmente la Web se asemeja a un grafo formado por nodos del mismo tipo, y arcos (links) igualmente indiferenciados. Por ejemplo, no se hace distinción entre la página personal de un pintor y el portal de una escuela de pintura, como tampoco se distinguen explícitamente los enlaces a las pinturas que hizo el pintor de los enlaces a las escuelas a las que pertenece. Por el contrario en la Web semántica cada nodo (recurso) tiene un tipo (pintor, escuela, pintura, período, lugar), y los arcos representan relaciones explícitamente diferenciadas (pintor-escuela, pintor-pintura, escuela-lugar, etc). [Cas].

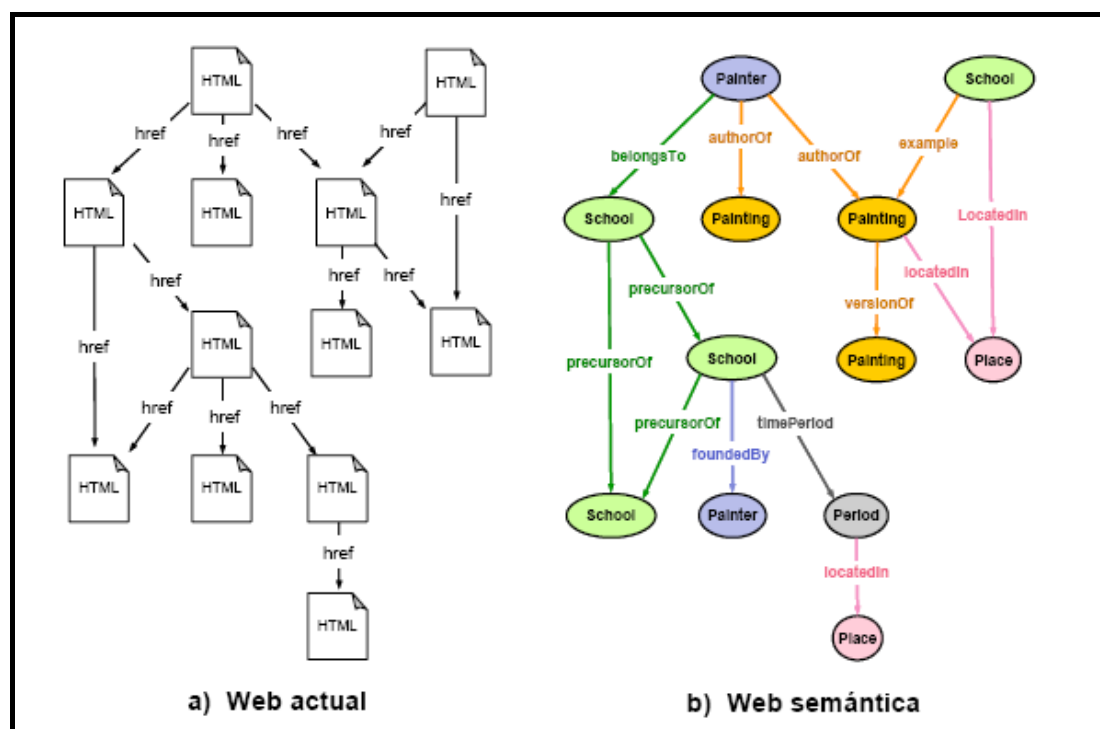


Figura 4: La Web actual v/s la Web semántica.¹²

La Web semántica mantiene los principios que han hecho un éxito la Web actual, como son los principios de descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución, o la apertura al crecimiento y uso no previstos de antemano. En este contexto un problema clave es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la Web: usuarios, desarrolladores y programas de muy diverso perfil. La Web semántica rescata la noción

¹² [Cas]

de ontología del campo de la Inteligencia Artificial como vehículo para cumplir este objetivo [Cas].

Se puede decir, por último, que la Web no solamente proporciona acceso a contenidos, sino que también ofrece interacción y servicios (comprar un libro, reservar un pasaje en un vuelo, hacer una transferencia bancaria, simular una hipoteca). Los servicios Web semánticos son una línea importante de la Web semántica, que propone describir no sólo la información sino definir ontologías de funcionalidad y procedimientos para describir servicios Web: sus entradas y salidas, las condiciones necesarias para que se puedan ejecutar, los efectos que producen o los pasos a seguir cuando se trata de un servicio compuesto. Estas descripciones, que pueden ser procesadas por máquinas, permitirán automatizar el descubrimiento, la composición y la ejecución de los servicios, así como la comunicación entre unos y otros [Can06].

La velocidad a la que esta nueva tecnología avanza es constante, pero aún hay algunos problemas importantes que entorpecen su adopción generalizada. Casi todos, relacionados con la reingeniería necesaria para conseguir que un sitio Web actual pueda integrarse en el mundo de la Web Semántica. Sin embargo, los primeros pasos para alcanzarla ya son completamente factibles y a juzgar por el desarrollo impresionante de Internet en los últimos 15 años es imposible predecir qué forma tomará la Web de aquí a otros 15 años más.

2.2 De la Web Actual a la Web Semántica

La primera pregunta que se debe responder es cómo encaja la Web semántica con la actual, es decir, cómo accederá el usuario a la Web semántica, y sobre todo, cómo hacer la transición de la Web actual a la Web semántica.

Para que la Web semántica pueda desarrollarse es importante que guarde, al menos al principio, una compatibilidad con la tecnología actual. Es deseable, por ejemplo, mantener el lenguaje HTML, u otros lenguajes compatibles con los navegadores actuales, como vehículo de comunicación con el usuario. La asociación entre las instancias de la Web semántica y el código HTML se puede establecer mediante distintos mecanismos [Can06].

Una de las formas posibles consiste en conservar los documentos actuales, y crear las instancias asociadas anotando su correspondencia con los documentos. Este enfoque se conoce como Botton-Up, es decir, un enfoque de abajo-arriba. Esta posibilidad es la más viable cuando se parte de un gran volumen de material antiguo, pero significa trasladar a los usuarios la tediosa tarea de introducir información adicional como por ejemplo, anotaciones RDF, además de problemas añadidos como el mantenimiento futuro cuando cambia la información. Otra forma consiste en generar dinámicamente páginas Web a partir de las ontologías y sus instancias. Este enfoque, se conoce como Top-Down y consiste en dejar la información tal y como está, y aplicar un conocimiento específico vertical para entregar resultados al usuario mediante una aplicación. En este caso no se necesita reanotaciones en RDF y OWL, sino que el reconocimiento es algorítmico. Esta última opción puede resultar factible cuando los documentos antiguos ya se estaban generando automáticamente a partir, por ejemplo, de una base de datos. No obstante, lo ambicioso del objetivo hace que las dos aproximaciones distintas: Botton-Up y Top-Down sean posibles y puedan convivir. De hecho los motores de búsqueda que utilicen el enfoque Top-Down, también podrían beneficiarse de las etiquetas RDF, aunque sean contrarios a la filosofía de etiquetas.

La transición de la Web actual a la Web semántica puede implicar un coste altísimo si se tiene en cuenta el volumen de contenidos que ya forman parte de la Web. Crear y poblar las ontologías supone un esfuerzo extra que puede resultar tedioso cuando se agregan nuevos contenidos, pero directamente prohibitivo en lo que respecta a

integrar los miles de *gigabytes*¹³ de contenidos antiguos. Las estrategias más viables combinan una pequeña parte de trabajo manual con la automatización del resto del proceso. Las técnicas para la automatización incluyen, entre otras, el mapeo de la estructura de las bases de datos a ontologías, el aprovechamiento, previa conversión, de los metadatos y estándares de clasificación presentes en la Web (y fuera de ella), y la extracción automática de los metadatos a partir de texto y recursos multimedia [Can06] .

Otra dificultad importante a la hora de establecer la Web semántica en la práctica, es la de consensuar ontologías en una comunidad, por poco amplia que sea. Convergencia a una representación común es una labor más compleja de lo que puede parecer, ya que cada parte del sistema conlleva, típicamente, ciertas peculiaridades necesarias, y un punto de vista propio que a menudo necesita incidir en la propia ontología. La representación del mundo no es neutra respecto al uso que se le va a dar: tanto un dietista como un biólogo tienen el conocimiento suficiente sobre las plantas, pero la representación que pueden hacer de esa materia es muy distinta, y probablemente no será adecuado imponer la misma representación para ambas perspectivas. Las vías para salvar esta dificultad consisten en compartir ontologías para las áreas comunes en que puede tener lugar una interacción o intercambio de información entre las partes, y establecer formas de compatibilidad con las ontologías locales, mediante extensión y especialización de las ontologías genéricas, o por mapeo y exportación entre ontologías [Can06].

¹³ Unidad de medida informática cuyo símbolo es GB. 1 GB equivale a 1024 MB (Megabyte)

2.3 Tecnologías que Involucra la Web Semántica.

La idea de la Web Semántica es establecer un marco a nivel mundial, capaz de abarcar tareas como la búsqueda teniendo en cuenta la semántica y aportar una tecnología de integración. Así, Berners-Lee, a través, del consorcio W3C¹⁴, pretende establecer una serie de estándares lo suficientemente consensuados como para que sean globalmente utilizados. Para ello, y partiendo de su dilatada experiencia Berners-Lee plantea que la arquitectura de la Web semántica se podría representar de la siguiente forma:

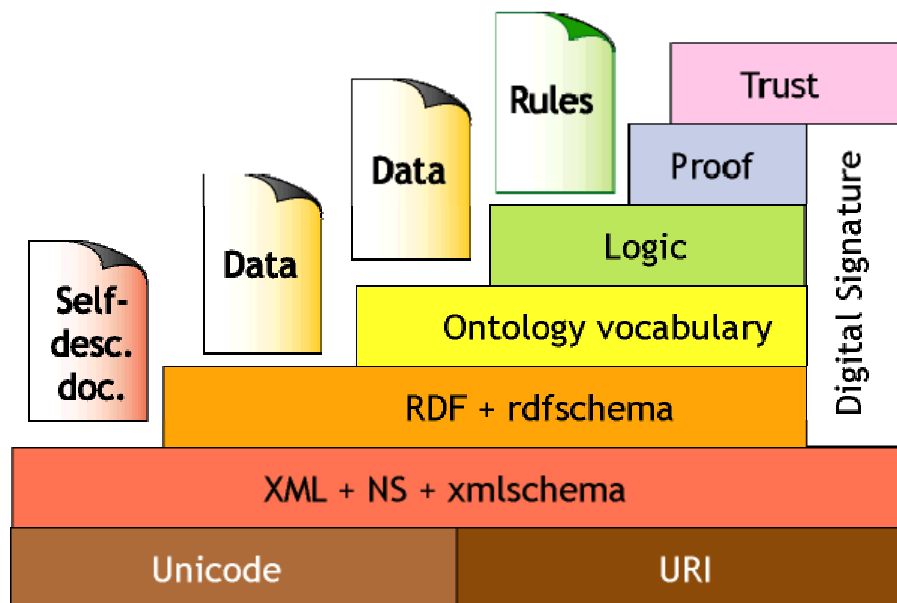


Figura 5: Arquitectura de la Web Semántica.¹⁵

Un agente inteligente **entiende** (lo que se le pide), **comprende** (el contenido de los sitios), **valida** (si lo encontrado corresponde a los pedido) y **deduce** (nueva información sobre la ya obtenida). Pero para entender una consulta, comprender sitios Web, validar información y deducir nueva información se precisa estandarizar. Y lo que se debe estandarizar es lo siguiente [Lam08]:

- El alfabeto
- Las referencias

¹⁴ Consorcio internacional donde las organizaciones miembro, personal a tiempo completo y el público en general, trabajan conjuntamente para desarrollar estándares Web.

¹⁵ [Ber00]

- El lenguaje
- El formato
- Las anotaciones sobre significado
- Los conceptos generales
- Las reglas y sistemas de deducción.

Estos son, precisamente, las funciones que aportan las distintas capas de la Web Semántica, las cuales se explicarán a continuación:

2.3.1 Unicode

El alfabeto. Se trata de una codificación del texto que permite utilizar los símbolos de diferentes idiomas sin que aparezcan caracteres extraños. De esta forma, se puede expresar información en la Web Semántica en cualquier idioma.

2.3.2 URI

Las referencias. URI es el acrónimo de “*Uniform Resource Identifier*¹⁶”, identificador único que permite la localización de un recurso (físico o abstracto) que puede ser accedido vía Internet. Se entiende por recurso: un documento, una imagen, un programa, un servicio, una dirección de correo electrónico, etc. Un URI no tiene que ser un link, ni tiene por qué tener contenido, sólo se utiliza para que el nombre sea único. Algunas URI pueden ser URL (*Uniform Resource Locator*¹⁷), que es el modo que tiene la Web de identificar cualquier tipo de archivo o recurso en cualquier parte del mundo, o URN (*Uniform Resource Name*¹⁸). Este último, es un sistema más amplio y preciso que URL pero, pese a que se espera que en el futuro sea el principal, aún no se usa realmente en la red.

¹⁶ Identificador Uniforme de Recursos

¹⁷ Localizador Uniforme de Recursos

¹⁸ Nombre Uniforme de Recursos, Descripción del espacio de nombre

2.3.3 XML + NS + XML Schema

Se trata de la capa más técnica de la Web Semántica. En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan entenderse entre ellos [Lam08].

XML¹⁹ representa una primera aproximación a la Web semántica, y aunque no está expresamente pensado para definir ontologías, es el estándar más extendido hoy día en las aplicaciones de esta línea pre-web semántica [Cas].

En su base, XML proporciona un conjunto de reglas para la creación de vocabularios que doten de estructura a los datos y documentos de la Web. XML da reglas claras para la sintaxis, ofreciendo un formato común para el intercambio de documentos [W3C04].

NS o *namespace*, sirve para cualificar nombres de elementos y atributos usados en documentos XML, asociándolos con espacios de nombres identificados por referencias URI [Lam08].

XML Schema (XMLS), por su parte, ofrece una plantilla para elaborar documentos estándar, es decir, define qué elementos puede contener un documento XML, cómo están organizados, y qué atributos y de qué tipo pueden tener sus elementos, además de otras posibilidades en el tratamiento de los documentos [Lam08b]. Por lo tanto, con XMLS se pueden acordar de antemano las estructuras que se van a utilizar, así como manejar tipos de datos primitivos y derivados [Cas]. Sirve como un método de composición de vocabularios Web. De esta forma, aunque se utilicen diferentes fuentes, se crean documentos uniformes en un formato común y no propietario [Lam08].

2.3.4 RDF + RDF Schema

Basada y apoyada en la capa anterior, esta capa define el lenguaje universal con el cual podemos expresar diferentes ideas en la Web Semántica [Lam08].

¹⁹ Extensible Markup Language, lenguaje de marcas extensible.

RDF²⁰, es un lenguaje para representar información acerca de recursos de la World Wide Web. Está especialmente destinado a la representación de metadata sobre recursos Web. Se basa en la idea de identificar las cosas mediante identificadores Web (o URI), y describir los recursos en términos de propiedades y valores de estas propiedades [W3C04b]. Es hoy el estándar más popular y extendido en la comunidad de la Web Semántica.

El elemento de construcción básica de este lenguaje, mediante el cual se definen sentencias, es en el formato de una 3-upla o triple, que consiste en dos nodos (Sujeto y Objeto, donde Sujeto es el recurso al que nos referimos y Objeto es un recurso o un literal que podría considerarse el valor de lo que se define) unidos por un arco (Predicado, que es el recurso que indica qué es lo que estamos definiendo). Por lo tanto, el modelo RDF o *Resource Description Framework* es un modelo común (*Framework*) que permite hacer afirmaciones sobre los recursos (*Description*) y que hace posible que estos recursos puedan ser nombrados por URIs (*Resource*) [Lam08].

El uso de identificadores, URIs, en RDF permite que una persona o grupo de personas convengan en otorgarle a un identificador un significado concreto con objeto de convertirlo en sujeto, objeto o predicado de cualquier sentencia RDF, articulando de este modo formas de comunicación en torno a él.

Se han definido diferentes formas sintácticas para la formulación escrita de RDF, entre ellas N3, N-Triples, RDF/XML, pero quizás la más extendida es la basada en XML. Es por ello que RDF se presenta a menudo como una extensión de XML [Cas]. Lo anterior, no significa que estas dos tecnologías son iguales, no hay mejor o peor, sólo son diferentes. Las especificaciones RDF proporcionan, una infraestructura potente para el intercambio de conocimiento en la Web. Es otras palabras, XML es a la sintaxis, lo que RDF es a la semántica [W3C04].

Por su parte, RDF Schema (RDFS) es una extensión semántica de RDF que proporciona primitivas adicionales. Las primitivas de modelado de datos proporcionadas por RDF pueden ser básicas (identificación de objeto, relaciones con otros recursos, mezclar definiciones y contenedores), por lo que RDFS tiene como objetivo enriquecer el modelo básico, proporcionando un vocabulario para RDF, que se asume tiene una

²⁰ Resource Description Framework, Infraestructura para la descripción de recursos.

cierta semántica. Permite a los diseñadores especificar una jerarquía explícita de clases/subclases de recursos y propiedades que describen estas clases, junto con las restricciones sobre las combinaciones permitidas de clases, propiedades y valores [Día03].

Se comentaran en detalle algunas características de RDFS:

- **Clases:** rdfs: Resource, rdf: Property y rdfs: Class.

Cualquier cosa descrita por una sentencia RDF se considera una instancia de la clase rdfs:Resource. La clase rdf:Property es la clase de todas las propiedades utilizadas en la caracterización de las instancias de rdf:Resource. Finalmente rdf:Class se usa para definir conceptos. Cada concepto debe ser una instancia de rdf:Class [Día03].

- **Propiedades:** rdf: type, rdfs:subClassOf, y rdfs: subPropertyOf .

La relación rdf:type modela interrelaciones del tipo **instancia-de** entre recursos y clases. Un recurso puede ser una instancia de más de una clase. Rdfs: subClassOf modela la jerarquía de clases, donde una clase puede ser subclase de otras subclases. Si una propiedad P2 es una subpropiedad (rdfs:subPropertyOf) de otra propiedad P1, y si un recurso R tiene una propiedad P2 con valor V, esto implica que el recurso R también tiene la propiedad P1 con valor V [26] [Día03].

- **Restricciones y Clases:** rdfs:ConstraintResource, rdfs: ConstraintProperty, rdfs:range y rdfs:domain.

Rdfs:ConstraintResource define la clase de todas las restricciones. Rdfs: ConstraintProperty es un subconjunto de Rdfs:ConstraintResource. Tiene dos instancias: rdfs:range y rdfs:domain que se usan para restringir el rango y el dominio de las propiedades. No se permite expresar más de una restricción de rango sobre una propiedad. En dominios sí se permite y se interpreta como una unión de dominios [Día03].

Por ejemplo, una sentencia RDF podría expresar el hecho de que el autor (predicado) del cuadro “*Starry Night*” (sujeto) fue el pintor *Vincent van Gogh* (objeto), como se ilustra en la figura 6. Encadenando estas tripletas se construyen grafos o redes semánticas para la Web.

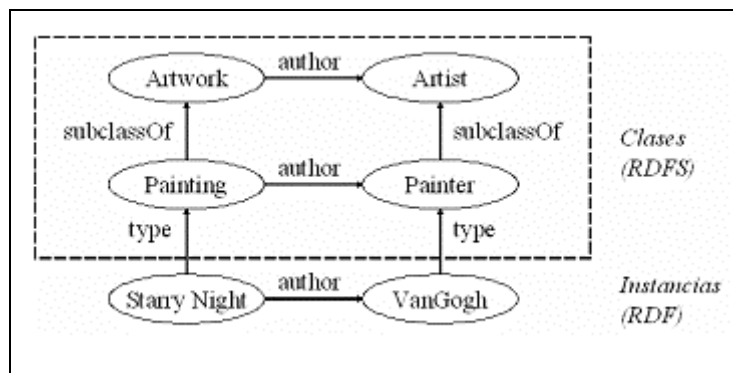


Figura 6: RDF y RDFS. ²¹

Con RDF Schema (RDFS) se pueden definir jerarquías de clases de recursos, especificando las propiedades y relaciones que se admiten entre ellas (ver figura 6). En RDF las clases, relaciones, y las propias sentencias son también recursos, y por lo tanto se pueden examinar y recorrer como parte del grafo, o incluso acertar sentencias sobre ellas [Cas].

Esta capa no sólo ofrece descripción de los datos, sino cierta información semántica. Tanto esta capa, como la anterior corresponden a las anotaciones de la información (metadatos) [Lam08].

2.3.5 Ontology Vocabulary

Lo siguiente que se necesita es un modelo por el que desarrollar vocabularios para un asunto (o dominio) específico. Ese es el papel de una Ontología [W3C04]. Heredada de la Inteligencia Artificial, una ontología es **“una especificación explícita formal de una conceptualización común”**, es decir, proporciona un modelo explícito obtenido por consenso descrito en un lenguaje que contiene a los conceptos, propiedades y relaciones más relevantes de un dominio [Gru93]

Una ontología, por lo tanto, ofrece un criterio para catalogar y clasificar la información [Lam08]. Las ontologías son utilizadas por personas, bases de datos y aplicaciones que necesitan compartir información específica sobre un determinado asunto o dominio, como la medicina, fabricación de herramientas, inmobiliarias, reparación de automóviles, etc. Las ontologías incluyen definiciones utilizables por

²¹ [Cas]

máquinas, sobre conceptos básicos del dominio y de las relaciones existentes entre los mismos. Codifican conocimiento en un dominio y también conocimiento que se expande a través de varios dominios. De este modo, hacen que este conocimiento sea reutilizable.

OWL²², el lenguaje de ontologías Web, proporciona un estándar para la definición de ontologías estructuradas, basadas en la Web, que ofrece una integración e interoperabilidad de datos más rica entre comunidades descriptivas. Los lenguajes anteriores de ontología, KIF²³, DAML²⁴, OIL²⁵, DAML+OIL²⁶, se utilizaron para desarrollar herramientas y ontologías para comunidades de usuarios específicas (particularmente en las ciencias y en aplicaciones de comercio electrónico de compañías específicas), pero no fueron definidos para ser compatibles con la arquitectura de la World Wide Web en general, y de la Web Semántica en particular.

OWL utiliza URIs para fijar nombres y la infraestructura para descripciones en la Web proporcionada por RDF para agregar las siguientes capacidades a las ontologías:

- Habilidad de ser distribuida por muchos sistemas.
- Escalabilidad a las necesidades de la Web.
- Compatibilidad con estándares Web para la accesibilidad y la internacionalización.
- Apertura y extensibilidad.

OWL se construye sobre RDF y RDF Schema y añade más vocabulario para la descripción de clases y propiedades: entre otras, relaciones entre clases (por ejemplo, inconexión), cardinalidad (por ejemplo, “exactamente uno”), igualdad, mayor riqueza de tipos en las propiedades, características de propiedades (por ejemplo, simetría) y clases numeradas [W3C04]. Es así, como OWL se suele considerar una extensión de RDF, pues incluye toda la capacidad expresiva de RDF(S) y la extiende con la posibilidad de utilizar relaciones lógicas entre recursos y relaciones [Cas].

OWL supone un gran paso adelante en la representación y organización de conocimiento en la World Wide Web. Se posiciona como ganador entre las necesidades

²² Ontology Web Language.

²³ Knowledge Interchange Format.

²⁴ DARPA's Agent Markup Language.

²⁵ Ontology Inference Layer.

²⁶ Precursor de OWL, unifica los lenguajes DAML y OIL.

de la industria de un lenguaje que dirija sus casos de usos Web actuales, y las restricciones de desarrollar un lenguaje de ontologías que sirva de engranaje a principios científicos establecidos y experiencia de investigación [W3C04].

En resumen, esta capa permite extender la funcionalidad de la Web Semántica, agregando nuevas clases y propiedades para describir los recursos y junto con ello da la posibilidad de realizar inferencia sobre la información [Lam08]. Sin embargo, es importante señalar que se pueden desarrollar aplicaciones Web semántica sin usar Ontologías. Muchas aplicaciones útiles pueden ser construidas sin esto, apoyándose en lo más fundamental y conceptos simples de la Web semántica [W3C08b].

La verdadera dificultad cuando se desarrolla una ontología, es **entender** el problema que tiene que ser modelado y **encontrar un acuerdo** en un nivel de comunidad. Esquemas RDF, para representar el nivel más simple de ontología, y/o OWL para una mayor expresividad, proporcionan un marco para formalizar ontologías en un lenguaje específico; el tiempo y la energía necesarios para aprender y usar estos, es sólo una fracción del tiempo necesario para desarrollar una ontología propia. Lo anterior, se refiere al hecho de entender los términos y las relaciones de estos términos dentro un área dada de conocimiento y lograr ponerse de acuerdo con los compañeros en relación a los conceptos y relaciones utilizadas [W3C08b].

2.3.6 Logic

Además de ontologías se precisan también reglas de inferencia. Es posible construir una ontología con las clases Persona, Perro, Mamífero, Rabia y Enfermedad, y las relaciones entre ellas (por ejemplo, un perro es un mamífero y la rabia una enfermedad). A su vez, las palabras “Luis” y “Fido” pueden asociarse a las clases de la ontología, definiendo por ejemplo que “Luis es una persona” y “Fido es un Perro”. Aunque estos metadatos pueden aclarar el sentido en que se usan estos conceptos, la utilidad de estos será mayor si la capa anterior (*Ontology Vocabulary*) se extiende a la capa lógica [Abi07].

La capa lógica se enfoca al razonamiento, es más general que los metadatos o las ontologías y permite descubrir conocimientos y relaciones ocultos entre “marañas” de datos. Un motor de inferencia o razonador (programa que genera deducciones siguiendo reglas lógicas) que procese los metadatos del documento mencionado en el ejemplo anterior deducirá automáticamente que Fido puede contraer la Rabia, pues Fido es una instancia de la clase Perro y esta es una subclase de Mamífero. Por lo tanto, lo más adecuado sería vacunarlo cuanto antes [Abi07].

Para la Web semántica, la gran ventaja de la lógica reside en que proporciona explicaciones para las deducciones. Éstas son muy importantes para la Web semántica porque los agentes inteligentes de la nueva Web las usarán para tomar decisiones, ejecutar acciones y colaborar entre ellos [Abi07].

Es importante mencionar que RDF-Schema y OWL tienen limitaciones para el razonamiento lógico. En este sentido, la W3C tiene en proceso de estandarización el lenguaje SWRL²⁷, lenguaje que propone eliminar estas limitaciones. Una ontología OWL contiene una secuencia de hechos y axiomas. Los axiomas pueden ser de varias clases, por ejemplo, axiomas de subclase, axiomas de equivalencias de clases o restricciones de propiedades. SWRL propone extenderlos con axiomas de reglas. Un axioma de regla consiste en un antecedente y un consecuente unidos por una relación de implicación, de tal manera, que una regla puede interpretarse como una expresión que indica que si el antecedente es cierto, entonces el consecuente también lo es. La especificación de SWRL ofrece una sintaxis abstracta que extiende a su vez la sintaxis abstracta de OWL para incluir la nueva relación en el lenguaje de ontologías.

2.3.7 Proof

Esta capa permite definir una infraestructura adecuada para que los agentes alcancen la máxima fiabilidad en sus razonamientos. Los agentes deben ser capaces de realizar demostraciones o, lo que es lo mismo a efectos prácticos, justificar el motivo por el cual tomó (o aconsejó tomar) una decisión [Cod06].

²⁷ Semantic Web Rule Language

Es la capa que permite definir una infraestructura para que los agentes puedan establecer relaciones lógicas complejas con otros agentes, e intercambiar pruebas sobre las inferencias obtenidas en una búsqueda.

2.3.8 Trust

Esta capa es la capa de confianza, es decir, aquella en que los agentes debieran ser muy escépticos acerca de lo que leen en la Web Semántica hasta que hayan podido comprobar de forma exhaustiva las fuentes de información. Pues una deducción es cierta si se basa en datos fiables. Por otro lado, los agentes debieran asegurarse de trabajar con agentes que sean autenticados por alguna “autoridad de confianza”. Además, la información entregada por un usuario u organización a un agente de un tercero debe ser manipulada de manera correcta, es decir, el agente no debiera usar esta información en beneficio del propietario del agente, pues de ser así, en el mejor de los casos, el uso de este agente puede resultar inútil para quien lo usa; en el peor, puede resultar contraproducente e incluso perjudicial. Por último, los agentes deberán estar sujetos a políticas de seguridad que garanticen que no transferirán datos confidenciales

2.3.9 Digital Signature

Bloque encriptado de datos que serán utilizados por los computadores y los agentes para verificar que la información adjunta no ha sido alterada y que ha sido ofrecida por una fuente específica confiable. Es decir, este bloque permitirá a los agentes comprobar que el documento es el original y que el autor de la información es quién dice ser.

Las capas de *Logic*, *Proof* y *Trust* se encuentran actualmente en fase de investigación, siendo las capas *Proof* y *Trust* las menos consolidadas, y sólo se han construido demostraciones de aplicaciones simples [Sam05]. Por lo tanto, y al menos durante un tiempo, XML, RDF y OWL serán, con mucha diferencia, los pilares más importantes del proyecto de Web Semántica. [Cod06]

Luego del análisis de cada una de las capas que componen la arquitectura de la Web semántica, se puede determinar que el objetivo principal que propone este

ambicioso proyecto, es que la Web pase de ser una Colección de documentos a una base de *Conocimientos*.

2.3.10 Lenguajes de Consulta

2.3.10.1 SPARQL

Aún con las distintas capas de la arquitectura de la Web Semántica, URIs, RDF, OWL, etc, esta Web es deficiente si no se establece un lenguaje de Consultas. Pues “ Pretender usar la Web Semántica sin un lenguaje de consulta es como pretender usar una base de datos relacional sin SQL²⁸”, explicó Berners-Lee. Es así, como la *World Wide Web Consortium*, el 15 de enero del 2008 anunció la publicación del estándar SPARQL (pronunciado "sparkle"), un lenguaje de recuperación basado en RDF; su nombre es un acrónimo recursivo del inglés SPARQL Protocol and RDF Query Language [W3C08c].

Existen muchos lenguajes de consulta con éxito, incluyendo estándares como SQL y XQuery²⁹, que estaban diseñados originariamente para consultas limitadas a un producto, formato, o tipo de información sencillo, en un repositorio de datos local. Tradicionalmente, ha sido necesario formular las mismas consultas de alto nivel de forma diferente, dependiendo de la aplicación o del convenio específico elegido para la base de datos relacional. Además, en las consultas de varios orígenes de datos ha sido necesario realizar declaraciones lógicas para combinar los resultados. Estas limitaciones han impuesto mayores costos de desarrollo y han creado barreras a la hora de incorporar nuevas fuentes de datos [W3C08c].

El objetivo de la Web Semántica es permitir a las personas compartir, combinar y reutilizar los datos de forma global. SPARQL ha sido diseñado para un uso a escala de Web, así permite hacer consultas sobre orígenes de datos distribuidos, independientemente del formato, pues las consultas son independientes de la

²⁸ Structured Query Language, Lenguaje de Consulta Estructurado, es un lenguaje declarativo de acceso a base de datos relacionales que permite especificar diversos tipos de operaciones sobre las mismas.

²⁹ XML Query, Lenguaje de consultas diseñado para consultar colecciones de datos XML. Es semánticamente similar a SQL, pero incluye algunas capacidades de programación.

representación física de los datos RDF (la estructura de la base de datos; su representación en un archivo RDF/XML, N-Triples; su representación virtual al ser creados en memoria en respuesta a una consulta; etc). Si la consulta se hiciera a través de, por ejemplo, XQuery, la aplicación tendría que conocer la forma en que la data RDF está representada exactamente.

Es más sencillo crear una consulta sencilla sobre diferentes almacenes de datos que crear múltiples consultas, además de tener un costo menor y de ofrecer resultados mejores [W3C08c].

El grupo de trabajo de Acceso a Datos RDF del W3C produjo 3 Recomendaciones de SPARQL: el **Lenguaje de Consulta SPARQL para RDF**, el **Formato XML de Resultados de Consultas**, y el **Protocolo SPARQL para RDF**. La primera especificación define la sintaxis y la semántica del lenguaje de consulta SPARQL para RDF. Puede utilizarse para expresar consultas sobre diversas fuentes de datos, tanto si los datos son almacenados nativamente como RDF, como si son visualizados como RDF mediante un *middleware*³⁰. La segunda especificación define un formato XML para el resultado de las consultas SPARQL. Estos resultados, pueden ser conjuntos de resultados o grafos RDF. La tercera especificación utiliza WSDL³¹ 2.0 para describir un medio que transmita consultas SPARQL a un servicio de procesamiento de consultas SPARQL y devuelva los resultados de las consultas a la entidad que las realizó. El protocolo SPARQL se describe de dos maneras: primero, como una interfaz abstracta independiente de alguna realización concreta, implementación o vinculación a otro protocolo; segundo, como HTTP³² y SOAP³³.

Por lo tanto, SPARQL es la clave para exponer los datos en la Web semántica. Con esta tecnología las personas se pueden centrar en lo que realmente quieren encontrar, en lugar de concentrarse en la tecnología de la base de datos o el formato utilizado para almacenar esos datos. Debido a que las consultas SPARQL expresan

³⁰ Es un software que funciona como una capa de abstracción de software distribuida, que se sitúa entre las capas de aplicaciones y las capas inferiores (Sistemas Operativos, Red, etc).

³¹ Web Services Description Language, formato XML que se utiliza para describir servicios Web.

³² HyperText Transfer Protocol

³³ Simple Object Access Protocol, es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML.

objetivos de alto nivel, son fáciles de extender a fuentes de datos heterogéneas superando las limitaciones de búsquedas locales y los formatos de datos únicos [W3C08c].

Sin embargo, presenta una limitación. Actualmente, la versión estándar de SPARQL se refiere sólo a la recuperación de los datos seleccionados de los grafos RDF. No existe un equivalente de las declaraciones SQL como *insert*, *update* o *delete*. Sin embargo, existe un significativo trabajo activo que tiene como fin extender SPARQL para que pueda apoyar operaciones de actualización [W3C08b].

2.3.10.2 Otros Lenguajes

Además de SPARQL han surgido una serie de lenguajes de consulta, similares al SQL de las bases de datos relacionales, que permiten expresar búsquedas complejas sobre grafos RDF mediante una sintaxis sencilla.

Este proyecto de tesis se basó en los estándares que plantea la W3C para el desarrollo de aplicaciones semánticas, esto por la calidad y visión de futuro que estas tecnologías ofrecen. Por lo anterior, el lenguaje de consulta a utilizar será SPARQL, sin embargo, a continuación se mencionan algunos de los lenguajes de consulta sobre documentos RDF más conocidos.

- **RDQL**

RDF Data Query Language, implementado por Hewlett Packard y posiblemente el más conocido, es un lenguaje de consultas para RDF desde un enfoque totalmente declarativo, derivado del lenguaje de consultas Squish. Por otro lado, RDQL da la base para la construcción de lenguaje SPARQL.

Considera un modelo RDF como un conjunto de tripletas (objeto, propiedad, valor). Una consulta RDQL puede especificar un conjunto de restricciones de los valores de las variables, además de listar las variables que requiere en el conjunto de respuestas.

Se pueden realizar consultas en RDQL desde una aplicación Java, PHP y Perl.

- **RQL**

RDF Query Language desarrollado en el instituto ICS-Forth, es un lenguaje de consulta declarativo para RDF/RDFS que de manera explícita captura la semántica de estas primitivas.

Su potencia semántica está basada en la evaluación de caminos de expresiones sobre grafos RDF. Permite el uso de variables tanto para denotar nombres de nodos(es decir, clases), como arcos(es decir, propiedades). Permite consultar esquemas RDF y descripciones RDF (es decir, instancias) en una misma consulta.

RQL, está definido por medio de un conjunto de consultas básicas, e iteradores que permiten construir otras consultas a través de una composición funcional.

- **SeRQL**

SeRQL (Sesame RDF Query Language, pronunciado como "circle") es un lenguaje de recuperación para RDF/RDFS, desarrollado por Aduna como parte del software Sesame. Combina características de otros lenguajes (principalmente RQL, RDQL, N-Triples y N3) y añade otras propias.

Sesame es un framework para RDF de código libre, con soporte para inferencias en RDF Schema y recuperación mediante queries. Originalmente desarrollado por Aduna, ahora es mantenido en cooperación con NLnet Foundation, y un grupo de desarrolladores voluntarios.

Dentro de las características más importantes de SeRQL están: la transformación de grafos, soporte de RDF Schema, soporte de los tipos de datos XML Schema y emparejado de caminos opcionales.

El diseño de SeRQL tuvo el objetivo de unificar algunas de las ventajas que ofrecían otros lenguajes, como parte de un lenguaje de recuperación ligero, pero potente.

Existen otros lenguajes de consulta RDF, que no se verán en este documento de tesis como Triple, N3QL, Versa y RxPath.

2.4 Proyectos realizados con las tecnologías de la Web Semántica

2.4.1 Vocabularios RDF

Dentro de los proyectos más conocidos de aplicación de Web semántica, se encuentran los vocabularios RDF, expresados en RDF/XML, entre ellos: FOAF, RSS y Dublin Core. Cada uno de los cuales puede reutilizarse y extenderse.

2.4.1.1 FOAF

The Friend Of A Friend (FOAF) es un proyecto creado para describir a las personas, los vínculos entre ellas, los intereses, las cosas que crean y hacen, a través de páginas Web fácilmente procesadas por máquinas [FOA]. Para distinguir de quién se está hablando, es decir, identificar de modo unívoco a una persona si se sabe que en el mundo hay cientos de miles de personas que comparten los mismos atributos, como el nombre, color de ojos, pelo, altura, intereses, etc, FOAF buscó un atributo propio de los usuarios de Internet, el email. El email es considerado una propiedad no ambigua acerca de una persona, en el sentido que toma a lo sumo una sola persona como destinatario, a lo largo del tiempo.

FOAF es una tecnología sencilla que hace más fácil el poder compartir y utilizar información de las personas y sus actividades (por ejemplo, fotos, calendarios, blogs) al transferir la información entre sitios Web, y automáticamente extender, fusionar y reutilizarla en línea [FOA]. Al igual que la versión de HTML, los documentos FOAF pueden ser vinculados para formar una red de información, pero con semánticas bien definidas. Además, al ser una aplicación RDF, FOAF aprovecha los beneficios de ser fácilmente recogidos y agregados. Y como todos los vocabularios RDF pueden ser fácilmente combinados con otros vocabularios, permite generar un rico conjunto de metadatos [FOA04].

```

<rdf:RDF [... ]>

<foaf:Person rdf:nodeID="me">
  <foaf:name>Diego Berrueta Muñoz</foaf:name>
  <foaf:nick>berrueta</foaf:nick>
  <foaf:mbox_shalsum>98a99390f2fe9395041bddc41e933f50e59a5ecb
  </foaf:mbox_shalsum>
  <foaf:homepage rdf:resource="http://www.berrueta.net/" />

  <foaf:knows>
    <foaf:Person>
      <foaf:name>Jose Emilio Labra</foaf:name>
      <foaf:mbox_shalsum>8114083efd55b6d18cae51f1591dd9906080ae89
      </foaf:mbox_shalsum>
      <rdfs:seeAlso rdf:resource=
        "http://di002.edv.uniovi.es/~labra/labraFoaf.rdf" />
    </foaf:Person>
  </foaf:knows>

</foaf:Person>

</rdf:RDF>

```

Figura 7: Ejemplo FOAF.³⁴

Fundada por *Dan Brickley* y *Libby Miller*, FOAF es una iniciativa que aborda de frente el objetivo de la Web Semántica, de crear una máquina que procese los datos de la Web, a través de aplicaciones que entiendan esta ontología. Esta apuesta aparenta ser bastante segura ya que el éxito temprano de la Web semántica radica en aplicaciones centradas en las personas.

2.4.1.2 RSS

Existen 3 tipos de formato RSS y sus siglas adquieren un significado diferente según la especificación usada. Estas son, **Really Simple Syndication**, **Rich Site Summary** y **RDF Site Summary**.

Un archivo RSS facilita el acceso a la información Web que se actualiza con regularidad, logrando que el usuario recupere al momento las novedades producidas en la información de su interés [Per06].

³⁴ [Pol05]

Los archivos RSS comúnmente se llaman feeds RSS o canales RSS y contienen un resumen de lo publicado en el sitio Web de origen. Se estructura en uno o más ítems. Cada ítem consta de un título, un resumen de texto y un enlace a la fuente original en la Web donde se encuentra el texto completo. Además puede incluir información adicional como el nombre de autor o la fecha y hora de publicación del contenido. Por lo tanto, cualquier fuente de información susceptible de poder ser “troceada” en ítems (los mensajes de un foro, por ejemplo) puede distribuirse utilizando RSS [Mic04]. Cada feeds o canal de información dispone de su propia dirección de Internet o URL del mismo modo que las páginas HTML convencionales. Sin embargo, a diferencia de estas no se pueden visualizar directamente con el navegador, de modo que es necesario utilizar un *Lector de Feeds* o *Agregadores*.

Los lectores de feeds funcionan de forma similar a los programas de correo electrónico. Pero en lugar de consultar un buzón de correo cada cierto tiempo para recoger los mensajes, consultan periódicamente las direcciones de los feeds para obtener la última versión disponible de su archivo RSS. Los lectores feeds pueden ser aplicaciones Web o aplicaciones locales que se instalan en el computador del usuario [Mic].

Por otro lado, cualquier usuario puede suscribirse a un feed y obtener las últimas noticias enviadas a su agregador o lector RSS, el cual le alertará cuando haya nueva información para leer. Esto le permite obtener los datos que necesita de forma rápida y precisa, pues no tiene que comprobar los múltiples sitios que ofrecen los contenidos que le interesan sin saber si se ha producido algún cambio en ellos o no.

Pero el RSS no sólo le sirve al usuario para recibir la información que otros le ofrecen, si no que también es de utilidad para mostrar los contenidos novedosos de su Web a otros internautas.

El proceso de sindicación de contenidos se convierte, de este modo, en un circuito en el que todos acceden y crean información con mayor facilidad y de forma más rápida [Per06].

```

<?xml version="1.0" encoding="iso88591"?><?xmlstylesheet
      href="/css/rss.css" type="text/css"?>
<rdf:RDF xmlns:dc="http://purl.org/dc/elements/1.1/"

  xmlns:rdf="http://www.w3.org/1999/02/22rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/"

  <channel rdf:about="http://www.w3c.es/noticias.rss">
    <title>W3C Oficina Española</title>
    <description>Guiando la Web hacia su máximo
      potencial...</description>
    <link>http://www.w3c.es</link>
    <dc:date>20051123</dc:date>
    <items> <rdf:Seq>
      <rdf:li rdf:resource="http://www.w3c.es/Noticias/
        2005/#L2005110774"/>
      <rdf:li rdf:resource="http://www.w3c.es/Noticias/
        2005/#L2005112382"/>
    </rdf:Seq> </items>
  </channel>

  <item
rdf:about="http://www.w3c.es/Noticias/2005/#L2005110774">
    <title>El W3C pone en marcha el Grupo de Trabajo de
      Formato de Intercambio de Reglas</title>
    <description>20051107:W3C se complace en anunciar la
      creación del Grupo de Trabajo de Formato de
      Intercambio de Reglas. "Después de
      [...]</description>

    <link>http://www.w3c.es/Noticias/2005/#L2005110774</link>
      <dc:date>20051107</dc:date>
  </item>

</rdf:RDF>

```

Figura 8: Ejemplo RSS. ³⁵

2.4.1.3 Dublin Core

Nació como un pequeño conjunto de descriptores, en principios pensados para que fuera el propio autor el que los incluyera en el documento o recurso, pero que rápidamente adquirieron alcance global porque también se interesaron en ellos numerosos y variados proveedores de información pertenecientes a diferentes sectores como el de las artes, las ciencias, la educación, los negocios y la administración. Hoy, los metadatos Dublin Core se han convertido en uno de los estándares más extendidos

³⁵ [Pol05]

para la recuperación de información en la World Wide Web y el DC se ha convertido en un vocabulario muy utilizado no sólo en el ámbito bibliotecario y documental, si no en otros muchos sectores. Además, este conjunto de metadatos se puede utilizar no sólo con HTML, si no sobre otros lenguajes estructurados como XML y conjuntamente con otros lenguajes de descripción como el planteado por la W3C, RDF.

El conjunto de elementos Dublin Core está compuesto por 15 descriptores como resultado de un consenso y un esfuerzo interdisciplinar e internacional. Ya existen transcripciones a 20 idiomas y es el estándar oficial del WWW Consortium y el estándar del Z39.50³⁶. Los metadatos Dublin Core han sido aprobados por el organismo nacional de estandarización norteamericano (ANSI/NISO Z39.85) y los utilizan como base tanto gobiernos como agencias supranacionales y muchas otras iniciativas de metadatos, pertenecientes a comunidades específicas como bibliotecas, archivos, educación, negocios, etc.

Los metadatos Dublin Core tratan de ubicar, dentro de Internet, los datos necesarios para describir, identificar, procesar, encontrar y recuperar un documento introducido en la red. Si este conjunto de elementos se lograra aceptar internacionalmente supondría que todos los procesos que indizan documentos en Internet encontrarían, en la cabecera de los mismos, todos los datos necesarios para su indización y además estos datos serían uniformes. Si el DC lograra estandarizar los metadatos de la cabecera de los documentos se facilitaría su indización automática y mejoraría la efectividad de los motores de búsqueda [Lam08b].

La *Dublin Core Metadata Initiative* (DCMI) es la responsable del desarrollo, estandarización y promoción de este conjunto de elementos de metadatos. Su objetivo es elaborar normas interoperables sobre metadatos y desarrollar vocabularios especializados en metadatos para la descripción de recursos que permitan sistemas de recuperación más inteligentes [Lam08b].

El conjunto de elementos de metadatos Dublin Core es un conjunto de metadatos previsto para describir documentos. Los elementos poseen etiquetas descriptivas que pretenden transmitir un significado semántico a los mismos [Lam08b].

³⁶ Es una norma para consultar catálogos de bibliotecas de Internet que utiliza las mismas reglas para todos los catálogos.

Cada elemento es opcional y puede repetirse. Además, los elementos pueden aparecer en cualquier orden. Aunque algunos entornos, como HTML no diferencian entre mayúsculas y minúsculas, es recomendable escribir correctamente cada metadato, según su definición, para evitar conflictos con otros entornos, como SGML³⁷ y XML [Lam08b].

Podemos clasificar el conjunto de elementos Dublin Core en 3 grupos que indican la clase o el ámbito de la información que contienen:

1- Elementos relacionados principalmente con el contenido del recurso. Sea título, tema, descripción, fuente, etc.

2- Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual. Entre ellos, autor, editor, derechos.

3- Elementos relacionados principalmente con la instanciación del recurso. Como fecha, tipo de recurso, formato, etc [Lam08b].

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22rdfsyntaxns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://media.example.com/
    audio/guide.ra">
    <dc:creator>Rose Bush</dc:creator>
    <dc:title>A Guide to Growing Roses</dc:title>
    <dc:description>Describes process for planting and nurturing
      different kinds of rose bushes.</dc:description>
    <dc:date>20010120</dc:date>
    <dc:language>enUK</dc:language>
    <dc:rights>(C) Rose Bush</dc:rights>
    <dc:type>moving image</dc:type>
    <dc:format>application/video</dc:format>
  </rdf:Description>
</rdf:RDF>
```

Figura 9: Ejemplo Dublin Core.³⁸

³⁷ Standard Generalized Markup Language, Lenguaje de Marcación Generalizado. Consiste en un sistema para la organización y etiquetado de documentos.

³⁸ [Pol05]

Es preciso destacar que se pueden realizar combinaciones de vocabularios para sacar otros más extensos. Es decir, se pueden combinar por ejemplo, los tres vocabularios anteriores y formar un vocabulario más extenso y potente.

2.4.2 Aplicaciones Semánticas

La implementación y adopción de aplicaciones Web Semánticas es aún una carrera lenta. Se listan las tecnologías ya estandarizadas y con ello variadas herramientas, pero es necesario desarrollar aplicaciones reales que pongan en práctica los principios de la Web semántica, que pueblen la Web con ontologías y que hagan que la Web semántica adquiriera la masa crítica imprescindible para hacerse realidad. Lo que sí está realmente claro, es que los campos donde son aplicables las tecnologías de la Web Semántica hoy son muchos y a continuación se mencionarán algunos de ellos.

- **Aplicaciones de búsqueda:**

Las tecnologías de web semántica permiten desarrollar buscadores avanzados aplicados a distintos dominios. Los dominios con mayor potencial son la administración pública, la medicina, las búsquedas sociales, el empleo, las artes, la ciencia, la política, la economía, la gestión del conocimiento, y otros [Can06].

- **Aplicaciones de asistencia al usuario:**

Estas aplicaciones están relacionadas con los agentes personales de usuario que permiten realizar búsquedas avanzadas en Internet, descubrimiento de servicios, composición de servicios, etc [Can06].

- **Aplicaciones de integración de fuentes de datos heterogéneas**

Estas aplicaciones permiten obtener, agrupar y correlacionar información dispersa en Internet sobre un dominio, como, por ejemplo, deportes o noticias [Can06].

- **Aplicaciones de anotación semántica de contenidos multimedia**

Estas aplicaciones permiten catalogar los contenidos multimedia de forma semántica, pudiéndose realizar catálogos de contenidos personalizados para los usuarios, descubrimiento de nuevos recursos multimedia de interés para el usuario, etc [Can06].

- **Aplicaciones de adaptación automática de contenidos en base a la anotación semántica de los mismos**

La idea que subyace en estas aplicaciones es que los contenidos web sean adaptados dinámicamente teniendo en cuenta su semántica y la personalización asociada al usuario. Actualmente existen mecanismos automáticos de adaptación de contenidos, pero basados únicamente en aspectos sintácticos de las páginas [Can06].

- **Aplicaciones para las empresas**

Bajo este punto aparecen todas aquellas aplicaciones de la web semántica encaminadas a mejorar los mecanismos actuales de gestión de las empresas, explotando al máximo el nuevo abanico de posibilidades que ofrecen las tecnologías y plataformas de web semántica [Can06].

2.5 Entorno de la Telefonía

El año 2006 constituyó el principio de una fase de lentitud en el crecimiento del sector de las telecomunicaciones. Después de tres años de evolución ascendente de los crecimientos anuales, comienza un periodo en el que se pronostican crecimientos moderados y en descenso. Según EITO³⁹, en el 2006 la tasa anual de crecimiento fue del 3,6%, y las previstas para los años 2007 y 2008, 2,8% y 2,4% respectivamente, revelan que entra en una fase de desaceleración. La estabilidad del mercado de servicios de telecomunicaciones se ve contrarrestada por la progresiva disminución de las tasas de crecimiento del sector de equipos de telecomunicaciones, que tras el despegue del año 2004, comienza a crecer a menor velocidad [Ora07].

Los cambios en tecnología, en las instituciones de mercado y en las políticas regulatorias y de competencia están transformando el sector de las telecomunicaciones, dirigiéndolo hacia la convergencia de mercados. Este escenario está determinando que continúe la tendencia observada el año 2006 en cuanto al aumento de acuerdos y alianzas entre empresas, lo cual lleva a la concentración del mercado TIC⁴⁰ en grandes compañías que traspasan fronteras. En todo este proceso no hay que pasar por alto la presión que ejercen los mercados emergentes que amenazan a los occidentales. En diciembre de 2006, Alcatel y Lucent completaron su fusión dando lugar al líder mundial en el suministro de soluciones de comunicación con presencia local en 130 países. Nokia y Siemens reaccionaron a esta fusión de sus rivales y firmaron en junio una alianza para liderar la convergencia de redes fijas y móviles. La alianza de AT&T y Bellsouth anunciada en marzo de 2006 y aprobada definitivamente en diciembre, crea un nuevo gigante de las telecomunicaciones que podría convertirse en la compañía dominante en EE UU, con el control de más de la mitad de líneas de teléfono y de acceso a Internet en el país. Todos los expertos confirman que las adquisiciones del sector están teniendo un efecto positivo e impulsan a las empresas de telecomunicaciones en la Bolsa [Ora07].

³⁹ Observatorio Europeo de las Tecnologías de la Información.

⁴⁰ Tecnologías de la Información y Comunicación.

Dentro del mercado de las telecomunicaciones, el sector de los servicios ha superado en el 2006 el billón de euros⁴¹ y se encuentra en una fase de estabilidad. La principal causa de ello hay que buscarla en que el avance de la telefonía móvil y el crecimiento sostenido de la banda ancha equilibran, en buena medida, el paulatino descenso de la telefonía fija, provocado por una disminución de líneas, de tráfico y de las tarifas (Gráfico Figura 10) [Ora07].

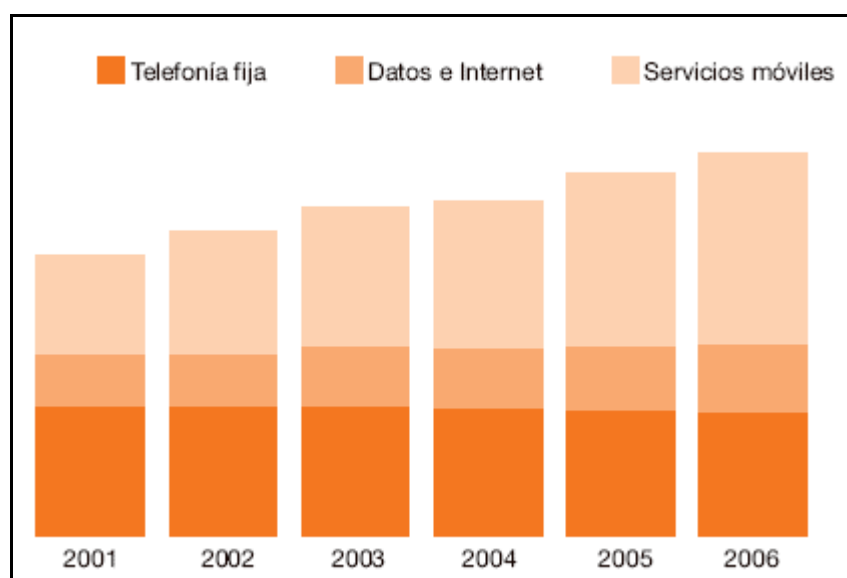


Figura 10: Mercado mundial de los servicios de telecomunicaciones por segmento. Años 2001-2006.⁴²

Los servicios móviles generaron en 2006 el mayor incremento en el mercado de los servicios de telecomunicaciones, con un 17% y unos ingresos en torno a 470.000 millones de euros. Así mismo, el número de usuarios pasó de 2.000 millones en 2005 a 2.600 en 2006. Sin embargo se contrarresta el auge de la telefonía móvil en los países en desarrollo, con la ralentización en los países más avanzados, como consecuencia de la maduración de sus mercados. Así, en el 2006, el crecimiento en los países en desarrollo ha sido mucho mayor, un 25%, que en los países avanzados, un 5,3%. Esta tendencia se acentúa en Europa y Japón, donde la maduración de sus mercados de telefonía móvil lleva a que los números de usuarios móviles sigan creciendo, pero los ingresos por usuario reflejan una tendencia a la baja, ya que el aumento de la competencia ha hecho disminuir los precios de los servicios de telecomunicaciones [Ora07].

⁴¹ 1 euro = 786, 789555 pesos chilenos.

⁴² [Ora07], Fuente IDATE.

El año 2006, es el año en que la telefonía móvil se consolida frente a la fija. El éxito de la telefonía móvil radica en que está sustituyendo a la fija en los países en desarrollo y complementándola en los desarrollados. Es también el año del despegue de la tercera generación (3G), que abre nuevos escenarios: la oferta de servicios y contenidos de Internet móvil. En los países menos adelantados la telefonía móvil ha sido, de todas las TIC, la que más ha contribuido a la disminución de la brecha digital, representando una vía de acceso importante a la sociedad de la Información. Además, las comunicaciones móviles inciden de manera positiva sobre la economía de los países menos favorecidos [Ora07].

La telefonía móvil es uno de los principales motores de convergencia tecnológica y la movilidad un factor cada vez más demandado. En cualquier lugar, en cualquier momento y en cualquier formato son las nuevas coordenadas de este sector. El teléfono móvil, con sus más de 2.600 millones de usuarios, está transformando el mercado rápidamente. Televisión móvil, Internet, vídeos y *e-mail* son algunos de los servicios que ofrece esta tecnología. Los operadores de telefonía móvil se convierten cada vez más en proveedores de servicios multimedia integrados, ofreciendo soluciones de voz y datos. Esta tendencia hace necesario que los operadores móviles replanteen su modelo de negocio. Según un estudio realizado por KPMG, el 40% de los usuarios de móvil no está dispuesto a pagar más por recibir estos nuevos servicios. Este estudio refleja que el “consumidor global” pertenece a una generación surgida de la era de Internet, poco acostumbrada a pagar por un contenido habitualmente gratuito de ahí su reticencia a pagar por los nuevos servicios [Ora07].

Del mismo modo, los nuevos servicios móviles plantean unos nuevos patrones de comportamiento en los usuarios. Según un estudio realizado por Point Topic para el Reino Unido, disminuyen las llamadas de voz y el envío de *sms*⁴³ y, por el contrario, aumentan el acceso a Internet y el envío de mensajes multimedia (*mms*⁴⁴). Otro dato relevante lo constituye el interés creciente que manifiestan los usuarios en la movilidad. Al 60% le gustaría usar el teléfono móvil para el envío de *e-mail*, a más del 45% le

⁴³ Short Message Service, Servicio de Mensajes Cortos. Es un sistema de mensajes de texto para teléfonos móviles.

⁴⁴ Multimedia Menssaging Service, Servicio de mensajería multimedia. Es un sistema para enviar mensajes multimedia entre teléfonos móviles.

gustaría buscar información y comprar en Internet y a más del 30% le gustaría interactuar con su banco a través del terminal móvil [Ora07].

Las operadoras se dan cuenta de la necesidad del usuario de ser autosuficiente y se empeñan en ofrecer servicios, tanto a través de canales Web como Wap⁴⁵, que permitan cubrir esta necesidad. Con el objetivo de aumentar los ingresos y disminuir los costos, las operadoras buscan que el usuario se auto atienda a través de su computador o teléfono móvil, es decir, que el cliente pueda comprar equipos móviles ofrecidos por la empresa telefónica, contratar nuevos planes, pagar sus cuentas con la operadora, entre otras, sin la necesidad de ser atendidos por una ejecutiva en una sucursal [Ora07].

El sector de la telefonía móvil está volcado, entonces, en la innovación y el lanzamiento de nuevos productos. El objetivo es diferenciarse de la competencia y ofrecer a los usuarios un cada vez más amplio abanico de servicios de calidad que satisfaga sus crecientes necesidades de comunicación [Ora07].

Ejemplos de esta apuesta decidida por la innovación son el lanzamiento de nuevas formas de comunicación (mensajería instantánea, video-llamada, “pulsar para hablar”) así como el empuje por parte de los operadores de nuevos servicios de datos (descarga de vídeos, TV en el móvil) y de redes que permiten el acceso a Internet a alta velocidad [Ora07].

En la industria de la telefonía móvil se apuesta decididamente por la continua evolución de las tecnologías de red, que permitan una mayor transmisión de datos y así ofrecer mejores servicios de conectividad a los usuarios. Si la tecnología GSM⁴⁶ fue conocida como 2G y a UMTS⁴⁷ se le dio el nombre de 3G o tercera generación, a día de

⁴⁵ Wireless Application Protocol, Protocolo de aplicaciones inalámbricas. Es un estándar abierto internacional para aplicaciones que utilizan las comunicaciones inalámbricas.

⁴⁶ Global System for Mobile Communications, Es un estándar digital para comunicaciones móviles.

⁴⁷ Universal Mobile Telecommunications System, Sistema Universal de Telecomunicaciones Móviles. Es una de las tecnologías usadas por los móviles de tercera de generación. Sucesor de GSM.

hoy ya se comercializan teléfonos 3,5G (HSDPA⁴⁸) y se empieza a hablar de su evolución, 3,75G (HSUPA⁴⁹) [Ora07].

Por otro lado, los servicios que están experimentando un mayor crecimiento en los últimos años están siendo los basados en la conectividad. Entre dichos servicios se englobarían tanto los servicios de navegación a través del móvil (acceso a correo, descarga de música o juegos, consulta de noticias, etc) como el acceso a Internet desde un ordenador utilizando el teléfono como un *módem*. Además, la navegación se está viendo favorecida por una política de tarificación más agresiva y fácil de entender por parte de los usuarios, en la que caben destacar las “*tarifas planas*”, que permiten navegar por un precio fijo hasta llegar a un máximo de información descargada [Ora07].

Estas mejoras han provocado que el número de usuarios que utilizan conexiones a Internet a través de *WAP* hayan pasado del 4,4% al 7,8% en tan sólo dos años [Ora07].

Los usuarios de servicios de datos en el teléfono móvil consideran que las principales barreras para el uso de estos servicios siguen siendo el costo, la escasez de contenidos de calidad y la lentitud de acceso [Ora07].

Finalmente, es importante mencionar, que los operadores de telecomunicaciones están apostando decididamente por la convergencia de sus distintos servicios de telecomunicaciones. Esto significa que a los clientes se les ofrecerán servicios empaquetados (telefonía móvil y fija, TV digital e Internet), en función de su perfil sociodemográfico y de consumo [Ora07].

El camino hacia la convergencia tiene un impacto significativo en la oferta de los operadores y en la forma en la que se dirigen al cliente. Mientras anteriormente un cliente tenía que contratar cada servicio de forma independiente y en muchos casos a través de canales de distribución y venta distintos, la convergencia implica una atención más directa a las necesidades de los usuarios [Ora07].

⁴⁸ High Speed Downlink Packet Access. Acceso descendente de paquetes a alta velocidad. Es un protocolo de acceso de datos para redes de telefonía móvil con alta tasa de transferencia de bajada. Es una evolución de UMTS.

⁴⁹ High Speed Uplink Packet Access, Acceso ascendente de paquetes a alta velocidad. Es un protocolo de acceso de datos para redes de telefonía móvil con alta tasa de transferencia de subida. Es una evolución de HSDPA.

Los grandes operadores ya han avanzado en este camino y los que no lo han hecho, están en camino. Por su lado, los que carezcan de algún servicio deberán firmar alianzas con otros proveedores [Ora07].

2.6 La Telefonía móvil en Chile

En Chile la telefonía móvil ha alcanzado una alta penetración. Y a pesar de los altos niveles que este segmento alcanzó hace años (niveles de países desarrollados), destaca el hecho que sigue aumentando considerablemente. La gran aceptación de esta tecnología, se debe en gran parte a que ha sido especialmente atractiva en segmentos de bajos ingresos por la posibilidad de acceder a tarjetas de prepago, lo que permite un mayor control de gasto sin tener que afrontar el pago de un cargo fijo, de alta incidencia sobre el presupuesto familiar [Min06].

La figura 11 muestra el aumento del número de clientes desde el año 2000 al año 2007. Al cierre del año 2007 casi 14 millones de chilenos tiene un teléfono móvil, lo que corresponde a que alrededor de 84 de cada 100 chilenos tiene un celular. Lo anterior, hace presagiar que este año se alcance un porcentaje de penetración del 100% o más.

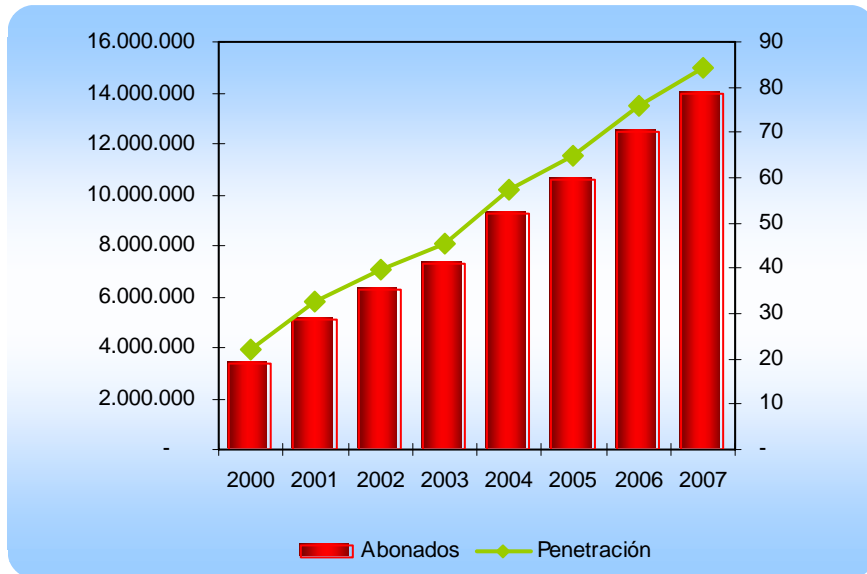


Figura 11: Número de clientes móviles y penetración cada 100 habitantes.⁵⁰

La telefonía móvil tuvo tasas de crecimientos espectaculares: 134,5% en el año 1999 y 55% en el 2001, las que sobrepasaron fuertemente las expectativas de crecimiento del mercado móvil que se tenían en el año 1999. Pese al aumento

⁵⁰ [Gob07]

experimentado en el año 2005 (10,5 %) y en los últimos años, el menor crecimiento estaría indicando el grado de madurez alcanzado por este mercado [Min06].

El mercado de la telefonía móvil en Chile, hoy se encuentra concentrado en tres compañías: Movistar, ENTEL PCS y Claro. Quiénes el año 2007 en su conjunto tenían ya 13.955.202 clientes [Min06]

Se estudiaron las 3 operadoras móviles existentes en Chile; sus propósitos, los servicios VAS⁵¹ que entregan y el número de clientes de cada una de ellas. A continuación, este análisis.

Movistar

Visión:

Quiere mejorar la vida de las personas, facilitar el desarrollo de los negocios y contribuir al progreso de las comunidades donde opera, proporcionándoles servicios innovadores basados en las Tecnologías de Información y las Comunicaciones. Espíritu de Progreso [Tel07].

Misión:

La misión de Movistar es ser líder en comunicaciones móviles, ofreciendo a los clientes un servicio de excelencia y un amplio rango de productos, desde comunicación de voz, larga distancia, data e Internet móvil. Además de apoyar a la comunidad en su desarrollo, privilegiar la búsqueda de soluciones para sus clientes, velar por los intereses de sus accionistas y generar un ambiente de trabajo que propicie el desarrollo de sus empleados" [Acc06].

ENTEL PCS

Visión:

Pretende ser la empresa de mayor valor económico dentro de la industria de telecomunicaciones móviles y sus desarrollos y desafíos, a través del reconocimiento de nuestros clientes, empleados, accionistas y sociedad en general [ENT].

Misión:

⁵¹ Servicios de Valor Agregado. Servicios de telefonía móvil que no incluyen voz.

Hacer que sus clientes y empleados sean los primeros en vivir el futuro, gracias a la imaginación y capacidad de nuestra empresa y quienes trabajamos en ella [ENT].

Claro

Visión:

Pretende Consolidarse como un grupo global de telecomunicaciones a través de operaciones internacionales que busquen satisfacer las necesidades y expectativas en comunicación de sus clientes. Buscan alcanzar los objetivos de crecimiento y financieros de sus accionistas, así como contribuir al desarrollo de sus recursos humanos y bienestar del entorno social de sus operaciones [CLA06].

Misión:

Se define como un grupo empresarial en expansión con enfoque a la internacionalización, primordialmente en el continente americano, e integración de sus negocios en el desarrollo económico y tecnológico en telecomunicaciones, centralmente inalámbricos, de los diversos países en los que tiene presencia [CLA06].

Al analizar las propuestas de los tres operadores móviles, se puede concluir que todos ellos persiguen el mismo propósito, ser una empresa líder en comunicaciones móviles. Y pretenden lograr esa meta cumpliendo distintos objetivos, pero que a su vez están muy vinculados. Estos objetivos son, lograr ser líder económico dentro de la industria de telecomunicaciones, ser líder en el desarrollo tecnológico y líder en satisfacer las necesidades y expectativas de sus clientes, ofreciendo servicios de calidad y una gran variedad de productos.

ENTEL PCS, Movistar y Claro participan en un mercado muy competitivo, el constante desarrollo tecnológico y con ello la innovación de nuevos servicios y productos junto con la baja de precios, lo prueban. Las tres operadoras móviles ofrecen continuamente nuevos productos y servicios a sus clientes, entendiendo como cliente personas y empresas. Intentan ser pioneras en el desarrollo de estos, sin embargo, a menudo y con el tiempo las tres operadoras terminan ofreciendo productos y servicios similares a sus clientes. Descarga de contenidos, tales como *ringtones*, *realtones*, *backtones*, canciones completas, videos, fondos, juegos; acceso a comunidades como *fotolog*, *blog*, *facebook*, *messenger*, salas de *chat*; recepción de información como

noticias, esoterismos, deportes o entretenimiento; acceso al correo electrónico (servidor de *email*, *email push*); reconocimiento de música; posicionamiento móvil (GPS, aGPS); y acceso a portales como mapcity, wikipedia y google son algunos de los tantos servicios que estas operadoras día a día ofrecen a los clientes chilenos. Por otra parte EPCS, Movistar y Claro se han preocupado también de satisfacer las necesidades de los clientes empresa, permitiéndoles entregar y recibir la información exacta en el momento oportuno. Para ellos han desarrollado servicios como transacción en línea, mensajería empresa, oficina móvil y localización, entre varios otros.

Las operadoras móviles buscan también potenciar sus canales virtuales, tanto Web como Wap, ofreciendo servicios que permitan al cliente auto atenderse en relación a todas sus necesidades. Es decir, intenta hacerle la vida más fácil al usuario al permitirle, por ejemplo, comprar equipos o accesorios móviles, pagar las cuentas de Internet, telefonía fija y móvil, contratar nuevos planes, etc, a través de su computador o teléfono celular. Lo anterior, permite una mayor eficiencia en la empresa pues se generarán mayores ingresos al ofrecer mejores y más servicios que satisfagan las necesidades de los clientes y por otro lado, se reducirán los costos al prescindir de ejecutivos y sucursales.

A pesar de las similitudes entre las tres empresas de telefonía móvil en cuanto a sus propósitos, servicios que ofrecen, etc. existen diferencias cuando se hacen estadísticas referentes a, por ejemplo, el número de clientes de cada una de ellas. A continuación se mostrará un gráfico que explicará de manera visual lo mencionado anteriormente.

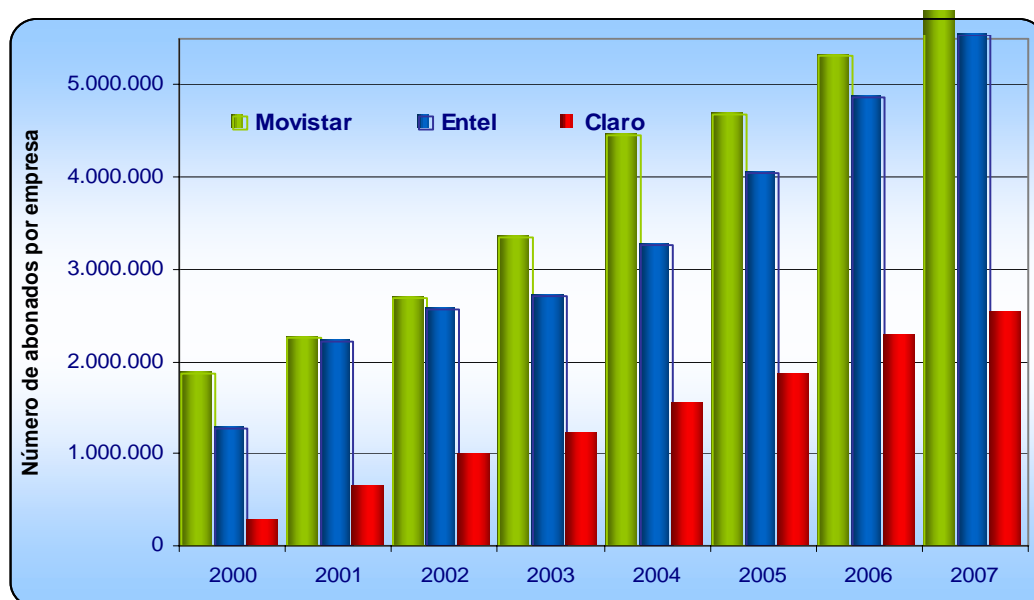


Figura 12: Evolución del número de clientes por empresa.⁵²

La notoria diferencia del número de clientes de Claro versus Movistar - EPCS, se debe principalmente a que Claro se inició con redes CDMA⁵³, a diferencia de las otras dos operadoras que trabajaban con redes GSM, un estándar europeo cada vez más usado por las operadoras del mundo. Este estándar, al ser ampliamente usado, ha provocado la creación de muchas herramientas compatibles con GSM, se han fabricado una gran variedad de equipos, tiene mejor cobertura al poder hacer roaming de un operador a otro, etc, lo que provoca que una mayor cantidad de usuarios se decida por operadoras que usan esta tecnología. Sin embargo, hace un par de años Claro Chile cambió sus redes a GSM y a pesar de que aún no alcanza el número de clientes de ENTEL PCS y Movistar, debido a la fidelización de los usuarios con sus operadoras, se espera que este año 2008, logre atraer más clientes, debido a ser la primera compañía en Chile en distribuir y vender *iPhone* de *Apple*.

Las compañías de telefonía celular, sin embargo, no sólo buscan aumentar el número de clientes, sino que sus consumidores utilicen sistemas de pospago o contrato,

⁵² [Gob07]

⁵³ Code Division Multiple Access, Acceso Múltiple por división de código. Tecnología de acceso de comunicaciones celulares e inalámbricas, que permite que un grupo de clientes comparta un mismo canal, distinguiéndolos por un código único para cada cliente.

pues más que dar importancia a la cantidad de usuarios que tenga una operadora, la clave para estas, es el ARPU⁵⁴.

Aunque el segmento de la telefonía móvil aún tiene espacio para crecer, el camino que estas tres operadoras siguen y deberán seguir, será “quitarse” clientes unos a otros y junto con ello lograr mantener a los clientes actuales, tanto antiguos como nuevos, centrándose específicamente en los clientes de pospago. Las operadoras, por lo tanto, de manera constante compiten por incorporar nuevas tecnologías que les permitan desarrollar nuevos servicios y productos de calidad y mejorar con ello la experiencia de usuario, de tal forma de fidelizar a sus clientes y evitar que estos sean tentados a cambiarse a las operadoras de la competencia.

3. Propuesta del Proyecto

3.1 Propuesta de prototipo de servicio a implementar

Luego de analizar tanto el entorno Web como el de la telefonía móvil en el mundo y en Chile, sus necesidades y objetivos, se decidió implementar un buscador de equipos móviles usando las tecnologías que propone la web semántica.

Actualmente las tres operadoras chilenas, entregan a sus clientes la posibilidad de encontrar, a través de su página web, los equipos móviles que cada una de estas compañías ofrece. El mecanismo de estos buscadores es muy similar en las tres operadoras, se basa principalmente en el uso de filtros. Esto es que dada una serie de categorías (marca, tipo, modalidad de compra, etc) el cliente puede escoger dentro de cada categoría las opciones que más se acerquen a lo que él busca.

⁵⁴ Average Revenue Per User, ingresos medios por usuario. Es la media o promedio de ingresos por usuario que obtiene, en un período de tiempo, una compañía de servicios con amplia base de usuarios.

Con la idea de proponer un servicio que se diferencie de la competencia, en su tecnología y forma de uso, y que permita al cliente una mejor experiencia de usuario, se propuso la idea de implementar un buscador Web de equipos móviles que le permita al cliente ingresar en lenguaje natural lo que desea encontrar, sin la necesidad de estar restringido a elegir opciones dentro de categorías.

3.2 Tecnologías a Utilizar

A continuación se mencionarán las tecnologías usadas y la justificación de su elección.

3.2.1 JAVA

La aplicación fue desarrollada en lenguaje Java, pues esta plataforma ofrece muchas ventajas.

Java se utiliza en los principales sectores de la industria de todo el mundo y está presente en un gran número de dispositivos, equipos y redes. La versatilidad y eficiencia de la tecnología Java, la portabilidad de su plataforma y la seguridad que aporta, la han convertido en la tecnología ideal para su aplicación a redes. De portátiles a centros de datos, de consolas de juegos a súper equipos científicos, de teléfonos móviles a Internet, Java se encuentra en todas partes.

Java se ha convertido en un recurso inestimable ya que permite a los desarrolladores, entre otras cosas [Jav]:

- Desarrollar software en una plataforma y ejecutarlo en prácticamente cualquier otra plataforma.
- Crear programas para que funcionen en un navegador Web y en servicios Web.
- Desarrollar aplicaciones para servidores como foros en línea, tiendas, encuestas, procesamiento de formularios HTML, etc.

- Desarrollar potentes y eficientes aplicaciones para teléfonos móviles, procesadores remotos, productos de consumo de bajo coste y prácticamente cualquier tipo de dispositivo digital [Jav]

El hecho que Java sea multiplataforma permite que la aplicación desarrollada en este proyecto de tesis pueda ser ejecutada en realmente cualquier sistema operativo como Windows, Linux, Mac, Unix, etc y sobre equipos que no tienen que ser necesariamente computadores como son los distintos dispositivos móviles que hoy existen en el mercado.

3.2.2 JENA

JENA es un *framework* de Java desarrollado por HP y disponible como software *Open Source*, para la construcción de aplicaciones Web Semántica. Proporciona un ambiente de programación para RDF, RDFS y OWL; lectura y escritura de documentos RDF en formato RDF/XML, N3 y N-Triples; además de un API para los lenguajes de consulta SPARQL y RDQL; incluye motores de inferencia basados en reglas, además de ofrecer mecanismos para añadir nuevos razonadores; y permite el almacenamiento persistente o en memoria de la información. La clase fundamental de JENA para los usuarios es la clase *Model*, una API para lidiar con un conjunto de tripletas RDF, es decir, modelos RDF.

3.2.3 Tomcat

Llamado también *Jakarta Tomcat* o *Apache Tomcat*, es un servidor Web Java con soporte de *servlets* y *JavaServer Pages* (JSP). Incluye el compilador Jasper, que compila JSPs convirtiéndolas en *servlets*.

- **JavaServer Pages:** es uno de los componentes más básicos empleados para aplicaciones de Servidor en Java. Su composición consta de dos grandes partes: HTML y lenguaje Java. Mediante HTML se especifica el contenido estático de despliegue y es mediante fragmentos del lenguaje Java, entre los símbolos `<%` y `%>`, que se genera contenido dinámico en efecto cumpliendo la definición de aplicación de Servidor. Todo JSP es convertido a un *servlet*, la razón de esto es que el surgimiento de JSP's se debió a la necesidad de facilitar un formato

programático sencillo, esto es, el uso de JSP's permite eliminar visualmente funciones de arranque, ejecución y terminación. La inclusión de éstas se lleva a cabo en la conversión de JSP a Servlet [Osm08].

- **Servlet:** Es la clase que se ejecuta en el servidor Web y a diferencia de un JSP, es un componente escrito puramente en Java. En términos Java esto equivale a una Clase y como cualquier otra Clase Java ésta se encuentra compuesta por diversos métodos y funciones. Un servlet, a diferencia de JSP no interactúa directamente con HTML [Osm08b].

3.2.4 RDF

Resource Description Framework (RDF) [Infraestructura para la Descripción de Recursos] es una base para procesar metadatos; proporciona interoperabilidad entre aplicaciones que intercambian información legible por máquinas en la Web.

Modelo RDF Básico

El modelo de datos de RDF es una forma de sintaxis-neutral para representar expresiones RDF y consiste en tres tipos de objetos:

- **Recursos**

Todas las cosas descritas por expresiones RDF se denominan *recursos*. Un recurso puede ser una página Web completa, tal como el documento HTML "<http://www.w3.org/Overview.html>"; una parte de una página Web, por ejemplo, un elemento HTML o XML específico dentro del documento fuente; una colección completa de páginas, como un sitio Web completo; un objeto que no sea directamente accesible vía Web, como un libro impreso, etc. Los recursos se designan siempre mediante URIs más un identificador opcional. Cualquier cosa puede tener un URI y la extensibilidad de estas permite la introducción de identificadores para cualquier entidad imaginable [SID01].

- **Propiedades**

Una *propiedad* es un aspecto específico, característica, atributo, o relación utilizado para describir un recurso. Cada propiedad tiene un significado específico, define sus valores permitidos, los tipos de recursos que puede describir, y sus relaciones con otras propiedades. En este documento no profundiza en cómo se expresan las características de las propiedades; para tal información, dirigirse a la especificación del esquema RDF [SID01].

- **Sentencias**

Un recurso específico junto con una propiedad determinada, más el valor de dicha propiedad para ese recurso es una **sentencia RDF**. Estas tres partes individuales de una sentencia se denominan, respectivamente, **sujeto, predicado y objeto**. El objeto de una sentencia (es decir, el valor de la propiedad) puede ser otro recurso o puede ser un literal; es decir, un recurso (especificado por un URI) o una cadena simple de caracteres (*string*) u otros tipos de datos [SID01].

A continuación se presenta un ejemplo simple, de lo explicado anteriormente.

Sea la sentencia:

“IMate es la marca del recurso http://www.epcs.cl/rdf/3.0#EQUIPO_12576.”

Se puede observar que comprende las siguientes partes:

Término lingüístico	Sujeto	Predicado	Objeto
Término lógico	Recurso	Propiedad	Valor (recurso o literal)
Ejemplo	http://www.epcs.cl/rdf/3.0#EQUIPO_12576	Marca	“IMate”

Figura 13: Partes de una declaración RDF.

Lo anterior, se puede representar gráficamente usando grafos etiquetados, también llamados “diagramas de nodos y arcos”. En este grafo, el nodo, dibujado como óvalo, representa un recurso; los arcos representan propiedades y los rectángulos, cadenas de literales [SID01].

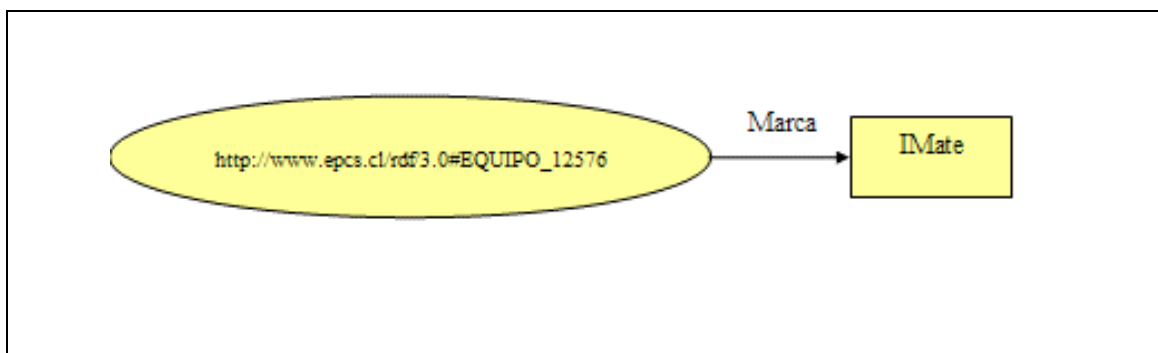


Figura 14: Diagrama de nodos y arcos

La dirección de la flecha es importante. El arco siempre empieza en el sujeto y apunta hacia el objeto de la sentencia. Este diagrama simple puede también interpretarse o leerse "*http://www.epcs.cl/rdf/3.0#EQUIPO_12576 tiene como marca IMate*", o en general "*<sujeito> TIENE <predicado> <objeto>*".

Un ejemplo de documento RDF, se presenta a continuación. La información mostrada, corresponde a un extracto de la descripción de un equipo móvil.

```
<rdf:RDF
  xmlns:epcs="http://www.epcs.cl/rdf/3.0#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <rdf:Description
rdf:about="http://www.epcs.cl/rdf/3.0#EQUIPO_12576">
  <epcs:NOMBRE_KIT
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Tarjeta
  SIM</epcs:NOMBRE_KIT>
  <epcs:JUEGOS_JAVA
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">No</epcs:
```

```

JUEGOS_JAVA>
<epcs:CARACTERISTICA
  rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ringtones
  Polifónicos</epcs:CARACTERISTICA>
<epcs:A2_NCUOTAS
  rdf:datatype="http://www.w3.org/2001/XMLSchema#string">18</epcs:
  A2_NCUOTAS>
<epcs:RINGTONES_MONOFONICO
  rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ringtones
  Monofónicos</epcs:RINGTONES_MONOFONICO>
<epcs:NOMBRE_MARCA>IMate</epcs:NOMBRE_MARCA>
<epcs:BACKTONES
  rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Backtones
  </epcs:BACKTONES>
</rdf:Description>
</rdf:RDF>

```

Figura 15: Fragmento RDF

3.2.5 SPARQL

SPARQL define un lenguaje de recuperación para RDF/RDFS. Esta tecnología de consulta SPARQL permite que las personas puedan centrarse en la información que quieren, sin tener en cuenta la tecnología de la base de datos o el formato utilizado para almacenar esos datos. Debido a que las consultas en el lenguaje SPARQL expresan objetivos de alto nivel, es fácil extenderlos a orígenes de datos inesperados, o incluso transferirlos a nuevas aplicaciones.

Componentes

El Lenguaje SPARQL posee tres componentes importantes: URIs, literales y variables procedentes del lenguaje RDF.

- Las URIs, se escriben entre los símbolos < y > y se les asocia a un prefijo (“PREFIX”) a fin de que no se tenga que escribir el namespace en su totalidad, cada vez que se haga referencia.

- Los literales, que se describen como una cadena de caracteres encerradas entre “”.
- Las variables. Estas son globales y son precedidas por los símbolos “?” o “\$” no formando parte del nombre de la variable. Estos símbolos sólo indican que una cadena de caracteres precedida por uno de ellos, es una variable.

Sintaxis

La sintaxis de SPARQL es muy similar a la de SQL y así como RDF se compone de tripletas, consistentes en sujeto-predicado-objeto, también lo hace este lenguaje. Las tripletas en SPARQL representarán patrones mediante los cuales se buscarán tripletas coincidentes en los datos.

El lenguaje SPARQL tiene cuatro tipos de consultas, estas con sus respectivas descripciones y ejemplos asociados al fragmento de datos de la Figura 15, se pueden observar a continuación.

- **SELECT:** Devuelve todo, o un conjunto de, las variables que coinciden con el patrón de búsqueda.

```

PREFIX   epcs: < http://www.epcs.cl/rdf/3.0# >
SELECT   ?equipo_id
WHERE {
    ?equipo_id epcs:NOMBRE_MARCA ?marca
    FILTER (regex (?marca, "iMate", "i"))
}

```

Figura 16: Sintaxis del operador SELECT

- **CONSTRUCT:** Devuelve un grafo RDF construido por la sustitución de variables en un conjunto de plantillas de tripletas. En otras palabras, permite la creación de un nuevo grafo RDF a partir de los resultados obtenidos en las consultas.

```

PREFIX   epcs: < http://www.epcs.cl/rdf/3.0# >
PREFIX   epcs2: < http://www.epcs.cl/rdf/3.0-V.2# >
CONSTRUCT
{
    ?equipo_id epcs2:MELODIA ?sonido
    ?equipo_id epcs2:KIT_INCLUYE ?kit
}
WHERE {
    ?equipo id epcs:RINGTONES MONOFONICO ?sonido
}

```

```
?equipo_id  epcs:NOMBRE_KIT ?kit
}
```

Figura 17: Sintaxis del operador CONSTRUCT

La consulta anterior, devolvería un nuevo grafo RDF con la siguiente información:

```
@prefix  epcs2: < http://www.epcs.cl/rdf/3.0-V.2# >
<http://www.epcs.cl/rdf/3.0#EQUIPO_12576">epcs2:MELODIA
"Ringtones Monofónicos"
<http://www.epcs.cl/rdf/3.0#EQUIPO_12576">epcs2:KIT_INCLUYE
"Tarjeta SIM"
```

Figura 18: RDF generado por consulta CONSTRUCT

Se crea un nuevo grafo, con propiedades epcs2 en base a la información de epcs.

- **ASK:** Devuelve un booleano indicando si existe una coincidencia con el patrón de búsqueda o no.

```
PREFIX  epcs: < http://www.epcs.cl/rdf/3.0# >
ASK WHERE {
    ?equipo_id  epcs:JUEGOS_JAVA ?juego
    FILTER (regex (?juego,"Si", "i"))
}
```

Figura 19: Sintaxis del operador ASK

La respuesta en este ejemplo sería "False", pues no existe una tripleta en el fragmento del documento, tal que JUEGOS_JAVA sea Si. A diferencia de esto, especifica que el equipo móvil No tiene Juegos Java.

- **DESCRIBE:** Retorna un grafo RDF que describe los recursos encontrados.

```
PREFIX eps: < http://www.eps.cl/rdf/3.0# >
DESCRIBE ?equipo_id
WHERE {
    ?equipo_id eps:NOMBRE_MARCA ?marca
    FILTER (regex (?marca,"IMate", "i"))
}
```

Figura 20: Sintaxis del operador DESCRIBE

La consulta anterior, devolverá un grafo RDF que contenga las tripletas del modelo, asociadas al recurso que cumple con la condición establecida en el WHERE. En este ejemplo, se devolvería el mismo grafo RDF mostrado en la Figura 15.

4. Desarrollo del Buscador de Equipos Móviles.

En este capítulo se presentará la especificación de requisitos, su análisis, diseño de arquitectura e implementación de la aplicación. Para lo anterior, se usarán los diagramas de UML utilizados con mayor frecuencia en los proyectos de software, considerados más útiles y que entregan información adicional en relación con los demás diagramas.

4.1 Especificación de requisitos

4.1.1 Requisitos Funcionales

Las funcionalidades que debe proporcionar el sistema al usuario son:

Requerimiento	Categoría	Prioridad
1- El sistema debe ser capaz de entregar el servicio de búsqueda de equipos móviles mediante interfaz Web.	Evidente	Alta
2- El sistema debe permitir al usuario hacer búsquedas mediante lenguaje natural, entregando una respuesta coherente a la frase de búsqueda ingresada.	Evidente	Alta
3- El sistema debe ser capaz de mostrar la descripción general de los equipos.	Evidente	Alta
4- El sistema debe ser capaz de detallar las descripciones del móvil elegido por el usuario.	Evidente	Media

Figura 21: Requisitos Funcionales

4.1.2 Requisitos No Funcionales

Requerimiento	Categoría	Prioridad
1- El sistema debe permitir que varios usuarios puedan realizar búsquedas de equipos de manera concurrente.	Oculto	Media
2- El sistema debe ser fácil de extender, es decir, se debe poder agregar nuevas	Oculto	Media

funcionalidades sin dificultad.		
3- El proceso de implementación debe producir código legible, de tal forma, que se puedan realizar modificaciones fácilmente.	Oculto	Media

Figura 22: Requisitos No Funcionales

4.2 Análisis de Requisitos

4.2.1 Casos de Uso Buscar Equipos Móviles

Con los casos de Uso, se describirá **qué** es lo que debe hacer el sistema. Es decir, se describirá bajo la forma de acciones y reacciones el comportamiento del sistema, desde el punto de vista del usuario.

Se especificará tanto el diagrama de casos de uso, como el documento de descripción de este. Este último detallará, el formato de alto nivel, que describirá de manera clara y concisa el proceso; y el formato expandido, que describirá en mayor detalle la interacción entre el sistema y sus actores.

Buscar Equipos Móviles.

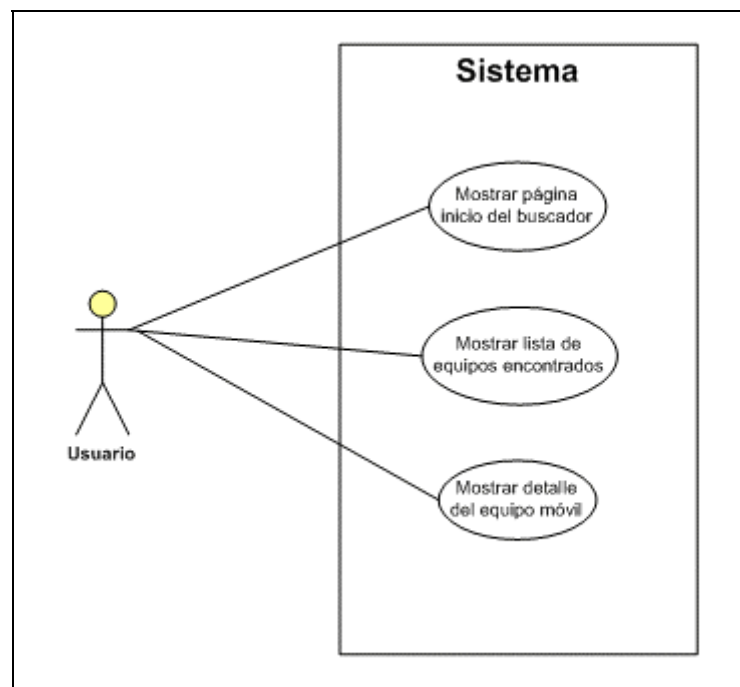


Figura 23: Caso de Uso Buscar equipos móviles

Formato de Alto Nivel

Caso de Uso: Buscar Equipos Móviles.

Actores: Usuario Web y Buscador.

Tipo: Primario

Descripción: El usuario busca en la Web el equipo móvil que desea y el sistema despliega una lista con los equipos encontrados, ofreciéndole la posibilidad de ver en detalle cada uno de ellos.

Formato Expandido

- **Primera Sección**

Caso de Uso: Buscar Equipos Móviles.

Actores: Usuario Web y Archivo RDF.

Propósito: Capturar la búsqueda y entregar resultado de esta.

Precondición: Escribir en lenguaje natural lo que se quiere buscar.

Resumen: El usuario escribe en el buscador el tipo de móvil que quiere encontrar, y el sistema le devuelve la lista de equipos que cumplan con las características planteadas por el usuario.

Tipo: Primario

Post Condición: Permitir que el usuario pueda ver en detalle la descripción del equipo.

- **Segunda Sección: Curso normal de eventos.**

Acción de los actores	Respuestas del sistema
<p>Este caso de uso comienza cuando:</p> <ol style="list-style-type: none">1. El usuario se conecta a Internet vía Web.2. El usuario escribe en la página, lo que quiere buscar.3. El usuario hace clic en botón Buscar.	<ol style="list-style-type: none">4. El sistema da inicio a la búsqueda.

<p>6. El usuario selecciona, haciendo clic, el equipo que más le llame la atención.</p>	<p>5. El sistema muestra una lista con los equipos móviles encontrados.</p> <ul style="list-style-type: none"> - Una descripción general. (marca-modelo) <p>7. El sistema muestra la descripción en detalle del equipo móvil.</p>
---	--

Figura 24: Curso normal de eventos de Caso de Uso Buscar Equipos Móviles.

- **Ultima sección: Curso alternativo de los eventos**

<p>En cualquier momento el Proceso búsqueda falla: Para dar soporte a la recuperación del proceso, el sistema debe asegurar que todos los estados y eventos significativos de una transacción puedan recuperarse desde cualquier paso del escenario.</p>	
<p>Acción de los Actores</p>	<p>Respuestas del Sistema</p>
<p>3. El usuario solicita la búsqueda de información no relacionada con los equipos.</p>	
<p>1. El usuario hace clic en Buscar.</p>	<p>2. El sistema no reconoce las palabras de la oración ingresada por el usuario.</p> <p>3. El sistema le envía un mensaje al usuario, indicándole el contexto de búsqueda.</p>

Figura 25: Curso alterativo de eventos de Caso de Uso Buscar Equipos Móviles.

4.2.2 Diagrama de Secuencia

A continuación se muestra el diagrama de secuencia, que permitirá conocer en más detalle, el diseño sobre el comportamiento del sistema. En este diagrama, se muestra para el caso de uso “Realizar búsqueda de equipos móviles”, los eventos que los actores generan (Usuario y Buscador), además del orden en que ocurren estos eventos.

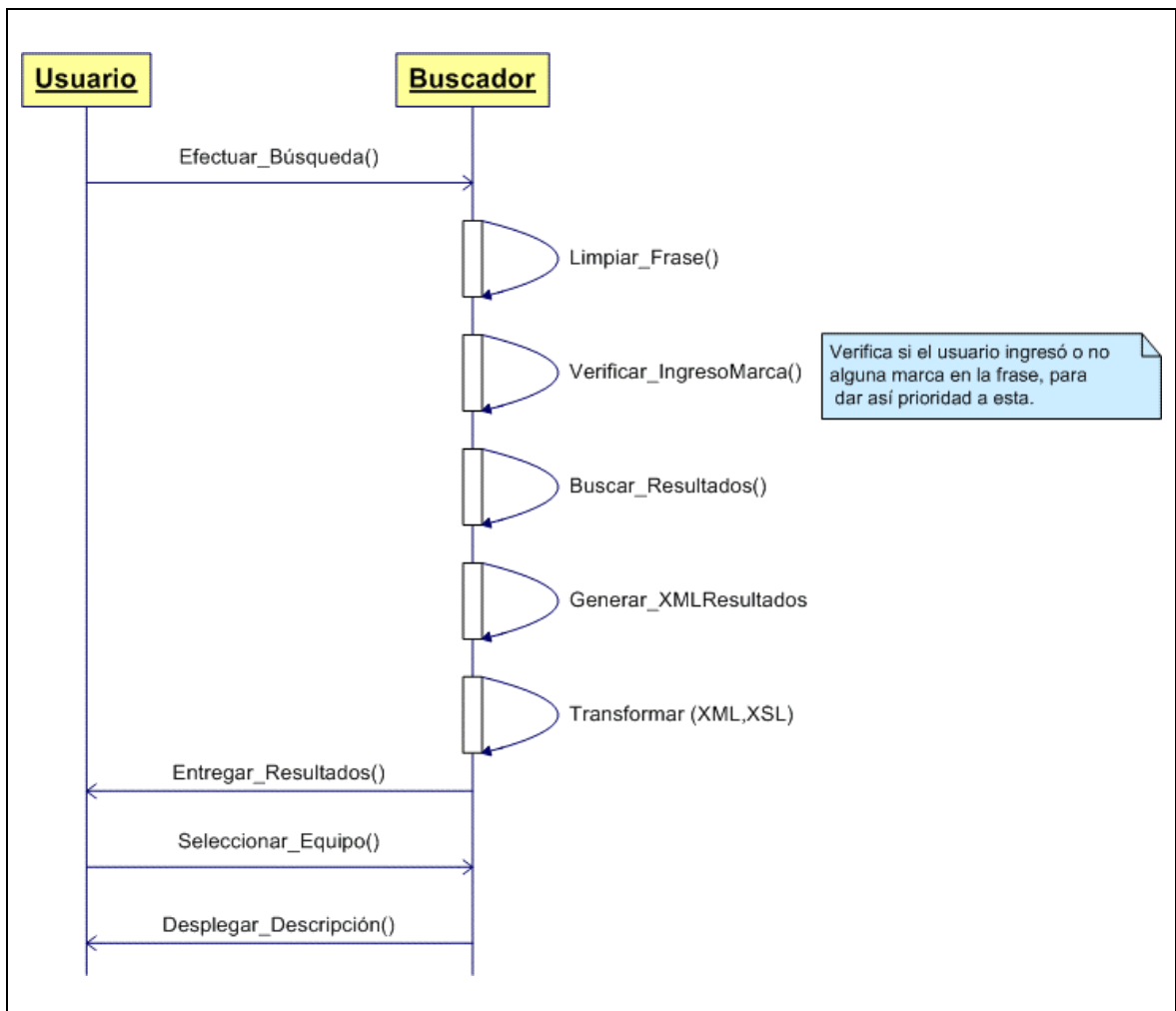


Figura 26: Diagrama de secuencia

4.2.3 Diagrama de Estados

El diagrama de estados, permitirá modelar el comportamiento del sistema al mostrar la secuencia de eventos permitida que pueden ser reconocidos y tratados por el sistema.

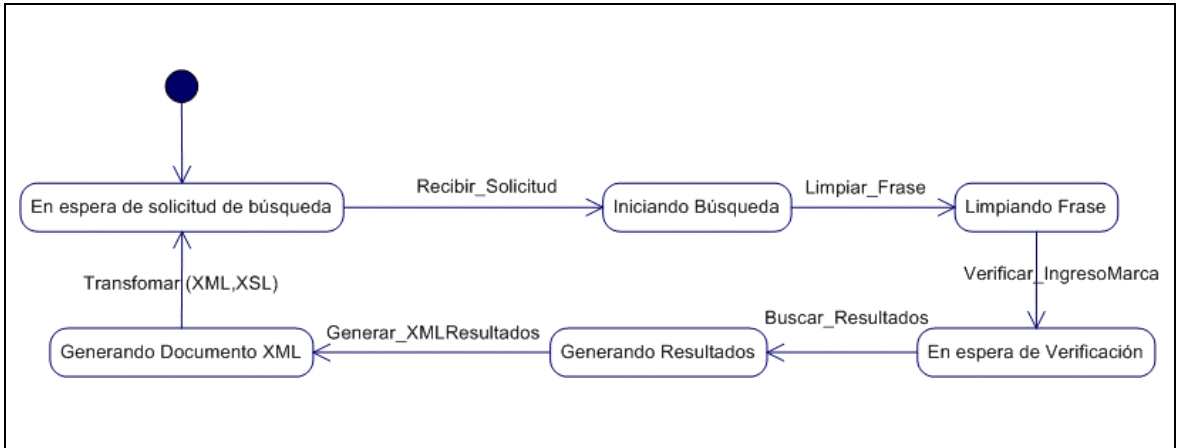


Figura 27: Diagrama de estados

4.3 Diseño

4.3.1 Diagrama de Clases

El diagrama de a continuación, permitirá visualizar las cinco clases, junto con sus métodos y atributos, que componen el sistema y cómo estas se relacionan unas con otras.

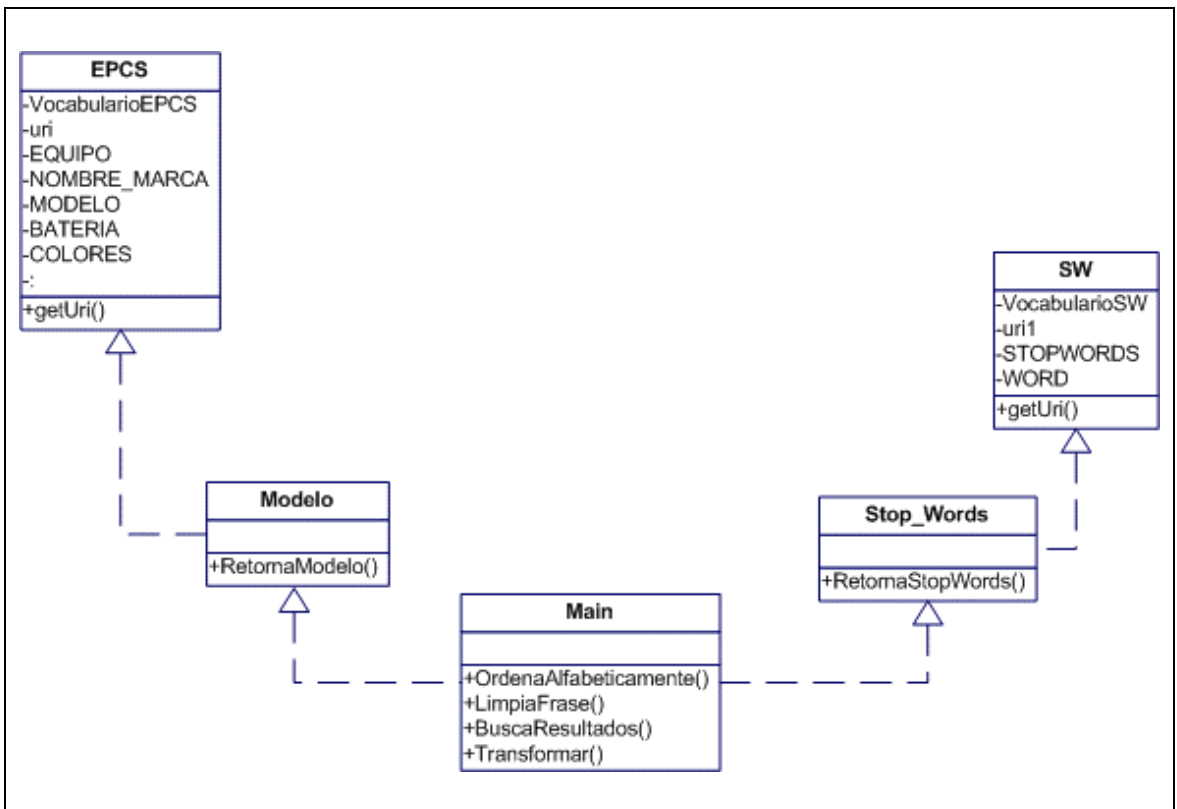


Figura 28: Diagrama de clases

4.4 Implementación

Luego de la toma de requisitos, su análisis y posteriormente el diseño de la aplicación, se comenzó con la implementación. Esta etapa, básicamente, estuvo formada por dos subetapas:

- La implementación del “core” de la aplicación buscador.
- La implementación de la interfaz Web.

4.4.1 Implementación del Buscador

Como se mencionó en el apartado “Tecnologías Usadas”, el desarrollo de la aplicación se hizo en lenguaje Java utilizando la API Jena. Se trabajó hasta el nivel RDF de la arquitectura de la Web Semántica, especificando un vocabulario que permitiera codificar conocimiento concreto acerca de equipos móviles. A cada equipo móvil se le asociaron piezas de información, a través de predicados que describieran marca, modelo, sonidos, precios, colores, entre otras características.

El desarrollo de la aplicación consta de 5 clases, cada una de las cuales se explicará brevemente a continuación:

- **Clase EPCS:** Clase que define el vocabulario para el namespace <http://www.epcs.cl/rdf/3.0#>, es decir, recursos y propiedades que se usará en el modelo de datos RDF de EPCS.
- **Clase Modelo:** Clase que define el modelo de datos RDF, especificando las tripletas RDF en función de los recursos y propiedades definidos en la clase EPCS. En esta clase se define la información asociada a cada equipo móvil.
- **Clase SW:** Clase que define el vocabulario para el namespace <http://www.stopwords.cl/#>, es decir, recursos y propiedades que se usará en el modelo de datos RDF de Stop_Words.

- **Clase Stop_Words:** Clase que define el modelo de datos RDF, especificando las tripletas RDF en función de los recursos y propiedades definidos en la clase Stop_Words. En esta clase se definen todas aquellas palabras que obstaculizan la búsqueda y que son ignoradas por la gran mayoría de motores de búsqueda cuando se utilizan en conjunto con otras.
- **Clase Main:** Clase que define toda la lógica de búsqueda en función de las clases Modelo y Stop_Words y los métodos OrdenaAlfabeticamente(), LimpiaFrase(), BuscaResultados() y Transformar(). Este último, se encarga de transformar el archivo XML, generado en el método BuscaResultados(), con un archivo XSL definido de acuerdo a los requerimientos de interfaz establecidos inicialmente.

4.4.2 Implementación de la Interfaz Web

Para la aplicación Web, se implementó una página JSP (página Web con fragmentos de HTML y código Java). Esta página instancia las clases desarrolladas para el “core” de la aplicación buscador a través de la definición de JavaBeans y se ejecuta en el contenedor Web, Tomcat 6.0.

4.4.3 Interfaz de Usuario

A continuación se muestran las pantallas de la interfaz de usuario involucradas al realizar una búsqueda en la aplicación.

La primera pantalla a la que se enfrentará el usuario para poder realizar su búsqueda es la mostrada en la figura 29.

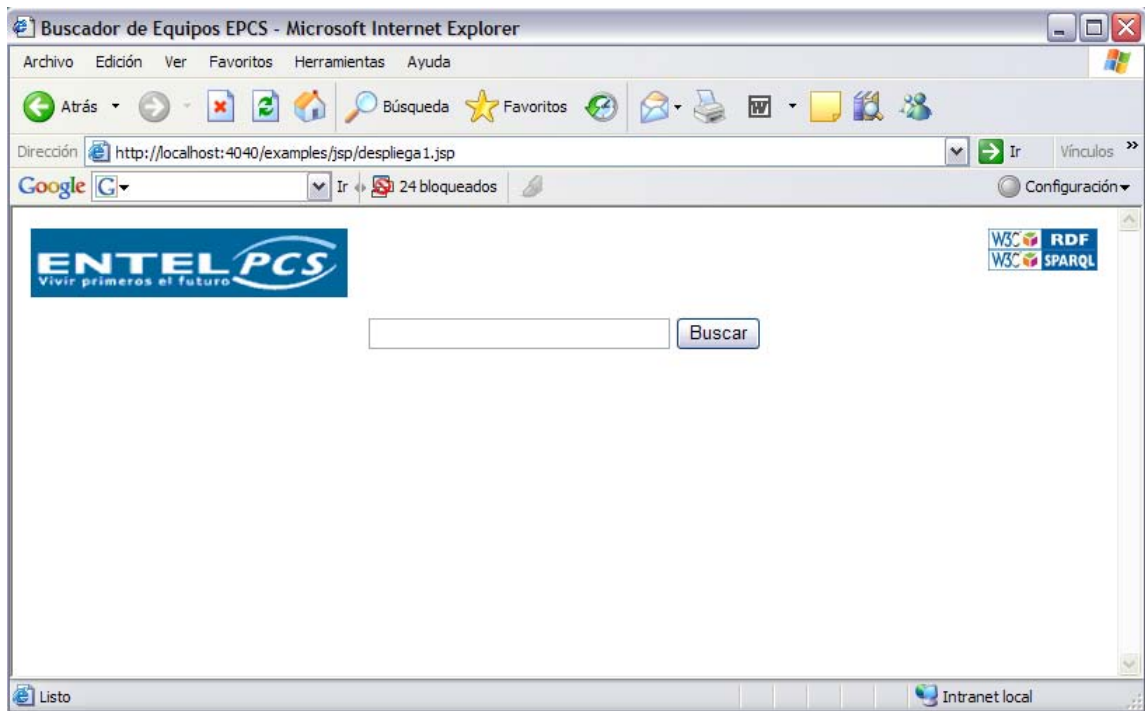


Figura 29: Interfaz inicial del buscador

A través de esta interfaz, el usuario podrá ingresar en lenguaje natural lo que desea buscar. Por ejemplo, si el usuario desea encontrar todos los celulares que tengan sonido midi, puede escribir en el textbox algo como esto:

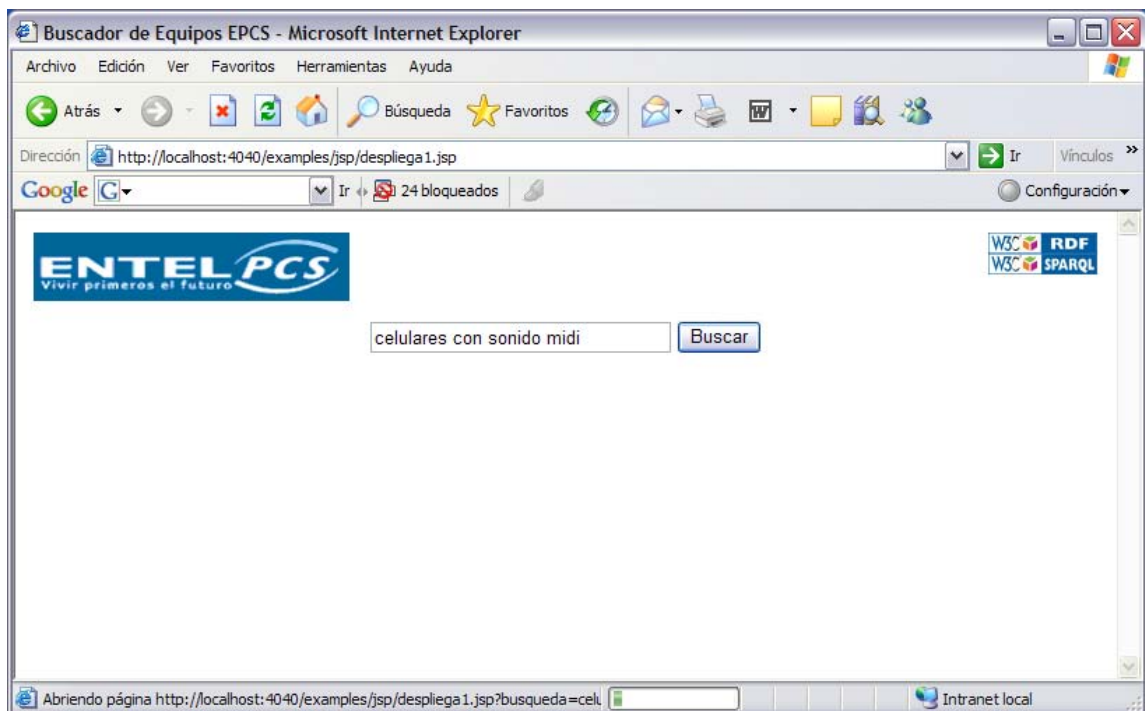


Figura 30: Ejemplo de búsqueda

Una vez que el buscador encuentre lo que el usuario busca, se desplegará en pantalla los resultados. Estos resultados mostrarán la marca del equipo junto con su modelo.

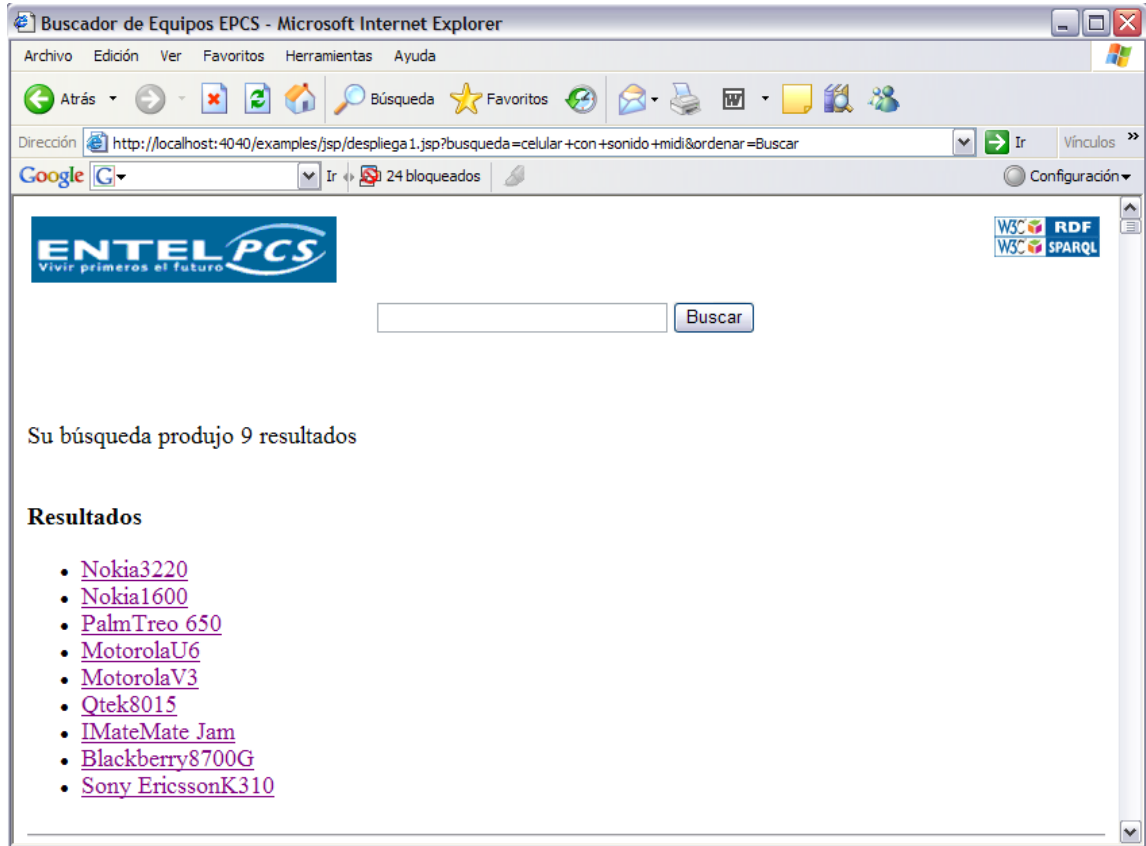


Figura 31: Interfaz de usuario que muestra resultados de la búsqueda

El usuario podrá elegir la alternativa que desee, haciendo clic sobre el link. Y el sistema responderá con la descripción en detalle del equipo escogido. Si por ejemplo, el usuario desea ver la especificación del equipo Nokia 3220, la interfaz desplegada al usuario será la que sigue.

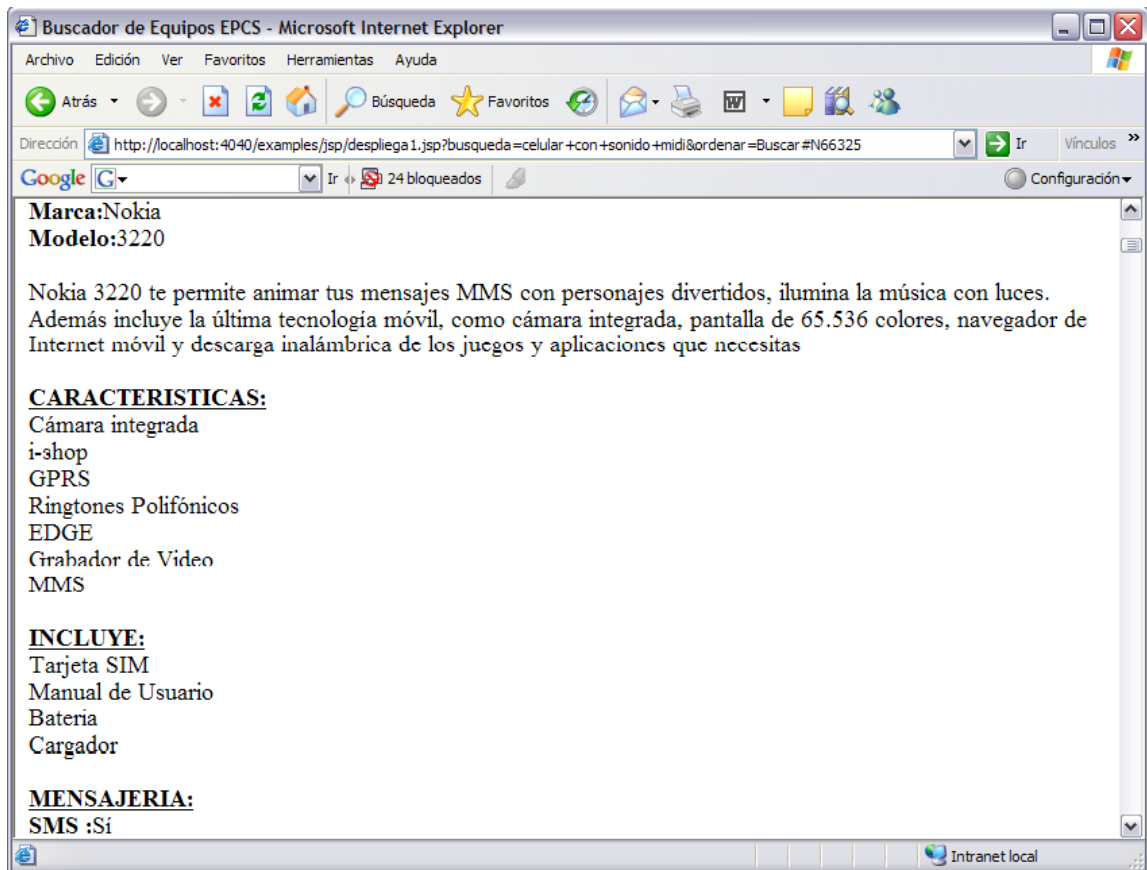


Figura 32: Descripción del equipo móvil escogido por el usuario.

4.4.4 Tiempos de Respuesta

Una vez que se implementó el buscador de equipos móviles, se evaluaron los tiempos de respuestas que este tiene en relación a la cantidad de equipos almacenados en el modelo RDF.

4.4.4.1 Hardware

Antes de realizar el análisis es necesario describir cuáles son las características de hardware sobre el que se realizaron estas mediciones, debido a que según las características de éste, los tiempos de respuesta de la aplicación varían.

Característica	Descripción
CPU	AMD Mobile Sempron 2800+ , 1.6GHz
Memoria	512 MB
Marca	Compaq
Modelo	Presario M2217LA
Sistema Operativo	Microsoft Windows XP Profesional versión 2002

Figura 33: Tabla de Características del hardware utilizado

4.4.4.2 Software

En relación a las herramientas software utilizadas para la implementación de este prototipo de servicio, se describe la siguiente tabla:

Característica	Descripción
Java	J2SDK 1.6.0
Jena	Versión 2.5.4
Tomcat	Versión 6.0

Figura 34: Tabla de características de software utilizado

4.4.4.3 Almacenamiento de la Información

Es importante mencionar que para el desarrollo de la aplicación se utilizó un almacenamiento en memoria del modelo RDF, es decir, el modelo RDF no se encuentra

almacenado de forma persistente⁵⁵ en una base de datos. Una vez que la aplicación requiere de la información contenida en el modelo, se instancia la clase que lo define y este se aloja en memoria mientras se realiza la consulta.

4.4.4.4 Medición de los Tiempos

Los datos obtenidos, se obtuvieron midiendo los tiempos desde que el cliente hace clic en buscar hasta que los resultados son desplegados en pantalla y se presentan en la tabla de continuación. En la primera columna se especifica el número de equipos móviles contenidos en el modelo RDF, en la segunda columna el número de resultados obtenidos por el buscador y en la tercera columna el tiempo en segundos requerido para encontrar lo solicitado en la búsqueda.

Para realizar esta prueba, se ingresó la siguiente consulta: Celular con midi

Equipos	Resultados	Tiempo (segundos)
5	5	1.734
10	9	2.422
15	14	2.969
20	18	3.234
25	23	3.781
30	27	4.172
35	32	4.672
40	36	4.953
45	41	5.156
50	45	5.425

Figura 35: Tabla para prueba de consulta

Gráficamente, lo anterior se representa como sigue:

⁵⁵ Sinónimo de durabilidad y permanencia

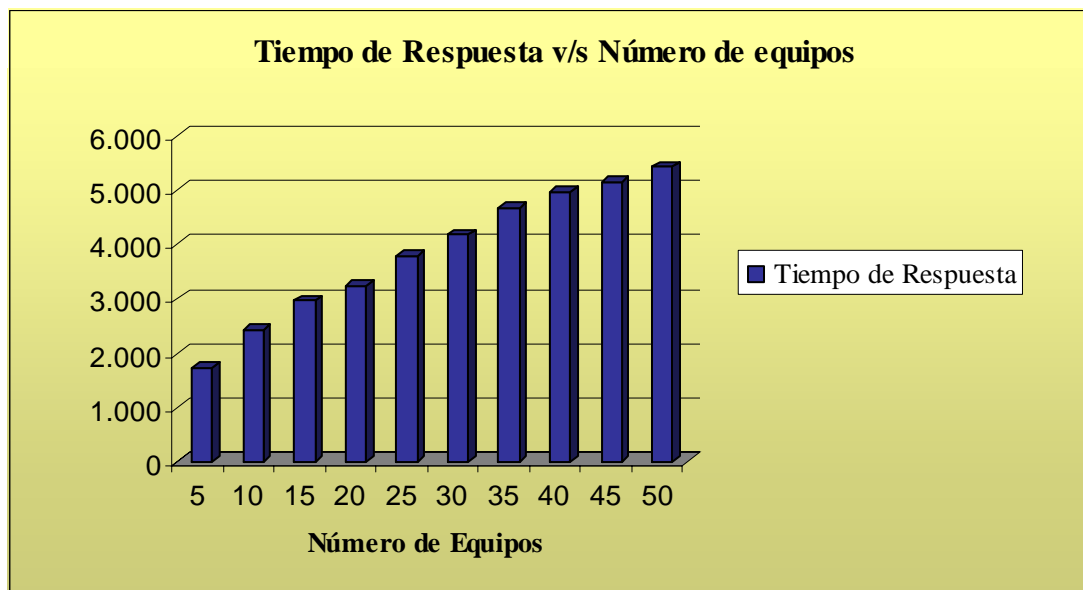


Figura 36: Gráfico Tiempo de respuesta v/s Número de equipos.

Considerando el hardware, el software y la forma de almacenamiento usados en la implementación del prototipo de servicio y a partir de los datos observados en el gráfico de la Figura 36, se puede concluir que la aplicación no es escalable, pues a medida que se aumenta la cantidad de equipos móviles a consultar, los tiempos de respuestas crecen, en vez de mantenerse en algún punto o comenzar a descender. Lo anterior es claramente una desventaja si el objetivo de la implementación del buscador es mejorar la calidad del servicio entregado al cliente y si a ello se suma que el número de dispositivos móviles ofrecidos por las operadoras aumentan día a día, dada la inminente evolución tecnológica.

A pesar que no se usó una Base de Datos para la implementación del buscador, es importante mencionar que su uso ya es factible en aplicaciones desarrolladas con tecnologías de la Web Semántica. En el caso de Jena, el componente SDB, permite el almacenamiento RDF y consultas SPARQL. Este almacenamiento es proporcionado por una base de datos SQL, la que actualmente puede ser, Oracle 10g, Microsoft SQL Server 2005, DB2 9, Postgre SQL, MySQL, Apache Derby, H2 y HSQLDB.

Obviamente, el almacenamiento persistente en bases de datos supone grandes ventajas sobre el almacenamiento en memoria y el almacenamiento tradicional en sistemas de archivos y esto porque a diferencia de estos dos últimos tipos de

almacenamiento, una base de datos o mejor dicho un DBMS⁵⁶ presenta las siguientes ventajas:

- Mejora la **integridad** de los datos, en cuanto a la validez y consistencia de los datos almacenados.
- Mejora en la **seguridad**, en cuanto a la protección de la base de datos frente a usuarios no autorizados.
- Mejora en la **accesibilidad** de los datos, pues muchos DBMS proporcionan lenguajes de consultas o generadores de informes que permiten al usuario hacer cualquier tipo de consulta sobre los datos.
- Mejora en la **productividad**, pues el DBMS proporciona muchas funciones y herramientas que permiten bajar drásticamente los tiempos de desarrollo y aumentar la calidad del sistema desarrollado si son bien explotados por los desarrolladores.
- Mejora en el **mantenimiento** gracias a la independencia de los datos, pues los DBMS separan las descripciones de los datos de las aplicaciones, lo cual permite simplificar el mantenimiento de las aplicaciones que acceden a las bases de datos.
- Aumento de la **conurrencia**, pues la mayoría de los DBMS gestionan el acceso concurrente a la base de datos y garantizan que no ocurran problemas como pérdida de información o de integridad.
- Mejora en los **servicios de copias de seguridad y de recuperación ante fallos**, pues los DBMS funcionan de modo que minimizan la cantidad de trabajo perdido cuando se produce un fallo, al proporcionar una forma eficiente de realizar copias de respaldo de la información almacenada en los DBMS, y de restaurar a partir de estas copias los datos que se hayan podido perder.

Por lo anterior, si bien la aplicación desarrollada no es escalable, pues a medida que se aumenta la cantidad de equipos móviles a consultar, los tiempos de respuestas crecen, en vez de mantenerse en algún punto o comenzar a descender, el uso de base de datos probablemente mejoraría este comportamiento.

⁵⁶ Database Management System, Sistema de Gestión de Base de Datos.

5. Pautas

5.1 Respecto al Diseño

De acuerdo al prototipo de servicio implementado y al estudio de las capas que componen la arquitectura de la web semántica, es posible dar pautas que permitan a los desarrolladores entender cuándo es conveniente ocupar una u otra tecnología. Estas pautas se referirán exclusivamente a las capas más consolidadas de la web semántica, es decir, XML, RDF/RDFS, Ontologías y una menos consolidada como la capa lógica.

Partiendo por las primeras capas de la arquitectura de la Web Semántica, XML y RDF, es importante mencionar que en este nivel fundamental, lo único común entre estos dos mundos es el uso de esquemas XML de datos para caracterizar los objetos literales en una tripleta RDF y también, como se mencionó anteriormente, el formato serialización de RDF basado en XML.

RDF es ideal para la integración de la información heterogénea en la Web, pues como se explicó, consiste de un amplio conjunto de relaciones (tripletas) y debido al uso de URIs es muy fácil fusionar a la perfección conjuntos de tripletas, es decir, datos descritos en RDF dentro de una misma aplicación. Lo anterior marca una diferencia con XML, pues la combinación de diferentes jerarquías XML en una misma aplicación puede llegar a ser muy compleja, por lo que no es una herramienta fácil para la integración de datos.

Se pueden desarrollar aplicaciones semánticas usando únicamente RDF y basándose en la lógica de la aplicación del programa. Lo anterior sólo depende de la aplicación y los objetivos que esta tenga. Por otro lado, el hecho de utilizar sólo RDF no limita a la aplicación, a que posteriormente, pueda extenderse a ontologías e inferencia.

En relación a la capa de Ontologías, algunas aplicaciones pueden requerir usarlas, y estas pueden ser muy simples (describiendo sólo uno o dos conceptos) o más complejas (con varios miles de términos) dependiendo de los requerimientos y metas de la aplicación. Un ejemplo para el papel que las ontologías cumplen sobre la Web Semántica es ayudar a la integración de los datos cuando, por ejemplo, puedan existir

ambigüedades en términos usados en diferentes conjuntos de datos, o cuando un poco de conocimiento extra puede llevar al descubrimiento de nuevas relaciones. Un ejemplo general puede ayudar. Una biblioteca puede querer integrar datos provenientes de diferentes editoriales. Los datos pueden ser importados en un modelo RDF común. Sin embargo, una base de datos puede usar el término “autor”, mientras que otra puede usar el término “creador”. Para hacer la integración completa, debería agregarse a la data RDF Schema la descripción del hecho que una relación descrita como “autor” es la misma que “creador”. Esta pieza extra de información es, de hecho, una ontología, aunque sea extremadamente simple.

En la Web Semántica, las reglas al igual que las ontologías son usadas para expresar restricciones adicionales y relaciones lógicas entre recursos. Estas dos formas permiten que procedimientos automáticos, basados en los datos y en la información adicional expresada en una ontología o conjuntos de reglas puedan generar nuevas relaciones, es decir, permiten que se de paso a la Inferencia. Por lo tanto, si en los requerimientos de una aplicación, la definición de restricciones y relaciones lógicas entre recursos no basta, pues además es necesaria una capa de razonamiento, entonces se entiende que la aplicación debe extenderse a la capa lógica de la arquitectura de la web semántica. La capa lógica se enfoca al razonamiento, es más general que los metadatos o las ontologías y permite descubrir conocimientos y relaciones ocultos entre “marañas” de datos. Es necesario, entonces, un motor de inferencia o razonador que procese los metadatos del documento.

Otro tema importante en relación a la definición y uso de ontologías, es que la Web Semántica no plantea regirse por una gigante y predefinida ontología, pues obviamente esto no sería factible. En lugar de ello, la Web Semántica se construye a partir de pequeñas ideas a fines a comunidades que pueden encontrar acuerdo sobre un conjunto de términos. Típicamente las aplicaciones Web Semánticas mezclan ontologías desarrolladas por distintas comunidades en la Web, pues el objetivo de esta es poder compartir y reutilizar como sea posible, facilitando el hecho de poder compartir datos e integrarlos. Obviamente, todas las personas no se pondrán de acuerdo sobre cómo definir exactamente todos los conceptos. Por ejemplo, mientras la mayoría de las personas tienen un concepto bastante estándar de un “perro” o un “gato”, no todos pueden distinguir entre “escalar” o un “vector”. Cualquier aplicación informática que intenta estandarizar su ontología, necesariamente distorsionará lo que al menos algunas

personas entienden de algunos conceptos; como consecuencia, habrá desajustes ontológicos en partes de la Web diseñadas por diferentes personas. Sin embargo, las aplicaciones pueden interactuar sin tratar de lograr un consenso global.

Evidentemente, algunos de los vocabularios pueden llegar a ser más ampliamente conocidos y utilizados, como Dublin Core, FOAF, entre otros y lo más probable es que sean grandes organizaciones (estructuras gubernamentales, compañías multinacionales, etc) las que difundan más rápidamente sus ontologías con el objeto de hacerlas útil y exclusivamente compatibles con sus intereses a escala global.

5.2 Respetto a las Herramientas

En relación a las herramientas se puede afirmar que hoy la disponibilidad de estas no es ninguna razón para no desarrollar aplicaciones Web Semánticas. En general, la mayoría de las herramientas existentes, ya son de buena calidad. Herramientas para definir detalladamente la clasificación de las relaciones en la información o para casos más complejos como la definición de relaciones lógicas entre recursos y relaciones, herramientas de consulta de información, herramientas para extraer y agregar datos a fuentes de datos tradicionales, herramientas para el almacenamiento de a información y razonadores, son algunas de las herramientas que la Web Semántica hoy ofrece para poder implementar aplicaciones.

- **Respetto a los lenguajes de representación.**

Estos ya se encuentran definidos por la W3C y son: RDF para la definición de recursos, RDF-Schemas y OWL para la definición de ontologías. RDF-Schema y OWL tienen limitaciones para el razonamiento lógico. En este sentido, SWRL , es una propuesta (es proceso de estandarización por parte de la W3C) que trata de eliminar estas limitaciones. El lenguaje SWRL propone definir reglas (cláusulas de Horn⁵⁷) que permitan inferir nuevo conocimiento sobre los elementos pertenecientes a una ontología.

⁵⁷ Es una regla de inferencia lógica con una serie de premisas (cero, una o más), y un único consecuente.

Dentro de los lenguajes de ontologías se ofrecen variaciones para describir terminologías simples o complejas: RDF Schemas, o varios dialectos de OWL (OWL Lite, OWL DL, OWL Full). Estas tecnologías difieren en expresividad, pero también en complejidad. Las aplicaciones pueden elegir a lo largo de un rango desde RDF Schemas para representar el nivel más simple de ontología al OWL Full para una máxima expresividad.

De acuerdo a lo anterior existen editores que soportan las tecnologías RDF/RDFS y OWL como Protegé, Swoop, WebODE, NeOn toolKit y Altova que al ocultar la mayor parte de la sintaxis compleja, facilitan al desarrollador la declaración de sentencias RDF como la construcción de las ontologías.

Por otro lado, existe Jena. Un entorno para Web Semántica de código libre programado en Java. Este proporciona una API para extraer y escribir datos de un grafo RDF, definir ontologías en OWL y realizar consultas SPARQL. Proporciona varios razonadores internos además de dar la posibilidad de utilizar otros.

- **Respecto a los lenguajes de consultas**

Los lenguajes de consulta que hoy existen para la Web Semántica son varios, ya más de 8. Entre ellos, SPARQL, RDQL, RQL, SeRQL mencionados en el apartado “Lenguajes de Consulta” de este proyecto. Sin embargo, SPARQL es el estándar que la W3C, organización precursora en la idea de la Web Semántica, propone como tecnología de consulta para exponer los datos en la Web. Se plantea como “el lenguaje y protocolo de búsquedas que realmente corresponde a la Web Semántica”.

- **Respecto a herramientas para extraer y unir fuentes de datos**

Referente a herramientas para extraer y unir fuentes de datos tradicionales y así asegurar el intercambio de datos con otras fuentes, existe **convertidores SQL para RDF; GRDDL y RDFa**. El desafío es concentrarse en desarrollar herramientas que puedan “exportar” datos que residen en bases de datos relacionales, hojas de cálculo, páginas wiki o páginas Web tradicionales en la forma de RDF. Esto no significa que los datos deben ser **físicamente** convertidos en la forma RDF y almacenados en, por

ejemplo, un documento RDF/XML. En lugar de ello, se proponen procedimientos automáticos que puedan producir datos RDF simultáneamente como respuesta a, por ejemplo, consultas.

Exportar los datos desde una base de datos relacional a RDF es un de los trabajos que aún queda por hacer en la W3C. En general, existen métodos que convierten consultas RDF (por ejemplo, consultas SPARQL) en consultas SQL sobre la marcha; es decir, mira la base de datos relacional como un repositorio RDF cuando este es consultado por una herramienta RDF. Comienzan a emerger soluciones generales, pero aún hay trabajo por hacer.

GRDDL, es un mecanismo que extrae declaraciones RDF a partir de contenido XHTML y XML compatible, utilizando programas como transformaciones XSLT [W3C07]. Es decir, se presenta como un puente que permite extraer semántica formal a partir de documentos XML estructurados, pero sin semántica, o débilmente anotados (por ejemplo, mediante microformatos).

Mediante RDFa se pueden representar los datos estructurados visibles en las páginas Web (eventos en calendarios, información de contacto personal, información sobre derechos de autor, etc.), a través de unas anotaciones semánticas incluidas en el código e invisibles para el usuario, lo que permitirá a las aplicaciones interpretar esta información y utilizarla de forma eficaz [W3C08]. Es un conjunto de extensiones XHTML propuestas por W3C para introducir semántica en los documentos.

Un ejemplo de extracción de datos usando RDFa, GRDDL y SPARQL:

- Se desea establecer una reunión entre tres personas, que tienen publicados en sus sitios Web los calendarios de sus citas y eventos. Estos datos están expuestos en páginas XHTML de forma gráfica, pero además se incluye información en RDFa.
- Una herramienta nos permite extraer, mediante GRDDL, los datos de sus calendarios en un formato homogéneo y fácil de tratar (RDF), para poder procesarlo posteriormente.

- Se realiza una consulta sobre la disponibilidad de las personas para un cierto día a una hora concreta. Los datos consultados están en formato RDF y la consulta se podría realizar mediante SPARQL.
- La herramienta procesa y analiza el resultado obtenido, concluyendo si las personas están disponibles en el instante que se había elegido previamente.

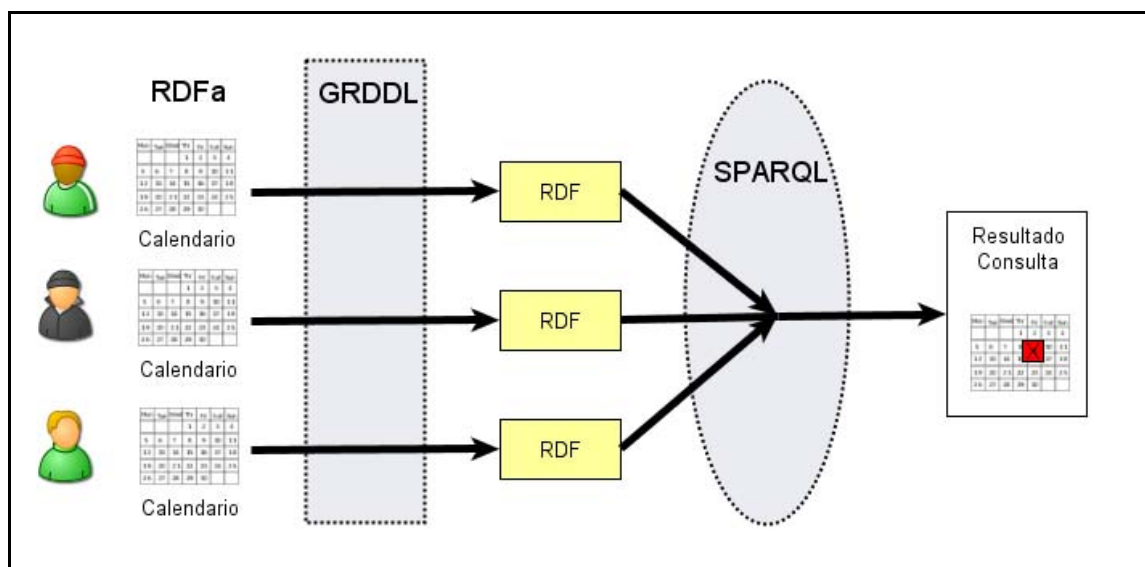


Figura 37: Ilustración del ejemplo de consulta de eventos de calendario.⁵⁸

El objetivo es reutilizar, en la medida que sea posible, datos existentes en su forma actual, y minimizar los datos RDF que tienen que ser creados manualmente.

- **Respecto al Almacenamiento**

Las tripletas representadas mediante RDF, bien provengan de un archivo RDFS, OWL, o cualquier otro derivado, se pueden almacenar de diversas formas, aunque la serialización más habitual se basa en la utilización de archivos XML. Este modo de representación presenta inconvenientes cuando se trabajan con volúmenes de información grandes, que sin embargo, es un punto donde precisamente las bases de datos relacionales llevan muchos años dando excelentes resultados. Es por esto que desde hace unos años existen diversos productos que permiten almacenar las tripletas RDF en bases de datos, principalmente relacionales, aunque también existen algunos productos con modos de almacenamiento específicos [Ech08].

⁵⁸ [W3C08]

Existen diferentes alternativas, como Jena, Sesame, 3Store, Kowari, y el reciente SDB de Jena. Esta última se ejecuta sobre bases de datos SQL y está diseñada específicamente para soportar el lenguaje de consulta SPARQL, desarrollado por la W3C. Las bases de datos que soporta son Microsoft SQL Server 2005 incluida SQL Server Express, Oracle 10gR2 incluida Oracle Express, IBM DB2 incluida DB2 Express, PostgreSQL v8, MySQL 5.0 (versiones iguales o superiores a 5.0.22), HSQLDB 1.8, H2 1.0.73 y Apache Derby 10.2 [Jen]. Una vez que la información está almacenada en la base de datos es posible acceder a ella a través del API de Jena, utilizado en este proyecto, o mediante la ejecución de consultas SPARQL vía línea de comando. Este componente de Jena proporciona un almacenamiento escalable y las herramientas de bases de datos para el balanceo de carga, seguridad, clustering, respaldo y administración pueden ser usadas para administrar la instalación.

- **Respecto a Motores de Inferencia**

Actualmente los principales razonadores semánticos toman como entrada una ontología, realizando un proceso de inferencia sobre ella. Ya existen varios, entre ellos, **Jena**, el framework utilizado en este proyecto, incluye un conjunto de razonadores tales como, OWL Reasoner, DAML Reasoner, RDF Rule Reasoner y Generic Rule Reasoner; **Bossam**, razonador para ontologías en OWL, en SWRL y reglas RuleML; **Hoolet**, razonador para ontologías OWL-DL; **Pellet**, razonador Java de código abierto para OWL-DL y que puede ser conectado directamente a Jena; **KAON2**, infraestructura de razonamiento para ontologías OWL-DL, SWRL Y F-Logic.

En resumen, se tiene lo siguiente:

Función	Tecnología /Herramienta
Lenguajes de representación	<ul style="list-style-type: none"> - RDF - RDFSCHEMA/OWL(Lite,DL, Full) - SWRL
Editores de RDF- RDFSCHEMA-OWL	<ul style="list-style-type: none"> - Protegé - Swoop - WebODE - NeOn toolKit - Altova

Framework Web Semántica	- Jena
Lenguaje de consulta	- SPARQL
Extracción y unión de fuentes de datos	- Convertidores SQL para RDF - GRDDL - RDFa
Almacenamiento	- Jena - SDB de Jena - Sesame - 3Store - Kowari
Razonadores	- Jena (OWL Reasoner, DAML Reasoner, RDF Rule Reasoner, Generic Rule Reasoner) - Hoolet - Pellet - KAON2

Figura 38: Tabla de Tecnologías/Herramientas según su función.

6. Conclusiones

El presente proyecto de tesis, permitió identificar el conjunto de estándares que conforman el marco de la tecnología Web Semántica, conjunto que si bien, hoy sigue siendo desconocido para muchos, para la W3C y sus seguidores no lo es. Junto con esto se plantean una serie de recomendaciones, dejando en claro que la Web Semántica no es una Web radicalmente diferente a la actual, con estándares distintos incompatibles a esta. Más bien, es una Web que pretende incorporar semántica a la Web existente y motivar, por otro lado, a los desarrolladores a que sus nuevas implementaciones se hagan utilizando esta tecnología.

Luego de la identificación y estudio de los estándares que propone la tecnología Web Semántica y por otro lado, el análisis exhaustivo de la evolución de la telefonía móvil en los últimos años en el mundo y en Chile, se analizó, diseñó e implementó un prototipo de buscador móvil basado en un modelo RDF. Un prototipo de servicio que pretende diferenciarse de la competencia nacional, hoy principal propósito de las operadoras móviles, ofreciendo una mejor experiencia de usuario con el objetivo de colaborar en la fidelización de los clientes con su operadora y evitar la tentación de estos a cambiarse a la competencia.

Finalmente, se elaboraron pautas para la implementación de sistemas que empleen Web Semántica. Si bien se encuentran explicadas de manera global a lo largo del proyecto de tesis, se detallan de forma clasificada al final de este. Lo que hace posible, tener una visión más clara y precisa de las tecnologías a utilizar, en relación al diseño y a las herramientas.

Desde que se planteó la idea de la Web Semántica en el año 2001, han pasado muchos años, tampoco existe total consenso acerca de los alcances y lo que realmente significa. Lo que si está claro es que la arquitectura de estándares que conforman esta tecnología ya es lo suficientemente consensuada como para que sea utilizada de manera global. Junto con ello, el desarrollo de las herramientas utilizadas para el despliegue de esta Web ha sufrido un gran avance.

No cabe duda que la propuesta de evolucionar la Web actual a una Web Semántica, es muy interesante, pero bastante ambiciosa al pensar en la profundidad y escala a la que esta se quiere aplicar, la idea de establecer un marco a nivel mundial supone un consenso global en la adopción de las tecnologías que esta futura Web propone. La dificultad de esta idea ha provocado que este concepto se haya ido posponiendo muchas veces, hasta el punto de levantar una cierta incredulidad en la comunidad científica.

Hasta ahora, se han desarrollado muchos proyectos en relación a esta tecnología. Empresas como *Oracle*, *IBM*, *Adobe*, *Software AG* son sólo algunas de las grandes corporaciones que han recogido esta nueva idea. Así también, empresas del sector de las telecomunicaciones como Telefónica I+D, que trabaja en el ámbito de la movilidad desarrollando el concepto de Semantic Mobile Web; Vodafone R&D con su portal Vodafone Live⁵⁹ y Nokia, con su portal⁶⁰ de Internet para sus proyectos de software Open Source, no son la excepción y ya han iniciado diversas líneas de investigación.

Se han desarrollado buscadores semánticos, como *Powerset*⁶¹, lanzado con grandes expectativas durante el año 2008 y que hoy al realizar pruebas comparativas con *Google* se comprueba que, al menos ahora, la utilización de tecnologías semánticas no mejora la calidad de las respuestas con respecto a los resultados ofrecidos por este último. Sin duda alguna, el campo de los buscadores semánticos es el que más expectativas ha generado en los últimos años, dado al rápido incremento en la cantidad de información en la Web, y es por eso, que en esta línea se concentran mayormente los desarrollos con esta nueva tecnología.

Los resultados que se han alcanzado hasta la fecha en la adopción universal de la Web semántica son muy preliminares. Lo que la Web Semántica necesita para prosperar y lograr su objetivo final, es el desarrollo de aplicaciones **reales** que pongan en práctica los principios que esta tecnología propone. Entre ellos, lograr una Web dotada de

⁵⁹ Portal Móvil de Vodafone: integra descripciones de contenidos (juegos, imágenes, ringtones, etc) de múltiples proveedores vía RDF y con el que se han incrementado de manera significativa el número de descargas

⁶⁰ Portal que está construido bajo uno de sus mismos proyectos, una ambiciosa arquitectura de la web semántica Open Source basada en URIQA (URI Query Agent, es un modelo para descubrir conocimiento de diversas fuentes basado en el uso de Uris, los cuales definen recursos.)

⁶¹ <http://www.powerset.com/> Buscador Semántico, que trabaja sobre Wikipedia. Lo que permite trabajar sobre un entorno infinitamente más homogénea y controlable que la Web en su conjunto.

estructura, con una semántica explícita que permita a las máquinas ser capaces de analizar todos los datos – contenidos, enlaces y transacciones entre la gente y los computadores-, una Web tratada como una base de datos global, una Web más inteligente, intuitiva, eficaz y hasta con mayor sentido común que permita reducir la mediación de operadores humanos en los procesos inteligentes de flujo de información.

Si bien la Web semántica aún no está asentada a nivel global por las limitaciones que aún presenta, con casi toda probabilidad dentro de la próxima década, será el futuro de la Web, por lo que es importante que las empresas sigan prestando interés a esta nueva propuesta. Mientras tanto y durante los próximos años las tecnologías semánticas irán penetrando en casi todos los ámbitos de los sistemas de información, habrá muchas soluciones parciales que convivirán, pero que no se podrán extrapolar a algo más global. Lo que si es seguro es que no habrá una ontología que “entienda” todo y que logre manejar toda la Web.

7. Referencias

[W3C08] W3C World Wide Web Consortium, (2008). Guía Breve de Web Semántica. Disponible en <http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica>.

Consultado el 19 de Mayo del 2007.

[Cas] Castells P. (n.d.). *La Web Semántica*. España: Universidad Autónoma de Madrid.

[Ber00] Berners-Lee T., (2000). Semántic Web – XML 2000. Disponible en <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html> . Consultado el 15 de Noviembre del 2007.

[Can06] Cantera J, Hierro J & Romo P. (2006). *Web Semántica: Tecnologías y Arquitectura*. España: Comunicaciones de Telefónica I+D.

[Lam08] Lamarca M., (2008). Hacia la Web Semántica. Disponible en http://hipertexto.info/documentos/web_semantica.htm. Consultado el 8 de Noviembre del 2007.

[Gru93] Gruber T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*. 5,199-220. London: Academic Press Ltd.

[W3C04] W3C World Wide Web Consortium, (2004). El Consorcio World Wide Web publica las Recomendaciones RDF y OWL. Disponible en <http://www.w3c.es/prensa/2004/nota040210.html> . Consultado el 5 de Noviembre del 2007.

[Lam08b] Lamarca M., (2008). Hacia la Web Semántica. Disponible en <http://hipertexto.info/documentos/dtds.htm>. Consultado el 8 de Noviembre del 2007.

[W3C04b] W3C World Wide Web Consortium, (2004). RDF Primer. Disponible en <http://www.w3.org/TR/rdf-primer/>. Consultado el 8 de Noviembre del 2007.

[W3C08b] W3C World Wide Web Consortium, (2008). W3C Semantic Web Frequently Asked Questions. Disponible en <http://www.w3.org/RDF/FAQ>. Consultado el 23 de Mayo del 2008.

[Abi07] Abián M., (2007). La Web Semántica: Además de metadatos (ontologías), lógica y confianza. Disponible en <http://www.wshoy.sidar.org/index.php?2007/01/30/37-la-web-semantica-metadatos-ontologias-logica-y-confianza>. Consultado el 20 Junio del 2007.

[Día03] Díaz P., (2003). Problemática y Tendencias en la Arquitectura de Metadatos en Web. *Anales de Documentación*, 006, 35-58. España: Universidad de Mursia.

[Cod06] Codina L & Roviro C., (2006). La Web Semántica. *Tendencias en Documentación Digital*, 9-54. Guijón: Trea

[Sam05] Samper J., (2005). Ontologías para Servicios Web Semánticos de Información de Tráfico: Descripción y Herramientas de Explotación. España: Universidad de Valencia.

[W3C08c] W3C World Wide Web Consortium, (2008). El W3C expone los datos en la Web con SPARQL. Disponible en http://www.w3c.es/Prensa/2008/nota080115_sparql. Consultado el 14 de Marzo del 2008.

[FOA] FOAF Project., (n.d.). The Friend of a Friend (FOAF) project. Disponible en <http://www.foaf-project.org/index.html>. Consultado el 13 de Noviembre del 2007.

[FOA04] FOAF Project., (2004). An Introduction to FOAF Danbri EDIT. Disponible en <http://www.foaf-project.org/2004/us/about.html> . Consultado el 13 de Noviembre del 2007.

[Per06] Pérez M., (2006). ¿Para qué sirve el RSS?. Disponible en http://es.geocities.com/rss_guia_facil/para_que_sirve_rss.html . Consultado el 15 de Noviembre del 2007.

[Mic04] Microsiervos., (2004). ¿Qué es RSS Y XML, RDF, ATOM,...?. Disponible en <http://www.microsiervos.com/archivo/internet/que-es-rss-y-xml-rdf-atom.html> . Consultado el 15 de Noviembre del 2007.

[Mic] Microsiervos., (n.d.).¿Qué es un feed RSS?. Disponible en <http://www.microsiervos.com/archivo/internet/que-es-un-feed-rss.html>. Consultado el 15 de Noviembre del 2007.

[Lam08b] Lamarca M., (2008). Metadatos Dublin Core. Disponible en http://www.hipertexto.info/documentos/dublin_core.htm. Consultado el 16 de Noviembre del 2007.

[Pol05] Polo L & Berruela D., (2005). *FOAF, RSS y Dublin Core: Los Vocabularios de la Web Semántica*. España: Fundación CTIC, Departamento de I+D+i.

[Ora07] Fundación Orange.,(2007). *Informe Anual sobre el Desarrollo de la Sociedad de la Información en España*. España: Fundación Orange.

[Gob07] Gobierno de Chile. Subsecretaría de Telecomunicaciones., (2007). Número de abonados móviles y Penetración cada 100 habitantes. Disponible en http://www.subtel.cl/prontus_subtel/site/artic/20070212/pags/20070212182348.html . Consultado el 16 de Junio del 2008.

[Min06] Ministerio de Economía, Fomento y Reconstrucción., (2006). Indicadores de Regulación. Servicios de Utilidad Pública, 2005. Chile: Ministerio de Economía.

[Tel07] Telefónica.,(2007).Informe RC 2006. Disponible en http://www.telefonicachile.cl/rc2006/telefonica/informe_rc/pagea02d.html?context=2886287&instanceid=2886288&lang=es .Consultado el 10 de Junio del 2008.

[Acc06] Acción RSE., (2006). Movistar. Disponible en <http://www.accionrse.cl/app01/home/DesCaso.asp?id=86>. Consultado el 10 de Junio del 2008.

- [ENT] ENTEL PCS., (n.d.). Visión/Misión. Disponible en <http://www.entelpcs.cl/compania/mision.iws> . Consultado el 10 de Junio del 2008.
- [CLA06] Claro Chile., (2006). Misión/ Visión. Disponible en <http://www.clarochile.cl/interior/quienes/mision.php> . Consultado el 10 de Junio del 2008.
- [Jav] Java., (n.d.). Conozca más sobre la Tecnología Java. Disponible en <http://www.java.com/es/about/> . Consultado el 19 de Junio del 2008.
- [Osm08] Osmosis Latina., (2008). JSP (Java Server Page). Disponible en <http://javaweb.osmosislatina.com/curso/jsp.htm> . Consultado el 19 de Junio del 2008.
- [Osm08b] Osmosis Latina., (2008). Servlets. Disponible en <http://javaweb.osmosislatina.com/curso/servlets.htm> . Consultado el 19 de Junio del 2008.
- [SID01] Sídár Traducciones.,(2001). Resource Description Framework (RDF). Especificación del Modelo y la Sintaxis. Recomendación del W3C 22 de Febrero del 1999. Disponible en <http://www.sidar.org/recur/desdi/traduc/es/rdf/rdfesp.htm> . Consultado el 23 de Junio del 2008.
- [W3C07] W3C World Wide Web Consortium, (2007). Introducción a GRDDL: Nota de Grupo de Trabajo. Disponible en <http://www.w3c.es/Noticias/2007/06/28/introduccion-a-22/> . Consultado el 16 de Julio del 2008.
- [Ech08] Echarte F., (2008). Almacenes de Tripletas RDF. Disponible en <http://www.eslomas.com/index.php/archives/2008/05/20/almacenes-de-tripletas-rdf/> . Consultado el 21 de Julio del 2008.
- [Jen] Jena., (n.d.). SDB- A SPARQL Database for Jena. Disponible en <http://jena.sourceforge.net/SDB/> . Consultado el 22 de Julio del 2008.