

**UNIVERSIDAD GABRIELA MISTRAL
FACULTAD DE INGENIERIA**

**APLICACIÓN DE WEB SEMANTICA PARA
FACILITAR LA GESTION DE RECURSOS
HUMANOS EN TESORERIA GENERAL DE LA
REPUBLICA**

Memoria para optar al título de Ingeniero de Ejecución en Informática

Autor : Raúl Geordy Pérez Peredo.
Profesor Guía : Jorge Tapia C.
Profesor Integrante : Roberto Carú C.

Santiago – Chile

Mayo, 2009

INDICE

I. INTRODUCCION	4
1.1 Motivación	7
1.2 Hipótesis	7
1.3 Objetivo General	7
1.4 Objetivos Específicos	8
1.5 Alcances.....	8
1.6 Antecedentes	8
II. MARCO TEORICO.....	9
2.1 Semántica	10
2.2 ¿Qué es la Web Semántica?	11
2.3 Un poco de Historia.....	14
2.4 Concepto Esencial para La Web Semántica	16
2.4.1 Ontologías	16
2.5 Procesamiento del Lenguaje Natural	22
2.6 Cómo Funciona la Web Semántica	25
2.7 Ventajas y Desventajas de la Web Semántica.....	27
2.8 La Web Semántica Hoy.....	27
III. APLICACIÓN DEL CONCEPTO SEMANTICO	29
3.1 Descripción del problema	29
3.2 Contexto de la investigación	30
3.3 Lógica del Negocio.....	35
3.4 Situación Actual	38
3.5 Aplicación del concepto.....	41
3.5.1 Primeros pasos hacia el objetivo	41
3.5.2 Tecnología para la aplicación	43
3.5.3 La aplicación.....	44
3.5.4 Uso del concepto Semántico	49
3.5.5 Ventajas y Desventajas	53

IV. CONCLUSION	55
V. GLOSARIO	57
VI. BIBLIOGRAFIA.....	59

I. INTRODUCCION

En muy pocos años la web se ha convertido en una herramienta de uso cotidiano en nuestra sociedad, en comparación con otros medios tan importantes, como lo son la radio, la televisión o el teléfono, por sobre los cuales posee algunas grandes ventajas en muchos aspectos. Hoy, la Web es un medio extraordinariamente flexible y económico en ámbitos como la comunicación, el comercio, el entretenimiento, el acceso a los servicios y a la información, la difusión de la cultura, etc.

Paralelo al crecimiento acelerado de la web, las tecnologías que la hacen posible han experimentado un rápido desarrollo, estas permiten una web mejor, potente, amplia, flexible y fácil de mantener.

Toda esta evolución influye y es afectada al mismo tiempo por la propia transformación de lo que hoy conocemos como la WWW¹.

Algunas de las tendencias evolutivas más notorias de los últimos años son:

- La generación dinámica de páginas Web.
- La conexión con Bases de Datos.
- La mayor interactividad con todo tipo de usuarios.
- El aceptar a la web como una plataforma universal para el acceso a la información.
- La adaptación al usuario.

¹ Forma abreviada para referirse a la web o al internet.

Esta evolución es continua en el tiempo, y es susceptible a mejoras. Entre las últimas tendencias que pueden repercutir en la web del futuro esta lo que se ha llamado la **web semántica** o **La Semántica aplicada a los datos**, conceptos inducidos por el inventor de la web y presidente del consorcio **W3C²**, **Berners Lee**.

La meta principal de la “**Web Semántica**” es lograr que las máquinas puedan “**entender**” la información existente en la internet y por lo tanto utilizar de una manera más óptima lo que esta contiene. Esta nueva internet estaría poblada por pequeños programas de software capaces de navegar por la internet y realizar algunos tipos de tareas por nosotros para ahorrarnos trabajo, tiempo y facilitar las búsquedas de información.

Para conseguir esta meta, la web semántica propone describir el contenido de la web con información entendible no solo por las personas, sino también para los programas que puedan llegar a existir, representar o reemplazar a las personas en tareas rutinarias o demasiado complejas para un ser humano. Las tecnologías involucradas en la web semántica buscan desarrollar una web más cohesionada, donde sea más fácil localizar, compartir e integrar información y servicios, para obtener un mejor resultado de los recursos disponibles en internet.

En la actualidad se piensa que la web del futuro será “inteligente”, es decir, que contará con información que las máquinas podrán “**entender**”, y a partir de la cual se podrán extraer un conjunto de datos que poseerán cierta unión lógica y de interés para los usuarios. Esto es una necesidad por cuanto la web actual está constituida por millones de páginas que implican demasiada información para manejar lo que hace tedioso el trabajo para los usuarios.

² Organismo internacional que define normas y reglas para regular los estándares de la web, necesarios para asegurar así el acceso universal a la información.

Los buscadores más usados en el mundo, entre ellos **Google, Yahoo y MSN**, si bien, en cierto modo permiten relacionar ciertas frases, no ofrecen aun búsquedas inteligentes que puedan adaptarse a las necesidades de los usuarios.

Los procedimientos actuales de búsqueda obedecen a criterios estadísticos, no a aquellos que el usuario intenta definir. Un ejemplo de esto es que en cualquier buscador puede encontrar miles de páginas que contengan palabras como **“Pérez”, “Chile”,** etc., pero no poseen aun la capacidad de resolver una petición de búsqueda como **“Busco a Raúl Pérez que vive en la calle Teatinos de Chile”**.

La capacidad que posee la internet actualmente, nos presenta ciertas limitaciones, un ejemplo de ello son los **motores de búsqueda**, que en este sentido, a pesar de que en la actualidad han mejorado mucho sus capacidades en los resultados de la información que nos entregan, son débiles en ciertos aspectos. Por ejemplo; si en algún buscador usted ingresa alguna de las siguientes frases **“¿Qué Presidentes ha tenido Chile en su historia?”** o **“¿Cuál es el nombre del General de Carabineros que murió en un accidente en Panamá?.”**, los resultados serán muchos y variados, lo que a la larga nos hará perder tiempo preciado muchas veces, en comenzar el trabajo de filtrar esta información.

Se explicará de la forma más clara posible, la aplicación del concepto existente en la **“Web Semántica”** en un sistema informático real y en funcionamiento, asociado al control de la Información de los funcionarios de La Tesorería General de la República, las ventajas que este concepto puede entregar al proceso de obtención y gestión de los datos, la tecnología básica involucrada, y la dinámica que entregará posterior a su implementación al sistema en cuestión. Este proceso se llevara a cabo sobre los datos ya existentes en las bases de datos de este sistema para lo cual deberá realizarse una estandarización para la definición del concepto semántico asociado a cada dato contenido en las bases de datos.

Todo lo anterior, con el gran objetivo de facilitar al usuario común y corriente, sin conocimientos avanzados de informática, la obtención y gestión de la información de forma más amigable.

1.1 Motivación

Hoy en día los métodos para la búsqueda de información en la mayoría de los sistemas son poco estandarizados, en ese contexto la motivación de esta investigación dice relación, en que la **Semántica aplicada a los datos**, supone superar las limitaciones en este sentido, mediante la introducción de descripciones explícitas de los datos, concepto que facilitará su obtención.

1.2 Hipótesis

La hipótesis de esta investigación pretende demostrar, que utilizando parte de los medios disponibles y **el concepto en sí**, de la semántica aplicada a los datos, se puede enriquecer la información asociada al Sistema de Gestión de Personal de la Tesorería General de la República.

1.3 Objetivo General

Dotar de semántica a los datos almacenados en el sistema de Gestión de Recursos Humanos de la Tesorería General de la República, esto, con el objetivo de facilitar el proceso de obtención de la información asociada a los funcionarios, durante los procesos de Selección de personal, o procesos de calificaciones y otros afines.

1.4 **Objetivos Específicos**

- Realizar una introducción a la realidad de la **web semántica**, su importancia en el internet de hoy, Experiencias, **concepto** y parte de la tecnología básica y esencial involucrada.
- Aplicar el **concepto semántico**, al Sistema Gestión de Recursos Humanos de la Tesorería General de la República y describir las ventajas que este concepto entregara.

1.5 **Alcances**

- Lograr el entendimiento del concepto **semántico**, su realidad actual, los ámbitos involucrados en su implementación.
- Finalmente se quiere lograr facilitar la **búsqueda y gestión de la información** para los usuarios no expertos, por medio de una aproximación al lenguaje natural de las personas.

1.6 **Antecedentes**

La División de Personal, es una de las unidades fundamentales, dentro de la Tesorería General de la República. A su cargo está la gestión y administración de los datos personales y laborales de todos los funcionarios pertenecientes al servicio. Es acá donde esta puesto el foco para la aplicación del concepto y la esencia de esta investigación, el de aplicar la semántica a la información existente en el Sistema de Gestión de Personal.

II. MARCO TEORICO

Actualmente es sumamente complicado medir el tamaño actual de la web, aunque se estima que es utilizada por unos **mil millones de usuarios**³, la cual está constituida por al menos **cuatro mil millones de páginas web**⁴. Estas cifras incluyen solo lo que se ha dado a conocer como la **Web Superficial**, constituida solo por páginas de **contenido estático**⁵, y la llamada **Web Profunda**, formada por páginas web cuya información que se despliega, es obtenida desde bases de datos.

En la actualidad, la información de casi todos los ámbitos de la sociedad se encuentra disponible en la web, y con la ayuda de un buen buscador se puede acceder a la información que se necesita. La web se ha convertido en una gran enciclopedia a nivel mundial, de todo tipo de conocimientos. Otro factor que la ha convertido en lo que es, es la facilidad para llevar a cabo diferentes actividades de la vida diaria, con la comodidad y economía que esto significa. Sin moverse del hogar, hoy se pueden llegar a realizar compras de productos y servicios de cualquier tipo, gestionar una cuenta bancaria, buscar locales de comida, matricularse en la universidad e incluso trabajar desde el hogar.

Pensando en todas las ventajas asociadas a la web de hoy en día, aún queda mucho espacio para mejoras. El enorme tamaño que ha alcanzado internet, que es la principal clave para su éxito, ha transformado algunas tareas que años atrás podían resultar relativamente sencillas, en un trabajo que hoy consume demasiado tiempo, como por ejemplo, la búsqueda de información de algún tipo específico.

Desarrollar programas de software que puedan ejecutar tareas de búsqueda, es una labor muy complicada, ya que es muy difícil emular, y aun más, crear en una

³ fuente www.w3c.org.

⁴ fuente www.w3c.org.

⁵ Páginas de internet que solo contienen código HTML básico, y que no obtienen de bases de datos la información que despliegan por pantalla.

máquina la capacidad de un ser humano para comprender los contenidos de la información, y los datos alojados en la web, tal y como se encuentran actualmente.

Los buscadores de hoy en día a pesar de poseer una gran eficacia, también poseen limitantes al momento de realizar búsquedas con un nivel un poco más elevado de complejidad, es decir, ingresadas en un buscador con un lenguaje similar al natural de las personas.

Finalmente podemos decir que la falta de capacidad de los buscadores para poder expresar y almacenar el significado de los datos, es uno de los mayores problemas que se tienen para la obtención de la información.

2.1 Semántica

Lo primero que se debe conocer, es el significado del término que se empleara, es para eso es que a continuación se entrega un pequeño resumen de lo que significa el término semántica. Esto nos ayudara a comenzar a entender el concepto en sí.

Semántica; según la Real Academia Española: se refiere al estudio de los signos lingüísticos y de sus combinaciones.

Ampliando un poco más esta explicación, podemos mencionar que se refiere al estudio del significado lingüístico; palabras, expresiones y oraciones. Los que estudian la semántica tratan de responder a preguntas como por ejemplo, “¿Cuál es el significado de X palabra?”, para ello, deben estudiar qué signos existen y cuáles son los que poseen significación, que significan para las personas normales, como los designan, de qué forma se refieren a ideas y cosas. Finalmente es importante como lo interpretan los seres humanos. La finalidad de la semántica es definir el significado de los signos o palabras y establecer que significan dentro de un determinado contexto.

2.2 ¿Qué es la Web Semántica?

La **Web** ha cambiado profundamente la manera en que las personas se comunican, negocian y realizan sus labores a diario, hoy se puede tener acceso a infinidad de recursos, en gran variedad de idiomas y diferentes formatos. Estos factores han proporcionado el éxito que la web tiene hoy, y a su vez han ocasionado contradictoriamente uno de los problemas más importantes; el exceso de información.

De acuerdo a estudios, se dice que la **Web semántica** será una extensión de la Web actual, en la que a la información existente se le otorgará una semántica bien definida, permitiendo que sistemas informáticos, de distinto tipo, descubran y utilicen esa información disponible en la Web. El mecanismo que ya se utiliza para poder crear el conocimiento en la web semántica son las **Ontologías**⁶.

Explicando en forma simple que es la Web Semántica, tenemos el siguiente ejemplo; si quisiéramos buscar información sobre un conocido animador de televisión, lo normal sería introducir en cualquier buscador de internet el siguiente texto; “**Lagos Animador**”, sin embargo casi siempre ocurre que no obtenemos el resultado deseado. Continuando con este ejemplo podríamos mejorar mucho más la búsqueda si pusiéramos en el buscador el siguiente texto; “**Sergio Lagos Televisión**”, pero los resultados que entregue ¿a quién corresponderán?, a ¿**Sergio Lagos, el animador?**, ¿a **Juan Lagos?**, o a cualquier persona desconocida, cuyo apellido por coincidencia es igual al de esta persona.

El problema radica en que no está clara la **semántica de la búsqueda**, ya que los buscadores casi siempre realizan sus búsquedas en base a la coincidencia de las palabras, y no en base al significado de estas. En este sentido en el consorcio de la **W3C** pensaron en la siguiente solución; agregar información acerca de la información, es decir, **metadatos**. En otras palabras, poder especificar que cuando se habla de

⁶ Termina que será desarrollado en capítulos posteriores.

Lagos Animador, la palabra “**lagos**”, es el apellido de una persona quien es animador de Televisión y no una gran masa de agua rodeada de tierra.

¿Y cómo se puede hacer esto?, una manera es usando un lenguaje llamado **RDF**⁷. Como primer paso se necesita identificar la información disponible en la Web y saber si un determinado dato es una **cosa, una persona u otro elemento.**, esto es importante, volviendo al ejemplo anterior, para poder identificar a **Lagos (animador)** de **lagos (agua)**.

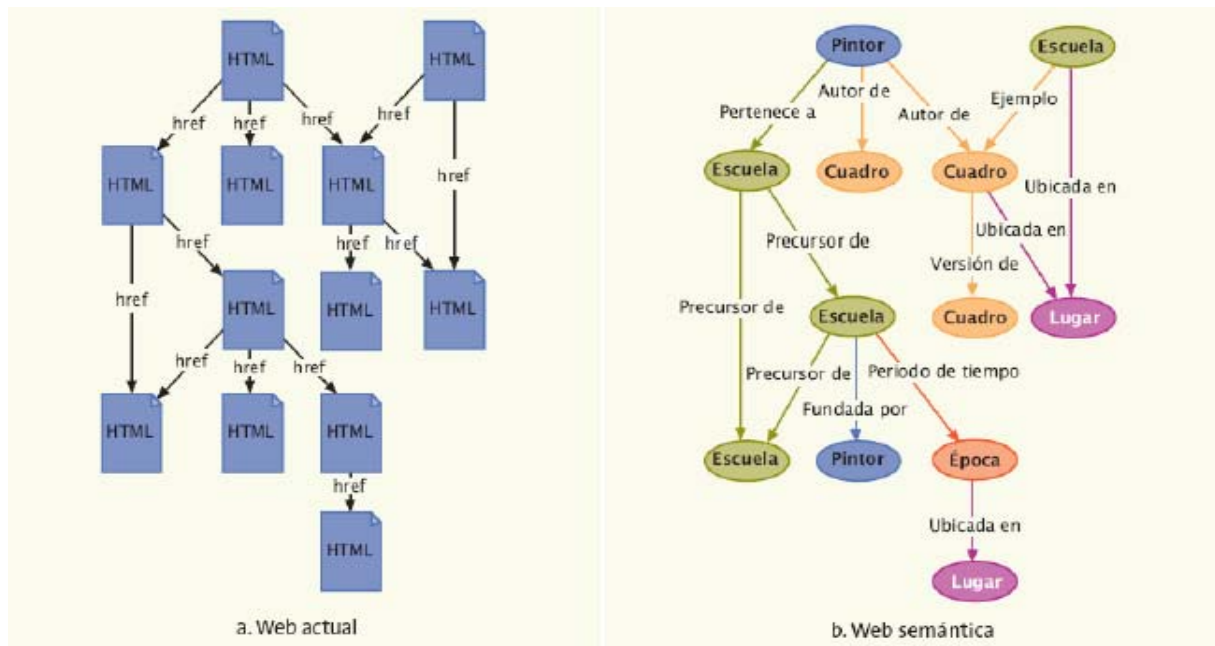


Figura 1: Comparación entre la Web actual y la Web Semántica.

¿Y cómo se puede hacer esto?, Cuando queremos decir algo o describir un dato o un recurso (**recordemos que un recurso puede ser una persona, un modelo de auto, o un edificio**), decimos que ese algo tiene cierta propiedad. Para describir los datos podemos usar identificadores los cuales usamos en nuestro lenguaje a diario, de

⁷ Método general para el modelamiento de la información, permite describir los datos en la web, proveyendo interoperabilidad entre las aplicaciones que intercambian información en lenguaje máquina por la web.

forma natural, como por ejemplo: “**su nombre es**” o “**es de color**”. Para llegar a tener algo así todavía falta algún tiempo para aplicarse a la información disponible en la Web, pero ya se están dando los pasos necesarios. La idea final es poder utilizar la web como algo parecido a una base de datos, donde la gente pueda realizar consultas un poco más complejas que las actuales.

2.3 *Un poco de Historia*

La Web semántica ha sido impulsada por el **Sr. Tim Berners Lee**, como se menciona, creador de la **WWW**, además de otras personas relacionadas con la **W3C**. El primer avance en este sentido fue la publicación en septiembre del año 1998, por parte del señor **Berners** de los documentos denominados: **Semantic Web Road Map** (El mapa carretero de la Web Semántica) y **What the Semantic can represent** (Lo que la semántica puede representar).

En el año 2000, el señor **Berners** ofreció una conferencia en el marco de la W3C, donde propuso: *“La nueva información debe ser reunida de forma que un buscador pueda “comprender”, en lugar de simplemente listar la información. La Web semántica sería una red de documentos “más inteligentes” que permitan, a su vez, búsquedas un poco más precisas. La idea sería aumentar la inteligencia de los contenidos de la información que hay en las páginas web dotándolas de contenido semántico.*

La Web actual posee una gran capacidad para almacenar datos y puede leer y visualizar los contenidos, pero no es capaz de “pensar”, ni de “entender”, todo lo que contiene”. Por lo anterior es necesaria una ampliación de ella, la **Web Semántica**, que hará posible no solo almacenar los datos, sino que entender e interpretar el sentido de esta información. Es así como el Sr. **Berners** presenta la nueva forma en que se basará la **Web Semántica**.

En mayo del año 2001, los señores **Tim Berners Lee, James Hendler y Ora Lassila**, popularizan la idea de la web semántica, al publicar un artículo en la revista **Scientific American** titulado “**The Semantic Web**”, donde explican de una forma sencilla su idea de la web semántica y los primeros pasos que hay que dar para llevarla a cabo.

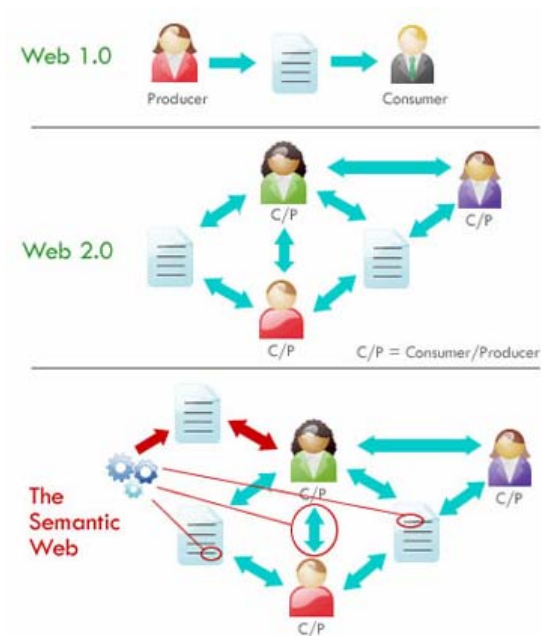


Figura 2: La Evolución de la Web.

2.4 Concepto Esencial para La Web Semántica

2.4.1 Ontologías

Con las ontologías los usuarios podrán ordenar u organizar la información de forma que los **agentes** (Pequeños programas) podrán interpretar el significado de la información, y por lo tanto, podrán buscar e integrar la información disponible en el internet.

Se espera que debido al “conocimiento” que podrán contener las ontologías, los programas puedan extraer automáticamente información de las páginas Web, procesarlos y extraer conclusiones de ellos.

Las ontologías provienen del campo de la inteligencia artificial; son vocabularios comunes para las personas y para los programas que trabajan en un área en común o contexto determinado. Según la **W3C** las ontologías definen los términos usados para describir ciertas áreas de interés. (Entiéndase como área de conocimiento por ejemplo: Física, aeronáutica, medicina, contabilidad, fabricación de productos etc.), las ontologías representan cierta visión del mundo con respecto a un área. Por ejemplo, una ontología que defina “**ser humano**” como “**espécimen vivo o muerto correspondiente a la especie Homo sapiens**”, define una perspectiva del mundo totalmente distinta a una ontología que lo defina como “**sujeto consiente y libre, centro de todo lo que existe**”.

Así como las **Ontologías** (ojo con la mayúscula inicial), estudia los tipos de objetos que pueblan nuestra realidad, las **ontologías**, clasifican y definen los tipos de cosas que existen en cierta área, así como sus relaciones y características. Por ejemplo, una ontología del mundo empresarial usa conceptos como **Venta, Compra, Transferencia, Pago** etc., y relaciones como “**Una transferencia corresponde a una**

Venta o a una Compra”, “Un pago corresponde a una o varias Transferencias”, etc.

Hoy en día existen ontologías específicas (de términos médicos, empresariales, etc.) como también ontologías de carácter general (que definen terminologías útiles para varias áreas).

Cualquier persona tiene en su cabeza ontologías, mediante las cuales entiende el mundo que nos rodea. Estas ontologías no se detallan en un documento ni se organizan de alguna forma, de manera natural, todos usamos ontologías. Por ejemplo, en las que un automóvil representa un medio de transporte y tiene 4 ruedas. ¿Concretamos u oficializamos de alguna forma estas ontologías? **NO**, no es necesario debido a que en el ejemplo recién explicado, los automóviles son tan comunes que todos compartimos la información de lo que son. Lo mismo nos sucede cuando pensamos en el área familiar, sabemos que una familia está compuesta por varias personas, que un hijo no puede tener más de un padre y una madre biológicos, que nuestros padres tienen o han tenido sus propios padres, por lo tanto, no es necesario que explicitemos nuestro conocimiento en este ámbito, pues forma parte de lo que todo el mundo sabe. Sin embargo, cuando se habla de términos poco comunes o cuando necesitamos que estos términos puedan ser procesados por máquinas, se precisa explicar las ontologías, es decir, desarrollarlas en un documento o darles una forma que pueda ser “**entendible**” por las máquinas.

Usando ontologías como metadatos

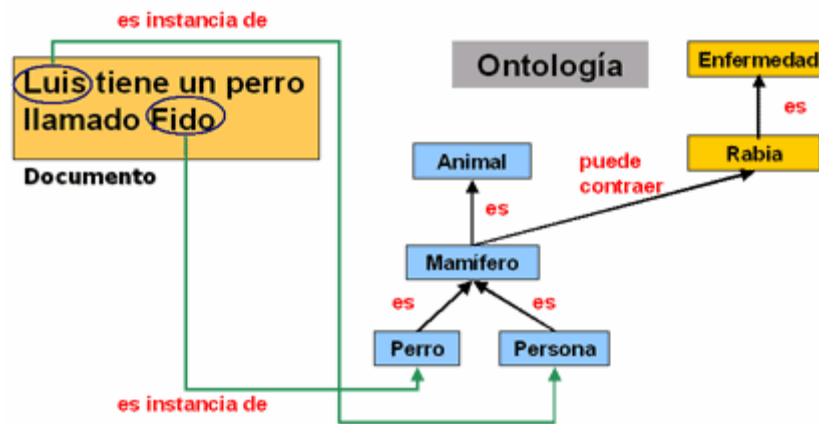


Figura 3: ontologías como Metadatos.

Las máquinas o computadores carecen de las ontologías con las que nosotros contamos para entender el mundo y poder comunicarse entre ellas, es por eso que necesitan que les explicitemos las ontologías. Cuando dos sistemas necesitan comunicarse, por ejemplo un **ERP**⁸, con un Sistema de bases de datos, aparecen de inmediato los problemas **semánticos** que dificultan o imposibilitan la comunicación entre ellos (no considerar problemas técnicos de conexión).

Los problemas semánticos son de dos tipos; **de Área** y **de Nombre**. Los conflictos de **Área** aparecen cuando conceptos similares en cuanto a significado, pero no idénticos, se representan en distintas áreas. Por ejemplo, el concepto representado por **Trabajador** en una base de datos puede corresponder a un trabajador bien calificado, mientras que en otra base de datos, puede usar el concepto **Trabajador** para identificar o referirse a cualquier tipo de trabajador, es decir, no bien calificado. Ambos conceptos están muy relacionados, pero no son equivalentes ni deberían mezclarse. Si usamos ontologías, podría especificarse que el primer concepto corresponde a una especialización del segundo, y un sistema de razonamiento automático basado en una

⁸ Sistemas integrales de gestión empresarial que están diseñados para automatizar la mayoría de los procesos en la empresa (finanzas, comercial, logística, producción etc.)

ontología, podría impedir que se contratara para labores que deben ser desempeñadas por un trabajador bien calificado a uno que no lo es.

Los conflictos de **Nombre** son de dos tipos; **sinónimos** y **homónimos**. Los **sinónimos** ocurren cuando los sistemas usan distintos nombres para referirse al mismo concepto. Por ejemplo, una Base de Datos puede usar **Trabajador** para el mismo concepto que otra usa **Empleado**. En este caso una ontología podría definir como idénticos los dos términos, así dos programas que deseen comunicarse sabrían como llevar datos de uno al otro.

Los **homónimos** aparecen cuando los sistemas usan el mismo nombre para referirse a conceptos distintos, Por ejemplo, para el software de una compañía de seguros, el **Conductor** representa a una persona que tiene contratada una póliza de seguros, mientras que para el software de una compañía de taxis, el **Conductor** representa a un trabajador que conduce un taxi. Como es obvio si se intentara **integrar automáticamente** ambos programas de software basándose en que ambos usan el mismo término (**Conductor**) para significar lo mismo, se produciría un problema muy grande; porque por ejemplo al dar de baja a un conductor de taxi se le quitaría su póliza de seguros, y al dar de alta un asegurado, se le daría de alta como taxista, aunque no tuviera licencia de taxista.

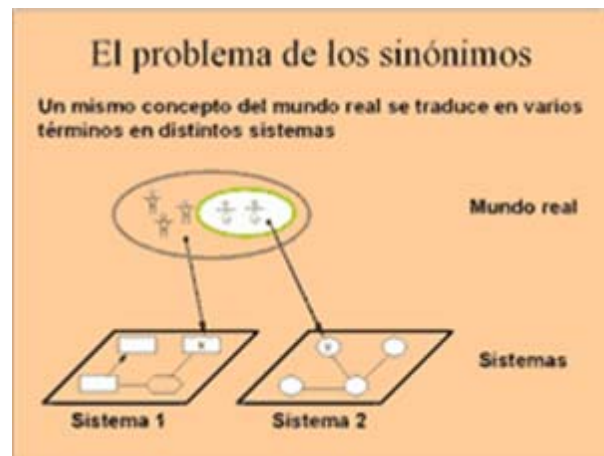


Figura 4: Problema de los sinónimos.

Las ontologías se pueden expresar de muchas maneras, como mínimo deben incluir un vocabulario de términos, con la definición de cada uno. Por ejemplo, una ontología para el mundo empresarial define el concepto **Venta** de la siguiente manera: **“Una venta es un acuerdo entre dos Entidades Legales para el intercambio de un Producto o Servicio por un precio de venta”**.

Las ontologías favorecen la comunicación entre las personas, empresas y sistemas debido a que proporcionan una comprensión común de un área específica, de modo que se eliminan confusiones conceptuales.

El mundo empresarial tampoco está ajeno a los problemas derivados de la falta de entendimiento común, algunas empresas usan el término **“recursos”**, para referirse a las máquinas, para otras en cambio los **“recursos”**, son las materias primas utilizadas en los procesos de producción. Mediante las ontologías, se puede favorecer la gestión de la información electrónica. Las ontologías también favorecen la comunicación entre los programas de software y el entendimiento común de la información entre ellos. Es importante mencionar que las ontologías serán imprescindibles en la aplicación de la semántica en la web y en los futuros sistemas de

gestión empresarial, pues permitirán que los programas de software estén de acuerdo en los términos que usan cuando se

comuniquen. Mediante las ontologías será mucho más fácil recuperar información relacionada.

Las ontologías también sirven para que **los sistemas interoperen**. Los sistemas son interoperables si pueden trabajar conjuntamente de una forma automática, con muy poco esfuerzo por parte del usuario. Por ejemplo, dos teléfonos celulares de distintas compañías interoperan para que los usuarios puedan hablar entre sí. En el área de la informática, las ontologías servirán para traducir los términos usados por distintos programas de software. Consideremos un programa de software empresarial que usa el término “**materia prima**”, y otro que utiliza el término “**suministro**”, ambos programas no podrían trabajar juntos. Para lograr la comunicación, una ontología tendría que hacer las veces de traductora entre ambos programas de software (la ontología tendría que usar el término “**recurso**”). La ontología tendría que actuar como un puente entre ambos sistemas, actuaría como una especie de lengua común.

Finalmente, podemos decir que en la base de la **Ingeniería de Software**, las ontologías ayudan a la especificación de los sistemas software, la falta de un entendimiento común nos lleva a dificultades en el proceso de identificar y especificar los requisitos del sistema que se desea desarrollar, las ontologías facilitan el acuerdo entre desarrolladores de sistemas y usuarios que utilizan estos sistemas.

2.5 *Procesamiento del Lenguaje Natural*

El “**Procesamiento del Lenguaje Natural**” (PLN) es una disciplina con larga trayectoria, nace en la década de 1960, como una sub área de la Inteligencia Artificial y la lingüística, con el objeto de estudiar los problemas derivados de la generación y comprensión automática del lenguaje natural.

El lenguaje natural, entendido como la herramienta que utilizan las personas para expresarse, posee propiedades que disminuyen la efectividad de los sistemas de recuperación de información textual. Estas propiedades son la **variación lingüística** y la **ambigüedad lingüística**. Cuando hablamos de la **Variación Lingüística** nos referimos a la posibilidad de utilizar diferentes palabras o expresiones para comunicar una misma idea. En cambio la **Ambigüedad lingüística** se produce cuando una palabra o frase permiten más de una interpretación.

Ambos fenómenos inciden en el proceso de recuperación de información aunque de forma distinta. La **variación lingüística** provoca el silencio documental, es decir, el no encontrar resultados relevantes que puedan cubrir la necesidad de información de los usuarios, ya que no han utilizado los mismos términos en los datos de búsqueda que los existentes en la información. En cambio la **ambigüedad lingüística** implica el ruido documental, es decir, se encuentran datos que no son significativos, ya que en los resultados de las búsquedas se puede encontrar el mismo termino buscado, pero con un significado diferente al requerido. Estas dos características dificultan considerablemente el tratamiento automatizado del lenguaje. A continuación se muestran algunos ejemplos que ilustran de mejor forma la ocurrencia de estos fenómenos en el proceso de recuperación de la información:

A **nivel morfológico** una misma palabra puede adoptar diferentes roles **morfo-sintácticos** en función del contexto en el que aparece, ocasionando problemas de ambigüedad (Ejemplo 1).

Ejemplo 1. ***Deja la comida que sobre, sobre la mesa de la cocina, dijo llevando el sobre en la mano.***

La palabra **sobre** es ambigua morfológicamente ya que puede ser un sustantivo masculino singular, una preposición, y también la primera o tercera persona del presente de subjuntivo del verbo sobrar.

A **nivel sintáctico**, centrado en el estudio de las relaciones establecidas entre las palabras para formar unidades superiores (frases y sintagmas), se produce ambigüedad a consecuencia de la posibilidad de asociar a una frase, mas de una estructura sintáctica. Por otra parte, esta variación supone la posibilidad de expresar lo mismo pero cambiando el orden de la estructura sintáctica de la frase. (Ejemplo 2).

Ejemplo 2. ***María vio a un niño, con un telescopio en la ventana.***

La interpretación de las dependencias de los dos sintagmas (frases) preposicionales, ***con un telescopio*** y ***en la ventana***, otorga diferentes significados a la frase: (1) María vio a un niño que estaba en la ventana y que tenía un telescopio, (2) María estaba en la ventana, desde donde vio a un niño que tenía un telescopio, y (3) María estaba en la ventana, desde donde miraba con un telescopio, y vio a un niño.

A **nivel semántico**, donde se estudia el significado de una palabra y el de una frase a partir de los significados de cada una de las palabras que la componen. La ambigüedad se produce porque una palabra puede tener uno o varios sentidos, es el caso conocido como **polisemia** (Ejemplo 3)

Ejemplo 3. *Luis dejó el periódico en el banco.*

El término banco puede tener dos significados en esta frase, (1) entidad bancaria y (2) asiento. La interpretación de esa frase va más allá del análisis de los componentes que forman la frase, se realiza a partir del contexto en que es formulada.

También hay que tener en cuenta la variación léxica que hace referencia a la posibilidad de utilizar términos distintos para representar un mismo significado, es decir, el fenómeno conocido como **sinonimia** (Ejemplo 4)

Ejemplo 4. *Coche / Vehículo / Automóvil*

A nivel pragmático, basado en la relación del lenguaje en el contexto en el cual es utilizado, en muchos casos no se puede realizar una interpretación literal y automatizada de los datos en determinadas frases. El sentido de las palabras que forman una frase debe interpretarse en un nivel superior recurriendo al contexto en el que es elaborada. (Ejemplo 5)

Ejemplo 5. *Se moría de risa.*

En esta frase no puede interpretarse literalmente el verbo morir, si no que debe entenderse en un sentido figurado.

Con los ejemplos expuestos queda de manifiesto la complejidad del lenguaje natural y que su tratamiento en cualquier área no resultara fácil.

2.6 *Cómo Funciona la Web Semántica*

Para incorporar semántica a los documentos y páginas que se encuentran en internet, es necesario definir una **ontología** y representarla con **metadatos**⁹ e incorporar estos a los documentos, de tal manera que describan, al menos, el contenido, el significado y la relación de los datos allí contenidos. Estos **metadatos** deben estar bien estructurados y escritos en algún **metalenguaje**¹⁰ que pueda ser interpretado y entendido por un computador.

Para la definición de los datos, la Web Semántica utiliza varios elementos, entre ellos **XML**, **XML Schema**, **el Framework de Descripción de Recursos**, **RDF (por sus siglas en ingles, Resource Description Framework)** y **el lenguaje de ontología Web, OWL (por sus siglas en ingles, Web Ontology Language)**, estos dos últimos estándares ayudan a convertir la Web en una infraestructura global en la que es posible compartir y reutilizar datos y documentos entre diferentes tipos de usuarios.

Cada uno de estos elementos tiene una función específica, que se describe a continuación:

- **XML**, es utilizado para definir la estructura de los documentos o páginas en internet, pero no define el significado de la información.
- **XML Schema (Esquemas XML)**, es el lenguaje utilizado para definir la estructura de los documentos XML.
- **RDF**, proporciona información descriptiva simple sobre los recursos (páginas, documentos etc.) que se encuentran en la web. Aporta una semántica básica para los datos que pueden ser definidos en XML.
- El **OWL** permite desarrollar temas o vocabularios específicos a los que se asocian los recursos definidos en el RDF. OWL proporciona un lenguaje para definir ontologías estructuradas, que pueden ser utilizadas por diferentes

⁹ Los metadatos son información que describe otra información, o datos sobre los datos.

¹⁰ Es el lenguaje utilizado para hacer referencia a otros lenguajes o para describir a un sistema de lenguaje de programación.

sistemas. Las ontologías, se encargan de definir los términos utilizados para describir y representar un área de conocimiento, son utilizados por los usuarios, las bases de datos y los sistemas que necesiten compartir información específica. Las ontologías incluyen definiciones de conceptos básicos en un campo determinado y la relación entre ellos.

Finalmente cabe mencionar que estos lenguajes son los que permiten a los usuarios incorporar significado semántico a los documentos que existen en la Web, permitiendo que los computadores puedan “**entenderlos**” e interpretarlos.

2.7 Ventajas y Desventajas de la Web Semántica

La principal ventaja que aporta la web semántica es que incorpora contenido semántico a los documentos o paginas que se encuentran o que se suban al internet, permitiendo organizar por determinados conceptos la información que ofrecemos, garantizando búsquedas sobre esta, **por el significado y no por el contenido textual**.

Entre las desventajas, en primer lugar se encuentra el hecho de que se hace necesario reestructurar todos los documentos que ya se encuentran en la internet, para adaptarlos al formato que propone la web semántica, de manera que puedan ser interpretados de forma correcta por los computadores. Esto incluye los problemas de idioma, ya que cada uno tiene su semántica diferente, lo que nos lleva a tener que crear buscadores semánticos por idiomas, y más aun, por área de conocimiento.

La segunda desventaja es en el caso de las ontologías para representar la semántica, ya que en ellas es necesario proveer relaciones de equivalencia entre conceptos, por ejemplo, en el caso del código postal, se debería establecer que el término **CP** equivale a **código postal**, y a **zip code**, en el caso del idioma inglés.

2.8 La Web Semántica Hoy

Hoy por hoy, los interesados en la web semántica están incursionando en variadas áreas de interés, desarrollando diversos proyectos de prueba con el objetivo de querer llevar a la práctica el concepto de la Web Semántica. En este sentido ante la gran cantidad de información existente es importante mencionar que, a nivel mundial, son pocas las empresas interesadas en realizar inversiones en este ámbito.

En el mundo de los navegadores web también se ha incursionado con el ímpetu por desarrollar el **concepto y la tecnología de la web semántica**.

Algunos representantes mundiales de bases de datos como Oracle, que ofrece una plataforma abierta de gestión de bases de datos basadas en **RDF**, que sirve como base para una nueva generación de sistemas de negocios a través de la semántica, ya en su versión 10G incluye búsquedas semánticas.

Otro ámbito pujante es la tecnología de telefonía móvil. En los últimos años se están produciendo importantes avances en distintas áreas relacionadas con la web semántica y la web móvil. No en vano, el principal organismo de estandarización en este ámbito, W3C, está invirtiendo importantes recursos en potenciar y desarrollar formalismos y tecnologías clave a través de la **Mobile Web Initiative** y la **Semantic Web Activity**.

De la misma forma podemos mencionar que una de las compañías que más ha apostado por comenzar a aplicar esta tecnología es el **Banco Bankinter**, éste lleva varios años con este proyecto, apoyado por ayudas financieras de la unidad europea, institución que ya ha financiado bastantes proyectos de la web semántica.

En la actualidad la mayoría de las empresas se muestran reticentes a invertir en una tecnología en la cual no ven “negocio” a corto plazo, y muchas de ellas se echan para atrás por este motivo, aun viendo en un futuro sus inmensas posibilidades. ¿Quién será la pionera que se anime y despierte a este prolifero mercado?, la respuesta de los expertos parece ser unánime. Será Google una vez más, la que dará el primer paso, la empresa ya está desarrollando proyectos con esta tecnología, y no se tardará mucho en ver los resultados.

Otros antecedentes del claro e insipiente interés de grupos a este ámbito es **El Real Instituto Alcano**, fundación Española, que presentó hace poco tiempo un buscador semántico para Relaciones Internacionales. Se define como un buscador de información especializada, que ya funciona en la actualidad, y que está disponible en su web www.realinstitutoelcano.org.

III. APLICACIÓN DEL CONCEPTO SEMANTICO

3.1 Descripción del problema

Con los años, La información ha llegado a convertirse en uno de los activos más importantes de cualquier compañía. Gracias a esta las empresas son capaces de hacer gestión, proyecciones, estadísticas, estudios, etc.

Una de las mayores trabas en la gestión de la información es el proceso para su obtención, tarea que en la mayoría de los sistemas requiere cierto nivel de experticia. Es justamente en esta etapa, en donde se detecta el problema existente en el actual sistema de Gestión de Personal, perteneciente a la Tesorería General, que aunque más que un problema es una oportunidad de mejora asociada al proceso.

Hoy por hoy, el sistema antes mencionado, maneja gran y variada cantidad de información, lo que dificulta la tarea para su obtención, en base a las metodologías conocidas, donde el usuario debe ser capaz de obtener la información con procedimientos muy poco similares al lenguaje natural de las personas y más que nada enfocado a un lenguaje informáticamente estructurado y poco flexible en la elaboración de las consultas. En la actualidad éste sistema permite la extracción de información mediante consultas “**dinámicas**”, en las cuales el usuario debe encontrar coincidencias en base a los parámetros ingresados en ciertas casillas predefinidas, casillas que al momento de ejecutar el proceso deben coincidir con los datos contenidos en la Base de Datos , casi exactamente.

3.2 Contexto de la investigación.

Esta Investigación se encuentra inserta en el ámbito del Departamento de Personal, perteneciente a la Tesorería General de la República, unidad encargada de gestionar el proceso de desarrollo laboral e integral de las personas, al servicio de la calidad y objetivos del Servicio.

En este sentido, cabe destacar que esta investigación tiene como objetivo principal, el ofrecer, a los actuales usuarios del sistema de gestión de personal, mejores condiciones de facilidad en la obtención de la información, para así gestionar de la mejor forma posible los datos asociados a los Recursos Humanos dentro del servicio, permitiendo con esto cubrir las necesidades de idoneidad deseadas por el funcionario y la Tesorería General.

Es un factor de alta calidad, el querer facilitar a los usuarios del sistema, el uso del concepto involucrado en este proceso, y de esta forma facilitar la obtención de datos asociados a los diferentes ámbitos que este involucra.

Es importante mencionar la relevancia de las funciones que desempeña el departamento de personal, entre las cuales se encuentran:

- Creación y Administración de los concursos públicos para la contratación de nuevo personal.
- Administración de toda la información referente al funcionario; Datos Personales, Datos Laborales.
- Tramitación de todas las Resoluciones¹¹ que involucran algún tipo de movimiento administrativo asociado a las labores del funcionario.
- Control de las licencias medicas.

¹¹ Documento formal y legal, cuyo objetivo es entregar determinada información a los funcionarios que sean afectados en modificaciones laborales, información de nuevas leyes, etc.

- Control de Vacaciones y Permisos.
- Calificaciones laborales.

Lo anteriormente expuesto, demuestra la importancia de esta investigación, cuyo objetivo principal es facilitar la gestión de la información asociada al ámbito personal y Laboral del funcionario.

A continuación se realiza una pequeña muestra gráfica de algunas de las opciones que controla el Sistema de Gestión de Personal, asociando una pequeña descripción a cada una de ellas.

Figura 5; Pantalla de acceso principal al sistema de Gestión de Personal, el usuario ingresando su Rut y su Clave, accede a los privilegios que por perfil tiene asignado.

Figura 6; Pantalla principal de datos del funcionario, en esta se despliega el conjunto de datos asociados al funcionario, esta pantalla es para aquellos funcionarios de perfil simple, es decir, solo tienen acceso a sus propios datos.

Figura 7; Pantalla principal de las resoluciones asociadas al funcionario, en esta se muestran todos los movimientos que afectan al funcionario, contratos, nombramientos, información de horas extras, etc.

Sistema de Gestión de Personal

Curso Evaluación del Desempeño

MISION VISION VALORES POLITICAS

POSTULA en Línea

Personal > Capacitación > Bienestar > Prevención > Jardín Infantil >

Martes, 16 de Septiembre de 2008 15:48:35

Control de Acceso

RUT

CLAVE

Entrar

NORMATIVA LABORAL

DOCUMENTACIÓN Y ESTUDIOS

PMG Y COMPROMISOS COLECTIVOS

PREGUNTAS FRECUENTES

- [PROCESO CALIFICATORIO](#)
- [DEBERES Y DERECHOS](#)

Bienvenidos al Boletín Inter@cción

Interacción

TITULARES

CAPACITACION A DISTANCIA

AGENDA DE TRABAJO

ISO 9001:2000

Entrevista a Rosalva Pérez Figueroa

Noticias y Comunicados

Instructivo presidencial sobre transparencia activa y publicidad de la información...

Más

Escalafón

Busqueda Funcionario

Cumpleaños

Actualizar

Temática de interés

- PLAN TRIENAL CODIGO BUENAS PRACTICAS LABORALES
- MANUAL DE SELECCION DE PERSONAS EN SERVICIOS PUBLICOS
- DIAGNOSTICO CODIGO BUENAS PRACTICAS
- COMPROMISOS 2006 - 2010 EN MATERIA DE GÉNERO
- LEY N°19.881.NUEVO TRATO LABORAL-MATERIAS RELEVANTES
- RESOLUCION N° 227.DELEGA FACULTADES Y AUTORIZA DELEGACION DE FUNCIONES
- DECLARACION DE INTERESES
- DECLARACION DE PATRIMONIO

825576 N° de visitas

Figura 5: Pantalla Principal Sistema de Gestión de Personal

Departamento
de Personal

Capacitación

SALIR DEL SISTEMA

Cambiar Clave




Abono

Señor : RAUL GEORDY PEREZ PEREDO

Rut: 12875857-7

Ubicación : TESORERIA GENERAL - ADMINISTRACION - RECURSOS MATERIALES



Dossier
Licencias Medicas
Registro Asistencia
Resoluciones
Calificaciones

Certificado Antiquedad
Datos Academicos

Certificado de Sueldo

Liquidacion de Sueldo

SOLICITUD DE DIAS

Administrativos
Feridos
Acumulados

Identificación	
Nombres	RAUL GEORDY PEREZ PEREDO
Función	ANALISTA
Título	ANALISTA DE SISTEMA
Institución	
Tesoreria	TESORERIA GENERAL - ADMINISTRACION - RECURSOS MATERIALES
Anexo	705
Grado	
Calidad jurídica	
Estamento	
Cargo	
Fecha en Servicio	
Asignación	

Situación		
	Fecha Inicio	Nº Dias
Días Administrativos	--	--
Días Feriado Legal	--	--
Licencias Medicas	--	--
Estado		ACTIVO

Figura 6: Conjunto de los principales datos asociados al funcionario

Departamento
de Personal

Señor : RAUL GEORDY PEREZ PEREDO
Rut: 12875857-7
Ubicación : TESORERIA GENERAL - ADMINISTRACION - RECURSOS MATERIALES



[Dossier](#) [Licencias Medicas](#) [Registro Asistencia](#) [Resoluciones](#) [Calificaciones](#)

[Certificado Antiquedad](#) [Datos Academicos](#)

SOLICITUD DE DIAS

[Administrativos](#)
[Feriados](#)
[Acumulados](#)

[Certificado de Sueldo](#)

[Liquidacion de Sueldo](#)

Resoluciones

Rut	Evento	Sub Evento	Nº Resolución	Fecha Emision	Fecha Inicio Efectivo	Nº Evento
12875857	CONTRATO		3453	14/12/2007	01/01/2008	31305
12875857	CONTRATO		2720	13/12/2006	01/01/2007	23463
12875857	CONTRATO		69	05/04/2006	01/04/2006	18696
12875857	CONTRATO		2492	13/12/2005	01/01/2006	16062
12875857	CONTRATO		2097	03/12/2004	01/01/2005	9449
12875857	CONTRATO		2417	05/12/2003	01/01/2004	3347
12875857	CONTRATO		3380	17/12/2002	01/01/2003	3380
12875857	CONTRATO		35	11/02/2002	01/02/2002	35
12875857	NOMBRAMIENTO		21	28/01/2002	01/01/2002	21
12875857	NOMBRAMIENTO		215	13/08/2001	01/08/2001	215
12875857	NOMBRAMIENTO		194	13/07/2001	01/07/2001	194
12875857	NOMBRAMIENTO		83	08/01/2001	01/01/2001	83
12875857	NOMBRAMIENTO		355	20/10/2000	01/10/2000	355
12875857	NOMBRAMIENTO		255	20/06/2000	01/07/2000	255
12875857	NOMBRAMIENTO		179	27/03/2000	01/04/2000	179
12875857	NOMBRAMIENTO		94	27/01/2000	01/02/2000	94
12875857	NOMBRAMIENTO		387	06/10/1998	01/10/1999	387
12875857	NOMBRAMIENTO		404	22/09/1999	01/09/1999	404
12875857	NOMBRAMIENTO		276	28/07/1999	01/07/1999	276
12875857	NOMBRAMIENTO		208	10/06/1999	01/06/1999	208
12875857	NOMBRAMIENTO		138	23/04/1999	01/04/1999	138
12875857	NOMBRAMIENTO		49	29/01/1999	01/01/1999	49
12875857	CONTRATO		1384	30/06/1998	01/06/1998	1384

Figura 7: Control de Resoluciones asociadas a los movimientos laborales del funcionario

Como se puede apreciar en las imágenes desplegadas con anterioridad, la variedad de información que se maneja es sustancial, por lo que se hace necesario, aplicar el concepto semántico para facilitar su obtención.

3.3 *Lógica del Negocio*

A primera vista, en la lectura del apartado del contexto del proyecto, es posible apreciar a grandes rasgos la lógica del negocio y el flujo en la información que hoy se administra.

Por lo tanto es importante, explicar el flujo en la creación de los datos, que se asocian a cada uno de los procesos vinculados a los funcionarios, los cuales se describen a continuación:

- El primer evento, en la lógica y generación de la información, se produce al momento de la selección del nuevo personal para determinados cargos, es aquí donde se almacenan los primeros datos asociados a cada postulante, información, que tiene relación con determinadas habilidades asociadas al cargo solicitado. A lo anterior se suma la información de carácter personal, la que quedará almacenada en el sistema, para su posterior evaluación de acuerdo a las necesidades del servicio.
- En segundo lugar mientras se realiza el proceso de selección, todos los datos asociados a esta instancia, como por ejemplo, resultados de test psicológico o de entrevistas personales, son almacenados de igual forma en el sistema en cuestión, para su posterior evaluación, es importante mencionar que debe quedar registro de las entrevistas como tal y de los resultados de estas.
- Posteriormente, de pasar estos postulantes todos los procesos de selección a formar parte de la Tesorería, en su calidad de funcionarios, se genera toda la información de carácter personal y laboral, y a ésta se suma la información que ya había sido registrada referente de sus habilidades laborales y personales, más el correspondiente test psicológico. Esta información es de especial importancia para futuras postulaciones del funcionario a otros cargos, o redistribuciones según necesidades del servicio.

- El siguiente paso, es la revalidación de la información de tipo académico asociada al funcionario, esta es realizada debido a la importancia que posee, en el sentido de ventajas económicas, por ejemplo, la **Asignación de Título**, que es cancelada a aquellos funcionarios que posean un Título Universitario.
- Las habilidades de su área específica de estudios, debe ser comprobada por medio de pruebas que el servicio realiza, cada ciertos plazos de tiempo, y que ayudan a mantener un mínimo nivel de capacidades asociadas al cargo que desempeña.
- La siguiente información que se controla y almacena, es la que tiene relación con la experiencia laboral en otros ámbitos ya sean de la empresa privada o entidades públicas, se lleva a cabo el correspondiente chequeo, para descartar inconvenientes económicos con otras entidades del estado o privadas. De esta manera se obtiene la probidad que amerite el cargo que desempeñara el funcionario.

Los datos descritos con anterioridad, tienen relación con el proceso de incorporación de una nueva persona al servicio, una vez que este ya es parte del personal de La Tesorería General, se generan nuevos tipos de datos.

Los usuarios más avanzados y con mayores privilegios de acceso al sistema, administran la información antes mostrada, proceso que muchas veces resulta complicado dependiendo de las necesidades de este.

Uno de los módulos más utilizados para la obtención de datos en el sistema es el que se muestra en la siguiente imagen.

Departamento de Personal

REGIONES
CAMBIAR PERFIL
MENU PRINCIPAL
SALIR DEL SISTEMA

Seleccione el tipo de consulta a realizar

Consulta	Buscar
<input style="width: 90%;" type="text" value="Datos Académicos"/>	<input style="width: 80%; background-color: #003366; color: white;" type="button" value="Buscar"/>

<input style="width: 95%;" type="text" value="RUT FUNCIONARIO"/> <input style="width: 95%;" type="text" value="APELLIDO PATERNO"/> <input style="width: 95%;" type="text" value="APELLIDO MATERNO"/> <input style="width: 95%;" type="text" value="NOMBRES"/>	Seleccione de esta lista los datos que desea mostrar como resultado del reporte. Para seleccionar mantenga presionada la tecla CTRL y con el Mouse seleccione.
--	---

Críterios para la consulta

Nombre Campo	Operador	Criterio de Busqueda	Totalizar
RUT FUNCIONARIO		<input style="width: 95%;" type="text"/>	
APELLIDO PATERNO		<input style="width: 95%;" type="text"/>	
APELLIDO MATERNO		<input style="width: 95%;" type="text"/>	
NOMBRES		<input style="width: 95%;" type="text"/>	
TITULO		<input style="width: 95%;" type="text"/>	
INSTITUCION ACADEMICA		<input style="width: 95%;" type="text"/>	
TIPO ASIGNACION		<input style="width: 95%;" type="text"/>	
NUMERO DE SEMESTRES	Igual a <input style="width: 20px;" type="text"/>	<input style="width: 95%;" type="text"/>	<input type="checkbox"/>
ESTADO TITULO		<input style="width: 95%;" type="text"/>	
ESTADO CIVIL		<input style="width: 95%;" type="text"/>	
SEXO		<input style="width: 95%;" type="text"/>	
FECHA NACIMIENTO	Igual a <input style="width: 20px;" type="text"/>	<input style="width: 95%;" type="text"/>	<input type="checkbox"/>
N° REGION	Igual a <input style="width: 20px;" type="text"/>	<input style="width: 95%;" type="text"/>	<input type="checkbox"/>
UBICACION		<input style="width: 95%;" type="text"/>	
DEPARTAMENTO		<input style="width: 95%;" type="text"/>	
SECCION		<input style="width: 95%;" type="text"/>	
OFICINA		<input style="width: 95%;" type="text"/>	

Figura 8: Esta pantalla nos muestra una consulta dinámica, que permite seleccionar información por algunos criterios determinados, es una opción muy utilizada, claro que con las limitantes asociadas a una nomenclatura informática que solo permite buscar coincidencias con el dato específico y en un lenguaje rígido.

Se explicó parte del proceso de generación de la gran cantidad de información que contiene el sistema, lo que finalmente gatilla en la dificultad para su obtención.

3.4 Situación Actual

En la actualidad la información del sistema de Gestión de Personal, se encuentra almacenada en una **Base de Datos**, en formato **MS SQL 2000**¹². Esta está constituida por una cantidad determinada de tablas en las cuales se encuentran almacenados los datos de los distintos tipos descritos.

Con la siguiente figura de ejemplo se explicará lo que es una base de datos, y con esta se continuará el resto de los apartados para explicar la aplicación del concepto:

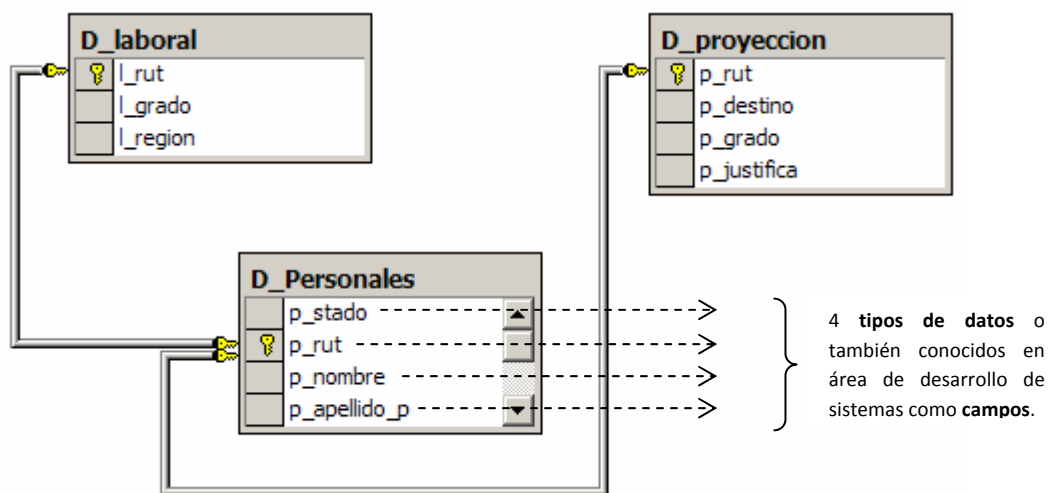


Figura 9: Diagrama de una base de datos de ejemplo constituida por 3 tablas.

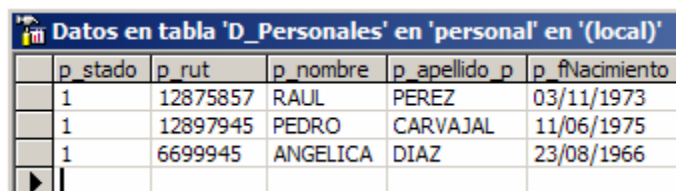
En la figura **N° 9** se puede apreciar una base de datos de ejemplo, que está constituida por 3 tablas, cada una de ellas con una cantidad determinada de **datos**, también conocidos como **campos** de distinto tipo, pero todos relacionados para este caso, con los de una persona perteneciente a la Tesorería General. En la parte superior de cada tabla se muestra el nombre que la identifica. A continuación para este ejemplo se describen cada una de ellas:

¹² Motor de base de datos, con el cual se administran los datos correspondientes al sistema.

- La tabla **D_Laboral**: almacena 3 tipos de datos o campos, el **Rut**, el **grado** y la **región**, estos tienen relación con los datos laborales del funcionario.
- La tabla **D_Personales**: almacena 4 tipos de datos o campos, el **estado**, el **Rut**, el **Nombre** y el **Apellido Paterno**, estos tienen relación con los datos personales del funcionario.
- La tabla **D_Proyeccion**: almacena 4 tipos de datos o campos, el **Rut**, el **destino**, el **grado**, y la **justificación**, estos tienen relación con los datos de proyección del funcionario, es decir, información que permite saber a qué lugar el funcionario desearía ser trasladado.

Como se muestra en la figura **N°9**, existe una línea que une las tres tablas. El objetivo de esta es demostrar que las 3 están relacionadas por medio del Rut del funcionario. Es importante mencionar que la figura **N°9** muestra la base de datos de ejemplo de forma completa, para un mejor entendimiento del lector.

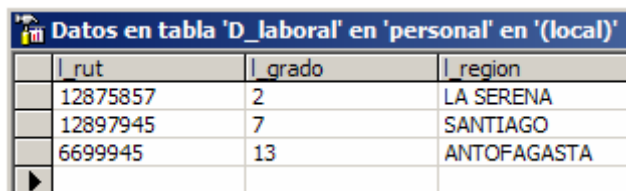
Pero si queremos visualizar los datos almacenados en cualquiera de las tres tablas, las figuras serían las siguientes:



	p_estado	p_rut	p_nombre	p_apellido_p	p_fNacimiento
	1	12875857	RAUL	PEREZ	03/11/1973
	1	12897945	PEDRO	CARVAJAL	11/06/1975
	1	6699945	ANGELICA	DIAZ	23/08/1966

Figura 10: Tabla de nombre **D_Personales** con 3 registros (filas).

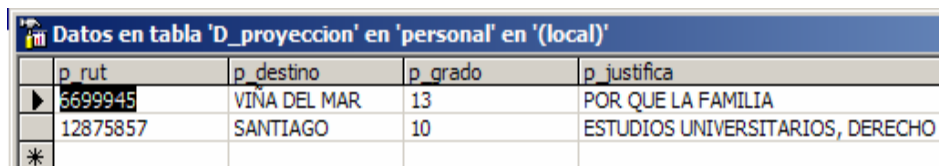
En la figura **N°10** se muestran **3 registros (filas)** almacenados, correspondientes a los datos de tres personas.



l_rut	l_grado	l_region
12875857	2	LA SERENA
12897945	7	SANTIAGO
6699945	13	ANTOFAGASTA

Figura 11: Tabla de nombre **D_laboral** con 3 registros (filas).

En la figura N°11 se muestran **3 registros (filas)** almacenados, correspondientes a los datos de tres personas.



p_rut	p_destino	p_grado	p_justifica
6699945	VINA DEL MAR	13	POR QUE LA FAMILIA
12875857	SANTIAGO	10	ESTUDIOS UNIVERSITARIOS, DERECHO

Figura 12: Tabla de nombre **D_proyeccion** con 2 registros (filas).

En la figura N°12 se muestran **2 registros (filas)** almacenados, correspondientes a los datos de dos personas.

En las tres tablas anteriores se repiten los mismos Rut, lo que explica **la relación** de los datos correspondientes a las tres personas del ejemplo.

Queda demostrado, en las figuras anteriores, que la información se encuentra almacenada en las tablas asociadas a cada tipo. Cuando se realizan búsquedas por medio de las opciones del sistema, de algún ámbito específico, los parámetros de las búsquedas deben coincidir con los datos almacenados en alguna de las tablas.

La rigidez que se aprecia en esta metodología, la mayoría de las veces es rechazada por los usuarios con poca experticia en sistemas informáticos, lo que conlleva a que la búsqueda de información de cualquier sistema, pase a convertirse en una tarea poco amigable.

Finalmente es importante mencionar que hoy en día, los sistemas de búsqueda de información deben adaptarse a las necesidades de los usuarios, por lo que cualquier proceso que logre conseguir este objetivo, y que funcione bajo las plataformas ya existentes, sin grandes cambios será bienvenido.

3.5 Aplicación del concepto

El primer paso para conseguir el objetivo que se persigue, que es aplicar semántica a los datos, es llevar a cabo este proceso de la manera lo menos invasiva posible para el sistema en cuestión, y sin grandes cambios en su estructura.

3.5.1 Primeros pasos hacia el objetivo

Para aplicar semántica a los datos almacenados en las bases de datos del sistema, es claro, en base al concepto, que se debe desarrollar la definición de cada uno de los datos que ahí se encuentran, pero no al dato como tal, sino que al campo, como se muestra en la siguiente figura.

	p_rut	p_destino	p_grado	p_justifica
▶	6699945	VINA DEL MAR	13	POR QUE LA FAMILIA
	12875857	SANTIAGO	10	ESTUDIOS UNIVERSITARIOS, DERECHO
*				

Primer campo llamado p_rut

Segundo campo llamado p_destino

Tercer campo llamado p_grado

Cuarto campo llamado p_justifica

Figura 13: Tabla de nombre D_Proyeccion con cuatro campos.

En el área de la informática, o mejor dicho del desarrollo de sistemas, se acostumbra por definición y regla a nombrar los **campos** de una **tabla** con nomenclaturas cortas y precisas, y con un orden definido que permita entender a primera vista el concepto del tipo de dato, y que en consecuencia, permiten definir el contexto de los tipos de datos que esa tabla almacena.

Por lo tanto, se puede concluir que la única forma de aplicar semántica a los datos contenidos, es desarrollar **una o más definiciones**, lo más estandar posible para cada uno de los **campos** de las **tablas** de la base de datos. Esta definición debe estar apegada al contexto del tipo de dato y de la organización donde el sistema es utilizado. Solo queda definir y aclarar la forma de llevar a cabo la definición antes señalada, definición que debe ir de la mano con el objetivo de no invadir la estructura del sistema en su esencia, y que hoy por hoy permita a los actuales usuarios, seguir trabajando de la manera que lo vienen haciendo, solo usando la estructura existente del sistema y sin caer en reestructuraciones gigantescas.

3.5.2 Tecnología para la aplicación

A medida que se desarrolló la investigación, se fue definiendo el mecanismo por medio del cuál se pretende llevar a cabo el proceso de aplicación del concepto. Analizando las herramientas y tecnología involucrada en este, resultando ser menos complejo de lo pensado, en el sentido de utilizar lo que ya existe en las bases de datos, y agregar **tablas nuevas**, cuyo objetivo será sólo **almacenar la o las definiciones** de los **campos** existentes.

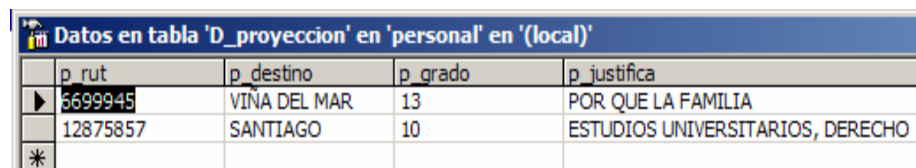
En este sentido se destaca, que en cuanto a la tecnología involucrada, no se agregará nada nuevo a lo ya utilizado actualmente y cuyo detalle se describe a continuación:

- El motor de bases de datos que actualmente utiliza el sistema es **MS SQL 2000**, en este contexto como la definición de los datos sera desarrollada y almacenada en esta misma base de datos, sólo se hace necesario **crear nuevas tablas**.
- En cuanto al sistema como tal, sólo sera necesario crear **nuevas opciones** y **métodos de búsqueda** que permitan acceder a la nuevas modalidades de extracción de la información.

3.5.3 La aplicación

A continuación se explicará paso a paso la aplicación del concepto, apoyado en imágenes que permitan a personas que no son especialistas en el área informática, entender el proceso de la mejor manera posible.

Como se explicó en párrafos anteriores, los datos que actualmente se encuentran almacenados en la base de datos del sistema en cuestión, están como se muestra en la siguiente figura:



p_rut	p_destino	p_grado	p_justifica
6699945	VINA DEL MAR	13	POR QUE LA FAMILIA
12875857	SANTIAGO	10	ESTUDIOS UNIVERSITARIOS, DERECHO

Figura 14: Tabla llamada **D_proyeccion**, con dos filas, corresponden a 2 registros.

En la figura N°14, se muestra la tabla llamada D_proyeccion, que contiene **4 campos (columnas)**, y almacena **2 registros (filas)**. Para este ejemplo, estos dos registros corresponden a los datos de dos personas distintas.

Cabe mencionar que los 4 campos que contiene esta tabla son “**pobres**”, semanticamente hablando, pues no poseen definición de lo que cada uno de ellos significa, lo que no permite identificar el contexto en el cual están involucrados, menos el tipo de sistema que los utiliza, es más, podríamos identificarlos mirandolos, y por conocimiento propio apuntar a que pertenecen a los datos de una persona aparentemente, pero no sabemos el contexto del sistema que los utiliza.

Por lo tanto se hace necesario aplicar una definición explícita a estos datos, definición que debe ser elaborada de tal forma que **no permita**, al sistema, perder consistencia,

en los datos almacenados. Para lograr este objetivo, se explicará paso a paso como llegar a él.

- Primero, para cada **campo** contenido, por ejemplo, en la tabla de la figura N°18, se creará una nueva tabla que contenga solo la o las definiciones de ese campo en particular, en el caso de este ejemplo, necesitaremos 4 tablas que nos permitan definir los campos llamados **p_rut**, **p_destino**, **p_grado** y **p_justifica**.
- Segundo, estas cuatro nuevas tablas, tendrán una relación directa con los 4 campos de la tabla de la figura N°15, como se explica en la siguiente figura;

Datos en tabla 'D_proyeccion' en 'personal' en '(local)'			
p_rut	p_destino	p_grado	p_justifica
6699945	VINA DEL MAR	13	POR QUE LA FAMILIA
12875857	SANTIAGO	10	ESTUDIOS UNIVERSITARIOS, DERECHO
*			

Datos en tabla 'p_rut'
def

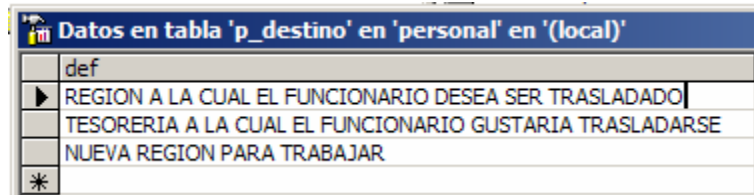
Datos en tabla 'p_destino'
def

Datos en tabla 'p_grado'
def

Datos en tabla 'p_justifica'
def

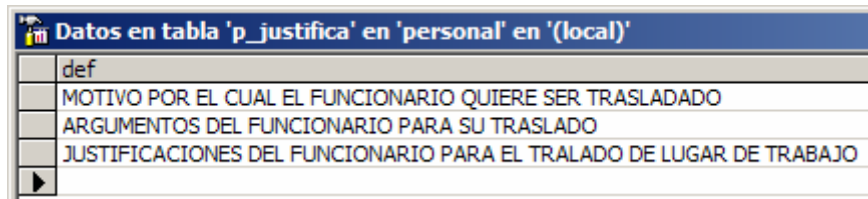
Figura 15: Tabla llamada **D_proyeccion**, cada uno de sus campos posee una tabla asociada.

- Una vez creadas las **tablas** necesarias para cada cada **campo** de las tablas del sistema que se definan para la aplicación del concepto, se desarrollarán las definiciones estándar, en conjunto con los usuarios, que más utilicen el sistema para la obtención de datos. La o las definiciones, se almacenaran en las nuevas tablas como se muestra en las siguientes dos figuras:



Datos en tabla 'p_destino' en 'personal' en '(local)'	
def	
▶	REGION A LA CUAL EL FUNCIONARIO DESEA SER TRASLADADO
	TESORERIA A LA CUAL EL FUNCIONARIO GUSTARIA TRASLADARSE
	NUEVA REGION PARA TRABAJAR
*	

Figura 16: Esta figura nos muestra las definiciones para el campo **p_destino**.



Datos en tabla 'p_justifica' en 'personal' en '(local)'	
def	
	MOTIVO POR EL CUAL EL FUNCIONARIO QUIERE SER TRASLADADO
	ARGUMENTOS DEL FUNCIONARIO PARA SU TRASLADO
	JUSTIFICACIONES DEL FUNCIONARIO PARA EL TRALADO DE LUGAR DE TRABAJO
▶	

Figura 17: Esta figura nos muestra las definiciones para el campo **p_justifica**.

- Y así sucesivamente, **tabla por tabla, campo por campo**, se desarrollarán **la o las definiciones** necesarias, de común acuerdo, por los usuarios del sistema que más utilicen estos datos. En este sentido es importante destacar que las definiciones que se realicen para todos los campos de las tablas del sistema deben ser desarrolladas con la colaboración de todos los usuarios involucrados en el uso avanzado del sistema, por supuesto, apuntando al contexto en el cual este se encuentra inserto.
- Finalmente, observando el diagrama completo asociado a la definición de los datos, el siguiente ejemplo ilustra el concepto como tal, implementado en la base de datos de ejemplo.

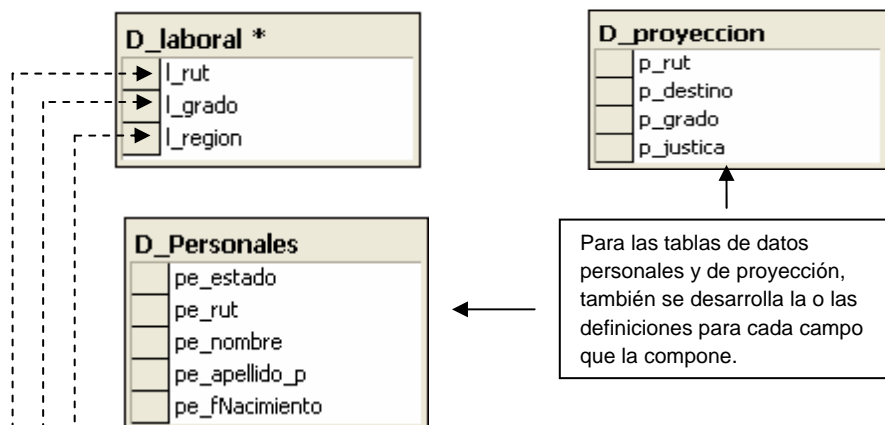


Figura 18: Esta figura muestra la base de datos de ejemplo, con sus tres tablas.

Datos en tabla 'l_region' en 'personal' en '192.168.3.60'	
def	
	LUGAR FISICO AL QUE PERTENECE EL FUNCIONARIO
	REGION EN LA CUAL ESTA UBICADO EL FUNCIONARIO
	CIUDAD EN LA CUAL TRABAJA EL FUNCIONARIO
	CIUDAD EN LA CUAL DESEMPEÑA LABORES EL FUNCIONARIO

Figura 19: en esta figura, se muestran todas las posibles descripciones o definiciones, asociadas al campo que almacena la **region** en donde trabaja del funcionario.

Datos en tabla 'l_grado' en 'personal' en '192.168.3.60'	
def	
	GRADO QUE EL FUNCIONARIO POSEE ACTUALMENTE ASOCIADO A UN ESCALAFON
	IDENTIFICADOR NUMERICO QUE PERMITE SABER EL GRADO QUE POSEE EL FUNCIONARIO ASOCIADO A UN SUELDO

Figura 20: en esta figura, se muestran todas las posibles descripciones o definiciones asociadas al campo que almacena el **grado** del funcionario.

Datos en tabla 'l_rut' en 'personal' en '192.168.3.60'	
def	
	IDENTIFICADOR NUMERICO UNICO ASOCIADO AL FUNCIONARIO
	RUT DEL FUNCIONARIO
	RUN UNICO NACIONAL DEL FUNCIONARIO

Figura 21: en esta figura, se muestran todas las posibles descripciones o definiciones, asociadas al campo que almacena el **rut** del funcionario.

Finalmente, los datos almacenados en sus respectivas tablas, las que poseerán una o varias definiciones acerca de su contenido y del contexto al que apuntan, se muestran de manera global y gráficamente en la siguiente figura.

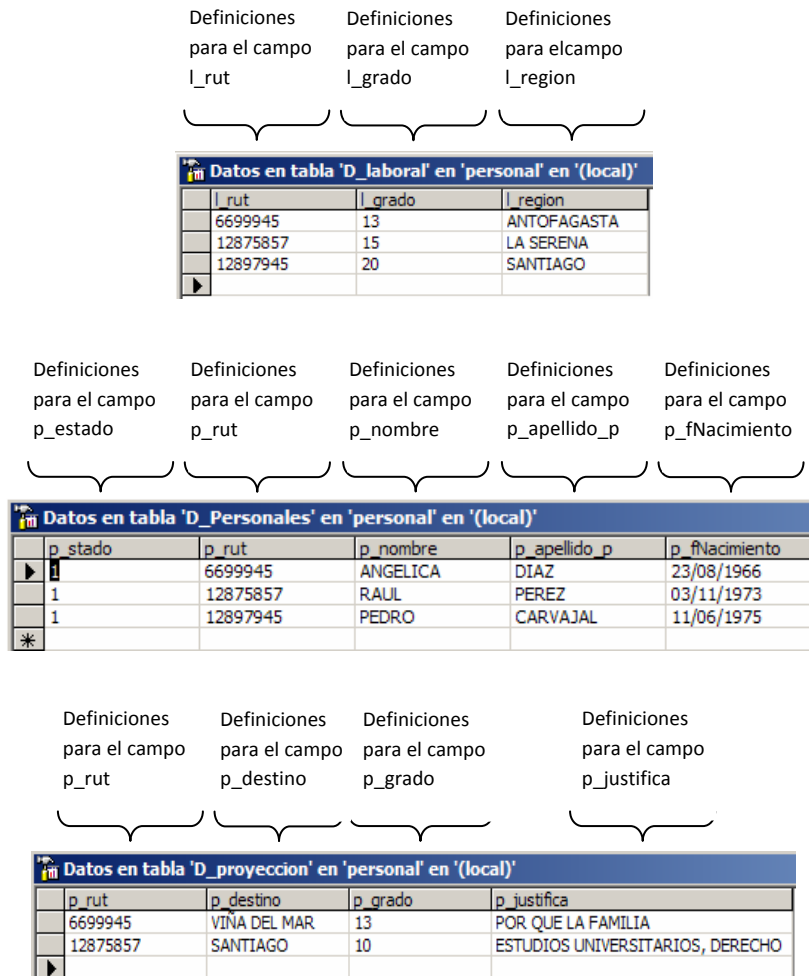


Figura 22: figura que muestra las 3 tablas de ejemplo, con algunos datos almacenados explicando con la figura de llave que cada campo debe tener asociada **la o las descripciones o definiciones** acordadas.

Se puede mencionar que el proceso de implementación necesitará gran cantidad de trabajo, y un periodo de comprensión y adaptación de los usuarios, además de crear un estándar en el proceso de incorporación de las definiciones para cada una de las tablas de la Base de Datos del Sistema. Lo importante es que una vez terminada esta etapa, las ganancias en facilidad del uso y ahorro de tiempo en las búsquedas recompensaran el cambio.

3.5.4 Uso del concepto Semántico

En el punto anterior sobre **la aplicación del concepto Semántico**, se explico la forma de enriquecer los datos, asociando a cada uno, **varios tipos de descripción o definición estándar**, en base a un acuerdo o establecimiento previo.

Por lo tanto, ya se tiene claridad del proceso, y solo falta aclarar la manera de cómo se accederá a los datos.

En este sentido es importante explicar y mencionar, que con este concepto inserto en nuestros datos, las búsquedas ya no serán realizadas accediendo directamente al dato, si no que se accederá a las **descripciones o definiciones** asociadas a cada uno de ellos, he ahí la importancia en que cada dato tenga la mayor cantidad posible de descripciones o definiciones, y que por supuesto estén en el contexto del sistema y del dato almacenado.

A continuación de la misma manera de cómo se detalló la aplicación del concepto, se explicará, la nueva propuesta para los nuevos métodos de búsqueda.

- Uno de los primeros pasos en este sentido será realizar un cambio en las metodologías de búsqueda de información. Actualmente, estas se realizan accediendo directamente al dato almacenado, es decir, si busco a los **funcionarios** cuyo lugar de trabajo es “**LA SERENA**”, y que su fecha de nacimiento sea igual a “**03/11/1973**”, internamente el sistema busca los datos

que **coincidan** con los **parámetros** ingresados por el usuario y en las tablas que sea necesario, como se explica en la siguiente imagen:

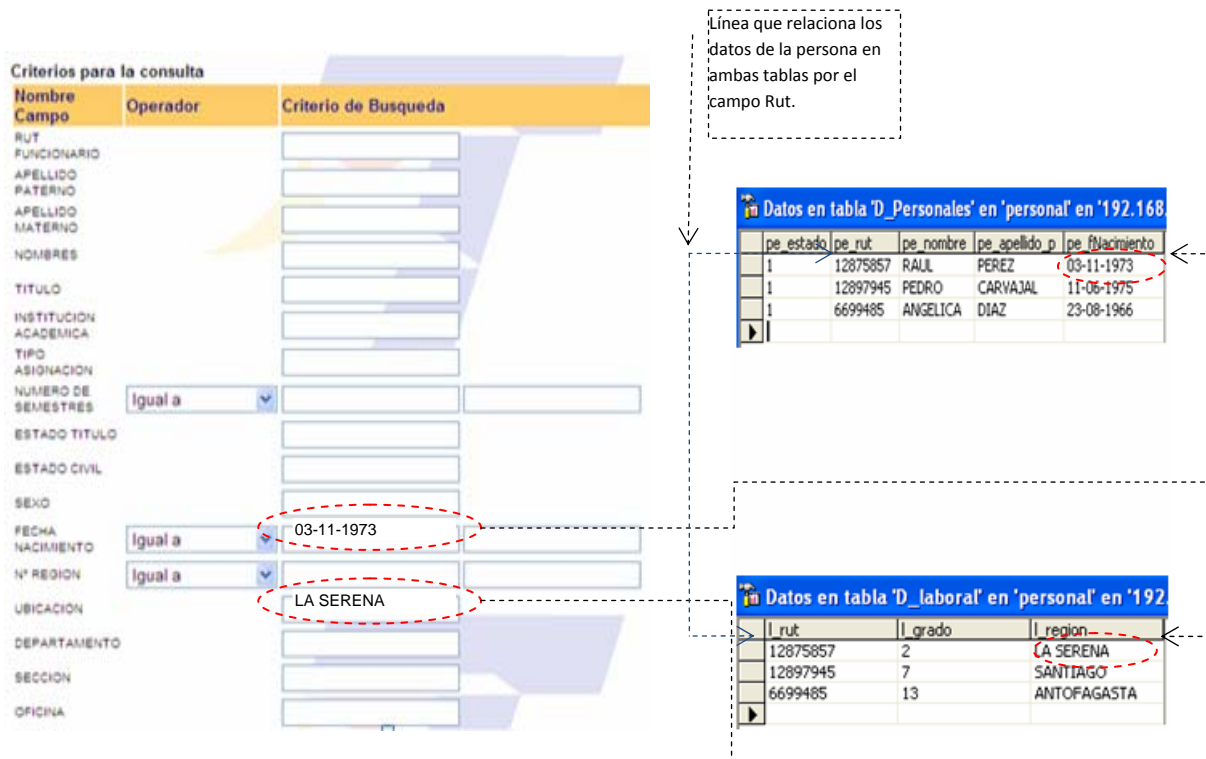


Figura 23: Figura que explica la actual metodología de búsqueda .

- Por lo tanto, la rigidez del proceso está dada por la cantidad de parámetros que posean las búsquedas. Hay que tener claro que si los parámetros ingresados por el usuario no coinciden con los datos almacenados, posiblemente no se obtengan resultados.
- Por lo anterior, es que se debe proponer un proceso de búsqueda más flexible, y fácil de usar. En base a la esencia del concepto semántico, cuyas búsquedas

proponen enfocarse en la **descripción o definición de los datos**, es que se desarrolla la siguiente metodología. Esta se basa en que las consultas se ingresen en un lenguaje parecido al **lenguaje natural de las personas**, la información descrita en estas consultas buscará al mismo tiempo las coincidencias **cercanas al dato (campo)** y **las coincidencias cercanas a la descripción o definición (del campo)**. Este proceso se debe implementar entregando más flexibilidad en la **igualdad** de los datos ingresados, es decir, que por menor que sea la coincidencia de lo buscado contra lo encontrado en la descripción del campo, se arroje como parte del resultado de la búsqueda. La siguiente imagen explica esta propuesta:

Nueva Forma de búsqueda.

Datos en tabla 'l_region' en 'personal' en '(local)'

def
REGION EN LA QUE TRABAJA EL FUNCIONARIO
UBICACION DE TESORERIA EN LA QUE TRABAJA EL FUNCIONARIO

(*)

Datos en tabla 'l_grado' en 'personal' en '(local)'

def
GRADO QUE POSEE EL FUNCIONARIO
VALOR NUMERICO DEL GRADO DEL FUNCIONARIO
GRADO DE LA ESCALA UNICA DE SUELDO EN QUE SE ENCUENTRA EL FUNCIONARIO
ESCALAFON DEL FUNCIONARIO

(*)

Datos en tabla 'D_laboral' en 'pers'

l_rut	l_grado	l_region
6699945	13	ANTOFAGASTA
12875857	15	LA SERENA
12897945	20	SANTIAGO

(*)

FUNCIONARIOS QUE TRABAJAN EN SANTIAGO Y QUE DESEA SER TRASLADADO A VINA DEL MAR POR PROBLEMAS CON LA JEFATURA

Datos en tabla 'D_proyeccion' en 'personal' en '(local)'

p_rut	n_destino	p_grado	p_justifica
6699945	VINA DEL MAR	13	POR PROBLEMAS CON LA JEFATURA
12875857	SANTIAGO	10	ESTUDIOS UNIVERSITARIOS, DERECHO

(*)

Datos en tabla 'p_justifica' en 'personal' en '(local)'

def
MOTIVO POR EL CUAL EL FUNCIONARIO QUIERE SER TRASLADADO
ARGUMENTOS DEL FUNCIONARIO PARA SU TRASLADO
JUSTIFICACIONES DEL FUNCIONARIO PARA EL TRALADO DE LUGAR DE TRABAJO

(*)

Datos en tabla 'p_grado' en 'personal' en '(local)'

def
GRADO QUE DESEA EL FUNCIONARIO QUE DESEA TRASLADARSE
IDENTIFICADOR NUMERICO QUE DESEA DEL FUNCIONARIO QUE QUIERE TRASLADARSE
VALOR DE LA ESCALA DEL FUNCIONARIO QUE QUIERE TRASLADARSE

(*)

Datos en tabla 'p_destino' en 'personal' en '(local)'

def
REGION A LA CUAL EL FUNCIONARIO DESEA SER TRASLADADO
TESORERIA A LA CUAL EL FUNCIONARIO GUSTARIA TRASLADARSE
NUEVA REGION PARA TRABAJAR

(*)

(*) Nuevas Tablas con definiciones.

(*) Tablas con datos del sistema.

- Explicando la imagen anterior, se puede mencionar que el usuario ingresará frases para la búsqueda, según el estándar definido. Estas buscarán información en las tablas de **descripción o definición** y en la de **datos**, como lo indican las líneas segmentadas.
- Esta es la propuesta, para la aplicación del concepto. Por supuesto el objetivo es entregar una idea de su aplicación y no entrar en las complejidades del lenguaje de desarrollo asociado.

3.5.5 Ventajas y Desventajas

En base a los conocimientos obtenidos durante la investigación y el desarrollo de la propuesta de implementación, se pueden rescatar ventajas y desventajas asociadas a la aplicación de la web semántica al mundo real.

- **Ventajas**
 - ✓ El enriquecimiento de los datos, permitirá entregar un contexto definido a cada uno de ellos.
 - ✓ La aplicación del concepto Semántico, permitirá obtener más flexibilidad en la elaboración de las consultas.
 - ✓ Se facilitará la adaptabilidad de los usuarios, a las nuevas metodologías de búsqueda.
 - ✓ Se obtendrán ganancias en tiempo, costos y usabilidad del sistema.
 - ✓ El enriquecimiento de los datos, permitirá a cualquier usuario entender rápidamente el contexto en el que este está involucrado.

- **Desventajas**

- ✓ La definición de los estándares y criterios, para el desarrollo de las definiciones de los datos, será uno de los procesos complicados a elaborar.

IV. CONCLUSION

La complejidad existente en los procesos de búsqueda, tanto en los buscadores de internet, como en los sistemas informáticos de distinto tipo, resultan finalmente en que el usuario consume demasiado de su tiempo en realizar los filtros necesarios que finalmente permitan llegar a un dato útil y que sea el que realmente necesita.

Los procesos de búsqueda deben ser primero que todo entendidos y comprendidos por cualquier tipo de usuario que los utiliza, ya sea experto o inexperto.

En este sentido la propuesta desarrollada, en base a la tecnología ya existente, modificando la lógica de las búsquedas y agregando definición a los datos, demostró la ganancia que se obtendrá en los procesos de obtención de la información, ganancia que se verá reflejada en ámbitos como, tiempo, costos, facilidad de entendimiento y por lo tanto de uso, flexibilidad, etc.

De la mano de estas ganancias se obtiene una muy importante, la adaptabilidad a las necesidades del usuario actual, factor que hoy en día es una ventaja en cualquier ámbito del quehacer diario.

La hipótesis se valida, en el sentido de que utilizando la tecnología actual disponible se podrán enriquecer semánticamente los datos del sistema y por lo tanto acceder a métodos más flexibles y precisos de búsqueda.

El beneficio de esta propuesta para la Tesorería General de la República, se podrá ver reflejado en muchos otros ámbitos de sus funciones como servicio público, ya que la información administrada es mucha y muy relevante. Muchas veces se obtienen conceptos tributarios complicados de entender para cualquier usuario, por lo tanto, aplicando semántica a los datos de los procesos, se podrá facilitar su uso, entendimiento y mejorar los tiempos de búsqueda, la ventaja estará en la experiencia ya adquirida en este proceso para su aplicación en otros sistemas del servicio.

La semántica aplicada a los datos ya es una realidad muy cercana, existen variados buscadores en internet que ya utilizan esta metodología y solo dependerá de nosotros el hacer que el resto de los sistemas usados a diario por los usuarios presten la mismas ventajas de usabilidad y facilidad.

V. GLOSARIO

- **Base de Datos.** Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. En éste sentido, una biblioteca puede considerarse una base de datos compuesta en su mayoría por documentos y textos impresos en papel. En la actualidad y debido al desarrollo tecnológico, la mayoría de las bases de datos están en formato digital electrónico que ofrecen un amplio rango de soluciones al problema de almacenar datos.
- **ERP.** Sistemas integrales de gestión empresarial que están diseñados para automatizar la mayoría de los procesos en la empresa (finanzas, comercial, logística, producción, etc.)
- **HTML.** siglas de **HyperText Markup Language** (*Lenguaje de Marcas de Hipertexto*), es el lenguaje que hace posible presentar información a través de páginas web, dicho de otra manera, hace posible construir dicha páginas.
- **Metadatos.** Son datos altamente estructurados, en un sentido genérico se definen como Datos de los datos, o datos que definen a otros datos. Surgen de la necesidad de describir, identificar o caracterizar cualquier tipo de cosa, objeto, etc.
- **Metalinguaje.** Un metalenguaje es un lenguaje usado para hacer referencia a otros lenguajes, en un aspecto más general, puede referirse a cualquier terminología o lenguaje usado para discutir acerca del mismo lenguaje, un texto sobre gramática, por ejemplo, o una discusión acerca del uso del lenguaje.
- **Motores de Búsqueda.** Un motor de búsqueda es un sistema informático que indexa u ordena los archivos almacenados en los servidores de los sistemas de distinto tipo. Un ejemplo son los buscadores de internet, algunos solo buscan

información en el internet y otros además buscan en servicios de noticias. Las búsquedas se hacen con palabras claves y los resultados son un listado de direcciones web en los se mencionan temas relacionados con las palabras claves buscadas.

- **Ontologías.** El termino ontología en la informática hace referencia a la formulación de un riguroso esquema conceptual dentro de un ámbito determinado, con la finalidad de facilitar la comunicación y la compartición de la información entre diferentes sistemas informáticos.
- **OWL.** Es un lenguaje informático que nos permite crear ontologías.
- **RDF.** Es un lenguaje que proporciona información descriptiva simple de cualquier tipo de recurso, permite crear y definir una sintaxis y semántica formal, permite estandarizar el vocabulario.
- **W3C:** Organismo internacional que define normas y reglas para regular los estándares de la web, necesarios para asegurar así el acceso universal a la información.
- **Web.** Termino abreviado para referirse a la internet.
- **XML.** Lenguaje informático utilizado para definir la estructura de las páginas de internet.

VI. BIBLIOGRAFIA

Internet

- <http://www.ii.uam.es/>
- <http://www.wshoy.sidar.org/>
- <http://www.w3.org/2001/sw/>
- <http://www.oracle.com/>

Libros

- Planeta Web 2.0: Inteligencia colectiva o medios fast food, Septiembre 2007, Cristóbal Cobo Romaní y Hugo Pardo.
- Hacia la Web Semántica, 2007, María García Hernández, Facultad de Comunicación y Documentación.
- Análisis de la web semántica, Estado actual y requisitos futuros, 2007, Eduardo Peis Redondo.