



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

## **INDEXADOR WEB A PARTIR DE UNA ONTOLOGÍA EN EL DOMINIO DE LA PYME**

**Juan Pablo Caballeros Torres**

Asesorado por el Ing. Freiry Javier Gramajo López

Guatemala, abril de 2008



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**INDEXADOR WEB A PARTIR DE UNA ONTOLOGÍA EN EL DOMINIO DE LA  
PYME**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**JUAN PABLO CABALLEROS TORRES**

ASESORADO POR EL ING. FREIRY JAVIER GRAMAJO LÓPEZ.

AL CONFERÍRSELE EL TÍTULO DE  
**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, ABRIL DE 2008



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Ing. Glenda Patricia García Saria
VOCAL II	Inga. Alba Maritza Guerrero de López
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
VOCAL V	
SECRETARIA	Inga. Marcia Ivonne Véliz Vargas

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Marlon Antonio Pérez Turk
EXAMINADORA	Inga. Virginia Victoria Tala Ayerdi
EXAMINADOR	Ing. César Augusto Fernández Cáceres
SECRETARIA	Inga. Marcia Ivonne Véliz Vargas



**HONORABLE TRIBUNAL EXAMINADOR**

Cumpliendo con los aspectos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**INDEXADOR WEB A PARTIR DE UNA ONTOLOGÍA EN EL DOMINIO DE LA  
PYME,**

tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, en agosto de 2006.



Juan Pablo Caballeros Torres





Guatemala, 18 de febrero 200

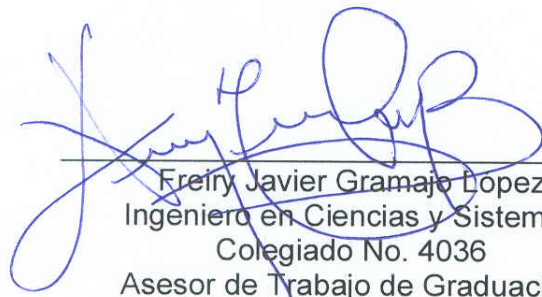
Ingeniero  
Carlos Azurdia  
Revisor de Trabajo de Graduación  
Escuela de Ciencias y Sistemas  
Facultad de Ingeniería

Respetable Ing. Azurdia:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante JUAN PABLO CABALLEROS TORRES, titulado "INDEXADOR WEB A PARTIR DE UNA ONTOLOGÍA EN EL DOMINIO DE LA PYME", y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Sin otro particular, me suscribo de usted.

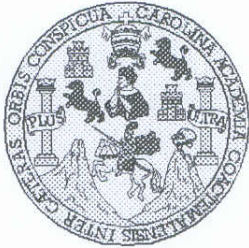
Atentamente,



Freiry Javier Gramajo Lopez  
Ingeniero en Ciencias y Sistemas  
Colegiado No. 4036  
Asesor de Trabajo de Graduación

*Freiry Javier Gramajo López*  
Ing. Ciencias y Sistemas  
Col. 4036  
DEA Intelectual





Universidad San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 12 de Marzo de 2008

Ingeniero  
**Marlon Antonio Pérez Turk**  
Director de la Escuela de Ingeniería  
En Ciencias y Sistemas

Respectable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **JUAN PABLO CABALLEROS TORRES**, titulado: **"INDEXADOR WEB A PARTIR DE UNA ONTOLOGIA EN EL DOMINIO DE LA PYME"**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,

  
**Ing. Carlos Alfredo Azurdia**  
Coordinador de Privados  
y Revisión de Trabajos de Graduación





E  
S  
C  
U  
E  
L  
A  
  
D  
E  
  
C  
I  
E  
N  
C  
I  
A  
S  
  
Y  
  
S  
I  
S  
T  
E  
M  
A  
S

UNIVERSIDAD DE SAN CARLOS  
DE GUATEMALA



FACULTAD DE INGENIERIA  
ESCUELA DE CIENCIAS Y SISTEMAS  
TEL: 24767644

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, de trabajo de graduación titulado **"INDEXADOR WEB A PARTIR DE UNA ONTOLOGIA EN EL DOMINIO DE LA PYME"** presentado por el estudiante **JUAN PABLO CABALLEROS TORRES**, aprueba el presente trabajo y solicita la autorización del mismo.*

**"ID Y ENSEÑAD A TODOS"**

  
Ing. Marlon Antonio Pérez Turk

Director, Escuela de Ingeniería en Ciencias y Sistemas



Guatemala, 16 de abril 2008



Universidad de San Carlos  
de Guatemala



Facultad de Ingeniería  
Decanato

Ref. DTG.111.08

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al trabajo de graduación titulado: **INDEXADOR WEB A PARTIR DE UNA ONTOLOGÍA EN EL DOMINIO DE LA PYME**, presentado por el estudiante universitario **Juan Pablo Caballeros Torres**, autoriza la impresión del mismo.

IMPRÍMASE.

  
Ing. Murphy Olympo Paiz Recinos  
DECANO

Guatemala, abril de 2008



/cc  
c.c. archivo.





## **ACTO QUE DEDICO A:**

<b>Dios</b>	Porque no me ha dejado solo en el caminar de mi vida, porque siempre a puesto a las personas indicadas a mi lado, sin él, nada de esto fuera posible.
<b>Mis Padres</b>	Arnoldo Caballeros, Verónica de Caballeros, porque siempre han estado ahí, cuando los he necesitado, siempre he recibido lo mejor de ellos, este logro es especialmente para ustedes.
<b>Mis Hermanos</b>	Erick y Claudia Caballeros, por sus consejos y apoyo incondicional.
<b>Mis Abuelitas</b>	Olga Tejeda, Isabel Escobar, por tenerme siempre en sus oraciones.
<b>Mónica Gabriela López</b>	Por todo su amor, y comprensión a lo largo de estos años.
<b>Mis Sobrinos</b>	Adriana Ximena y Erick Sebastián Caballeros, con mucho amor.
<b>Freiry Javier Gramajo López</b>	Por su amistad y enseñanzas durante estos últimos años.



## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	III
GLOSARIO.....	V
RESUMEN.....	IX
LÍMITES.....	XI
OBJETIVOS .....	XIII
INTRODUCCIÓN.....	XV
<b>1. PYME .....</b>	<b>1</b>
1.1. Definición.....	1
1.2. Clasificación .....	1
1.3. Importancia de PYME y TICs .....	6
1.4. Leyes en Guatemala .....	17
1.5. Instituciones y Organizaciones que brindan apoyo a la PYME .....	20
<b>2. ONTOLOGÍAS.....</b>	<b>29</b>
2.1 Web Semántica .....	29
2.2 Redes Semánticas .....	30
2.3 Ontologías .....	35
2.4 Herramientas.....	40
2.5 Búsquedas Inteligentes .....	41
<b>3. MACHINE LEARNING (APRENDIZAJE COMPUTACIONAL).....</b>	<b>47</b>
3.1 Definición:.....	47
3.2 Tipos de Aprendizajes .....	48
3.2.1 Clasificación .....	49
3.2.2 <i>Clustering</i> .....	54
3.3 Repositorios de Datos .....	56
3.4 Datos sintéticos .....	58
3.5 <i>Missing Values</i> (Valores Faltantes).....	60
3.6 Herramientas GPL.....	61

<b>CONCLUSIONES.....</b>	<b>63</b>
<b>RECOMENDACIONES .....</b>	<b>65</b>
<b>BIBLIOGRAFÍA.....</b>	<b>67</b>
REFERENCIAS ELECTRÓNICAS.....	69
<b>APÉNDICE .....</b>	<b>73</b>
A.1.    Utilización de OntoPyme.....	73
A.1.1    Requerimientos Mínimos.....	73
A.1.2    Instalación.....	73
A.1.3    Pantalla Principal.....	74
A.1.4    Cargar una Ontología .....	75
A.1.5    Guardar una Búsqueda .....	75
A.1.6    Cargar una Búsqueda.....	76
A.1.7    Mantenimiento de Atributos .....	77
A.1.8    Generar ARFF.....	78
A.1.9    Pantalla de Resultados.....	80
A.2.    Manual Técnico.....	82
A.2.1    UI.....	83
A.2.2    Clases.....	90

## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1	Número de personas que cuentan con una computadora agrupado por etnia.....	10
2	Diferentes eras del comercio electrónico.....	11
3	Red semántica del dominio vehículo.....	31
4	Resultados de la búsqueda realizada el 26 de abril de 2007, con el buscador www.google.com.....	32
5	Resultado de la búsqueda.....	33
6	Resultado de búsqueda.....	33
7	Representación gráfica de una ontología. OntoEdit versión 2.6.....	36
8	Ejemplo de funcionamiento de un árbol de decisión.....	52
9	Ejemplo de un archivo ARFF.....	57
10	Modelo entidad relación de la base de datos sintética para el ejemplo descrito.....	59
11	Pantalla que muestra el fin de la instalación.....	74
12	Pantalla principal de OntoPyme.....	74
13	Pantalla que muestra búsquedas previamente almacenadas.....	76
14	Pantalla que muestra el mantenimiento de conceptos.....	77
15	Selección de atributos para la generación del archivo arff.....	78
16	Selección de la ubicación y nombre del archivo a generar.....	80
17	Muestra de los resultados de la búsqueda.....	81
18	Almacenamiento de búsquedas.....	82

## TABLAS

I. Datos crecimiento.....	9
II. Tipos de aprendizaje computacional .....	48
III. Ejemplo de datos que conformarían un árbol de decisiones. ....	51

## GLOSARIO

<b>Clicks and Bricks</b>	Término utilizado para todas aquellas empresas que poseen un sitio web, en el cual comercializan sus productos, pero que de igual forma poseen un lugar físico donde las personas pueden acudir a comprar.
<b>E-Commerce</b>	Comercio Electrónico
<b>Empresa Dot.com</b>	Término utilizado para todas aquellas empresas que solo existen a través de una página web. Es decir que éstas no cuentan con una tienda física, en la cual puedan vender sus productos.
<b>Html</b>	Que por sus siglas en inglés significa <i>HyperText Markup Language</i> o lenguaje de marcación de hipertexto. Este lenguaje es el que actualmente se utiliza en la Red Mundial (WWW). Fue creado con el fin de poder presentar textos como hipertexto.
<b>Inferencia</b>	Se entiende por Inferencia, al proceso en cual se razona con base a actividades u eventos anteriores, con el fin de concluir una proposición relacionada con el tema.
<b>Intranet</b>	Red local o privada que es utilizada usualmente, para el manejo de la comunicación interna dentro de

una empresa. Estas, por lo general, utilizan los mismos estándares y herramientas del Internet.

<b>Metadato</b>	Tipo de dato que se utiliza para describir información, contenido y alguna otra peculiaridad o característica especial de algún dato o contenido en una página Web.
<b>Metalífero</b>	adj. Que contiene metal
<b>PHP</b>	Hypertext Preprocessor, lenguaje de programación que permite crear contenido dinámico que puede interactuar con una o varias base de datos a la vez.
<b>RDF</b>	<i>Resource Description Framework</i> , lenguaje que especifica un modelo de metadatos, implementado como una aplicación XML.
<b>RDFS</b>	El <i>Resource Description Framework Schema</i> , es un lenguaje utilizado para la descripción de un vocabulario en RDF.
<b>Repositorio</b>	Sitio centralizado que almacena diferentes tipos de información. Existen Repositorios de imágenes, de música etc.
<b>Semántica</b>	Visto desde el punto de vista de la informática, se considera a la Semántica como la información



contenida en los datos. En otras palabras, el sentido de los datos.

<b>Taxonomía</b>	Estructura jerárquica, orientada a la clasificación de objetos de un mismo dominio
<b>Telemática</b>	Es el estudio, gestión así como el diseño de todas las redes y servicios de comunicación de datos. Facilitando así, el transporte de texto video y audio.
<b>Telepresencia</b>	Estar ser transportado “físicamente” de un lado a otro. Es decir, que por medio de las telecomunicaciones y la robótica, una persona pueda estar en otro lugar representado por una máquina o bien un robot [1].
<b>Workflow</b>	Flujo de trabajo, se entiende como flujo de trabajo un diagrama en el cual se indiquen los pasos o bien procesos a ejecutar, para poder llegar a un objetivo en específico.
<b>TICs</b>	Tecnologías de la información y de la comunicación.
<b>XML</b>	Que por sus siglas en inglés es <i>Extended Markup Language</i> , o bien Lenguaje de Marcas Extensible, desarrollado por World Wide Web Consortium (W3C). El XML es un lenguaje práctico que se utiliza primordialmente para la creación de lenguajes.



## **RESUMEN**

Los medios de comunicación, la forma de almacenar la información, la interpretación y la representación de ésta, han ido evolucionando con el tiempo. Actualmente, es tanta la importancia de ello que tanto la pequeña y la mediana empresa PYME, deben de aprovechar el alcance de los medios de comunicación actuales, como la Web, con el propósito de vencer los obstáculos que trae consigo la globalización.

Es por ello que se implementa una aplicación, la cual es capaz de tener un repositorio con información o datos, de las pequeñas y medianas empresas en Guatemala. Dicha aplicación tiene clasificadas a la pequeña y mediana empresa por su actividad económica, utilizando una ontología, la cual es capaz a su vez, de indexar documentos relacionados con las actividades económicas de cada empresa. De tal forma que al momento de realizar una búsqueda, conforme a la ontología, también se pueda mostrar cualquier documento publicado en la Web, en formato TXT, PDF, PPT, XLS o bien HTML relacionado, que contenga información primordial acerca de las PYMEs en Guatemala.



## LÍMITES

1. La ontología no es incremental de forma automática.
2. Para la clasificación de las PYMEs se utilizará una ontología ya existente.
3. Las búsquedas en el internet están limitadas por el Motor de Búsqueda Google.



## **OBJETIVOS**

### **GENERAL**

Manejar una ontología en el dominio de las PYMEs, y realizar búsquedas de documentos previamente indexados a partir de la ontología, siendo éstos una instancia de la ontología.

### **ESPECÍFICOS**

1. Definir los conceptos básicos de las PYMEs y su relación con las TICs en Centroamérica.
2. Manejo de una ontología, previamente desarrollada, del dominio de las PYMEs y sus actividades económicas.
3. Utilización de la clasificación por sector económico de la PYME, propuesta por la ONU utilizada en una ontología.
4. Realizar búsquedas en el internet, dependiendo de los resultados obtenidos a partir de la ontología.
5. Manejo de una base de datos para almacenar los índices generados por cada documento indexado.
6. Implementación de una aplicación, que sea capaz de realizar búsquedas a la ontología, luego realizar búsquedas en el Internet, para finalmente indexar, y clasificar los documentos (links) encontrados.
7. Utilizar como lenguaje de programación, para la aplicación a desarrollar, Java.





## INTRODUCCIÓN

Las formas de comunicación han ido evolucionando con el paso del tiempo, en conjunto con las formas de hacer negocios. Hoy en día, las pequeñas y medianas empresas, tanto de Guatemala como de Centroamérica, deben de estar preparadas para lo que se avecina, es decir, la globalización. Por lo general, las PYMEs son un factor relevante en la economía de los países antes centroamericanos, debido a que, está registrado que éstas generan más del 20% del Producto Interno Bruto en sus respectivos países<sup>1</sup>. Es común pensar que el desarrollo de una PYME se vea afectado directamente por su tamaño y el capital que esta posee o bien, el capital que ésta genera. Sin embargo, uno de los factores principales que limitan el crecimiento, o bien el sostenimiento de una PYME, es que éstas, por lo general, se encuentran excluidas, y no cuentan con un soporte o colaboración que las ayude.

La tecnología es un factor que puede definitivamente, ayudar a que la pequeña y mediana empresa se logre desarrollar, no solo en un mercado local, sino que también en los mercados internacionales. Cabe resaltar que la tecnología como tal, no es directamente la solución, para esto debe de haber un cambio cultural en las personas, de tal forma que éstas puedan ver a la tecnología como un medio o herramienta para el desarrollo.

Por otra parte, existen ramas de la Inteligencia Artificial, las cuales pueden ser de gran apoyo al momento de construir sistemas informáticos, por ejemplo, sistemas completos de búsqueda de empresas, o bien sistemas de comercio electrónico. El uso de Ontologías, así como el uso de técnicas de Aprendizaje Computacional para la mejor clasificación de las categorías de la

---

<sup>1</sup> La PYME en Centroamérica y su vinculación con el sector externo, CEPAL.

PYME, o como es el caso del presente trabajo, la clasificación por actividad económica, según la ONU.

El presente documento proporciona información que resalta la importancia de las PYMEs en conjunto con la utilización de TICs, implementando así, una aplicación llamada "OntoPyme", la cual es capaz de indexar, por medio de una Ontología, documentos que contengan información de actividades y servicios relacionados con cualquier pequeña o mediana empresa, facilitando así, la búsqueda y clasificación de los resultados. De tal forma que, teniendo los resultados de las búsquedas almacenados y clasificados en una base de datos, es posible la generación de un archivo de tipo ARFF, el cual puede ser analizado por cualquier herramienta que implemente algoritmos de Aprendizaje Computacional y que, por supuesto, sea capaz de interpretar dicho formato de archivo. OntoPyme es idea original del Ing. Javier Gramajo López, por lo que dicha idea es propiedad intelectual del autor de acuerdo a las leyes Guatemaltecas relacionadas con Propiedad Intelectual, Artículos 5 y 6, Capítulo I, Artículos 11, 30 a 35, Decreto 33-98, del 28 de abril de 1998.

# **1. PYME**

## **1.1. Definición**

La PYME, que por sus siglas significa Pequeña y Mediana Empresa, no posee una definición, la cual es aceptada de manera global, ya que en cada país se pueden adoptar criterios diferentes en torno a las PYMEs y su clasificación. Para algunas personas o bien países, lo importante al momento de clasificar a la PYME son, ya sea el número de empleados con ésta cuenta, la cantidad de capital que ésta pueda generar y/o que posea etc [2]. Además las PYMEs pueden pertenecer a diversos sectores, tales como el agrícola, el agroindustrial, manufactura y diversos servicios, por lo que hace un poco más difícil su clasificación y su definición. Por lo tanto, podemos decir que su clasificación está basada tanto en términos cualitativos como cuantitativos, es decir características y cantidades.

En Centroamérica se considera a una empresa como una PYME si y solo si, ésta esta formada por menos de 100 empleados<sup>2</sup>.

## **1.2. Clasificación**

Para efectos del presente trabajo, las PYMEs serán clasificadas de acuerdo a la división que maneja la ONU en relación con la Clasificación Industrial Internacional Uniforme (CIIU) de todas las actividades económicas. Las cuales han sido desarrolladas a través de los años por un grupo de expertos en clasificaciones económicas internacionales y sociales. La parte encargada, en la ONU, de dar seguimiento a esto es, en la actualidad, la División de

---

<sup>2</sup> El Entorno Financiero de la Microempresa en Guatemala, Universidad Rafael Landívar, Instituto de Investigaciones Económicas y Sociales.

Estadística de las Naciones Unidas y la Clasificación Central de Productos CCP<sup>3</sup>. Este es un trabajo que se ha llevado a cabo ya desde hace varios años, y que constantemente se ha ido actualizando, por ejemplo en el año 2002 se realizó una nueva actualización, agregando así nuevas categorías correspondientes al sector de la información, específicamente al sector de las tecnologías de la información y comunicaciones [3]. Todos estos cambios y actualizaciones que son realizadas de forma periódica, así como también son revisados de forma estricta por un grupo de expertos en clasificaciones Económicas y Sociales de la ONU [4]. De modo que no es un proceso simple, que además a sido un esfuerzo en conjunto ya de varios años atrás. Es por eso que se decidió tomar dicha clasificación para la elaboración de la ontología.

La clasificación propuesta por las Naciones Unidas es la siguiente:

- Actividad
- Categoría de la actividad
- División de la actividad
- Grupo de la actividad
- Producto

Por actividad se refiere, como su nombre lo indica, a lo que la empresa se dedica a hacer, su mercado, su propósito etc. Por ejemplo fabricación de alimento para animales, transporte de carga, cría de animales entre otros.

La categoría de la actividad es la agrupación de tipo lógica, en términos generales, de las actividades. El detalle no es importante en esta sección. Por ejemplo, se puede mencionar la Agricultura, Construcción, Enseñanza, entre otras.

---

<sup>3</sup> Clasificación Central de Productos (CPC); Versión 1.0;

Al decir división de la actividad, se refiere a una agrupación siempre lógica de las actividades, pero aun nivel más detallado que el nivel descrito anteriormente. Por ejemplo en este nivel podemos encontrar categorías como Extracción de minerales de uranio, Extracción de minerales metalíferos, Reciclamiento, Suministro de Gas, entre otros [5].

Por Grupo de la Actividad se entiende a las mismas categorías pero mas específicas aún de lo que el nivel anterior muestra. Es decir, si seguimos con las categorías antes mencionadas, un ejemplo sería el siguiente, Extracción de minerales metalíferos no ferrosos, excepto los minerales de uranio y torio, entre otros.

El último elemento de la clasificación es el producto como tal.

En esta clasificación se han incluido todas aquellas actividades relacionadas con las TICs, según la ONU, lo cual ha sido siempre un tema de debate, ya que las clasificaciones y divisiones creadas, pueden ser interpretadas de diferente forma, dependiendo siempre del punto de vista desde el cual se esté analizando, es por eso que la Organización de Cooperación y Desarrollo Económico ha impulsado, en los últimos años, la estandarización de todo aquel contenido que referente al sector de las TICs. De tal forma que, la ONU tomó la decisión de la creación de los siguientes principios.

Si se trata de actividades de fabricación, sus productos deben tener por función del procesamiento y la comunicación de información, incluidas su transmisión y su presentación visual, o deben utilizar medios de procesamiento electrónico para detectar, medir y/o registrar fenómenos físicos o controlar un proceso físico. En cambio si se tratara de actividades de servicios, sus

productos deben tener por objeto posibilitar el procesamiento de información y su comunicación por medios electrónicos [6].

De modo que en base a dichos principios se crearon las siguientes divisiones, con el fin de evitar conflictos, ambigüedades y malas interpretaciones.

- FABRICACIÓN DE MAQUINARIA DE OFICINA, CONTABILIDAD E INFORMÁTICA
  - Fabricación de maquinaria de oficina, contabilidad e informática
- FABRICACIÓN DE MAQUINARIA Y APARATOS ELÉCTRICOS N.C.P.
  - Fabricación de hilos y cables aislados
- FABRICACIÓN DE EQUIPO Y APARATOS DE RADIO, TELEVISIÓN Y COMUNICACIONES
  - Fabricación de tubos y válvulas electrónicos y de otros componentes electrónicos
  - Fabricación de transmisores de radio y televisión y de aparatos para telefonía y telegrafía con hilos
  - Fabricación de receptores de radio y televisión, aparatos de grabación y reproducción de sonido y vídeo, y productos conexos
- FABRICACIÓN DE APARATOS E INSTRUMENTOS MÉDICOS Y DE APARATOS PARA MEDIR, VERIFICAR, ENSAYAR, NAVEGAR Y OTROS FINES, EXCEPTO INSTRUMENTOS DE ÓPTICA
  - Fabricación de aparatos e instrumentos médicos y de aparatos para medir, verificar, ensayar, navegar y otros fines, excepto instrumentos de óptica

- Fabricación de equipo de control de procesos industriales
- COMERCIO AL POR MAYOR Y EN COMISIÓN, EXCEPTO EL COMERCIO DE VEHÍCULOS AUTOMOTORES Y MOTOCICLETAS
  - Venta al por mayor de ordenadores, equipo periférico y programas de informática
  - Venta al por menor de partes y equipo electrónicos y de comunicaciones
- CORREO Y TELECOMUNICACIONES
  - Telecomunicaciones
- ALQUILER DE MAQUINARIA Y EQUIPO SIN OPERARIOS Y DE EFECTOS PERSONALES Y ENSERES DOMÉSTICOS
  - Alquiler de maquinaria y equipo de oficina (incluso ordenadores)
- INFORMÁTICA Y ACTIVIDADES CONEXAS
  - Consultores en equipo de informática
  - Edición de programas de informática
  - Otras actividades de consultoría y suministro de programas de informática
  - Procesamiento de datos
  - Actividades relacionadas con bases de datos y distribución en línea de contenidos electrónicos
  - Mantenimiento y reparación de maquinaria de oficina, contabilidad e informática
  - Otras actividades de informática

### 1.3. Importancia de PYME y TICs

El tema de las TICs se ha vuelto muy importante en los últimos años, por lo que día a día se ha vuelto mas actual el estudio de éstas y el impacto que se tiene sobre la sociedad y sobre todo, sobre la forma de hacer negocios. Ya que está comprobado que un número creciente de empresas están aumentando su competitividad al aplicar las TICs en los procesos empresariales [7].

Las TICs no son más que las tecnologías que son utilizadas para la gestión y transformación de la información. Es decir que éstas habilitan el acceso a los diferentes medios de comunicación, así como también facilitan la capacidad de poder contribuir a dichos medios con el fin de poder generar mayor conocimiento o bien brindar información más actualizada de forma rápida. Por lo tanto, es posible generar y/o gestionar conocimiento, permitiendo así el desarrollo de diversos tipos de actividades económicas, sociales, educativas, sanitarias, culturales, científica entre otras [8].

Miratia menciona en su artículo "La Tecnologías de la Información y la Comunicación en la Educación" publicado en el año 2005, por la revista nº4 de Infobit, hace referencia a Garcias (1996), Bartolomé (1989) y Cabero (1996), quienes se dieron la tarea de agrupar a las TIC en tres grandes sistemas de comunicación: el video, la informática y la telecomunicación [9]. Esta clasificación abarca un gran número de diferentes tecnologías y medios de comunicación, entre los cuales podemos mencionar:

- El video interactivo
- El videotexto
- El teletexto
- La televisión por cable y satelital



- La Web
- Los sistema multimedia
- Las teleconferencia en sus distintos formatos (audio conferencia, videoconferencia,
- Los sistemas expertos
- La realidad virtual
- La telemática
- La telepresencia.

La aceptación de dichas tecnologías, ha aportado eficacia a la exportación de bienes y servicios (un claro ejemplo de esto es la implementación del sistema SEADDEX, el cual se habla más adelante en éste mismo capítulo). De igual forma han abierto nuevos mercados para la pequeña y la mediana empresa, creando a su vez nuevas oportunidades de empleos, especialmente para el sector informático.

Como vemos, el uso de TICs conlleva grandes beneficios. Según un artículo publicado en el fórum del comercio [7], el impacto digital que se obtiene es bastante grande y de gran ayuda para los pequeños exportadores de los países en vías de desarrollo como el nuestro. En dicho artículo se mencionan ciertos aspectos y ventajas del uso de las TICs. A continuación se enumeran los puntos considerados como los más importantes.

- Estudios de Mercado
  - Análisis de Mercado en Línea: Es decir, que por ejemplo, que es posible realizar una encuesta de forma rápida y económica teniendo mayor un mayor alcance.
  - Búsqueda de compradores y vendedores en línea
- Procesos Empresariales

- Uso de TICs para agilizar las cadenas de abastecimiento
- Digitalización de procesos de producción
- Mercadotecnia digital por distintos canales
- Gestión en línea del Transporte.
- Documentos de Exportación en línea
- Pagos/Transacciones
  - Ventas y pagos totalmente en línea
  - Comprobación de solvencia de compradores.
- Relaciones con el cliente
  - Servicios postventa en línea las 24 horas (de ser posible)
  - Bases de datos para gestión de clientes.
- Nuevos Mercados
  - Poder participar en mercados virtuales de empresas y realizar negocios en línea tipo B2B o bien B2C.

El atraso tecnológico, o en otras palabras, la Brecha Digital (Indicador del nivel de atraso a nivel tecnológico de los países en vías de desarrollo, en comparación de los países desarrollados), limita de forma significativa que éstas tecnologías, se apliquen de forma eficiente y rápida. Aunque es de reconocer que, por ejemplo, el número de usuarios de Internet y en sí, la utilización de las TICs en América Latina, crece año con año, según un informe publicado por la CEPAL<sup>4</sup>. La siguiente figura muestra los datos relacionados con la telefonía móvil en Centroamérica y el Caribe.

---

<sup>4</sup> Los Caminos hacia una Sociedad de la Información en América latina y el Caribe, CEPAL

**Tabla I. Datos crecimiento**

Teléfonos Fijos		Teléfonos Móviles		Computadores		Usuarios Internet	
Tasa Penetración	Crec Ac % 1999-2004	Tasa Penetración	Crec Ac % 1999-2004	Tasa Penetración	Crec Ac % 1999-2004	Tasa Penetración	Crec Ac % 1999-2004
17,8	38,5	31,6	304,6	8,9	135,5	11,4	467,1

**Fuente:**

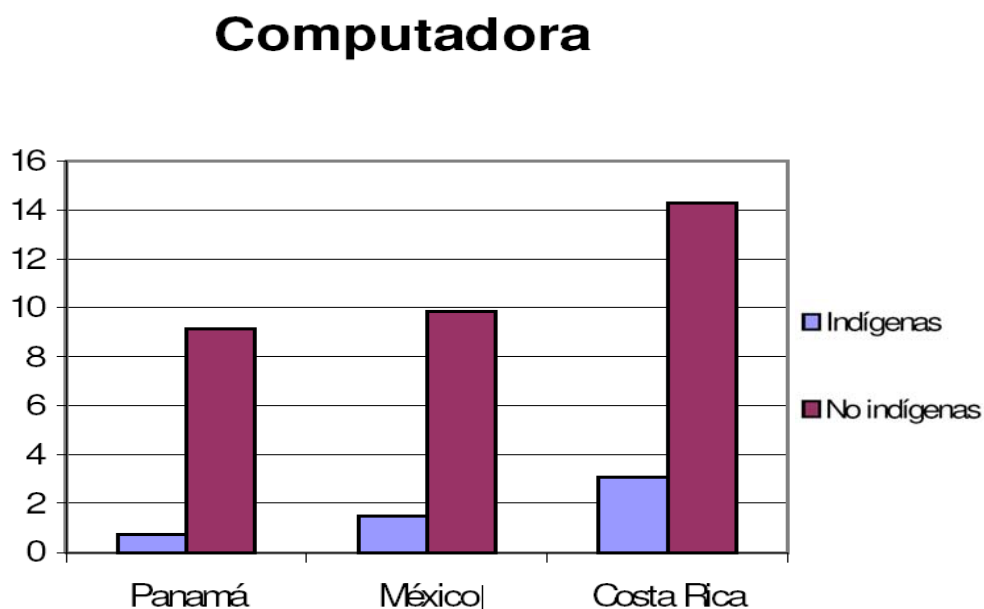
<http://www.eclac.cl/prensa/noticias/comunicados/4/22964/HojasInformativaAccesoBrechaDig02.pdf>

En la tabla se muestran los 4 indicadores básicos con los cuales se mide el acceso a las TIC's, lo que de una u otra forma, contribuye al cálculo de la brecha digital en un país. Los datos son exclusivamente para Centroamérica y el Caribe, y se muestra un crecimiento del 38.5% en el uso de telefonía fija, así como un crecimiento acumulado del 304% en telefonía móvil. De igual forma el uso de computadoras y los usuarios de internet crecieron significativamente en los años de 1999 al 2004.

El grado de urbanización de los países es un factor importante que afecta la brecha digital, ya que está comprobado que aquellos países que cuentan con un grado de urbanización alto, tienen un uso mucho mayor de TICs. Esto se puede observar fácilmente, como se menciona en el documento de CEPAL antes mencionado, donde la continuidad de la tendencia a la urbanización se ve reflejada directamente en las cifras sobre el uso y por supuesto la aplicación de las tecnologías de la información y comunicación en las comunidades modernas. Uno de los factores importantes, donde se explica este fenómeno es que en las economías de escala, las cuales se derivan primordialmente de la producción y el uso de la infraestructura (que por lo general los costos fijos, así como los precios son un poco altos) pueden ser aprovechadas de mejor manera en los centros urbanos, más que en los centros rurales.

Otro factor importante que afecta directamente la Brecha Digital, el cual es importante mencionar, es el pertenecer a un tipo de etnia diferente al predominante en el país. Ya que como lo muestra la gráfica siguiente, la probabilidad de tener una computadora en un hogar no indígena (en este caso) es cinco veces mayor que en el conjunto indígena<sup>5</sup>.

Figura 1 Número de personas que cuentan con una computadora agrupado por etnia



**Fuente:** Bases de micro datos de Panamá, México y Costa Rica, ronda de censos 2000  
Los Caminos hacia una sociedad de la información en América Latina y el Caribe, CEPAL.

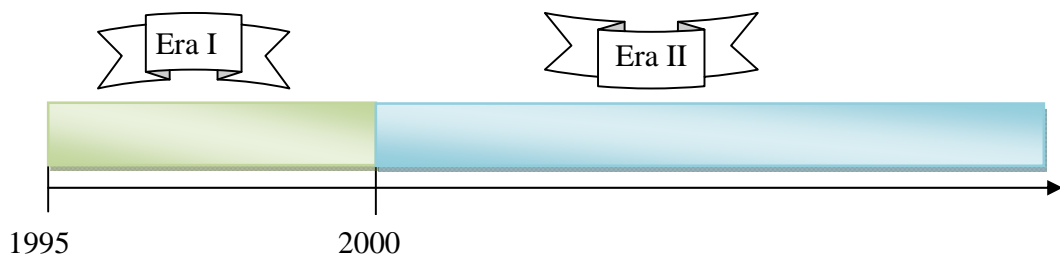
Una consecuencia o ventaja, si se le puede llamar de esa forma, del uso correcto de las TICs, es el comercio electrónico.

El comercio electrónico en sus comienzos no fue como lo es ahora. Se entiende por *Era I del Comercio electrónico*, el período comprendido de finales

<sup>5</sup> Informe 2000 (LC/G.2125-P), Santiago de Chile, 2001. Publicación de las Naciones Unidas, N° de venta: S.01.II.G.12

del año 1995, hasta aproximadamente el año 2000 con el colapso de las acciones en el mercado de valores de las empresas *dot.com*.

**Figura 2 Diferentes eras del comercio electrónico**



Esta es conocida también como la era oscura del comercio electrónico. Ya que, durante estos 6 años, se realizaron muchas estafas por parte de los proveedores de servicio a sus clientes, cobrando así grandes cantidades de dinero por sitios que no eran 100% funcionales o peor aún, 100% seguros, además fue en estos primeros años cuando se realizaron las mayores estafas y robos de información, tanto, dentro de los sistemas, como robo de información en línea, utilizando técnicas como el *Phishing* (término utilizado para denominar el uso de un tipo de ingeniería social, que se caracteriza por intentar adquirir una contraseña, número de tarjeta de crédito, cuenta bancaria o cualquier otra información importante de las personas) además no existían entes reguladores o certificadores de sitios (como el caso de *VeriSign* [10], ni mucho menos leyes internacionales que regularan o bien estandarizaran sus implementaciones. Fue en esta primer era, cuando se generaron las bases de datos que actualmente se encuentran en el internet con números de tarjetas de crédito así como también las bases de datos con contraseñas. Gracias a estos robos fue posible, posteriormente, crear herramientas que incluso generaran números de tarjetas de crédito. Un ejemplo de ello se puede encontrar en el siguiente sitio [[http://rapidshare.com/files/3697063/CC\\_Gen.zip.html](http://rapidshare.com/files/3697063/CC_Gen.zip.html)].

Otro factor que influyó de igual forma, fue que prácticamente el acceso al internet era bastante limitado (como se ha mencionado en la sección anterior).

Algunos aspectos positivos que se tuvieron durante la Era I del comercio electrónico, fueron entre otros, que se logró un gran éxito tecnológico, creando así la mayor parte de la infraestructura digital, la cual, fue lo suficientemente sólida para sustentar el rápido crecimiento que éste tendría en la su Era II.

En resumen, comienzo de los negocios en el internet, durante la Era I, no fue como se esperaba, hubo mucha pérdida de credibilidad por parte de los usuarios, debido al robo de información personal, y por parte de los empresarios también, pero aún así se lograron establecer los cimientos para lo que hoy conocemos y utilizamos a diario.

La Era II del comercio electrónico, que se podría decir, es la que estamos viviendo, han cambiado varias cosas, actualmente ya existen leyes concernientes al comercio electrónico, así como también estándares de implementación para sitios de dicha índole. De la Era II del comercio electrónico se espera lo siguiente<sup>6</sup>:

- La tecnología de E-Commerce continuará propagándose a través de todas las actividades comerciales y económicas de las empresas.
- Los precios manejados en el E-Commerce crecerán hasta cubrir los costos reales de estar haciendo negocios en la web.

---

<sup>6</sup> ECommerce: Bussines, Technology, Society. Kenneth C. Laudon, Carol Guercio Traver. Addison Wesley Editorial. Primera Edicion, 2002.

- El número de compañías online con éxito continuará declinando y el E-Commerce en las empresas, se seguirá utilizando, al adoptar una estrategia “clicks and bricks”.
- La regulación del E-Commerce y la Web por el gobierno crecerá tanto como en los Estados Unidos y en todo el mundo.

Pero con el paso del tiempo, el comercio electrónico, en el mundo y en Guatemala especialmente, va recobrado la confianza de los usuarios y empresarios, a tal punto que actualmente, en Guatemala, es posible realizar diferentes pagos de servicios en línea, tal es el caso del pago de la Luz Eléctrica, el cual puede ser realizado desde un cajero automático, o bien totalmente en línea a través del sitio de casi cualquier banco del sistema [<http://www.eegsa.com/facturacion.php>]. Además, ahora en Guatemala ya es posible realizar compras totalmente en línea, en tiendas en el territorio nacional, así como también participar en subastas de cualquier tipo de artículos [<http://www.gubiz.com/>].

Todas estas transacciones, se realizan, de una forma segura, según lo dictan los estándares internacionales, es decir, cuando se habla de realizar una transacción segura en línea, se habla de datos debidamente encriptados mediante algoritmos complejos, además se entiende que una conexión segura, garantiza la autenticidad, la confidencialidad, integridad así como también el no-repudio de los datos.

El protocolo, que es utilizado de forma común es el *Secure Socket Layer* o *SSL*. El *SSL*, encripta la información, además provee autenticación tanto del servidor, como de la máquina que se conecta al servidor. Este protocolo cuando es utilizado en conjunto con HTTP, es conocido como HTTPS (http seguro)<sup>7</sup>.

---

<sup>7</sup> López, Manuel, Desarrollo de Sistemas en Internet. Guatemala: Editorial Universitaria, 1999

Con el fin de resaltar la importancia del uso de las TICs en una PYME, dejando por fuera el tema de las transacciones en línea, podemos mencionar el uso de Intranets [11]. Una intranet no es más que un conjunto de sitios Web, a los cuales solo puede ser accesado dentro de la red local de una empresa, no puede ser visto desde el Internet como un sitio al público. Una intranet en una empresa ayuda a mejorar las comunicaciones, la coordinación y el intercambio de conocimiento a nivel interno, de igual forma reduce costos ya que evita la impresión de papeles para crear comunicados internos, entre otros. Las intranets son un medio efectivo y rápido de colocar al alcance del personal de una empresa, ya sea administrativo u operativo, la información necesaria, para que éstos puedan desarrollar su trabajo de mejor manera, resolviendo así los problemas de forma mas rápida, de igual forma mejorando procesos internos o incluso construyendo nuevos recursos. La intranet también cumple con el objetivo de divulgar información de manera rápida y convertir a los empleados en partes activas de una red corporativa de conocimiento, es decir, dar al usuario la información que este necesita para su trabajo y compartirla con los demás.

Una intranet pretende que cada cual tenga la información necesaria en el momento oportuno sin que tenga que recurrir a terceros para conseguirla.

Luego de analizar las ventajas intrínsecas del Comercio Electrónico o E-Commerce y sus aplicaciones, se puede analizar las ventajas que éste brindaría a una PYME. El hecho de que una PYME cuente con una página de Internet, es un gran paso, claro esta que el tener una página de internet no es Comercio Electrónico, se consideraría E-Commerce si y solo si, por medio de esta página



se realiza algún tipo de transacción en donde se involucre algún pago por cualquier tipo de servicio o bien adquirido.

Contar con un sitio Web es la única forma en que el mundo se pueda enterar de su existencia, de lo contrario podría decirse que dicha PYME tiene muy escasa su probabilidad de crecer y darse a conocer para el mercado externo o para posibles compradores que utilicen diferentes medios tecnológicos de comunicación.

Con el Internet, prácticamente se podría decir que no existen barreras, y si no se cuenta con un sitio en la Red, entonces las oportunidades de crecimiento de la empresa se ven aún mas limitadas, no se dice que es imposible que una PYME pueda desarrollarse sin contar con uno, pero sería de gran beneficio, si ésta tuviese como mínimo un sitio en la Web en el cual se describa la actividad de la empresa y la forma en la que cualquier persona o empresa o institución pudiera contactarla. Lo recomendable sería, que las PYMEs contaran con un sitio en el cual se puedan realizar negocios en línea y no solo una página Web. De ser así, sus oportunidades se incrementarían aún más, debido a que no solo tendrá la facilidad de ser conocido a nivel mundial, sino que además podrá realizar negocios, de forma instantánea. Pero no es tan sencillo llegar a realizar negocios en línea, se requiere de infraestructura, así como también una buena logística de parte de la empresa para poderse organizar y así poder cubrir su demanda. Para el caso de Guatemala, existe aún otro factor que impide que dichos cambios se den, y no se refiere tanto a infraestructura, sino que a los problemas culturales y la falta de información de muchos pequeños empresarios. Para poder llegar a pensar en infraestructura, es necesario antes que todo, lograr cambiar el pensamiento de las personas, más que todo en el ámbito rural, haciéndoles ver las grandes ventajas así como creándoles la necesidad de la utilización de tecnología, tanto para sus procesos

productivos como comerciales. Una vez hecho esto, es posible comenzar a pensar en infraestructura y apoyo, de lo contrario los esfuerzos serán en vano.

Como todos ya sabemos, Guatemala, como muchos otros, es un país en vías de desarrollo. En el año 2004 contaba ya con 3.168 millones de usuarios de telefonía celular y contaba con 756,000 usuarios de internet [12].

En el año 2005, el 94.8% de la telefonía celular en Guatemala era completamente Digital, en comparación con la telefonía en Costa Rica, donde el 100% de la telefonía era digital. El mismo cuadro se planteaba para Nicaragua, en cambio en Honduras, apenas se contaba con un 26% de líneas Digitales. Esto demuestra que Guatemala, a nivel Centroamérica no se encuentra tan atrasado de sus países hermanos [13].

Según datos de “Telefónica”, una de las compañías más grandes de telefonía celular en Guatemala, para el año 2007 contaban con un aproximado de dos millones de clientes, habiendo realizado una inversión de aproximadamente \$650 millones durante cuatro años de operaciones en el país. Además, tienen proyectada una inversión de \$80 millones para el año 2008, siendo esta no solo para líneas, sino que también otros servicios, como el internet inalámbrico [14].

#### 1.4. Leyes en Guatemala

En Guatemala, el Comercio Electrónico cuenta desde el 23 de enero del año 2001 con iniciativa de ley en el congreso (cabe mencionar que en esa época la iniciativa de ley, no contó con el suficiente apoyo por parte de los legisladores para que ésta fuera aplicada con rapidez), la cual es la encargada de velar por la regulación y control del Comercio Electrónico en Guatemala. Esta ley también es la encargada de velar para sentar las bases para su desarrollo y sostenimiento. No fue hasta principios del año 2006, cuando esta iniciativa comenzó a tomar fuerza y comenzó a ser apoyada mas que todo por el sector empresarial al ver el gran alcance y la gran oportunidad que existen en el comercio electrónico para las pequeñas y medianas empresas en Guatemala [15].

Esta iniciativa de ley, ya esta comenzando a dar sus frutos, el Banco Interamericano de Desarrollo *BID*, a través del *Fomin* (Fondo Multilateral de Inversiones), comenzó a impulsar, a finales del año 2005 un proyecto de comercio electrónico en conjunto con la Cámara de Comercio de Guatemala [16].

El proyecto [17] tiene como fin principal, poder incrementar la productividad y a su vez la eficiencia de las pequeñas y medianas empresas en Guatemala, utilizando claro esta, la gran gama de herramientas y tecnologías de comercio electrónico, así como también las tecnologías de la información y comunicación. El proyecto esta planteado desde el 17 de diciembre del 2004 y como uno de los objetivos del proyecto, se puede mencionar: El generar la confianza en la población al momento de realizar cualquier tipo de transacción en línea.

Esa confianza se va generar en la población creando, claro está, las condiciones reglamentarias y las operaciones favorables, tanto para los empresarios, como para los usuarios, de modo que todos puedan acceder y utilizar dichas tecnologías de forma fácil y sencilla. Además es necesario hacer conciencia y poder capacitar, no solo a las pequeñas y medianas empresas, sino que a los usuarios en general, en el uso de TICs, y en el comercio electrónico.

En marzo del 2007 fue entregada al Congreso una ley para autorizar las firmas electrónicas, con el único fin de poder hacer más fácil y seguras las transacciones monetarias [18].

En nuestro país, la Ley General de Telecomunicaciones, establece en su Artículo número uno, el establecer un marco legal para desarrollar las actividades de telecomunicaciones así como también normar el aprovechamiento y la explotación del espectro radioeléctrico con la finalidad de apoyar y promover el desarrollo eficiente de las telecomunicaciones [19].

Por lo que la Superintendencia de Telecomunicaciones *SIT*, es la encargada de:

- Crear, emitir, reformar y derogar sus disposiciones internas, las que deberán ser refrendadas por el Ministerio;
- Administrar y supervisar la explotación del espectro radioeléctrico.
- Administrar el Registro de Telecomunicaciones.
- Dirimir las controversias entre los operadores surgidas por el acceso a recursos esenciales
- Elaborar y administrar el Plan Nacional de Numeración

- Aplicar, cuando sea procedente, las sanciones contempladas en la ley de Telecomunicaciones.
- Participar como el órgano técnico representativo del país, en coordinación con los órganos competentes, en las reuniones de los organismos internacionales de telecomunicaciones y en las negociaciones de tratados, acuerdos y convenios internacionales en materia de telecomunicaciones.
- Velar por el cumplimiento de la ley y demás disposiciones aplicables.

En mi opinión, el generar confianza en las personas de la generación actual, es decir, las personas que actualmente están frente a las grandes empresas es un punto crítico y a su vez difícil de lograr, ya que estos fueron criados bajo una cultura de no-tecnología. Es común ver que muchas personas que se encuentran al frente de las empresas, vean la inversión en tecnología como un gasto innecesario o bien vean el comercio electrónico como un riesgo. En cambio, las personas que actualmente están en edad de 18 a 25 años, por ejemplo, han crecido en conjunto con la tecnología. Por lo que cuando éstos estén al frente de una empresa, es más fácil que promuevan soluciones tecnológicas que soporten los procesos empresariales y que éstos a su vez, transmitan y generen confianza tanto, a sus empleados, como a sus clientes en el uso de dichas soluciones.

Claro está, que las leyes son un factor importante para el desarrollo, la aceptación y el uso de dichas tecnologías en nuestro país, pero se debe reconocer que éstas, a mi criterio, aún están en una etapa de desarrollo.

### **1.5. Instituciones y Organizaciones que brindan apoyo a la PYME**

El encargado del desarrollo científico y tecnológico del país, así como también de promocionar y coordinar las actividades científicas y tecnológicas que realice el SINCYT (Sistema Nacional de Ciencia y Tecnología), es el Consejo Nacional de Ciencia y Tecnología, CONCTY [20].

En el informe de actividades del 2006 del CONCYT, se encuentra la elaboración de un plan Nacional de Ciencia, Tecnología e Innovación. Entre los diferentes objetivos de dicho plan podemos encontrar

- Impulsar el aumento de la inversión nacional en ciencia, tecnología e innovación, a niveles comparables con los de otros países de la región.
- Crear programas para desarrollar la investigación, la invención y la innovación en el país, que incluyan la formación, la capacitación y la actualización de recursos humanos capaces de producir y administrar el conocimiento con la máxima calidad.
- Promover las publicaciones científicas y técnicas, y estimular las actividades de los inventores e innovadores.
- Dar apoyo al desarrollo y a la aplicación de los conocimientos y de los recursos científicos así como tecnológicos que contribuyan a la productividad y a la competitividad.
- Apoyar el desarrollo y la aplicación de los conocimientos y de los recursos científicos y tecnológicos que contribuyan a resolver las necesidades básicas de la población.
- Apoyar el desarrollo de tecnologías, procedimientos y medios que contribuyan a reducir los riesgos para la población, en todos los campos

en que sea necesario, así como los que propicien y apoyen el manejo sostenible del ambiente.

- Apoyar y promover las actividades que preparen al país, en su conjunto, para la sociedad de la información y el conocimiento, con una dinámica que permita cerrar la brecha digital.

En dicho informe [21], también se menciona que se facilitará el financiamiento para atender y estimular la demanda de innovación tecnológica de las PYMEs además de brindar financiamiento para la mejora de la oferta los servicios tecnológicos, así como la creación de centros tecnológicos, en su mayoría en las Universidades, con el único fin de promover las transacciones comerciales entre pequeñas y medianas empresas y sus centros especializados.

Solo durante el año 2006 se beneficiaron 46 pequeñas y medianas empresas gracias al apoyo del CONCYT, el cual aportó un total de Q.2,370,000.00. Lamentablemente, no se encuentra registro alguno, al menos de forma pública, del detalle de las empresas que obtuvieron dicho apoyo, la cantidad recibida, y la forma en la que esta fue invertida.

Además, los extensionistas tecnológicos desarrollaron durante ese mismo año, actividades de divulgación del programa, brindaron asistencia técnica a unas 454 PYMEs del área rural, impulsando así, entre otras cosas, el uso y aplicaciones de nuevas tecnologías [22].

Con el fin de poder resolver el problema del acceso a la información y poder brindar a todas las Pequeñas y Medianas Empresas una asesoría, así como también formación tecnológica, se crearon los *Centros Comunitarios Digitales* [23]. Los cuales no fueron concebidos para ayudar directamente a las

PYMEs, sino que a las diversas comunidades del interior del país con el fin de innovar con nuevas tecnologías el desempeño de sus actividades económicas y sociales. Esta claro que las PYMEs pueden obtener provecho de dichos centros, donde se les debería de brindar la capacitación y atención necesaria.

Las ventajas, según el CONCYT, que brindarán dichos Centros Comunitarios Digitales, son, entre otras [24]:

1. Capacitación en software aplicativo (procesador de textos, hojas electrónicas, presentaciones, generación de páginas WEB).
2. Asesoría en el uso de los recursos multimedia, scanner, fotocopidora, etc.
3. Comunicación por correo electrónico y otros medios (Messenger, telefonía, etc)
4. Promoción del uso de los recursos informáticos y de comunicación para desarrollar actividades económicas y sociales.
5. Abrir canales de exhibición de los productos y/o servicios de la localidad para expandir el mercado de éstas comunidades y mejorar sus ingresos monetarios.
6. Crear oportunidades de realizar reuniones grupales en un área adecuada con recursos multimedia para planificar proyectos.
7. Fortalecer la creación de páginas WEB locales dando a conocer su propia cultura, economía y sociedad.
8. Mejorar la calidad de vida a través del acceso a las tecnologías de información y comunicaciones.

El compromiso de dichos Centros de Comunitarios Digitales se realizó en conjunto con el gobierno de China, para brindar el equipo necesario a las áreas departamentales. Actualmente, dichos Centros se encuentran operando en:



Santa Lucía Cotzumalguapa, Escuintla, desde el 11 de Septiembre del 2006, San Juan Chamelco, Alta Verapaz desde el 5 de noviembre del 2006 y San Lucas Tolimán, Sololá desde el 29 de enero del 2007. La Secretaría Nacional de Ciencia y Tecnología (SENACYT), piensa abrir otros 2 centros más durante el año 2007. De cualquier manera la forma en que se han implementado dichos centros no permite que la población se apropie de proyecto y terminan siendo utilizados como Café Internet.

Todos estos esfuerzos, comienzan a dar fruto, ya que el 29 de marzo del 2007, fue publicado que Guatemala ascendiera 19 posiciones en el ranking mundial del Informe Global de Tecnología Informática, el cual es publicado por el foro económico mundial. Aunque Guatemala ocupa la posición 79 de 122 países, todavía esta muy por debajo de Costa Rica, posición 56, y México 49 [24].

Las PYMEs cuentan también, con el apoyo del AGEXPORT o Asociación Guatemalteca de Exportadores. Su misión es *“Hacer de Guatemala, un país exportador, mediante la promoción del crecimiento de las exportaciones, basados en la competitividad, contribuyendo así al desarrollo económico y social de Guatemala en forma sustentable”* [25]

En otras palabras, se busca tanto promover como fomentar el desarrollo de las exportaciones del país, así como también el prestar servicios que contribuyan únicamente al sector exportador, motivando así a nuevos empresarios a dirigir sus actividades al mercado internacional. Cada pequeña y mediana empresa puede apoyarse en AGEXPORT, a través de su portal [www.export.com.gt](http://www.export.com.gt).

La Cámara de Comercio de Guatemala es otro punto de apoyo para las PYMEs, ya que ésta desde el año 2006 cuenta con varios portales sectoriales, los cuales están a la disposición de los miembros de la Cámara de la Industria, en especial las PYMEs. Actualmente existen 4 portales

- Portal del Proyecto [[www.negociosenguatemala.com/portal/ebusiness](http://www.negociosenguatemala.com/portal/ebusiness)]
- Portal de Turismo [ [www.negociosenguatemala.com/portal/turismo](http://www.negociosenguatemala.com/portal/turismo)]
- Portal de AgroNegocios [[www.negociosenguatemala.com/portal/agronegocios](http://www.negociosenguatemala.com/portal/agronegocios)]
- Portal de Comercio [[www.negociosenguatemala.com/portal/comercio](http://www.negociosenguatemala.com/portal/comercio)]

Entre los organismos Internacionales que colaboraron para la realización y mantención de dichos portales se encuentra el BID o Banco Interamericano de Desarrollo y el FOMIN o Fondo Multilateral de Inversiones, los cuales donaron un aproximado de 2.8 millones de dólares para el desarrollo de dichos portales, así como el desarrollo de otras actividades y proyectos en beneficio exclusivamente para las PYMEs. Proyectos que deben de contener soluciones que utilicen y aprovechen al máximo las TICs [26].

La creación de dichos Centros Comunitarios, además de ser un gran beneficio para las distintas comunidades, ayuda a disminuir la Brecha Digital del País, creando en las comunidades una conciencia en torno de la utilidad de las TICs, difundiendo de esa manera, la cultura Informática y sus diversas herramientas, haciéndole ver a las personas que con el uso de dichas tecnologías pueden llegar a poder cumplir de mejor forma sus propósitos, y por ende mejorar su nivel de vida.

El sector turístico, también cuenta con un apoyo importante para su desarrollo, la Fundación para el Desarrollo de Guatemala *FUNDESA*, busca elevar la competitividad, tanto de la mediana, como de la pequeña empresa

turística, brindando apoyo en forma de capacitaciones administrativas, asistencias técnicas, soporte en temas de comercialización y mejora de calidad de los servicios prestados, a través de los Centros Empresariales de Turismo CETS. Estos centros se encuentran actualmente ubicados en Antigua Guatemala, Cobán, Flores, Livingston, Panajachel, Retalhuleu, y Quetzaltenango [27]. Desafortunadamente no se cuenta con información clara de las empresas que han sido beneficiadas con dichos proyectos.

FUNDESA, se justifica su ayuda a los sectores turísticos debido a que éste, según ellos, es uno de los más importantes para el crecimiento económico del país. Este tipo de proyectos esta siendo apoyado por el INGUAT, y el Programa Nacional para la Competitividad PRONACOM.

Otro programa interesante con que cuenta FUNDESA, es el denominado *Eurocentro*<sup>8</sup>, cuyo objetivo principal es el apoyo a las empresa para que éstas se internacionalicen, así como también las empresas europeas puedan realizar negocios en Guatemala. El Eurocentro realiza actividades de diversos sectores no solo de turismo, entre los cuales podemos mencionar, alimentos, refrescos, construcción, energía renovable entre otros.

Al momento de visitar cada uno de los sitios, el panorama parecer ser bastante bueno, mucho apoyo, mucha inversión tanto en tecnología, así como en infraestructura, existen varios planes, cursos talleres de mercadeo, informática etc. Pero si realmente existe tanto apoyo y tanta inversión en Guatemala, porque es que no se ven los resultados, no se aprecia dicho avance tecnológico en las pequeñas y medianas empresas, es decir el cambio que se debería de tener no es tan palpable como debería.

---

<sup>8</sup> <http://www.fundesa.org.gt/eurocentro.php>

Una explicación es que dichas implementaciones de centros y demás, no se realizaron con la debida planeación previa. Como lo menciona Javier Gramajo presidente de fundaTICs en su propuesta “Plan Estratégico para la Instalación, Uso y Sostenibilidad, de Centros Comunitarios Digitales”. Según Javier Gramajo López, los Centros Comunitarios Digitales, están actualmente, siendo subutilizados, no están siendo aprovechados al máximo. El uso que actualmente tienen es de un Café Internet. FundaTICs es una organización no gubernamental que propone la integración de los sectores público, privado, académico y comunidad internacional para la puesta en marcha de proyectos relacionados con el uso, desarrollo y aplicación de las TICs en la PYMEs en Turismo, Patrimonio Cultural y Natural, Educación e Emigración.

Por lo tanto, la inversión mencionada anteriormente, en efecto se ha realizado, pero no de la forma en que ésta debió de haberse hecho. Esta se realizó sin antes concientizar a la población, identificando las actividades económicas y sociales de las comunidades, identificando necesidades, así como los productos locales con alto potencial de ser comercializados por medio de las TICs.

El tema del apoyo a la PYME, no es un tema que se esté tratando exclusivamente para Guatemala, o que sea un factor de importancia únicamente para nuestro país. El tema de la pequeña y mediana empresa, ha recorrido todo el continente americano, es tanto su impacto e importancia, que el BID o Banco Interamericano de Desarrollo cuenta con un programa llamado Innovación en Tecnología de la Información y Comunicación (TIC), para el Comercio Electrónico y el desarrollo de las Pequeñas y Medianas Empresas ICT4BUS, el cual está siendo financiado directamente por el FOMIN (Fondo Multilateral de Inversiones)<sup>9</sup>. El propósito general de este programa es poder

---

<sup>9</sup> <http://www.ict4bus.org/ict4bus/default.asp>

mejorar la eficiencia, competitividad y productividad de las PYMEs en toda América Latina y el Caribe.

Actualmente, se han aprobado 23 iniciativas de proyectos por medio de este programa, de los cuales únicamente un proyecto pertenece a Guatemala y fue gestionado y por AGEXPORT<sup>10</sup>. El proyecto se aprobó bajo el nombre de “Facilitación y Desarrollo del Comercio Electrónico en Guatemala”<sup>11</sup>. Este proyecto, busca mejorar el portal ya existente de la AGEXPORT, agregando opciones de Comercio Electrónico, bibliotecas virtuales y otra serie de funcionalidades. Además, busca la mejora y el crecimiento del sistema actual SEADDEX. Actualmente, brinda los servicios de Ventanilla Única para los exportadores, donde pueden llenar formularios aduaneros, declaraciones de exportaciones entre otros, todo realizándolo de una forma 100% electrónica, facilitando así los trámites con el fin de realizar la exportación mas rápida.

Entre otras instituciones que apoyan a nivel Latinoamericano a las PYMEs, podemos mencionar al BCIE o Banco Centroamericano de Integración Económica, el GTZ entre otros.

---

<sup>10</sup> <http://www.ict4bus.org/ict4bus/awards.asp>

<sup>11</sup> <http://www.ict4bus.org/ict4bus/files/FDCE-Guatemala.pdf>



## 2. ONTOLOGÍAS

### 2.1 Web Semántica

La Web Semántica, o bien *Semantic Web*, fue presentada al mundo por Tim Berners Lee en el año 2000 [28]. Esta propone básicamente lo siguiente; que toda la información que se desee publicar en la Web, debe de contener información adicional, la cual añade un sentido mas lógico a la información, es decir en otras palabras, agregar una explicación al contenido que se desea publicar, definido de tal forma, que un programa sea capaz de poder interpretar de forma rápida, sencilla y lo mas importante, sin supervisión o intervención de un humano, la información publicada.

Por lo tanto, el desarrollo de la Web Semántica pretende que en un futuro, se envíe un programa hacia la Web, en búsqueda de las páginas que contengan la información exacta de lo buscado. A éstos “programas” inteligentes encargados de realizar las búsquedas se les llama Agentes.

La Web Semántica pretende, por decirlo de una forma sencilla, darle sentido a toda la información disponible, de manera que cuando se desee buscar sobre un tema en específico, se obtengan resultados que concuerden al 100% con la descripción y sentido deseado, y no obtener una concordancia que sea puramente sintáctica (como lo realizan los buscadores de hoy en día).

Lo que se pretende es poder obtener una concordancia lógica (semántica) del tema buscado. Con esto no solo se logra obtener una búsqueda refinada, sino que una búsqueda acertada, gracias a que las aplicaciones de la Web Semántica van más allá, que simples búsquedas textuales en documentos.

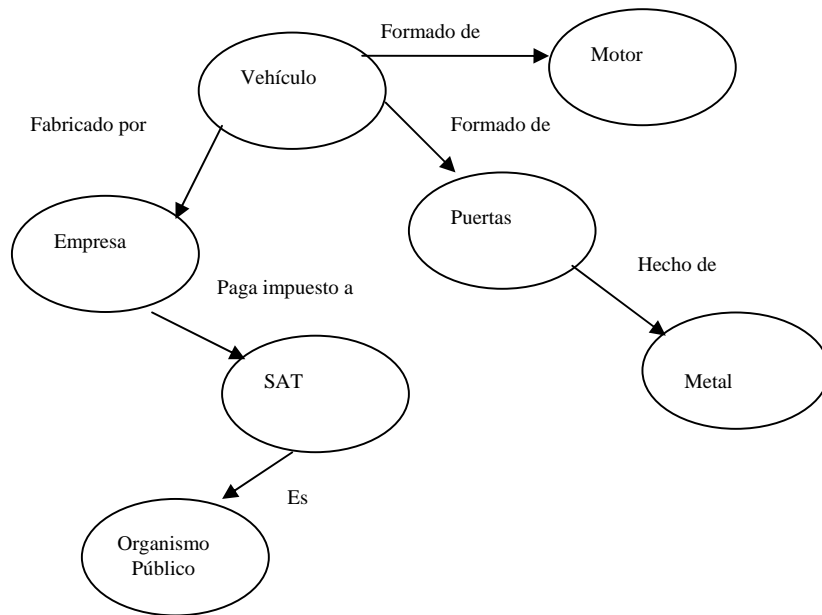
Porque teniendo búsquedas efectivas donde la basura sea eliminada, idealmente por dar un ejemplo, un 1%, las aplicaciones de comercio electrónico se verían beneficiadas de manera significativa. Para nuestro caso de estudio, las PYMEs, obtendrán también un gran beneficio de dicha tecnología. Debido a que si toda la información de la Web se encuentra bajo una misma estructura, basada en RDF o bien OWL, las empresas, tanto proveedores como clientes, podrán intercambiar sus datos de productos y servicios siguiendo los esquemas establecidos, de tal forma que se logra una fácil integración y transferencia de información entre empresas, facilitando así, las relaciones empresariales y por ende, promoviendo y mejorando las negociaciones.

## **2.2 Redes Semánticas**

En el inciso anterior, se menciona que la Web y la información contenida en ella deben de contener una estructura lógica, la cual debe de ser interpretada por los agentes. Esta estructura se llama Red Semántica, la cual no son es más que redes similares de conceptos interdependientes. Es decir, en otras palabras, una red semántica es un diagrama en el cual se relacionan y se van descomponiendo los conceptos en otros más, pero con diferentes relaciones entre sí. A continuación se presenta un ejemplo de una red semántica para un Vehículo.



**Figura 3 Red semántica del dominio vehículo**



Como se puede observar en el ejemplo, se parte de Vehículo, el cual está formado de motor y puertas, las cuales están fabricadas de metal. El vehículo es fabricado por una Empresa, la cual debe de pagar impuestos a la SAT, la cual es una entidad del gobierno. Esa sería la forma de leer o interpretar, la red semántica, ahora bien, este como se mencionó es un ejemplo pero se puede observar claramente en el como es que a partir de un concepto se van creando las relaciones con otros conceptos y poco a poco se va profundizando. El propósito es que en el Internet toda la información se encuentre almacenada de esta forma, de modo de poder crear una red de conocimiento general, la cual pueda ser recorrida e interpretada fácilmente.

El problema del Internet de hoy en día, es que la información solo se encuentra ahí, pero no se encuentra relacionada entre sí, ni tampoco se

encuentra almacenada en una estructura de conceptos lógicos. Esto es, debido a que las páginas se encuentran vinculadas una con la otra sin importar el contenido y sin hacer algún tipo de mención específica y/o estándar, de forma que pueda ser interpretado por cualquier programa.

Es por eso, que hoy en día, querer realizar una compra en línea, implica invertir bastante tiempo en la búsqueda en el Internet de alguna empresa candidata. Comenzando porque la búsqueda realizada en sí, no es muy confiable ya que los resultados mostrados probablemente no contengan información certera, o simple y sencillamente retorne lo que se conoce como “basura”, es decir, páginas con contenido inútil. Si tomamos de ejemplo a una empresa que fabrica envases de vidrio que desea buscar a una empresa a la cual pueda venderle sus productos, probablemente dicha empresa realizaría una búsqueda como la siguiente:

#### *Compra de envases de vidrio*

Ya que lo que dicha empresa desea es vender sus productos. Dicha búsqueda fue realizada utilizando el buscador [www.google.com](http://www.google.com), ingresando exactamente esa frase y el resultado fue el siguiente:

**Figura 4 Resultados de la búsqueda realizada el 26 de abril de 2007, con el buscador [www.google.com](http://www.google.com)**

Resultados 1 - 10 de aproximadamente 232.000 de compra de envases de vidrio. (0,29 segundos)

Como se puede observar, los resultados de la búsqueda fueron aproximadamente 232 mil páginas, ahora bien, si la búsqueda se realizara en un sistema con las características que se describen, entonces, probablemente ésta le retornará resultados mas certeros clasificando así las empresas por

actividad económica, ya que entre los primeros resultados que se obtuvieron al realizar dicha búsqueda, se obtuvo

**Figura 5 Resultado de la búsqueda**

[Vidrala compra a Owens-Illinois dos plantas en Cataluña e Italia ...](#)  
La **compra**, comunicada ayer a la Comisión Nacional del Mercado de Valores (CNMV), ... El fabricante alavés de **envases de vidrio**, que tras estas compras ...  
[www.cincodias.com/.../Owens-Illinois/plantas/](http://www.cincodias.com/.../Owens-Illinois/plantas/)  
[Cataluna/Italia/138/millones/cdsmerA00/20041119cdsdiemp\\_7/Tes/ - 49k -](#)  
[En caché - Páginas similares](#)

El cual es un artículo de una revista en internet, la cual contiene información sobre una compañía fabricante de envases de vidrio, la cual recientemente compró a otra compañía llamada “Owens-Illinois”. Incluso, se retornaron sitios como el que se muestra a continuación.

**Figura 6 Resultado de Búsqueda**

[PERFUMES ORIGINALES EN ENVASES RECICLABLES | DeRemate.com.ar](#)  
(Confirmarás tu **compra** en la página siguiente) ... si lo tengo en esta promo, los 30ml. originales fraccionados a un **envases de vidrio**, en 50 pesos, ...  
[www.deremate.com.ar/accdb/viewitem.asp?idi=13389343](http://www.deremate.com.ar/accdb/viewitem.asp?idi=13389343) - 76k -  
[En caché - Páginas similares](#)

El anterior es un sitio conocido debido a que las personas individuales, no empresas, pueden colocar cualquier tipo de objeto para ser subastado en el internet [[www.deremate.com.ar](http://www.deremate.com.ar)], y se observa claramente en la figura 5 los criterios por los cuales dicho sitio fue tomado en cuenta, ya que en la página existe un párrafo que incluye la palabra *compra*, y mas abajo en otro fragmento de algún otro párrafo esta la frase *envases de vidrio*. Aunque efectivamente la herramienta de búsqueda encontró la frase “compra de envases de vidrio”, en segmentos separados del documento, éstos no concuerdan lógicamente con el sentido de la búsqueda, es decir no satisfacen el fin de haber buscado en la

herramienta, ya que esperaba obtener resultados con páginas de empresas interesadas en la compra de dichos envases.

Actualmente existe un buscador, tipo *Google.com*®, el cual esta basado en ontologías, su nombre es *Swoogle*® [<http://swoogle.umbc.edu/>]. *Swoogle*®, es básicamente, un motor de búsqueda que utiliza los conceptos de la Web Semántica, lo que significa, a un nivel mas técnico, es que el sitio busca en la red mundial (WWW), sitios que estén creados en RDF.

Este proyecto esta siendo desarrollado por el grupo denominado *Ebiquity*, que pertenecen al departamento de Ciencias de la Computación e Ingeniería Eléctrica de la Universidad de Maryland [<http://www.umbc.edu/>]. Este buscador se encuentra actualmente en continuo desarrollo y es una muestra del futuro, hacia donde vamos, es una muestra clara de lo que se quiere hacer, con el fin único de obtener mejores resultados de búsquedas.

*Swoogle*® indexa la información con el fin de poderla encontrar rápidamente, al igual que *Google*®, pero con la diferencia, como se ha mencionado, que éste lo hace a un nivel semántico basándose siempre en las ontologías de cada sitio. Es decir, cada sitio web, esta formado, como se menciona, por una ontología almacenada en lenguaje OWL. Actualmente no muchos sitios están basados en RDF, y al realizar la misma búsqueda en este motor, pues retornó 0 resultados.

Aparte de *Swoogle*, existen otras herramientas que utilizan tanto la Web Semántica como técnicas de aprendizaje computacional para poder realizar mejores búsquedas así como también indexaciones de documentos para su mejor recuperación. De tal forma que el poder compartir el conocimiento contenido en dichos documentos sea un proceso sencillo y rápido. Un ejemplo

de dichos sitios es Freebase [[www.freebase.com](http://www.freebase.com)], el cual es un sitio pensado en compartir de forma gratuita información acerca de cualquier tipo de información, en donde las personas son capaces de poder editar y agregar información adicional a la existente. Este sitio clasifica de diferentes formas los temas, de la forma que si se está viendo información, por ejemplo, de un actor, éste a su vez puede ser un cantante, de tal forma que el sitio permite el poder navegar entre diferentes categorías, e ir consultando información relacionada. Lo interesante de este sitio, como ya se mencionó es que la ontología que es utilizada, puede ser modificada por el usuario desde el sitio, permitiendo así que cada vez se almacene más y más conocimiento.

Por otro lado, Twine [<http://www.twine.com>] es un sitio similar, el cual, a diferencia de los dos anteriores, permite compartir información a únicamente un cierto grupo de personas.

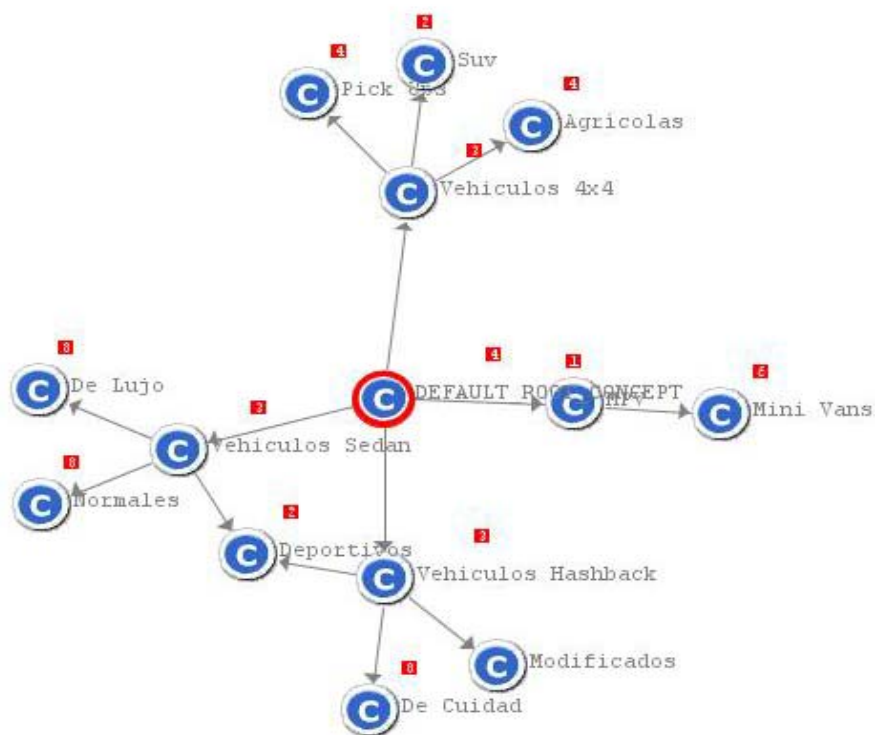
Otro vivo ejemplo es TrueKnowledge [<http://www.trueknowledge.com>], es un sitio con características similares a los descritos anteriormente, con la diferencia que éste está orientado a que los agentes realicen búsquedas y no una persona.

## **2.3 Ontologías**

Una ontología, se puede definir como una representación formal de las relaciones entre uno o varios conceptos, referentes a un dominio en particular. Las ontologías proveen información y/o conocimiento, sobre un tema en específico, el cual puede ser compartido de forma fácil entre otras personas o inclusive otras máquinas y/o programas. Por decirlo de otra manera, una ontología es una forma estructurada de representar el conocimiento, ya que se

podrían decir que éstas son el vocabulario con el cual se transmite el conocimiento. Por lo general una ontología contiene una descripción jerárquica, de los conceptos más importantes acerca del dominio seleccionado<sup>12</sup>.

Figura 7 Representación gráfica de una ontología. OntoEdit versión 2.6



La figura muestra un ejemplo sencillo de una ontología, para este caso, se está utilizando una ontología de dominio Vehículos. Como se puede observar en la figura todo parte de un concepto base o padre, luego se va desglosando por categorías o jerarquías, las cuales detallándose cada vez más en sus niveles inferiores.

<sup>12</sup> State of Art: Ontologies, Engineering Faculty, Computer Science and Systems Department Universidad San Carlos de Guatemala, Abril 2005, Javier Gramajo López

Una ontología, puede tener varias áreas de uso, tal como la ingeniería, la medicina, la educación etc. Las ontologías se encuentran clasificadas básicamente en 6 categorías descritas a continuación<sup>13</sup>:

- Ontologías de dominio: las cuales, como su nombre lo indican se especializan únicamente en un dominio específico.
- Ontologías de Metadatos: Ontologías que describen el contenido de publicaciones en línea. Es decir, un metadato no es más que un dato, o palabras por ejemplo, que sirve para poder describir información, contenido y alguna otra peculiaridad o característica especial de algún dato o contenido.
- Ontologías Genéricas: este tipo de ontologías, se basan en tratar de organizar o bien, representar el conocimiento en general, proveyendo así un tipo de noción básica sobre el tema a tratar.
- Ontologías representativas: no hacen referencia a un dominio en específico, solo como su nombre lo indica, representan algo, cualquier cosa que se desee ser representada por medio de una ontología.
- Ontologías de Métodos y tareas: las ontologías de tareas definen, como su nombre lo dice, tareas clave, sus entradas y sus salidas en un proceso, de modo que la construcción de un *Workflow*, se haga más sencilla. Y las ontologías de métodos proporcionan los términos base en para poder utilizar o aplicar los famosos PSM, que por sus siglas en inglés significa: Problem Solving Method, que no son mas que paquetes organizados de unidades PSK, los cuales se encuentran indexados por la meta o el problema que se desea resolver, (para mas información visite <http://www.cse.ohio-state.edu/~chandra/Ontology-of-Tasks-Methods.PDF>)

---

<sup>13</sup> State of Art: Ontologies, Engineering Faculty, Computer Science and Systems Department  
Universidad San Carlos de Guatemala, Abril 2005, Javier Gramajo López

- Ontologías de Aplicación: estas ontologías, podría decirse que son una combinación de las ontologías de métodos y las de dominio, ya que son utilizadas para modelar un dominio en particular.

Ahora bien, se han explicado las funcionalidades de las ontologías, los diferentes tipos existentes, a continuación se presentan las partes por las cuales está formada la ontología<sup>14</sup>.

- Conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser objetos, planes y/o estrategias.
- Relaciones: representan la interacción y enlace entre los conceptos del dominio. Por ejemplo: pertenece-a, es-un, sinónimo-de.
- Funciones: son cálculos que pueden ser ejecutados sobre los datos en la ontología, su resultado puede variar dependiendo de la cantidad de datos actuales en la ontología. Por ejemplo, pueden aparecer funciones como *número de vehículos, número\_de\_hijos* etc.
- Instancias: se utilizan para representar objetos reales representados por los conceptos.
- Axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Con ellos es posible inferir sobre los elementos existentes en la ontología.

---

<sup>14</sup> State of Art: Ontologies, Engineering Faculty, Computer Science and Systems Department  
Universidad San Carlos de Guatemala, Abril 2005, Javier Gramajo López



La creación de una ontología no es un tema difícil, pareciera serlo debido a todos los términos y teoría involucrada, pero no lo es, básicamente es necesario estar claro en el dominio que se va a desarrollar la ontología. A continuación se presentan una serie de pasos para la creación de una ontología de cualquier dominio propuestos por César Llamas [29].

- Definición del dominio y alcance
- Utilización de Ontologías existentes.
- Enumeración de conceptos importantes
- Definición de conceptos y su jerarquía.
- Definir las propiedades de las clases
- Crear las instancias

En la definición del dominio y el alcance se debe de definir para que se utilizará la ontología, cual es el dominio que cubre la ontología, su propósito.

En el segundo paso se aconseja consultar una ontología existente, ya que tomando como base una existente, se pueden tomar ideas o bien fragmentos de ésta, esto con el fin de la conceptualización de la misma, copiando o bien, observando y analizando el formato de otras ontologías similares.

Como tercer paso, se recomienda enumerar los conceptos clave para la ontología, no es necesario o bien recomendable en este paso relacionarlos, sino que solo realizar una lluvia de ideas y escribir todos aquellos conceptos o términos de relevancia para el dominio que se especificó.

Una vez escrito la lista de conceptos, se procede a seleccionar aquellos de mayor relevancia, y se clasifican o bien se van jerarquizando. Es aquí donde la ontología va tomando forma. Existen varias técnicas para poder jerarquizar los

conceptos, ya sea de abajo para arriba (Bottom-up), es decir comenzar agrupando las clases mas específicas para ir avanzando a las mas generales o bien, el método contrario Top-down, comenzando por los conceptos mas generales avanzando así, hacia los más específicos (totalmente opuesto al método anterior).

Una vez terminados los pasos anteriores se procede a crear propiedades a las clases para luego y como último paso, poder ingresar las instancias, es decir los objetos como tales. Se entiende por objeto o por instancia la representación de algo real, es decir, que si se tenía una clase Vehículo, que tiene como propiedad el número de cilindros, la instancia o el objeto será “Bmw M3 6 cilindros”.

Actualmente el lenguaje utilizado y reconocido como un estándar, para la representación de ontologías en la Web es el OWL (Web Ontology Language). Las ventajas de utilizar este lenguaje es que puede ser mejor interpretado por las máquinas, es decir el software encargado de entenderlas. Esta fácil interpretación de los datos se realiza gracias a que el OWL, está basado tanto en XML como en RDF y RDFS.

## **2.4 Herramientas**

Las herramientas más comunes donde se crean y se diseñan las ontologías son Protegé-2000, la cual es una herramienta desarrollada en la Universidad de Stanford [<http://protege.stanford.edu/>], OIL-Ed, el cual es un producto que ya no cuenta con el soporte del fabricante, ya que este ha sido oficialmente discontinuado [<http://oiled.man.ac.uk/>] y por último, OntoEdit, el cual es uno de los más populares desarrollado en Alemania. OntoEdit, forma parte del

OntoStudio cuya tecnología esta basada en el OntoBroker®  
[<http://ontoedit.com/>]

. La ontología a utilizar para el desarrollo del portal, se encuentra realizada y diseñada en esta herramienta.

## **2.5 Búsquedas Inteligentes**

Toda la estructura antes mencionada, ontologías, RDF, Web semántica en general, fue pensada claramente para el uso humano, para su beneficio, pero precisamente no solo los humanos los que directamente utilizan esta tecnología. Los encargados de lidiar directamente con las estructuras de la Web semántica son los Agentes, denominados “Inteligentes”. Los agentes no son más que programas diseñados para ir navegando en la Web, leyendo y analizando contenidos semánticos escritos en OWL, utilizando las ontologías como su base de conocimiento principal.

En una búsqueda inteligente, no solo se muestran los resultados exactos de la búsqueda, sino que también el agente se encarga de mostrar páginas o sitios relacionados con el tema seleccionado [30].

¿Pero cómo es que se logra realizar una búsqueda de este tipo, que presente los resultados deseados?

Como ya se mencionó en la sección anterior, una ontología está compuesta de conceptos, relaciones, funciones, axiomas y por último instancias. A esto se podría agregar los sinónimos.

Todos estos componentes son esenciales para poder realizar búsquedas en una ontología, ya que todos ellos de una u otra forma, describen, interrelacionan los elementos que existen dentro de ella. Los conceptos se encuentran amarrados con otros conceptos, por medio de las relaciones existentes entre ellos, donde a su vez una relación puede ser de varios tipos, entre ellos, “Esta-Formado-De”, “Pertenece-a”, “Trabaja-Con” [31]. Es decir, que un agente al momento de examinar dicha estructura, puede llegar a comprender la estructura lógica y jerárquica de los conceptos almacenados. Esto sin necesidad de la intervención humana. Por lo que con esos dos elementos ya es posible crear un proceso, o en otras palabras, como se ha estado mencionando un agente que de forma independiente vaya analizando ontología por ontología en la Web. Hasta el momento, contando con dichos elementos, es posible realizar búsquedas un poco más refinadas, teniendo en cuenta estructuras jerárquicas de conceptos.

Las funciones, no son más que un tipo especial de relación en donde se identifica un elemento mediante la realización de uno o varios cálculos. Éstas pueden llegar a ser un elemento muy útil, y mas si éstas se utilizan en ontologías incrementales, es decir ontologías que van creciendo con forme se va adquiriendo el conocimiento. Una función, por ejemplo, podría retornar el número de elementos asociados que posee un concepto (si se desea realizar una búsqueda con un mínimo de resultados), o bien realizar una categorización de las clases, asignaciones de fechas, o cualquier otro tipo de dato importante que se relacione con algún concepto. De tal forma que, ahora contamos con elementos dinámicos, que obviamente pueden cambiar en el tiempo, para realizar búsquedas mas eficientes sobre una ontología.

Una ontología también, como ya se mencionó, cuenta con Axiomas, los cuales son teoremas que son utilizados en conjunto con las relaciones que para

poder inferir datos o bien información que no esta almacenada de forma explícita dentro de la taxonomía de conceptos [30].

Por definición una taxonomía, en un sentido mas general se conoce como “La Ciencia de la Clasificación” [32], por lo que hablar de una taxonomía, no es mas que hacer referencia a una estructura jerárquica, orientada a la clasificación de objetos de un mismo dominio, partiendo del concepto mas genérico, hacia los conceptos mas específicos [33]. A diferencia de una taxonomía, una ontología cuenta con una estructura interna mucho más rica, en la cual es posible almacenar aún más información, de la que normalmente una simple taxonomía lo permitiría.

Al hablar de inferir conocimiento, se refiere al gran poder, de cierto modo, que poseen los axiomas en las ontologías, por ejemplo, si desearía realizar una búsqueda acerca de los odontólogos, en cierta área metropolitana. Suponiendo que se cuenta con un agente, el resultado de la búsqueda podría ser

- Dra. Rosa Mendía – Dentista – Ortodoncia
  - Consulta Q200.00
  - Hospital Clínico “La Paz” – Carretera a Masagua km. 59.

Este es un ejemplo del resultado de la búsqueda, pero, suponiendo que un agente haya realizado dicha búsqueda, y si la Dra Rosa tenga en su página de Internet una ontología, el agente Web, estaría en la capacidad de utilizar un atributo “*Trabaja en Hospital*”, para luego ir a buscar la dirección del hospital, al sitio dicho centro médico. De tal forma que la búsqueda sería un poco mas completa.[*Jornadas de Ingeniería Web, Adolfo Lozano Tello*].

Otro ejemplo, de búsqueda mas refinada, o bien, en este caso inteligente, sería si dentro de la ontología existe un axioma "*Si el médico trabaja en un hospital de mi seguro, entonces no tendré que pagar cargo extra*". El agente sería capaz de obtener dicha información y mostrarnos resultados de médicos que estén afiliados al seguro médico.

Un elemento más de las ontologías, que ya fue explícitamente mencionado, en el ejemplo anterior, son las instancias, las cuales son utilizadas para representar objetos específicos que se encuentran representados por los conceptos, para nuestro ejemplo, una instancia es la Dr. Rosa Mendía. Otra instancia es el Hospital. En otras palabras, la instancia, es el elemento del mundo real, mapeado a la definición ontológica.

Por otro lado, una búsqueda inteligente, se puede ayudar de los sinónimos, cuando se esta analizando una ontología, es común buscar sinónimos para los términos en ella. Y luego realizar búsquedas con los sinónimos, o bien buscar nuevamente en la ontología el sinónimo. Con esto la búsqueda se complementa aún mas, y el usuario podrá obtener mejor y mayores resultados de calidad.

Actualmente, una herramienta muy utilizada para la búsqueda de sinónimos en las ontologías es WordNet®, el cual es un proyecto que esta siendo desarrollado por la Universidad de Princeton Estados Unidos. WordNet® [<http://wordnet.princeton.edu/>], no es mas que una base de datos léxica, que contiene, adverbios, verbos, adjetivos, agrupados en sets de sinónimos cognitivos, donde cada uno expresa un concepto diferente. A un inicio WordNet, solo estaba disponible en idioma Inglés, pero actualmente existe el proyecto Euro WordNet [34], el cual busca crear bases de datos similares la WordNet, con la diferencia que éstas bases de datos estén disponibles en varios idiomas,

con el fin de no tener la limitante del idioma inglés. Actualmente el proyecto ha desarrollado aplicaciones en diferentes lenguajes, Holandés, Italiano, Español, Alemán, Francés, Eslovaco y Checo.





### 3. *MACHINE LEARNING* (APRENDIZAJE COMPUTACIONAL)

#### 3.1 Definición:

El aprendizaje es algo básico y clave en la formación del ser humano, éste se ha estudiado tanto en las personas como en los animales. Por definición, el aprendizaje se entiende como “ganar conocimiento a partir de la experiencia” [35], o simple y sencillamente el ganar conocimiento.

Hoy en día, los sistemas han ido evolucionando junto con las ambiciones del hombre, hasta tal punto que el hombre siempre ha tratado de crear o innovar máquinas cada vez más potentes e *inteligentes*, al decir máquinas inteligentes, se entiende que éstas tengan la capacidad de, al igual que nosotros, aprender cosas nuevas, ya sea a partir de la experiencia, o bien a partir de una base de conocimiento previamente establecida. Todo esto con el fin de que los sistemas nos ayuden, por ejemplo, a tomar decisiones, de valor, decisiones acertadas, decisiones “inteligentes”, luego de haber analizado una gran cantidad de datos. Lo cual, para una persona representaría la inversión de bastante tiempo de análisis y comprensión de datos. Es decir, que con el Aprendizaje Computacional o *Machine Learning*, una computadora es capaz de analizar una gran cantidad de datos en relativamente corto tiempo y además de eso, es capaz de darnos una solución coherente al problema.

Otra forma de describir el aprendizaje de las máquinas o *Machine Learning*, como un conjunto de técnicas, en su mayoría de tipo estadístico, para analizar una gran gama de posibilidades con el fin de poder determinar la mejor solución a cierto problema, o bien encontrar relaciones entre datos así como también clasificar la información, y lo mas importante poder establecer un regla para la interpretación de los datos. *Machine Learning* tiene como meta crear

sistemas que se estén mejorando continuamente debido a la experiencia que van a ir tomando, es decir, que con forme el paso del tiempo, los sistemas vayan aprendiendo mas y mas, con el fin de ser mas certeros en los resultados que éstos propicien. “La predicción es la clave” [36].

*Machine Learning* es un eslabón dentro de la Inteligencia Artificial, que a su vez se nutre del conocimiento de otras disciplinas, la estadística y probabilidad, un poco de filosofía y Ciencias de la Computación. Cada disciplina es igual de importante, por ejemplo, el uso de regresiones, taxonomía numérica, modelos estocásticos, series de tiempo etc.

### 3.2 Tipos de Aprendizajes

El aprendizaje computacional puede ser dividido en 2 categorías:

Tabla II. Tipos de aprendizaje computacional

Tipo de Aprendizaje	Método de Inferencia
Inductivo / Supervisado	Árboles de Decisiones Reglas de Clasificación
Conceptual / No Supervisado	Redes Bayesianas Algoritmos Genéticos

- Aprendizaje Inductivo: Se llama aprendizaje inductivo cuando una persona, o bien para nuestro caso una máquina, trata de aprender de alguna base de conocimiento muy grande, es decir, a partir de algún patrón previamente observado, o a partir del ambiente que lo rodea, obteniendo así reglas y patrones que interrelacionan a los datos entre sí.

Por decirlo de otra manera, a partir de los datos con los que ya se cuenta, se construye una teoría, la cual predice el comportamiento de los demás ítems, esta teoría que se construye, no es más que un criterio de agrupación. A este tipo de aprendizaje se dice que se necesita un “maestro”, es decir, alguien o algo, que se encargue de estar alimentando el sistema con la información necesaria así como también de crear definiciones de clases, con las cuales se pueda tener una base o referencia para la toma de decisión<sup>15</sup>.

- Aprendizaje conceptual o no dirigido: A diferencia del inductivo, aquí no es necesario un “maestro”, o una persona encargada de estar alimentando el sistema con grandes cantidades de información, ya que con solo tener la información que se desea analizar, se genera la clasificación a través de las observaciones. Aquí no hay objetos preclasificados, lo que se utiliza es una función, por decirlo de ese modo, que evalúe o bien genere las clases necesarias o convenientes para la información.

### 3.2.1 Clasificación

Básicamente la clasificación consiste, en asignar una etiqueta de una clase existente, a un conjunto de clases o datos que se encuentran sin clasificación alguna. Para poder lograr esto se necesitan como mínimo dos pasos a seguir. El primer paso, es algo que ya se mencionó en el párrafo anterior, “entrenar” el sistema con un juego de datos y clases previamente establecidas por el “maestro”. Una vez hecho esto, pues ya se da a la tarea de clasificar la información de acuerdo a su base de conocimiento. Dentro de la

---

<sup>15</sup> *Classification and Data Clustering Study in Incomplete Data Domain, Jorge Rodas, Javier Gramajo, University of Catalonia.*

clasificación, se encuentran lo que son los Árboles de Decisión, Redes Bayesianas, Redes Neuronales, entre otros.

Los Árboles de Decisión enfrentan el problema en forma de pasos a seguir, es decir, siguen un camino establecido a partir de diferentes instancias. Los nodos en un Árbol de Decisiones evalúan una característica en particular, éstos la comparan contra un valor fijo, preestablecido. Es decir que, conforme se va descendiendo del árbol, los atributos se van comparando con cada nodo del árbol, de tal forma que toda información que alcance dicho nodo, contiene las mismas características que cualquier otro dato que logró alcanzar ese nodo a ese nivel del árbol. De tal forma, que el dato que se esta evaluando va descendiendo por el árbol hasta llegar a lo que se conoce como las hojas del árbol, que no es mas que el fin del recorrido, es decir nodos dentro de la estructura que no contiene otros nodos hijos. En esas hojas, se encuentran las etiquetas, con las cuales se clasificaría el dato entrante. A continuación se presenta un ejemplo del funcionamiento de un árbol como el descrito anteriormente.

Los datos que se muestran a continuación en la tabla III, corresponden a la clasificación del Mal contra el Bien, es decir, clasificar a una persona de acuerdo a su apariencia si ésta es mala o buena. En la figura se observan los personajes previamente clasificados, por el maestro, tal es el caso de Batman, Robin, Alfred como personas “Buenas”, y el Pingüino, la Gatúvela y el Guasón como “Malos”. Estas clasificaciones se realizaron en base al conjunto de atributos o características que éstos tienen. Por ejemplo si utilizan máscara, capa, corbata, o si fuman.

Tabla III. Ejemplo de datos que conformarían un Árbol de Decisiones.

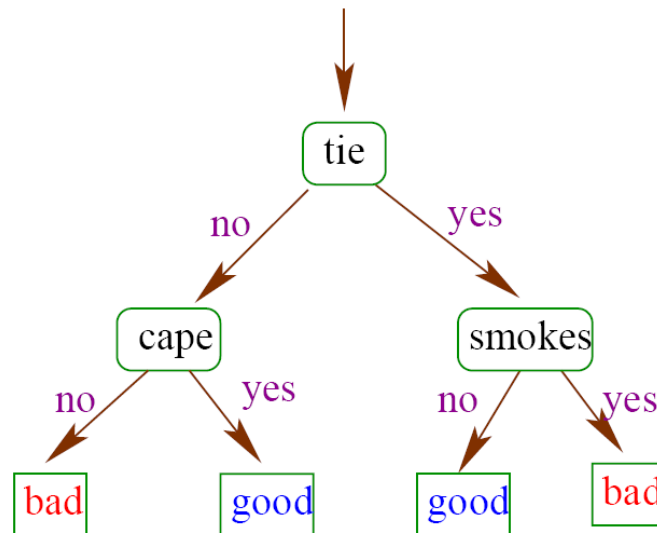
- problem: identify people as good or bad from their appearance

	sex	mask	cape	tie	ears	smokes	class
	<u>training data</u>						
batman	male	yes	yes	no	yes	no	Good
robin	male	yes	yes	no	no	no	Good
alfred	male	no	no	yes	no	no	Good
penguin	male	no	no	yes	no	yes	Bad
catwoman	female	yes	no	no	yes	no	Bad
joker	male	no	no	no	no	no	Bad
	<u>test data</u>						
batgirl	female	yes	yes	no	yes	no	??
riddler	male	yes	no	no	no	no	??

Fuente: <http://www.cs.princeton.edu/courses/archive/fall04/cos557/Articles/schapire-slides.pdf>

En la parte inferior de la figura se aprecia los datos de prueba, los que se pretenden clasificar a partir de los datos con que ya se cuentan. Como se observa en la siguiente figura (Figura 8), los personajes poco a poco se van clasificando de acuerdo a sus atributos y las definiciones.

Figura 8 Ejemplo de funcionamiento de un Árbol de Decisión



Fuente: <http://www.cs.princeton.edu/courses/archive/fall04/cos557/Articles/schapire-slides.pdf>

Por ejemplo, se establece que si no tiene corbata y si no tiene capa entonces se clasifica a la persona como “Mala”, de lo contrario la persona sería clasificada como “Buena”. De esta forma es básicamente como un árbol de decisiones trabajaría.

Un buen algoritmo de estos árboles el es ya conocido ID3. Este algoritmo es no incremental, es decir que el tamaño del árbol no crece con forme a la experiencia de forma automática y sin supervisión alguna. Como cualquier otro árbol, el algoritmo del ID3 también necesita atributos en los datos, así como clases predefinidas. La fórmula, por así decirlo, que utiliza el ID3, es una propiedad estadística llamada en inglés “*Information Gain*”, que traducido al español vendría siendo “Ganar Información”. Este algoritmo trabaja de mejor manera clasificando de forma más efectiva los datos. Si se desea obtener más información acerca del funcionamiento del ID3, se puede consultar la siguiente dirección. [<http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm>].

Los métodos Bayesianos, por otro lado, proveen una de las formas más antiguas para realizar clasificaciones, siempre hablando del aprendizaje supervisado. Este tipo de métodos, estiman las distribuciones, de las probabilidades condicionales, para cada atributo, teniendo siempre en cuenta la etiqueta de la clase, la cual se obtiene obviamente de la base de datos. La exactitud de los métodos Bayesianos, se ve afectado directamente por la base de datos, ya que si ésta no está completa, es decir no cuenta con todos los juegos de datos, entonces la precisión de la clasificación no sería la mejor<sup>16</sup>. Los modelos bayesianos, generalmente son modelos que expresan aspectos cualitativos del conocimiento [37]. Existen 3 formas de poder inferir a partir de métodos bayesianos.

- Comparando dos hipótesis
- Estimación por parámetros
- Selección a partir de modelos

La primera forma, pues es una forma muy estadística en la cual se comparan 2 hipótesis, es decir evaluar una posible solución estableciendo ciertos criterios de aceptación los cuales deben de ser satisfechos de modo que la hipótesis que se está evaluando sea aceptada como verdadera. Con esta forma únicamente se pueden comparar, como su nombre lo establece, dos hipótesis.

Por estimación de parámetros se pueden llegar a comparar el número de hipótesis que se desee, y por último con la selección a partir de modelos, aparte

---

<sup>16</sup> *Classification and Data Clustering Study in Incomplete Data Domain, Jorge Rodas, Javier Gramajo, University of Catalonia*

de poder comparar varias hipótesis a la vez, es posible comparar hipótesis que son diferentes de forma cualitativa, es decir no es necesario contar con hipótesis homogéneas para poder realizar análisis.

### **3.2.2 Clustering**

Por definición: *Clustering* son regiones continuas de un espacio que contiene relativamente una alta densidad de puntos, separados uno de otro por las regiones que contienen una densidad relativamente baja de puntos<sup>17</sup>. En otras palabras, es un conjunto de técnicas utilizadas generalmente para la agrupación de los datos de una forma congruente y lógica.

Dentro del *Clustering*, existen diferentes técnicas las cuales pueden ser agrupadas de la siguiente forma [38]:

- Técnicas Jerárquicas: en esta técnica, las clases o datos son clasificadas en grupos, para ello el proceso de clasificación es bastante recurrente en varios niveles, formado así una estructura de árbol. Esta técnica es de las de uso más común, y su algoritmo básicamente es, primero encontrar los 2 objetos mas cercanos y mezclarlos, con el fin de hacer uno solo, luego con este nuevo objeto, repetir nuevamente el procedimiento, de buscar los 2 objetos mas cercanos y para crear uno nuevo y así sucesivamente hasta que ya no existen objetos que mezclar.
- Técnicas de para la optimización y el particionamiento: con esta técnica, las agrupaciones se van formado a partir de un criterio previamente establecido. Con esta técnica, por lo general el resultado son N Clústeres. Donde cada Clúster esta representado por su centroide, dicho

---

<sup>17</sup> Cluster Analysis, Heinemann Ed. Books Ltd, London 1981.



en otras palabras, su dato representante, el cual no es más que un dato que contiene las características de todos los elementos del Clúster. A diferencia de la técnica anterior, con esta se necesita únicamente recorrer el set de datos una sola vez.

- Técnicas para documentos de texto: aquí se toman en consideración las palabras clave que son encontradas en los documentos, de modo que cuando se solicite una consulta acerca de un texto, en lugar de consultar toda la base de datos en búsqueda de lo solicitado, la técnica lo que haces es ir a buscar esa palabra o frase al Clúster que la contenga como su palabra clave. Cabe mencionar que el orden en el que aparecerían los documentos resultantes de la búsqueda, depende del número de veces el cual aparezca dicha palabra clave [39].

Una de las más comunes aplicaciones del *Clustering*, es la Minería de Datos, o *Data Mining*, que por definición se entiende como el proceso para descubrir, nuevos patrones, tendencias o correlaciones significativas a partir de una gran cantidad de datos, utilizando técnicas de reconocimiento de patrones, así como técnicas estadísticas y matemáticas.

La diferencia básica entre *Machine Learning* y *Data Mining*, es que *Machine Learning* aparte de buscar patrones y similitudes de entre los datos, crea reglas en base a dichos patrones, en base a lo observado, con el fin de poder realizar las clasificaciones de una forma más rápida y eficaz.

Cuando se habla de aprendizaje computacional, es importante mencionar a las *Redes Neuronales*, las cuales son capaces de trabajar tanto con un aprendizaje supervisado, como con uno no supervisado.

El uso de las redes neuronales es un poco complejo, ya que el verdadero problema podría decirse es el poder “entrenarlas” de la mejor forma, y es común que las redes neuronales, tengan o encuentren problemas al momento de resolver tareas muy simples. Pero éstas redes, poco a poco han ido ganando popularidad y lo que es importante es que son bastante aceptadas debido a que éstas, a pesar de su complejidad de entrenamiento, pueden llegar a resolver tareas tan complejas para las que no pueda existir nunca un algoritmo. Por ejemplo, si se habla del procesamiento de imágenes, una red neuronal bien construida, podría separar la imagen por sus diferentes texturas. Esto es un ejemplo del nivel de complejidad de los problemas que pueden ser resueltos por una red neuronal, ya que sería prácticamente imposible poder implementar un algoritmo el cual clasifique las texturas a partir de una imagen.

Otra ventaja de su uso, es que ellas se pueden adaptar fácilmente al comportamiento de su ambiente. Esto significa que si el comportamiento de una red neuronal fue definido desde su inicio, este puede volver a redefinirse, es decir su implementación no representaría grandes problemas [40].

### **3.3 Repositorios de Datos**

Un problema común, para una persona que se esta iniciando en el tema de Machine Learning, es el poder conseguir grandes cantidades de datos para poder ser analizados, ya que a menos que se posea una empresa con una base de datos enorme, es un poco complejo el poder obtener dichos. Actualmente en el internet existen sitios web con archivos en diferentes formatos. Un formato muy popular es el ARFF, que por sus siglas en inglés es Attribute-Relation File Format [<http://www.cs.waikato.ac.nz/~ml/weka/arff.html>]

. Un archivo ARFF, no es más que un archivo de texto, el cual describe una lista de instancias (datos), las cuales comparten una serie de atributos. Estos archivos fueron ideados por el departamento de Ciencias de la Computación en la Universidad de Waikato.

Básicamente un archivo con formato ARFF, tiene dos secciones, el encabezado y los datos. En el encabezado se almacena la información del nombre de la relación y su lista de atributos con su tipo de dato.

**Figura 9 Ejemplo de un archivo ARFF**

```
§ 1. Title: Iris Plants Database
§
§ 2. Sources:
§   (a) Creator: R.A. Fisher
§   (b) Donor: Michael Marshall (MARSHALL@PLU@io.arc.nasa.gov)
§   (c) Date: July, 1988
§
@RELATION iris

@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class      {Iris-setosa,Iris-versicolor,Iris-virginica}
```

The **Data** of the ARFF file looks like the following:

```
@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
```

**Fuente:** <http://www.cs.waikato.ac.nz/~ml/weka/arff.html>

En la figura se aprecian claramente las dos secciones por las cuales, esta formado un archivo ARFF.

Este formato de archivo es comúnmente utilizado, por la herramienta *Weka*, la cual se describe en la sección siguiente. El sitio de *Weka*, contiene

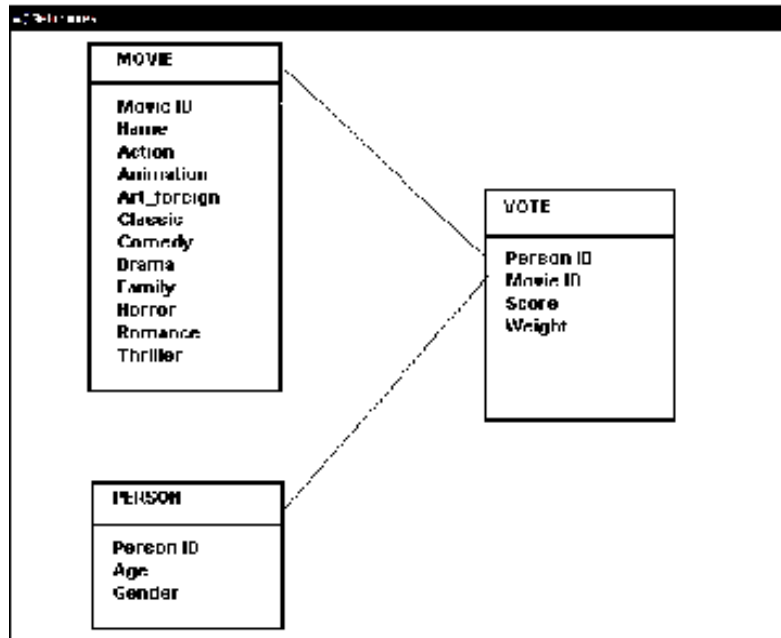
algunos repositorios de datos, bastante funcionales [<http://www.csc.lsu.edu/~ji/compbio/index.htm>]. Además se puede encontrar una gran variedad de archivos listos para análisis en [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. En dicho sitio existen archivos con datos de diferentes áreas de aplicación, de tal forma que una persona interesada pueda escoger entre diferentes dominios.

### **3.4 Datos sintéticos**

Los datos sintéticos surgen de la necesidad de los investigadores, los cuales deseaban explorar los diferentes efectos de ciertas características de los datos en los modelos que se realizaban. De tal forma que, construyeron sets completos de datos, los cuales contenían ciertas propiedades, por ejemplo de auto correlación. Con esto ellos fueron capaces de construir datos sintéticos con una o dos tipos de estructuras de gráficas, las están distribuidas de forma independiente e idéntica [41].

Los datos sintéticos almacenados en las denominadas bases de datos sintéticas. Estas bases de datos son utilizadas para realizar pruebas específicas de Clustering y Clasificaciones. Por ejemplo, si se tiene una base de datos sintética de ejemplo, en el cual se llevará un control de la votación de las personas y sus gustos en lo que respecta a tipos y géneros de películas. El set de datos quedaría así.

Figura 10 Modelo entidad relación de la base de datos sintética para el ejemplo descrito



Fuente: <http://www.research.digital.com/SRC/eachmovie/>

En la figura se aprecian las tablas, las cuales almacenan información, tanto de la persona, así como también almacenan bastante información acerca de los diferentes tipos de películas, con el fin de poder obtener la mayor información posible. De modo que cuando se realice la encuesta, y la tabla *VOTE*, se encuentre plenamente llena, va a ser posible correr los algoritmos de clasificación y clustering, para lograr encontrar algún patrón o características similares en gustos de películas entre personas de diferentes edades [42].

### 3.5 *Missing Values* (Valores Faltantes)

Por definición un Missing Value o Valor Faltante, es según el diccionario de Machine Learning de la Universidad de Stanford [<http://robotics.stanford.edu/~ronnyk/glossary.html>], un Valor Faltante es cuando un valor de un cierto atributo no es conocido o simplemente no existe.

Hasta el momento se ha hablado de diferentes formas de analizar, y clasificar la información, pero es común que no siempre se tenga la información completa, es decir, que con el ejemplo anterior, de la encuesta acerca de los gustos de películas, nos haya faltado una combinación entre gustos y edades. Aunque para nuestro caso no es imperante conocer absolutamente todos los datos, es decir todas las combinaciones posibles, pero para algunos otros casos de análisis si lo es.

Por ejemplo, encontrar el camino óptimo entre dos puntos. Para este ejemplo es necesario contar con todas las posibles combinaciones con el fin de poder analizarlas y poder estar seguro que al momento de evaluar los caminos, se obtendrá el camino mas corto. Otro ejemplo de la necesidad de poder contar con todos los datos es cuando se realizan aplicaciones estadísticas en la genética y biología molecular [<http://www.bepress.com/sagmb/vol4/iss1/art10/>].

Para ejemplos como los descritos anteriormente, procesar una gran cantidad de datos, los cuales pues, no estén completos, puede ser una tarea muy ardua y difícil de completar. Muchos programas estadísticos, lo que hacen simplemente es omitir las observaciones incompletas, es decir, como se mencionó anteriormente, si algún dato no cuenta con su juego de atributos completo, entonces, éste es completamente ignorado, el problema de suprimir

datos, toma importancia si la cantidad de ellos no es muy grande, por lo que se elimina significado a los resultados [45].

Uno de los algoritmos más comunes, que se puede adaptar para poder encontrar valores faltantes, son los de Kohonen. Kohonen es uno de los investigadores más famosos en la denominada “Neurocomputación”, cabe mencionar que el ha desarrollado varios tipos de redes neuronales, entre los cuales podemos mencionar 3 de los más importantes, VQ (Vector Quantization), SOM (Self Organizing Map) y por último LVQ (Learning Vector Quantization). Si se desea analizar a mayor detalle dichas redes se puede consultar [<http://www.faqs.org/faqs/ai-faq/neural-nets/part1/section-11.html>].

### **3.6 Herramientas GPL**

Una herramienta gratuita que permite realizar análisis como los antes descritos, es *Weka (Waikato Environment for Knowledge Analysis)* [<http://www.cs.waikato.ac.nz/ml/weka/>]. Esta herramienta de minería de datos, fue desarrollada por la universidad de Waikato en Nueva Zelanda. Dicha herramienta contiene una gran variedad de algoritmos (151 aproximadamente) de Machine Learning, los cuales pueden ser mandados a llamar directamente desde la aplicación hacia una base de datos, o bien éstos pueden ser utilizados directamente desde cualquier código escrito en java.

En general, Weka provee herramientas para realizar *Clustering*, regresiones, clasificaciones y todo esto lo permite analizar de forma totalmente gráfica [45]. La herramienta puede procesar grandes cantidades de datos en archivos de formatos ARFF, CSV. Como se mencionaba anteriormente, esta herramienta es capaz de analizar grandes bases de datos, y a su vez, proporcionar un resultado, ya sea lograr realizar un clasificación de los objetos,

encontrar relaciones entre ellos, crear índices, realizar regresiones y con ello lograr predecir comportamientos o bien tendencias en base a la información analizada.

El sitio oficial de Weka tiene algunos repositorios de datos gratuitos, los cuales pueden ser descargados desde el sitio, para poder ser utilizados por la herramienta. Por otro lado, es común encontrar sitios con este tipo de bases de datos, por ejemplo, en el laboratorio de computación de Berkley [<http://crd.lbl.gov>], es posible descargar este tipo de archivos para diferentes áreas de aplicación la medicina [<http://crd.lbl.gov/~cding/>], y muchos otros, por lo general es un poco difícil encontrar archivos de este tipo, para realizar análisis de prueba, ya que la cantidad información es grande.



## CONCLUSIONES

1. Las pequeñas y medianas empresas forman parte de un sector, que muchas veces, es excluido y olvidado en la mayoría de países de Latinoamérica, por lo que si a éstas se les prestara más atención y soporte, fueran un factor clave en el desarrollo de un país.
2. Una buena práctica, es adoptar estándares internacionales, si los hubiese, al momento de realizar cualquier clasificación, de tal forma que el intercambio de datos entre diferentes sistemas no sea un obstáculo que vencer.
3. Con forme el avance de la utilización e implementación de la Web Semántica en la Red Mundial (WWW), los negocios en línea crecerán y llegarán a ser algo común en las comunidades mundiales.
4. En Guatemala, aún no existe un marco legal, bien definido, que soporte el crecimiento y la aplicación de las nuevas Tecnologías de la Información y Comunicación, las leyes actuales no son los suficientes para brindar el marco legal necesario.
5. Las herramientas de análisis de datos, utilizando algoritmos de aprendizaje computacional, como Weka, pueden ser de gran ayuda al momento de realizar cualquier tipo de análisis sobre una gran cantidad de información, ya que éste tipo de herramientas además de ser confiables son gratuitas, lo cual evitaría que las PYME's tuvieran que gastar en licencias.

6. En Guatemala existen varias áreas de oportunidad en donde la economía del país puede crecer si se logra aprovechar al máximo los beneficios que brinda el comercio electrónico, si éste fuera aprovechado por las pequeñas y medianas empresas.

## RECOMENDACIONES

1. Que la aplicación OntoPYME sea ampliada y desarrollada en todas sus dimensiones para darle continuidad y que ésta permita ampliar las capacidades de las PYMEs.
2. Propiciar el apoyo nacional de las PYMEs por medio de programas de Incubadoras, que permitan atraer capitales de inversión que facilite a las PYMEs generar proyectos de Innovación y, por tanto, generar desarrollo para el país.
3. Crear un programa de Incubadoras de PYMEs con base tecnológica, en donde se apoye a los emprendedores con un plan integral.
4. Entender cuál es el clima apropiado para que las PYMEs puedan tener ventajas competitivas, respecto a las empresas con mayor capacidad de inversión en tecnología, Investigación e Innovación.
5. Es recomendable que PYMEs adopten las TICs como elementos trascendentales para su desarrollo y sostenibilidad a lo largo del tiempo.



## BIBLIOGRAFÍA

1. Naciones Unidas, **La PYME en Centroamérica y su vinculación con el sector externo**, Publicación de la Comisión Económica para América Latina y el Caribe. 1999.
2. Instituto de Investigaciones Económicas y Sociales, **El Entorno Financiero de la Microempresa**. Guatemala: Editorial Universitaria Rafael Landívar, 2002.
3. Naciones Unidas, **Clasificación Central de Productos .CPC. versión 1.0**. Estados Unidos 1999.
4. Naciones Unidas, **La inversión extranjera en América Latina y el Caribe**, Publicación de las Naciones Unidas
5. Laudon, Kenneth C, Traver, Guercio, **ECommerce: Bussines, Technology**. Estados Unidos: Addison Wesley, 2002.
6. López, Manuel, **Desarrollo de Sistemas en Internet**. Guatemala: Editorial Universitaria, 1999.
7. Asociación Gremial de Exportadores de Productos no Tradicionales, **La clave del Comercio Electrónico-Guía para pequeños y medianos exportadores**, 2000.
8. López, Javier Gramajo, **State of Art: Ontologies**. Guatemala: Editorial Universitaria, 2005.
9. López, Javier Gramajo, Rodas, Jorge, **Classification and Clustering in Incomplete Data Domain**. Spain, Technical University of Catalonia. 2000.

10. Chang, David Armando, **Análisis e Implementación de una Ontología de Actividades Económicas y Productos para PYMEs de Guatemala**, Guatemala: Editorial Universitaria 2007.

## REFERENCIAS ELECTRÓNICAS

1. El arte de la telepresencia en Internet.  
<http://www.ekac.org/telepresencia.html>
2. Dirección General de Política de la PYME  
<http://www.ipyme.org/IPYME/es-es>
3. Comisión de Estadística de las Naciones Unidas  
<http://unstats.un.org/unsd/statcom/doc01/2001-14s.pdf>
4. Comisión de Estadística de las Naciones Unidas  
<http://unstats.un.org/unsd/class/intercop/newsletter/issue5/newslet5s.pdf>
5. Comisión de Estadística de las Naciones Unidas – Clasificación Industrial Internacional Uniforme, Revisión 4  
<http://unstats.un.org/UNSD/cr/registry/docs/istic4s-070306.pdf>
6. Comisión de Estadística de las Naciones Unidas  
[http://unstats.un.org/unsd/cr/registry/docs/i31\\_ict\\_s.pdf](http://unstats.un.org/unsd/cr/registry/docs/i31_ict_s.pdf)
7. Fórum de Comercio Internacional  
[http://www.forumdecomercio.org/news/fullstory.php/aid/537/Evoluci%F3n\\_dde\\_mercado.html](http://www.forumdecomercio.org/news/fullstory.php/aid/537/Evoluci%F3n_dde_mercado.html)
8. Preguntas mas frecuentes sobre la sociedad de la información  
[http://www.telefonica.es/sociedaddelainformacion/html/faq\\_home.shtml#3](http://www.telefonica.es/sociedaddelainformacion/html/faq_home.shtml#3)
9. Fundación boliviana de Informática y Telemática  
[http://fundabit.me.gob.ve/index.php?option=com\\_content&task=view&id=191&Itemid=80](http://fundabit.me.gob.ve/index.php?option=com_content&task=view&id=191&Itemid=80)
10. Verisign – Certificados Digitales  
<http://www.verisign.com>
11. Intranet, Intranets, Extranets  
<http://www.ucm.es/info/multidoc/multidoc/revista/cuad6-7/eulalia.htm>
12. The World Factbook  
<https://www.cia.gov/cia/publications/factbook/geos/gt.html#Comm>

13. International Telecommunication Union  
[http://www.itu.int/ITU-D/icteye/Reporting/ShowReportFrame.aspx?ReportName=/WTI/CellularSubCellularSubscr&RP\\_intYear=2005&RP\\_intLanguageID=1](http://www.itu.int/ITU-D/icteye/Reporting/ShowReportFrame.aspx?ReportName=/WTI/CellularSubCellularSubscr&RP_intYear=2005&RP_intLanguageID=1)
14. Prensa Libre – Artículo “Sector que sigue creciendo”  
<http://www.prensalibre.com/pl/2007/septiembre/29/183729.html>
15. Siglo XXI – Artículo “Acuerdan impulsar comercio electrónico”  
<http://www.sigloxxi.com/index.php?link=noticias&noticiaid=5063>
16. Prensa Libre – Artículo “E-Commerce, la tendencia que viene”  
<http://www.prensalibre.com/pl/2006/abril/17/139341.html>
17. Banco Interamericano de Desarrollo  
<http://www.iadb.org/projects/Project.cfm?project=GU-M1002&Language=Spanish>
18. Prensa Libre – Artículo “Podrían aprobar firma electrónica”  
<http://www.prensalibre.com/pl/2007/marzo/09/165293.html>
19. Ley General de Telecomunicaciones  
<http://www.itu.int/ITU-D/treg/Legislation/Guatemala/leygen.pdf>
20. CONCYT  
<http://www.concyt.gob.gt/pag.php?id=15&idioma=1&sitio=concyt>
21. CONCYT – Plan Nacional 2005-2014  
<http://www.concyt.gob.gt/pag.php?id=17&idioma=1>
22. CONCYT – Informe de Actividades 2006  
<http://www.concyt.gob.gt/uploads/archivos/memoriaCONCYT06.pdf>
23. CONCYT – Información General sobre los Centros Comunitarios Digitales  
<http://www.concyt.gob.gt/uploads/archivos/CENTROSCOMUNITARIOSDIGICENTR.doc>
24. Web Forum  
[http://www.weforum.org/en/media/LatestPressReleases/gitr\\_2007\\_press\\_relrele](http://www.weforum.org/en/media/LatestPressReleases/gitr_2007_press_relrele)
25. Asociación Guatemalteca de Exportadores  
<http://housecall65.trendmicro.com/>



26. Prensa Libre – Artículo “Darán Financiamiento para Impulsar comercio electrónico  
<http://www.prensalibre.com/pl/2006/julio/08/146289.html>
27. Fundación para el Desarrollo de Guatemala  
<http://www.fundesa.org.gt/cets2.php>
28. Semantic Web RoadMap  
<http://www.w3.org/DesignIssues/Semantic.html>
29. Desarrollando una ontología sencilla  
<http://www.infor.uva.es/~cllamas/MAS/DesOntoSimpl.pdf>
30. Ontologías en la Web Semántica  
<http://www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf>
31. Uso de las ontologías en tareas de recupero de la información  
<http://usuarios.fceia.unr.edu.ar/~tallaric/tesina/presentacion.pdf>
32. Wikipedia  
<http://es.wikipedia.org/wiki/Taxonom%C3%ADa>
33. Taxonomías para la categorización y organización de la información en sitios web  
<http://www.hipertext.net/web/pag264.htm>
34. Euro WordNet  
<http://www.illc.uva.nl/EuroWordNet/>
35. Diccionario Clave  
<http://clave.librosvivos.net/>
36. MIT Open CourseWare  
<http://ocw.mit.edu/NR /rdonlyres/Electrical-Engineering-and-Computer-Science/6-867Machine LearningFall2002/7AF7673F-EF49-4F65-8A67-B9B58860E8DD/0/lecture1.pdf>
37. Bayesian Methods for Machine Learning  
<ftp://ftp.cs.utoronto.ca/pub/radford/bayes-tut.pdf>
38. Data Clustering Techniques  
<http://www.disi.unitn.it/~periklis/pubs/depth.pdf>

39. Data Clustering and Its Application's  
[http://members.tripod.com/asim\\_saeed/paper.htm](http://members.tripod.com/asim_saeed/paper.htm)
40. Introduction to Machine Learning  
<http://robotics.stanford.edu/people/nilsson/mlbook.html>
41. Synthetic Data Generation  
<http://kdl.cs.umass.edu/software/documentation/tutorial/ch06s09.html>
42. Research Digital  
<http://www.research.digital.com/SRC/eachmovie>
43. Glossary of Terms  
<http://robotics.stanford.edu/~ronnyk/glossary.html>
44. Missing Values: processing with the Kohonen algorithm  
<http://asmda2005.enst-bretagne.fr/IMG/pdf/proceedings/489.pdf>
45. Data Mining: Practical Machine Learning Tools and Techniques  
<http://www.cs.waikato.ac.nz/~ml/weka/book.html>

## APÉNDICE

### A.1. Utilización de OntoPyme

#### A.1.1 Requerimientos Mínimos

La aplicación fue desarrollada utilizando el entorno de desarrollo que proporciona Borland JBuilder 2005 Enterprise, utilizando de base del JDK 1.5.0\_13.

Los requerimientos mínimos para la correcta ejecución de la aplicación son:

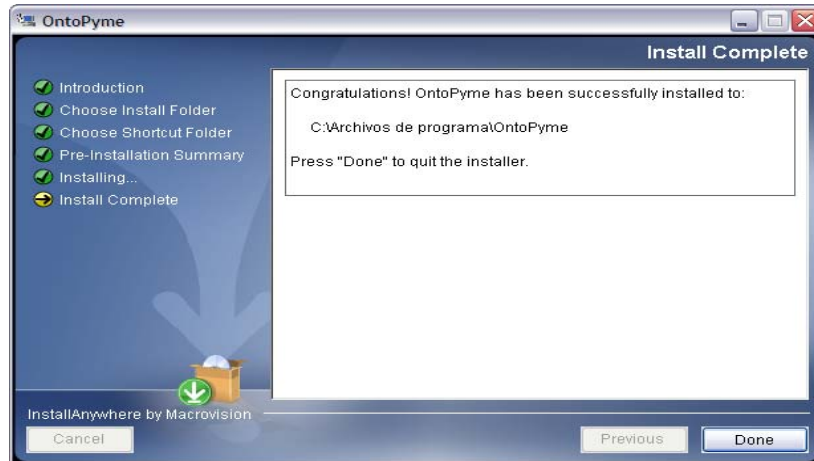
1. Java Runtime Enviroment versión 1.5.0\_13 o superior
2. Base de datos Mysql versión 5.0.45 o superior
3. Windows XP
4. Pentium III 750 Mhz
5. 256Mb de memoria RAM
6. Conexión a Internet de mínimo 64kbps.
7. Internet Explorer 5.0 o superior

#### A.1.2 Instalación



Para poder instalar la aplicación, primero se debe de hacer clic sobre el icono Install.exe que se proporciona en el CD. Luego debe de seguir las instrucciones que se presentan para poder realizar la instalación. Se recomienda no realizar ninguna modificación a las configuraciones estándar. La instalación finalizará cuando aparezca esta pantalla.

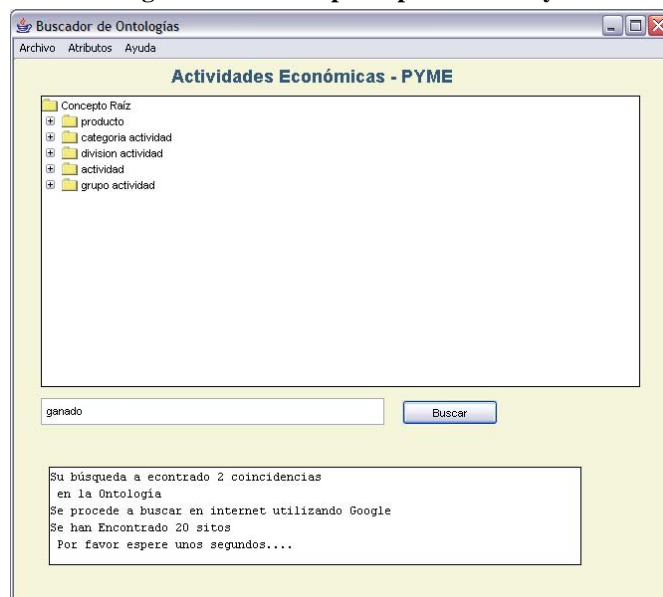
**Figura 11 Pantalla que muestra el fin de la instalación**



Una vez instalado el producto, debes de hacer clic en Inicio → Todos los Programas → OntoPyme (Este path puede variar dependiendo de la configuración y/o versión de Windows).

### A.1.3 Pantalla Principal

**Figura 12 Pantalla principal de OntoPyme**



La pantalla principal de la aplicación, esta compuesta básicamente de 5 elementos importantes, el primero la barra del menú, desde el cual el usuario podrá encontrar diferentes opciones entre ellas, Guardar y Cargar búsquedas, agregar atributos (todas estas opciones se explicarán mas adelante). La siguiente sección, la que se denota con el número 2, muestra las partes de la ontología que previamente se tuvo que haber cargado al sistema. En la sección número 3, el usuario debe de ingresar las palabras clave, las cuales servirán para realizar la búsqueda, primero en la ontología, para luego ir a buscar a Internet páginas relacionadas con lo encontrado en la Ontología. Y por último en la sección número 4, se muestran resultados de las búsquedas, tanto en la ontología, como en Internet.

#### A.1.4 Cargar una Ontología



En el menú, Archivo → Abrir RDF. Se encuentra la opción de cargar el archivo que contiene la ontología en formato RDF.

Una vez cargada la Ontología, te aparecerá, como lo muestra la figura anterior en la sección número dos, un árbol con cada elemento de la Ontología del dominio de la Pyme, la cual puedes encontrar dentro del folder “Ontología” que se incluye en la instalación. Luego, solo debes de ingresar una o varias palabras clave, o bien debes de hacer clic en ellas desde el árbol, para luego presionar el botón buscar.

#### A.1.5 Guardar una Búsqueda

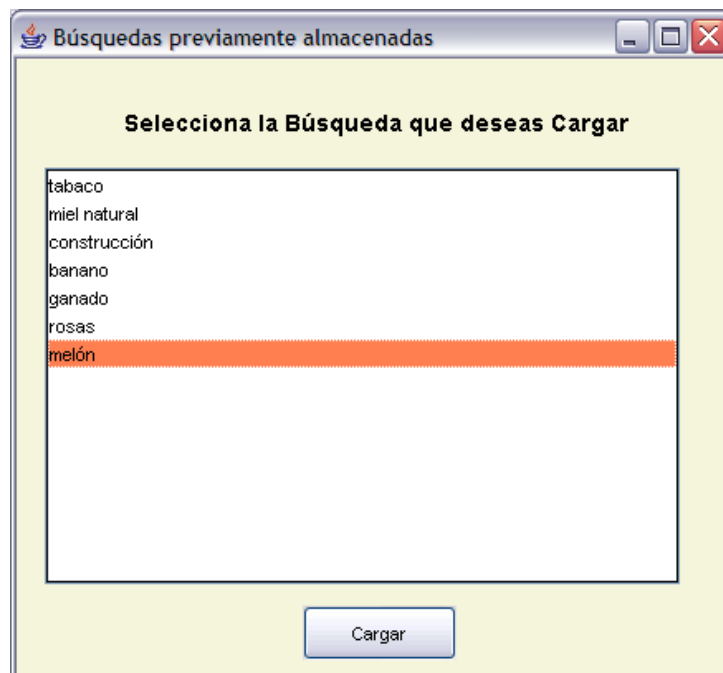
Una vez hayas efectuado una búsqueda, si esta resultó ser lo que se esperaba, existe la opción de almacenarla en la base de datos, para luego poderla consultar de nuevo, sin necesidad de realizar la carga del archivo que contiene

la ontología, o incluso sin necesidad de conexión a Internet. Esta opción se encuentra en el Menú Archivo → Guardar Búsqueda.

#### A.1.6 Cargar una Búsqueda

Puedes cargar búsquedas previamente almacenadas, la forma de identificar las búsquedas, es a través de las palabras clave que hallas ingresado

**Figura 13** Pantalla que muestra búsquedas previamente almacenadas



Como se puede observar en la imagen, solo se debe de seleccionar la búsqueda, y luego hacer clic en el botón de Cargar. Seguido, se mostrará la ventana de resultados, la cual se hablará mas adelante.

### A.1.7 Mantenimiento de Atributos

Los atributos son todos aquellos elementos que el usuario puede crear, con el fin de poder almacenar más información acerca de los links obtenidos. Es decir, un atributo puede ser una característica en común que contengan varios links, a los cuales se les deseará realizar un análisis más profundo utilizando algoritmos de aprendizaje computacional. Para poder crear atributos, se debe de ingresar al Menú Atributos y luego seleccionar la opción Mantenimientos.

**Figura 14** Pantalla que muestra el mantenimiento de conceptos



The screenshot shows a software window titled "Mantenimiento de Conceptos" with a close button in the top right corner. The main content area is titled "Mantenimiento de Atributos" and contains the following elements:

- A label "Atributos Acutales" followed by a dropdown menu currently displaying "Producto".
- A label "Nuevo Atributo" followed by an empty text input field.
- A label "Tipo" followed by a dropdown menu currently displaying "String".
- Two buttons at the bottom: "Guardar" (Save) and "Eliminar" (Delete).

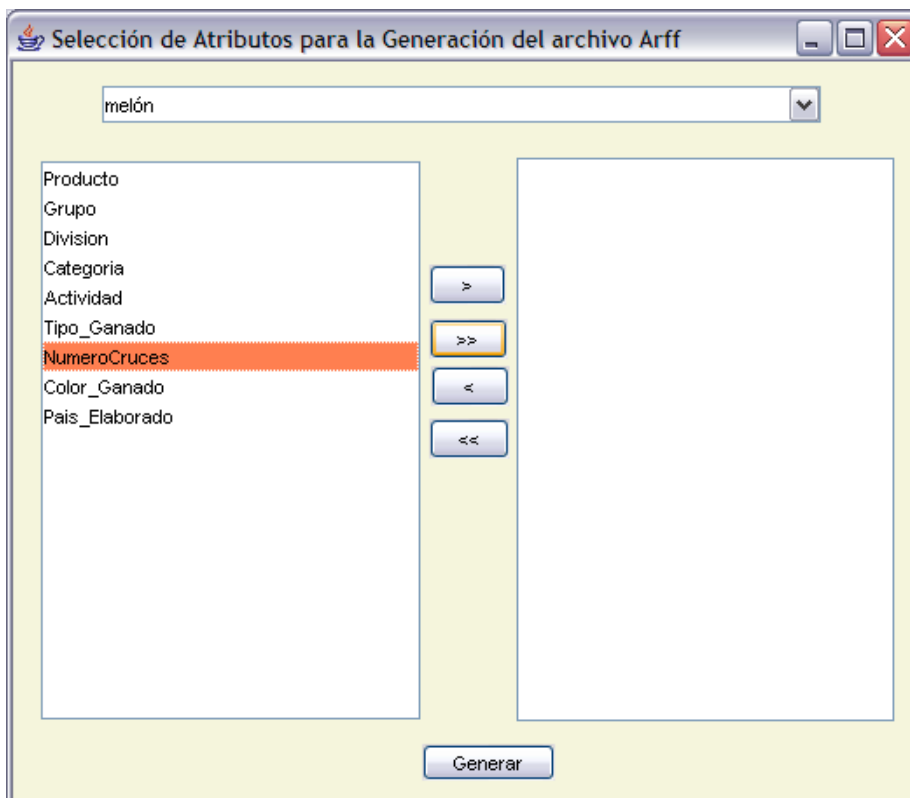
Como lo muestra la figura, primero se muestran los atributos actuales, seguido de un campo llamado Nuevo Atributo, en el cual debes de ingresar el nombre del atributo que se desee crear. Y por último, se encuentra el campo Tipo, en el cual se debe de indicar el tipo de dato que el atributo tendrá, únicamente están habilitados dos tipos de datos, "String" o cadena de caracteres, y "Numeric" o número. Debes de tener en cuenta estos tipos de datos, ya que al momento de

generar el archivo de tipo arff, los datos que hayas ingresado serán interpretados de la forma que indique el tipo de dato seleccionado.

#### A.1.8 Generar ARFF

En el Menú principal, siempre en la parte de Atributos, se encuentra la opción de Generar Arff, la cual permite, como su nombre lo dice, la generación del archivo Arff, para luego, ser procesado por alguna herramienta capaz de interpretar dichos archivos, tal es el caso de Weka. La pantalla previa a generar el archivo luce así.

**Figura 15 Selección de atributos para la generación del archivo arff**



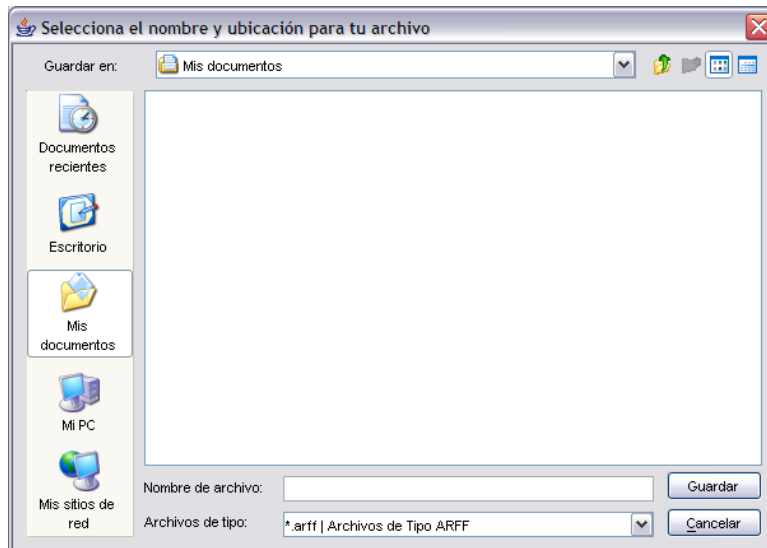


En el lado superior, se muestran todas las búsquedas que se han almacenado en la base de datos, nuevamente éstas salen por las palabras clave con las que éstas fueron realizadas. En el panel de la izquierda se muestran, los atributos disponibles para esa búsqueda, los cuales tuvieron que haber sido previamente creados y alimentados.

Para poder seleccionar los atributos que deseas que se incluyan en el archivo arff, únicamente debes de hacer clic sobre el nombre del atributo, luego presionar el botón ">", para así pasar el atributo NúmeroCruces, al panel de la derecha. De igual forma si ya habías seleccionado un atributo y ya no lo deseas, puede realizar la operación inversa. Ésta es, seleccionarlo en el panel derecho, y luego presionar el botón "<", o bien si simplemente deseas seleccionar todos los atributos puedes presionar el botón ">>" y automáticamente todos los atributos restantes del panel izquierdo se pasarán panel derecho.

Una vez seleccionados los atributos que se deseen incluir, puedes hacer clic en el botón Generar, y te aparecerá un cuadro de diálogo pidiendo que se seleccione la ubicación y el nombre del archivo a generar.

**Figura 16 Selección de la ubicación y nombre del archivo a generar**



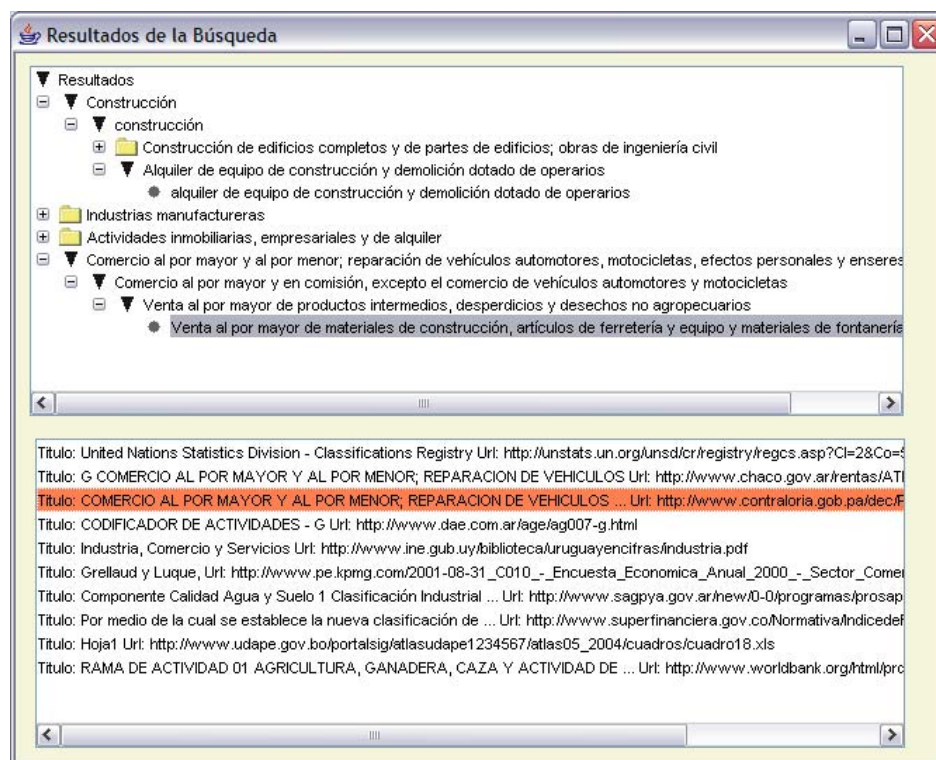
Una vez nombrado el archivo solo debes de presionar Guardar, y tu archivo se generará, y éste estará listo para ser analizado, como ya se mencionó anteriormente, por cualquier herramienta de procesamiento de archivos arff, tal es el caso de Weka.

#### A.1.9 Pantalla de Resultados

La pantalla de resultados, muestra, como su nombre lo indica, los resultados, tanto de la búsqueda realizada en la ontología, como los resultados de la búsqueda realizada en Internet para cada elemento que se encontró en la ontología. Esta pantalla esta formada básicamente en 2 partes, la primera, es el panel superior, en el cual se muestra en forma de árbol los resultados de búsqueda en la ontología. Tal y como está clasificada la ontología, primero aparecen las actividades, luego la categoría de la actividad, luego la división, seguido del grupo de la actividad y por último el producto. Es posible ir expandiendo y/o contrayendo los elementos en el árbol, dependiendo del nivel

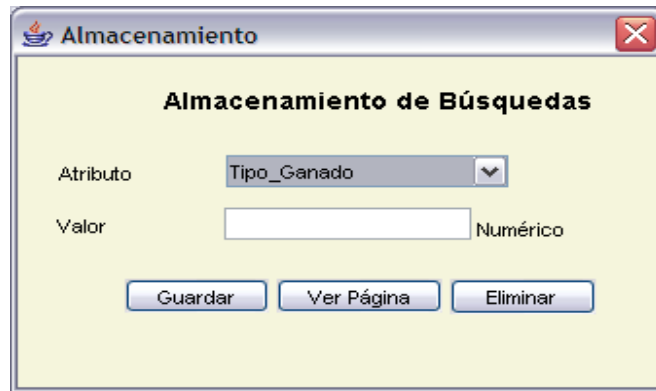
en el que éstos se encuentren, si el icono que aparece es un fólger amarillo, indica que dicho elemento contiene a su vez otros a un nivel jerárquico menor.

**Figura 17 Muestra de los resultados de la búsqueda**



En el panel inferior, dependiendo del elemento seleccionado en el panel superior, se muestran los links a los sitios Web, donde primero se encuentra el título del sitio seguido del Url del mismo. Por lo que, al momento de hacer clic sobre el Url, automáticamente abre un cuadro de diálogo, en el cual, es posible, almacenar el valor de algún atributo en la base de datos, relacionado a ese Url, eliminar un valor de algún atributo ingresado anteriormente, o bien simplemente abrir el Explorador de Windows, para navegar en el sitio

Figura 18 Almacenamiento de búsquedas



The image shows a software dialog box titled "Almacenamiento" with a close button in the top right corner. The main heading inside is "Almacenamiento de Búsquedas". There are two main input areas: "Atributo" with a dropdown menu currently showing "Tipo\_Ganado", and "Valor" with a text input field and the label "Numérico" to its right. At the bottom of the dialog, there are three buttons: "Guardar", "Ver Página", and "Eliminar".

Solo se debe de seleccionar el atributo a el cual se desea ingresar un valor, inmediatamente, se coloca el tipo de atributo a la par del campo valor, para que siempre se tenga en mente (en caso se haya olvidado). Si se desea eliminar algún valor, solo es de seleccionar el nombre en la lista desplegable, y luego presionar el botón Eliminar.

## A.2. Manual Técnico

La aplicación se compone básicamente en un paquete llamado Ontopyme, el cual esta organizado en dos carpetas UI, la cual contiene todos las clases gráficas de la aplicación, es decir, contiene las clases de las ventanas y cuadros de diálogo que dan forma a OntoPyme. El segundo directorio llamado Clases, contiene, como su nombre lo indica, todas aquellas clases que complementan la aplicación. A continuación se describen las clases más importantes contenidas en cada una de estas carpetas.

## A.2.1 UI

### A.2.1.1 Frm\_Main

Forma principal de la aplicación.

- Frm\_Main()

Constructor de la clase.

- jMenuItem1\_actionPerformed

Procedimiento que se activa al momento que el usuario ingresa al menú y selecciona la opción de Abrir un Rdf. El procedimiento muestra el cuadro de diálogo para la selección del archivo que se desea abrir. Luego se hace crea la una instancia de la clase Ontología. Para luego crear un DefaultTreeModel, el cual será alimentado dependiendo de las clases y conceptos existentes en la ontología previamente analizada. Por último se le indica elemento JTree que se alimente del árbol DefaultTreeModel previamente generado, con el fin que el usuario pueda visualizar las clases de forma jerárquica.

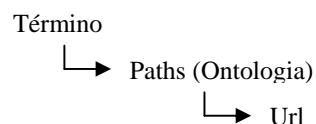
- btnGuardar\_actionPerformed

Procedimiento que se activa al momento que el usuario selecciona la opción de Guardar una búsqueda. Lo primero que realiza es buscar en la base de datos si ya existe dicha búsqueda, de no ser así, ingresa las palabras clave a la base de datos (tabla búsqueda), obteniendo así, un código que identifica a la búsqueda recién almacenada, con dicho llave, se procede a almacenar, los términos, de la búsqueda por separado (tabla término), esto debido a que por cada término, se realiza una búsqueda en la ontología, lo que brinda un “camino” diferentes dentro de la

ontología por lo que se almacena el término y luego en la tabla path\_onto, se almacenan los diferentes path o caminos encontrados. Una vez hecho esto, se obtiene un identificador único del término, con el cual se procede a almacenar cada Url (tabla path\_url). De tal forma que todos los elementos resultantes del proceso quedan debidamente almacenados en sus respectivas tablas en la base de datos.

- btnCargar\_actionPerformed  
Procedimiento que se activa al momento en que el usuario presiona el botón de Cargar Búsqueda. Este procedimiento no hace mas que crear una nueva instancia de la froma frm\_CargaBusqueda, enviando como parámetro el arreglo Resultados, en el cual se almacena en la memoria del programa lo que se encuentra en la base de datos.
- jMenuItem2\_mouseClicked2  
Crea una nueva instancia de la ventana Mantenimiento.
- btnArf\_actionPerformed  
Procedimiento que crea la instancia de la clase Frm\_SelAttr, la cual es la encargada de seleccionar los atributos a incluir en la generación del archivo de tipo arff.
- jButton1\_actionPerformed  
Procedimiento principal de la clase, ya que éste es el encargado de, primero, separar los términos que ingresa el usuario en el cuadro de texto, para ello se auxilia de la clase Términos, la cual será explicada mas adelante. Luego, por cada término ingresado,

realiza una búsqueda en la ontología, almacenando cada path encontrado en un vector llamado paths. Teniendo almacenadas, estos datos se procede a realizar la búsqueda por Internet, para ello se crea una nueva instancia de la clase WebSearch. Por medio del método doSearch, se envía como parámetro el path de la búsqueda en la ontología, para que éste sea analizado por el api de Google, de tal forma que éste retorne los diferentes Url resultantes. Los cuales se almacenan en un arreglo de datos llamado Url. Como paso final, se arma un arreglo de datos con la siguiente estructura.



Luego de haber armado dicha estructura en memoria, se procede a crear una instancia de la clase Frm\_Resultados, a la cual se envía como parámetro dicha estructura para que éste la interprete.

#### A.2.1.2 Mantenimiento

Cuadro de diálogo que permite crear y eliminar atributos.

- Mantenimiento()  
Constructor de clase, este método manda a llamar al procedimiento carga\_datos.
- Carga\_datos()  
Procedimiento que realiza una consulta a la base de datos, con el fin de obtener todas las descripciones de los atributos que

actualmente están creados, para luego ingresarlos en la lista desplegable.

- Existe\_atributo()  
Función de tipo booleana, que retorna verdadero si el atributo que se envía como parámetro ya existe en la base de datos, de lo contrario retorna falso.
- btnGuardar\_actionPerformed()  
Evento que se activa al momento en el que el usuario hace clic en el botón de guardar. Este procedimiento, valida que el atributo no exista en la base de datos. Si este no existe, se procede a insertarlo en la tabla Atributo. Como último paso, se limpia la lista desplegable para luego volverla a cargar, con el fin de poder actualizar los datos.

#### A.2.1.3 MessageBox

Venta de alerta para el usuario

- MessageBox()  
La clase MessageBox, es una clase que cuenta con 6 tipos diferentes de constructores, cada uno brindando una funcionalidad diferente, es decir brindar un nivel de configuración y personalización del cuadro de mensaje diferente. Lo básico que debe de ser enviado como parámetro, es, el título del mensaje, el cuerpo del mensaje y un elemento de tipo Frame. Además de eso pues en otros constructores se puede determinar el número de botones que se mostrarán o bien establecer la posición en la que



se desea que aparezca la descripción, es decir el mensaje de error.

#### A.2.1.4 Main\_AboutBox

Ventana que despliega la información en general de la aplicación.

- Main\_AboutBox()

Ventana que se muestra cuando el usuario hace clic en el menú “Acerca De”.

#### A.2.1.5 Ingreso\_Atributos

Forma, que cumple la función de darle al usuario una interfaz, la cual le permita manejar, es decir, insertar y eliminar, valores asignados a un atributo para una página de internet para cada búsqueda.

- Ingreso\_Atributos()

Constructor de la clase, que manda a llamar al método Carga\_datos, luego de haber inicializado todos sus componentes.

- Carga\_Datos()

Procedimiento que carga las descripciones de los atributos a la lista desplegable cmbAtributo. Este método únicamente carga aquellos atributos que fueron definidos por el usuario.

- btnGuardar\_actionPerformed()

Método que se activa al momento que el usuario hace clic sobre el botón de Guardar. Lo primero que realiza es validar que el usuario haya ingresado algún dato den el campo txtValor. Una vez hecho esto se valida si la cadena ingresada contiene caracteres

especiales, es decir contiene, diagonales, signos de interrogación etc. Luego dependiendo del tipo de dato del atributo, es decir, si es un atributo numérico o bien un atributo de cadena de caracteres. Una vez validado que el dato que fue ingresado esta correcto se procede a validar que la búsqueda exista en el sistema, de lo contrario ésta se graba automáticamente en él. Una vez almacenada la búsqueda, se procede a almacenar el valor del atributo en la tabla Búsqueda\_Atributo\_Valor.

- btnEliminar\_actionPerformed()  
Este es el opuesto del método descrito anteriormente, aquí se elimina el valor de un atributo previamente almacenado. El procedimiento primero busca y obtiene la llave con la que se almacenó dicho atributo, para luego eliminar el registro de la base de datos.
- btnExplorer\_actionPerformed()  
Procedimiento que ejecuta el Explorador de Windows enviándole como parámetro la dirección Url actual.
- cmbAtributo\_actionPerformed  
Evento que se activa al momento en el que el usuario selecciona un ítem de la lista desplegable. Este procedimiento realiza una consulta en la base de datos en donde obtiene el valor, si lo hubiese, del atributo, así como también el tipo de dato del atributo, con el fin de poderlo mostrárselo al usuario.

- `setQuery()`  
Procedimiento que almacena en la variable local la consulta realizada por el usuario.

#### A.2.1.6 `Frm_SelAttr`

Ventana que muestra las opciones disponibles para que el usuario seleccione los atributos que formarán el archivo Arff.

- `btnAdd_actionPerformed()`  
Procedimiento que se activa cuando el usuario ha seleccionado un atributo del panel de la derecha de la pantalla, hacia el panel de la izquierda.
- `btnAddAll_actionPerformed()`  
Procedimiento que selecciona todos los atributos del panel de la derecha, para trasladarlos al panel de la izquierda, es decir incluye todos los atributos para que sean tomados en la generación del archivo arff.
- `btnRemove_actionPerformed()`  
Realiza la operación inversa al método Add. Retorna al panel de la derecha el atributo seleccionado en el panel de la izquierda.
- `jButton1_actionPerformed()`  
Procedimiento que retorna todos los atributos seleccionados al panel en el cual no son tomados en cuenta.
- `btnGenerar_actionPerformed()`

Procedimiento que genera el archivo Arff, dependiendo de los atributos previamente seleccionados. Una vez seleccionada la ubicación y el nombre del archivo, se procede a realizar una consulta a la base de datos, la cual retorna las diferentes Url's que cumplen con las condiciones previamente establecidas. El procedimiento, una vez identificados los registros a generar, realiza una llamada a preparaArff.

- PreparaArff()  
Este método es el encargado de finalmente generar el archivo, éste obtiene de la base de datos los valores de cada atributo para luego imprimirlos en el archivo.

## A.2.2 Clases

### A.2.2.1 Arbol

Clase que como su nombre lo dice, contiene los métodos necesarios para poder armar un árbol en la memoria, el cual tiene la capacidad de almacenar n atributos asociados al nodo del árbol.

- Arbol()  
Constructor de la clase, el cual puede recibir como parámetro únicamente un nodo o bien, el nodo junto con un arreglo de datos, es decir la información asociada a dicho nodo.
- goNode()  
Método que mueve el puntero del nodo actual, hacia el nodo indicado.

- `addNode()`  
Agrega un nodo al árbol, a partir de la posición actual.
- `goRoot()`  
Coloca el puntero sobre la raíz del árbol.
- `countCurrentChilds()`  
Retorna cuantos hijos tiene el nodo actual.
- `getInfo()`  
Retorna la información asociada al nodo actual, la cual se encuentra dentro de un objeto tipo `ArrayList`.
- `removeNode()`  
Elimina el nodo que cumple con la descripción que se envía como parámetro.
- `getCurrentKey()`  
Retorna el la llave del nodo actual
- `getChilds()`  
Retorna una lista con la información de todos los nodos hijos, que se encuentran a partir de la posición actual.
- `NewNodeList()`  
Añade todos los nodos hijos, del nodo actual, a la lista de hijos.

#### A.2.2.2 Buscar

Procedimiento que comienza a buscar en la ontología. La búsqueda comienza a partir del concepto "Producto", es decir, que va desde abajo hacia arriba según el la clasificación de las actividades económicas, el procedimiento se vale del método busca\_nodos, el cual se describe a continuación.

- BuscaNodos()  
Método puente que lo único que realiza es obtener los nodos hijos, de la categoría que se va a analizar para luego realizar la llamada al procedimiento barre\_arbol.
- Barre\_arbol()  
La función básicamente lo que realiza es una búsqueda en cada instancia de la ontología.

#### A.2.2.3 Búsqueda

Clase que contiene que realiza una búsqueda en memoria de la ontología.

- Add()  
Método que agrega una instancia al concepto al cual este pertenece. Lo que realiza básicamente es ir buscando en memoria, los conceptos de la ontología, e ir comparando las llaves de las instancias con el fin de poder encontrar el path al cual el término clave pertenece, es decir, que se realiza una búsqueda en cada elemento y/o instancia de la ontología encontrando las coincidencias respecto al parámetro de búsqueda, con el fin de poder insertar una instancia mas a la ontología.

- toString()  
Función sobrecargada, que retorna el path encontrado como resultado de la adición de un elemento.

#### A.2.2.4 Conexión

Clase que brinda la interfaz con la base de datos.

- Conexión()  
Constructor de la clase, el cual realiza la comunicación inicial con el servidor de base de datos.
- Cerrar()  
Cierra la comunicación con la base de datos.
- Exec()  
Ejecuta una instrucción SQL, alguna operación que altere los datos.
- Consulta()  
Realiza una consulta a la base de datos, retornando así un tipo de dato ResultSet, el cual contiene los registros solicitados.
- getCon()  
Retorna el objeto que contiene la conexión activa.
- existeBusqueda()  
Realiza una consulta a la base de datos, realizando una consulta en la tabla búsqueda, verificando la existencia de algún registro.

- `getBusqueda()`  
Retorna el identificador único con el cual se almacenó un registro en la tabla.
- `getAttrSk()`  
Retorna el identificador único con el cual se almacenó el atributo en la base de datos.
- `getAttrTipo()`  
Retorna el tipo de dato de cierto atributo.
- `getTerminoSk()`  
Retorna el identificador único del término.
- `getPathSk()`  
Retorna el identificador único del path almacenado.

#### A.2.2.5 Info

Clase que realiza el manejo de la información adicional (atributos) de las instancias que se cargan a la memoria al momento de leer la ontología a partir del archivo. Esta clase es utilizada en los nodos del árbol.

- `addAtrib()`  
Añade un atributo a la instancia.
- `getAtrib()`  
Retorna el atributo dependiendo el nombre que se haya enviado como parámetro.



- `setAtrib()`  
Coloca un nuevo valor al atributo
- `removeAtrib()`  
Remueve un atributo de la lista.
- `countAtrib()`  
Retorna la cantidad de atributos que actualmente se encuentran asociados a la instancia.

#### A.2.2.6 MyFileFilter

Clase que contiene únicamente el filtro que se aplica a los cuadros de diálogo para, abrir o guardar un archivo.

#### A.2.2.7 Nodo\_Búsqueda

Clase en la cual se van almacenando la información de cada nodo que se va obteniendo a partir de la ontología. Esta clase forma parte de una estructura, en la cual se encuentra mapeada la ontología.

- `toString()`  
Retorna el path del miembro actual.

#### A.2.2.8 Nodo

Clase, que como su nombre lo indica, es el nodo del árbol que es crea en memoria.

- `addChild()`  
Agrega un hijo al nodo en la posición actual.

- getInfo()  
Retorna el objeto de tipo Info asociado al nodo actual.
- getKey()  
Retorna la llave del nodo.
- getChild()  
Retorna el hijo del nodo.
- removeChild()  
Remueve un nodo hijo del árbol.
- hasChild()  
Retorna verdadero si el nodo actual contiene hijos, de lo contrario retorna falso.
- countChilds()  
Retorna el número de nodos hijo, que contiene el nodo actual.

#### A.2.2.9 Ontología

Clase que, primero parsea el archivo RDF, y luego va armando el árbol de la ontología en la memoria.

- Load()  
Carga el archivo RDF y realiza su procesamiento llamando al método CreaLista(), luego con la ontología totalmente cargada a memoria realiza la llamada al método CrearArbol().
- CreaLista()

Procedimiento que contiene la lógica para la lectura del formato RDF. El procedimiento se auxilia de la librería Tidy, para el parseo de documento Xml. El procedimiento utiliza básicamente las siguientes etiquetas:

rdfs:label – La cual contiene la del elemento.

rdf:resource – La cual indica que el elemento contiene atributos asociados.

rdf:Description: La cual brinda los elementos necesarios que complementan la descripción del elemento.

- setNS()  
Coloca el valor de la llave de la instancia al objeto nodo que la contiene.
  
- CrearArbol()  
Método que finalmente consolida el árbol en la memoria.
  
- CrearNodo()  
Crea un nuevo nodo y lo prepara para su inserción en el árbol. Este procedimiento es recursivo.
  
- SiguienteNodo()  
Procedimiento recursivo utilizado para la creación del árbol.
  
- getTerminos()  
Función que retorna una lista con todos los términos hijos.

- `getJTreeNode()`  
Retorna el objeto de tipo `DefaultMutableTreeNode`, con el fin de poder dibujar en el panel central de la pantalla principal el árbol.
- `busca_onto()`  
Procedimiento que inicializa la búsqueda en la ontología,

#### A.2.2.10 `OntoPyme_Main`

Clase principal, la que se manda a llamar desde la línea de comando.

#### A.2.2.11 `Path_Onto`

Clase que contiene un arreglo de `Url` con sus respectivos resultados de búsqueda para cada path que se haya encontrado en la ontología.

- `getPath()`  
Retorna el path perteneciente al objeto, este path es el índice, según la ontología.
- `addURL()`  
Agrega una dirección `Url` al índice.
- `setUrl()`  
Colca el arreglo de datos, el cual contiene las direcciones encontradas.

- getUrl()
 

Retorna el arreglo de datos de las direcciones.
- getUrlsize()
 

Retorna la cantidad de Url almacenados actualmente.
- getUrlbyIndex()
 

Retorna una dirección Url, en específico, a partir de su índice.

#### A.2.2.12 RBusqueda

Estructura en la que se almacena el resultado de la búsqueda que realiza el usuario sobre la ontología.

- addPathOnto()
 

Agrega un nodo de tipo Path\_Onto a la estructura, el cual corresponde al resultado de la búsqueda de un término ingresado por el usuario.
- getTermino()
 

Retorna el término asociado al elemento actual
- addPath()
 

Agrega un path al término actual, creando así un nuevo nodo path\_onto.
- get\_elem()
 

Retorna el elemento de tipo Rbúsqueda a partir de un término en específico.

- `get_Url()`  
Retorna la lista de Url asociadas al path o índice dentro de la ontología.
  - `getOntoelem()`  
Retorna el objeto `path_onto` asociado al término actual.
  - `getPathElembyUrl()`  
Retorna el objeto `path_onto` que contenga cierto Url.
  - `devPathOnto()`  
Retorna la lista de objetos `path_onto` asociadas al término actual.
- A.2.2.13 Términos
- Clase que maneja los términos que ingresa el usuario. Contiene los métodos necesarios para la correcta interpretación de éstos.
- `Parsear()`  
Obtiene término por término de la búsqueda ingresada, esto debido a que el usuario puede ingresar caracteres como “ , +, - etc.
  - `Size()`  
Retorna la cantidad de términos encontrados.
  - `getTermino()`  
Retorna un término según su posición en la lista
  - `getTerminos()`  
Retorna todos los términos en forma de lista

#### A.2.2.14 Utils

Clase que contiene funciones que son de utilidad para cualquier elemento dentro de la aplicación, las funciones que contiene son puramente de validación de tipos de datos.

- `es_numero()`  
Función que valida que absolutamente todos los caracteres contenidos en la cadena sean números.
  
- `tiene_especiales()`  
Función que valida si la cadena contiene algún carácter considerado como especial, es decir comas, punto, diagonales etc, es decir cualquier carácter que no sea alfanumérico.

#### A.2.2.15 WebSearch

Clase que utiliza el API de Google, el cual provee una interfaz de conexión al motor de búsqueda de Google.

- `WebSearch()`  
Constructor de la clase el cual crea una nueva instancia del API de Google.
  
- `doSearch()`  
Función que realiza la búsqueda en internet, utilizando cada índice de búsqueda obtenido desde la ontología. La función retorna una lista de datos que esta formado por el título del sitio seguido de su Url.
  
- `dev_cant_sitios()`

Retorna la cantidad de sitios encontrados