

Universidad Autónoma Metropolitana Unidad Azcapotzalco

División de Ciencias Básicas e Ingeniería

Licenciatura en Ingeniería en Computación

Modalidad: Proyecto terminal

Sistema de comparación de ontologías utilizando un enfoque de medición basado en vectores

Ildefonso César Aparicio Orozco

Matrícula: 209203746

Asesora:

Maricela Claudia Bravo Contreras

Profesor Asociado

Departamento de Sistemas

Julio 2014

Yo, Maricela Claudia Bravo Contreras, declaro que aprobé el contenido del presente Reporte de Proyecto de Integración y doy mi autorización para su publicación en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.



Maricela Claudia Bravo Contreras

Firma.

Yo, Ildefonso César Aparicio Orozco, doy mi autorización a la Coordinación de Servicios de Información de la Universidad Autónoma Metropolitana, Unidad Azcapotzalco, para publicar el presente documento en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.



Ildefonso César Aparicio Orozco

Firma.

RESUMEN

En la actualidad la web es el medio de comunicación más grande en el cual se puede tener una gran cantidad de información a la mano, hoy en día la web es un universo de aplicaciones y páginas web interconectadas lleno de contenido digital (vídeos, fotos, juegos, etc.).

Uno de los problemas con que el usuario se enfrenta al interactuar con la red es la búsqueda de información, puesto que al momento en el que el usuario hace una búsqueda en el buscador (google, bing, yahoo, etc.), este le arroja miles de resultados candidatos a la información que él desea.

Con el paso del tiempo se han desarrollado tecnologías que hacen que este problema se reduzca al más mínimo, pero es imposible que un buscador le arroje la información 100% exacta al usuario.

Con el fin de darle solución a esta situación, se han diseñado diferentes métodos para encontrar la información semánticamente, de tal manera que pueda ser interpretada o clasificada por agentes inteligentes, las ontologías son ejemplo de ello.

Las ontologías permiten a los usuarios a organizar la información en taxonomías de conceptos, cada uno con sus propiedades y describir las relaciones entre conceptos.

La realización de este proyecto terminal contempla el diseño y la implementación de un sistema web que realice la comparación de modelos ontológicos utilizando un método de medición de la similitud basado en vectores. En este proyecto se propone realizar la comparación de ontologías para distinguir las diferencias y las semejanzas de las ontologías.

La comparación de ontologías es un proceso donde se identifican las similitudes que hay entre una y varias ontologías, en el cual se aplica un algoritmo de que permite la comparación de las características que conforman las ontologías (clases, propiedades de objetos, propiedades de datos e individuos).

El comparador al ser agregado a cualquier motor de búsqueda semántico arrojará mejores resultados proporcionándonos información relevante y más confiable, ya que hoy en día no existen motores de búsqueda (buscadores) que utilicen los comparadores de ontologías, por lo que los resultados que arrojan los buscadores no son muy exactas.

Es por ello que el comparador se convierte en un elemento importante para un buscador y gracias a ello el usuario puede tener información más fiel a lo que busca y necesita.

TABLA DE CONTENIDO

Resumen.....	2
Tabla de contenido.....	3
Índice de figuras.....	4
1 Introducción.....	5
2 Justificación.....	6
3 Antecedentes.....	6
4 Objetivos.....	7
4.1 Objetivo general.....	7
4.2 Objetivos específicos.....	7
5 Marco Teórico.....	7
5.1 La web semántica.....	7
5.2 Ontologías.....	8
5.3 OWL.....	10
5.4 Similitud entre ontologías.....	10
6 Desarrollo.....	11
6.1 Tecnologías usadas.....	11
6.2 Implementación.....	12
6.2.1 Módulo de extracción y representación de vectores.....	12
6.2.2 Módulo de comparación.....	13
6.2.3 Módulo de visualización de resultados.....	16
7 Resultados.....	17
8 Análisis de resultados.....	23
9 Conclusiones.....	24
10 Bibliografía.....	25

ÍNDICE DE FIGURAS

Figura 1. Elementos de las ontologías.....	9
Figura 2. Características de las ontologías.	9
Figura 3. Módulos del Sistema	12
Figura 4. Características de una ontología.....	12
Figura 5. Fórmula para obtener la similitud entre dos ontologías.....	13
Figura 6. Esquema de las ontologías que se quieren comparar.....	14
Figura 7. Fórmula para obtener la similitud entre la ontología A y B.....	14
Figura 8. Interfaz principal.....	17
Figura 9. Pantalla de ayuda.....	18
Figura 10. Información del sistema web.....	18
Figura 11. Botón para iniciar el sistema.	19
Figura 12. Botón buscar.	19
Figura 13. Selección de archivos OWL (ontologías).....	20
Figura 14. Vista de archivos abiertos en el sistema.....	20
Figura 15. Archivos subidos al sistema.	21
Figura 16. Mensaje de proceso de comparación terminado.	21
Figura 17. Resultados de la primera comparación.	22
Figura 18. Otra comparación de una iteración n.....	22
Figura 19. Botón de Terminar.....	23

1 INTRODUCCIÓN

Actualmente la web es un medio de comunicación masivo, donde se puede tener acceso a la información en cualquier parte del mundo. Pero al momento en que un usuario hace una búsqueda de un tema en específico se tropieza con un cumulo de información irrelevante. Esto es debido a que en la web se aloja gran cantidad de información disponible, que es difícil de organizar y clasificar de manera que los datos sean relevantes para cualquier individuo que se encuentre recopilando información.

Con el fin de minimizar este problema, se han desarrollado diferentes métodos para encontrar la información semánticamente, de tal manera que pueda ser interpretada o clasificada¹ por agentes inteligentes.

Nuevas tecnologías Web proponen nuevas técnicas y paradigmas para la representación de conocimiento que faciliten localizar, compartir e integrar recursos, este conocimiento semántico explícito que describe y estructura la información y los servicios disponibles.

Las ontologías son un conjunto de vocabularios de conceptos y relaciones suficientemente enriquecidos que permiten representar el conocimiento y la información relativa a cualquier dominio de manera formal, además son legibles y re-utilizables por programas informáticos [1].

Un enfoque estándar para el resultante problema radica en el uso de ontologías para la descripción de datos. Las ontologías permiten a los usuarios a organizar la información en taxonomías de conceptos, cada uno con sus propiedades y describir las relaciones entre conceptos [2].

El problema, es la operación de tomar dos ontologías distintas, la búsqueda de un conjunto de entidades con relaciones similares que existen en ambas ontologías y devolver las entidades similares.

En este proyecto terminal se propone diseñar un sistema web mediante un algoritmo de similitud que permita realizar la comparación de ontologías para distinguir las diferencias y las semejanzas de las ontologías.

Con la obtención de las similitudes de las ontologías se puede favorecer y obtener información más precisa y relevante para el usuario al momento de realizar una búsqueda.

¹Determinar la clase o conjunto a que corresponde dicha información.

2 JUSTIFICACIÓN

En la actualidad no existen motores de búsqueda (buscadores) que utilicen los comparadores de ontologías, por lo que los resultados que arrojan los buscadores no son muy relevantes. Es por ello que se plantea implementar un comparador para obtener datos mucho más precisos e información más relevante.

La comparación de ontologías es la tarea de identificar las similitudes que hay entre una y varias ontologías, mediante un algoritmo de similitud que permite la comparación de las características que conforman las ontologías². Gracias a la comparación se tendría una información más fiel a lo que busca el usuario.

El comparador al ser agregado a cualquier motor de búsqueda arrojará mejores resultados proporcionándonos información relevante y más confiable.

3 ANTECEDENTES

Proyectos Terminales

1. *Sistema clasificador de ontologías* [3]. Este proyecto terminal hace uso de Ontologías como unidad fundamental del desarrollo de un sistema que permite leer y clasificar un conjunto de ontologías utilizando métodos de máquinas de soporte vectorial, pero en el momento del análisis, no se concentra en las similitudes de las ontologías, solo en el filtro y la clasificación de las mismas.
2. *Extracción automatizada y representación de servicios web mediante ontologías* [4]. Este proyecto trabaja con ontologías para el desarrollo de un sistema que permita extraer los elementos más relevantes de múltiples descripciones de servicios Web y que los represente mediante un modelo ontológico, sin embargo no hace ninguna comparación de ontologías con alguna metodología.
3. *Sistema de recuperación de información de textos de investigación de la Web* [5]. Para recabar información, se hizo uso de una araña web³, y un analizador sintáctico (*parser*). La principal diferencia es que el proyecto citado en este punto en particular, carece de un módulo que le permita hacer la comparación entre ontologías y además mostrar al usuario los resultados.

Artículos

4. *Ontology matching using vector space* [6]. Este artículo hace mención sobre una forma de hacer una comparación de ontologías mediante vectores. En él se plantea una forma más sencilla y factible para llegar con los resultados y poder hacer un análisis sobre ellos, pero

²Clases, propiedades de objetos, propiedades de datos e individuos.

³Un programa que se encarga de visitar recursos, generalmente páginas web, con el fin de extraer información.

no se demuestra una implementación de la misma que pueda hacer la comparación por similitud.

5. *An Ontology-Based Information Retrieval Model* [7]. En este artículo se hace mención de un análisis de las ontologías y como se puede hacer una extracción y clasificación de las mismas mediante el análisis de sus características. Estas características son representadas en vectores y se hace un análisis en cuanto al peso que tiene en el documento, pero no se concentran en utilizar dicha información para la comparación de las ontologías, solo únicamente para su clasificación.

4 OBJETIVOS

4.1 OBJETIVO GENERAL

Diseñar e implementar un sistema web que realice la comparación de modelos ontológicos utilizando un método de medición de la similitud/diferencia basado en vectores.

4.2 OBJETIVOS ESPECÍFICOS

- Diseñar e implementar un módulo que permita extraer los elementos de la ontología (clases, propiedades de objetos, propiedades de datos e individuos) y su representación en vectores.
- Diseñar e implementar el algoritmo de comparación para los vectores de características de las ontologías.
- Diseñar e implementar un sistema web a través del cual se realice la comparación y se visualicen los resultados.

5 MARCO TEÓRICO

5.1 LA WEB SEMÁNTICA

En los últimos años, muchos investigadores están diseñando modelos para transformar la red desde un espacio de información a un espacio de conocimientos. Recientemente, Tim Berners-Lee, uno de los inventores de la Web, defiende el desarrollo de la Web con conocimientos [8], y organizaciones como SematicWeb⁴ se encargan de estandarizar lenguajes y herramientas para hacer efectiva la web semántica.

La idea de la web semántica es que los datos puedan ser utilizados y “comprendidos” por los ordenadores sin necesidad de supervisión humana, de forma que los agentes web puedan ser diseñados para tratar la información situada en las páginas web de manera semiautomática.

⁴ www.semanticweb.org

Se trata de convertir la información en conocimiento, referenciando datos dentro de las páginas web a metadatos con un esquema común consensuado sobre algún dominio. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que además podrán tener información adicional de cómo hacer deducciones con ellos, es decir, axiomas que podrán aplicarse en los diferentes dominios que trate el conocimiento almacenado [9].

Con ello, se mejorará la búsqueda de información y se potenciará el desarrollo de aplicaciones de comercio electrónico, ya que las anotaciones de información seguirán un esquema común, y los buscadores web compartirán con las anotaciones web los mismos esquemas. Empresas que traten con clientes y proveedores, podrán intercambiar sus datos de productos siguiendo estos esquemas comunes consensuados.

Los agentes web no sólo encontrarán la información de forma precisa, si no que podrán realizar inferencias automáticamente buscando información relacionada con la que se encuentra situada en las páginas, y con los requerimientos de la consulta indicada por el usuario.

5.2 ONTOLOGÍAS

El término ontología cuyo origen se encuentra en la Filosofía fue adoptado por la comunidad de investigadores de Inteligencia Artificial para describir formalmente dominios de conocimiento. Para la IA, lo que “existe” es aquello que puede ser representado.

Muchas definiciones del término ontología han sido propuestas en las últimas décadas. La definición más completa y utilizada es la de Gruber (1993) y extendida por Studer (1998):

En 1993, Gruber amplía esta definición incluyendo la especificación de restricciones de rango y dominio entre las relaciones, inclusión jerárquica, axiomas, etc. La definición que dio Gruber es la siguiente:

"Vocabularios de términos representacionales - clases, relaciones, funciones, constantes de objetos - con definiciones acordadas en forma de texto legible para los humanos y ejecutable por las máquinas, con restricciones declarativas sobre su uso bien formado".

En 1998 Studer la define de la siguiente manera:

"Una ontología es una especificación formal y explícita de una conceptualización compartida".

En el 2001 Lassila y McGuinness, dos de los autores promotores de la Web semántica, consideran que para que una especificación explícita de una conceptualización pueda ser considerada una ontología, ésta debe cumplir con tres propiedades:

- 1) Vocabulario finito y controlado.
- 2) Interpretación NO ambigua de clases y relaciones entre términos, y
- 3) Jerarquías estrictas de las relaciones de subclases entre clases.

La interpretación de estas definiciones es que las ontologías definen conceptos y relaciones de algún dominio, de forma compartida y consensuada; y que esta conceptualización debe ser representada de una manera formal, legible y utilizable por los ordenadores.

Para entender mejor del término ontología desde la perspectiva de la Computación, se distinguen los elementos constitutivos de las ontologías (ver Figura 1 y Figura 2) y las características que debe presentar una ontología.

- a. Una ontología se define como conjunto de conceptos o clases (vocabulario o conceptualización), relaciones taxonómicas (jerárquicas), relaciones entre conceptos, individuos o instancias, y axiomas.
- b. La ontología debe cumplir con las características de ser una especificación formal, explícita y compartida.

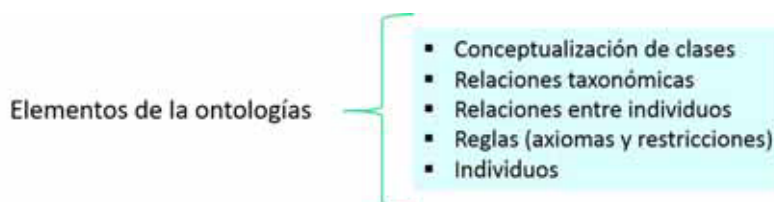


Figura 1. Elementos de las ontologías.

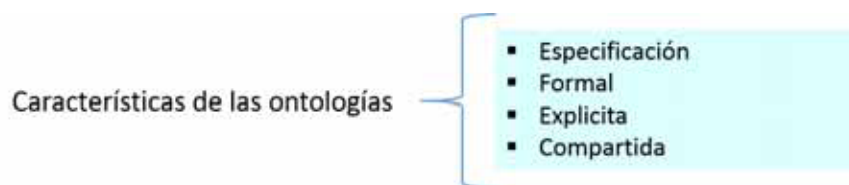


Figura 2. Características de las ontologías.

En resumen, una ontología es un sistema de representación del conocimiento que resulta de seleccionar un dominio o ámbito del conocimiento, y aplicar sobre él un método con el fin de obtener una representación formal de los conceptos que contiene y de las relaciones que existen entre dichos conceptos. Además, una ontología se construye en relación a un contexto de utilización. Esto quiere decir que una ontología especifica una conceptualización o una forma de ver el mundo, por lo que cada ontología incorpora un punto de vista. Además, una ontología contiene definiciones que nos proveen del vocabulario para referirse a un dominio.

Las posibles aplicaciones y usos de las ontologías son:

- Repositorios para la organización del conocimiento.
- Servir de herramienta para la adquisición de información.
- Servir de herramientas de referencia en la construcción de sistemas de bases de conocimiento que aporten consistencia, fiabilidad y falta de ambigüedad a la hora de recuperar información.
- Crear una red de relaciones que aporte especificación y fiabilidad.
- Permitir compartir conocimiento.

- Posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre organizaciones, comunidades científicas, etc.
- Permitir la reutilización del conocimiento existente en nuevos sistemas.
- Permitir la interoperabilidad entre sistemas distintos.
- Establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo entre diferentes contextos.
- Servir de base para la construcción de lenguajes de representación del conocimiento.

5.3 OWL

OWL: OWL Web Ontology Language⁵ o Lenguaje de Ontologías para la Web es un lenguaje de etiquetado semántico para publicar y compartir ontologías en la Web. Se trata de una recomendación del W3C, y puede usarse para representar ontologías de forma explícita, es decir, permite definir el significado de términos en vocabularios y las relaciones entre aquellos términos (ontologías). Se trata de un lenguaje diseñado para usarse cuando la información contenida en los documentos necesita ser procesada por programas o aplicaciones, en oposición a situaciones donde el contenido solamente necesita ser presentado a los seres humanos.

5.4 SIMILITUD ENTRE ONTOLOGÍAS

Un método de similitud es utilizado para la comparación de dos elementos del cual se requiere saber su semejanza, existe un resultado el cual indica el índice que diferencia o relevancia que existe entre un elemento y otro.

Una de tantas formas de obtener la similitud es aplicar esta comparación entre conjuntos como se presenta en la siguiente fórmula:

$$Sim C_{ab} = \frac{C_a \cap C_b}{C_a \cup C_b}$$

Donde C_a y C_b son dos conjuntos diferentes.

La similitud entre los dos conjuntos arroja un valor en un rango de 0 – 1, lo que indica que si existe un cociente con valor 1, significa que los conjuntos se parecen uno al otro, lo que indicaría un 100% de similitud. En caso contrario si se obtiene un cociente de 0, significa que no existe ningún tipo de similitud entre ellos.

Para un mejor entendimiento se presenta un ejemplo donde se observa detalladamente cómo interactúan los conjuntos.

⁵ <http://www.w3.org/TR/owl-features/>

Sea C_a y C_b dos conjuntos de números diferentes.

$$C_a = \{10, 50, 20, 45, 15, 60\}$$

$$C_b = \{15, 70, 80, 45, 60, 25\}$$

Aplicando la fórmula de la similitud, se tiene:

$$\begin{aligned} Sim O_{ab} &= \frac{C_a \cap C_b}{C_a \cup C_b} = \frac{\{10, 50, 20, 45, 15, 60\} \cap \{15, 70, 80, 45, 60, 25\}}{\{10, 50, 20, 45, 15, 60\} \cup \{15, 70, 80, 45, 60, 25\}} \\ &= \frac{\{45, 15, 60\}}{\{10, 15, 50, 20, 25, 45, 60, 70, 80\}} = \frac{3}{9} \cong 0.333 \end{aligned}$$

Lo que significa que se tiene un aproximado de 33% de similitud entre ambos conjuntos.

Es así como se obtiene una forma de comparar conjuntos y sacar su índice de similitud entre ellos. Este proyecto terminal contempla este método y su implementación en la comparación de ontologías.

6 DESARROLLO

6.1 TECNOLOGÍAS USADAS

Para la elaboración y funcionamiento de este proyecto terminal se usó el lenguaje de programación Java. Para tal fin, se empleó el paquete de desarrollo Java Platform (JDK) 7u21 de 64 bits. Como editor de código, el entorno de desarrollo integrado (IDE, por su sigla en inglés) llamado Netbeans en su versión Java EE 7.2.1 de 64 bits, ejecutado en la plataforma Microsoft Windows 8.1 de 64 bits.

El proceso de manejo de ontologías se ayudó mediante el uso de algunas bibliotecas externas. Principalmente, se hizo uso de OWL API versión 2.0 para la extracción de los elementos de las ontologías en formato OWL⁶. Se usó el marco de trabajo PrimeFaces Framework versión 5, para la visualización de los resultados. Así como, el Framework Bootstrap para los estilos visuales del sistema.

Para la edición de las ontologías se utilizó el editor de ontologías Protégé en su versión 3.5.

⁶Web Ontology Language

6.2 IMPLEMENTACIÓN

La realización de este proyecto terminal contempla el diseño y la implementación de un sistema web que realiza la comparación de ontologías utilizando un método de medición de la similitud basado en vectores. Dicho sistema estará conformado por los siguientes módulos (ver Figura 3):

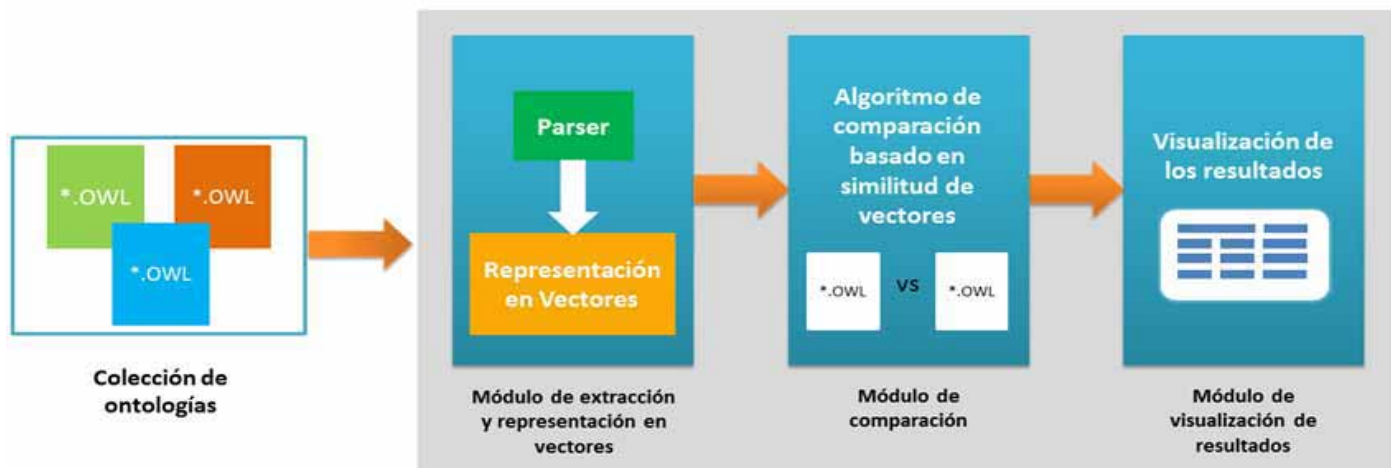


Figura 3. Módulos del Sistema

6.2.1 Módulo de extracción y representación de vectores

Este módulo se encargará de extraer los elementos de una colección de ontologías en formato OWL, mediante un analizador sintáctico (también conocido como *parser*), que posteriormente se almacenan de forma local. Esta información es representada de manera vectorial para que las características puedan ser procesadas (clases, propiedades de objetos, propiedades de datos e individuos).

El vector de características de una ontología $O = \langle C, P, D, I \rangle$ (ver Figura 4)

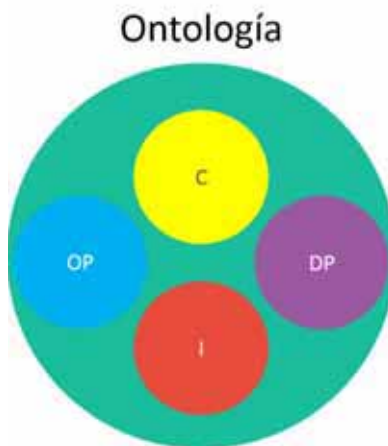


Figura 4. Características de una ontología.

Dónde:

- C: conjunto de clases.
- OP: conjunto de propiedades de objetos.
- DP: conjunto de propiedades de datos.
- I: conjunto de individuos.

Cada conjunto tiene sus respectivos elementos que la conforman, como por ejemplo las clases.

Sea $C = \{\text{nombreClase 1, nombreClase 2, nombreClase 3, \dots, nombreClase n}\}$

Lo que el *parser* hace es obtener todos los elementos de las clases, propiedades de objetos, propiedades de datos e individuos, y después almacenarlos en una lista para su futura utilización en el método de la comparación.

6.2.2 Módulo de comparación

Este módulo realizara la comparación de las ontologías mediante un método de similitud, donde se comparan las características que conforman cada una de las ontologías.

El método de similitud utilizado en el proyecto presente, se basa en obtener un cociente que representa la similitud de una ontología y otra.

Esta similitud se basa en extraer las propiedades de las ontologías (clases, propiedades de objetos, propiedades de datos e individuos) que se quieren comparar y mediante una división por medio de una intersección y unión para así obtener el cociente que representa la similitud que existe entre las ontologías comparadas.

Para la comparación de las ontologías se utiliza un método de similitud basado en la siguiente formula (ver en figura 5):

$$Sim O_{ab} = \frac{O_a \cap O_b}{O_a \cup O_b}$$

Figura 5. Fórmula para obtener la similitud entre dos ontologías.

Donde O_a y O_b son dos ontologías diferentes.

La fórmula se aplica a cada conjunto de elementos que conforman la ontología.

Para observar con mejor detenimiento como seria la implementación de este módulo, se propone el siguiente ejemplo:

Sea O_a y O_b son dos ontologías distintas (ver Figura 6), se requiere saber cuál es el índice de similitud entre ellas.

$$O_a = \{C_a, OP_a, DP_a, I_a\}$$

$$O_b = \{C_b, OP_b, DP_b, I_b\}$$

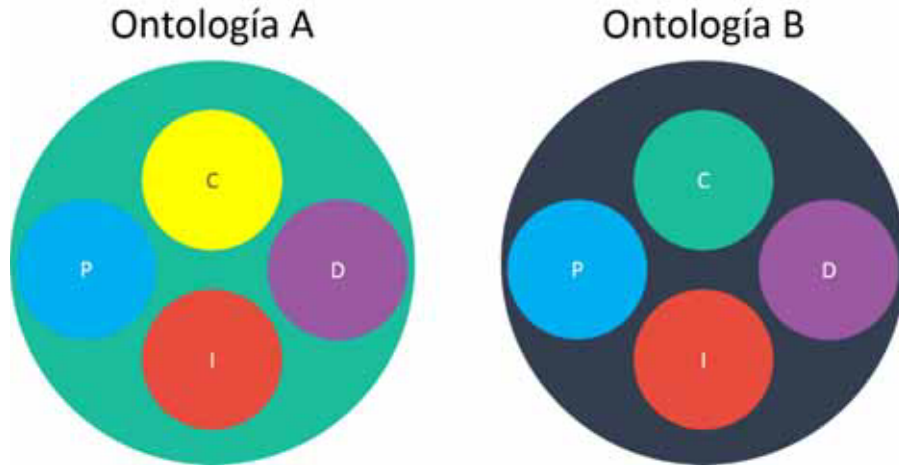


Figura 6. Esquema de las ontologías que se quieren comparar.

Dónde:

- C_a : conjunto de clases de la ontología A.
- OP_a : conjunto de propiedades de objetos de la ontología A.
- DP_a : conjunto de propiedades de datos de la ontología A.
- I_a : conjunto de individuos de la ontología A.

Y dónde:

- C_b : conjunto de clases de la ontología B.
- OP_b : conjunto de propiedades de objetos de la ontología B.
- DP_b : conjunto de propiedades de datos de la ontología B.
- I_b : conjunto de individuos de la ontología B.

Para este caso se proponen que las 2 ontologías sean a cerca de comida: comida mexicana (ontología A) y comida china (ontología B).

Para la obtención de la similitud, se tiene que extraer el conjunto de elementos que constituye a cada una de las ontologías, es decir, obtener el conjunto de clases, propiedades de objeto, propiedades de los datos e individuos de las dos ontologías, para así sacar un promedio y obtener la similitud entre ambas (ver Figura 7).

$$Sim O_{ab} = \frac{Sim C_{ab} + Sim OP_{ab} + Sim DP_{ab} + Sim I_{ab}}{4}$$

Figura 7. Fórmula para obtener la similitud entre la ontología A y B.

Por consiguiente tenemos primero que obtener la similitud entre las clases de las ontologías A y B, para después de la misma manera obtener la de los demás conjuntos de características.

Sea $Sim C_{ab}$:

$$\begin{aligned}
 Sim C_{ab} &= \frac{C_a \cap C_b}{C_a \cup C_b} \\
 &= \frac{\{\text{dulce, picante, agridulce, salado, amargo}\} \cap \{\text{insipido, dulce, agridulce, picante}\}}{\{\text{dulce, picante, agridulce, salado, amargo}\} \cup \{\text{insipido, dulce, agridulce, picante}\}} \\
 &= \frac{\{\text{dulce, picante, agridulce}\}}{\{\text{dulce, picante, agridulce, salado, amargo, insipido}\}} \\
 &= \frac{\{3\}}{\{6\}} = \frac{1}{2} = 0.5 \\
 \mathbf{Sim C_{ab} = 0.5}
 \end{aligned}$$

Sea $Sim I_{ab}$:

$$\begin{aligned}
 Sim I_{ab} &= \frac{I_a \cap I_b}{I_a \cup I_b} \\
 &= \frac{\{\text{mole, cochinita pibi, arroz, salsa, torta, tacos, pasta}\} \cap \{\text{arroz, rollos primavera, chop suey, pasta, din sum}\}}{\{\text{mole, cochinita pibi, arroz, salsa, torta, tacos, pasta}\} \cup \{\text{arroz, rollos primavera, chop suey, pasta, din sum}\}} \\
 &= \frac{\{\text{arroz, pasta}\}}{\{\text{mole, cochinita pibi, arroz, salsa, torta, tacos, pasta, rollos primavera, chop suey, din sum}\}} \\
 &= \frac{\{2\}}{\{10\}} = \frac{1}{5} = 0.2 \\
 \mathbf{Sim I_{ab} = 0.2}
 \end{aligned}$$

Y aplicando de igual manera para las propiedades de objeto y las propiedades de dato, suponemos que tenemos:

$$\begin{aligned}
 \mathbf{Sim OP_{ab} = 0.0} \\
 \mathbf{Sim DP_{ab} = 0.05}
 \end{aligned}$$

Ahora si entonces se procede al cálculo total de similitud:

$$\begin{aligned}
 Sim O_{ab} &= \frac{Sim C_{ab} + Sim OP_{ab} + Sim DP_{ab} + Sim I_{ab}}{4} \\
 &= \frac{0.5 + 0.0 + 0.05 + 0.2}{4} = \frac{0.75}{4} \cong 0.1875 \\
 \mathbf{Sim O_{ab} \cong 0.1875}
 \end{aligned}$$

Con esto ahora se conoce que el índice de similitud entre la ontología A y la ontología B es de 1.1875. Por consiguiente se sabe que se parecen un 18.75%.

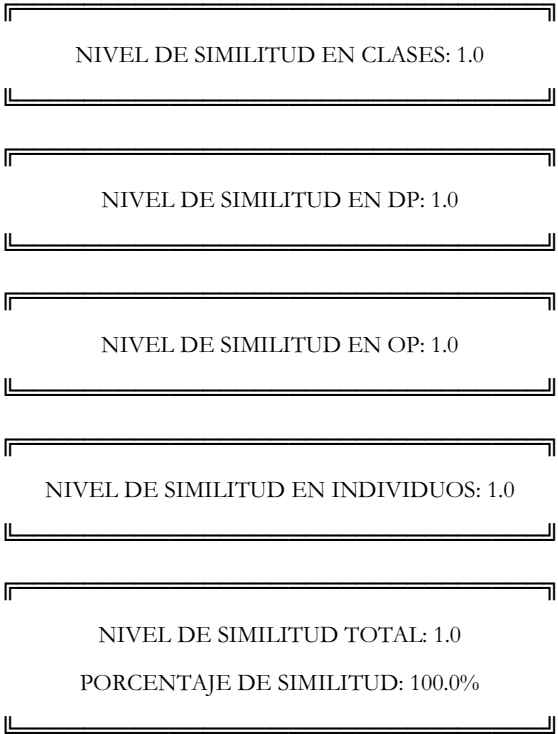
Es así como el módulo de comparación de ontologías mediante la aplicación de la fórmula de similitud, obtiene un índice de relevancia entre las ontologías.

El módulo ya implementado el sistema compara todas las ontologías que se cargaron en el módulo anterior, haciendo así una comparación de todas contra todas, arrojando así los resultados en pantalla.

6.2.3 Módulo de visualización de resultados

Este módulo permite al usuario visualizar los resultados finales de la comparación, donde se muestran los índices de similitud que hay entre cada una de las propiedades de las ontologías, así como el promedio y el porcentaje de cuanto hay de similitud entre las ontologías que se compararon.

Un ejemplo de la salida de datos sería el siguiente:



Véase de mejor manera en la sección de resultados.

7 RESULTADOS

El sistema implementado se muestra a continuación mediante una serie de pasos y visualización de los resultados obtenidos.

Cabe mencionar que este sistema acepta ontologías cuya extensión sea únicamente **.owl** y que sus características no estén referenciadas a otro sitio que no sea el local.

A continuación se muestra el sistema completo y su funcionalidad (ver figura 8).



Figura 8. Interfaz principal.

El sistema cuenta con 3 principales secciones:

1. Sección de ayuda
2. Sección de Acerca de
3. Sección de Resultados

En la sección de **ayuda** se presenta una guía resumida para los usuarios de cómo se emplea el sistema (ver figura 9).

AYUDA



Guía rápida para el uso del comparador de ontologías.

- INICIO
 1. Dar click en el botón **Iniciar** para comenzar la ejecución del sistema.
 2. Cargar Archivos: OWL
 1. Dar click en el botón **Buscar** para buscar los archivos a procesar.
 2. Seleccionar por bloques las ontologías(extension .owl) y de **aceptar** en su explorador de archivos, en el caso de no ser extensión .owl mandara el sistema un mensaje de error.
 3. Si desea **Cancelar** la subida de los archivos, dar click en el botón **Cancelar** para aliminar los archivos seleccionados.
 4. Dar click en el botón **Subir** para cargar las ontologías que se desean comparar.
- ANALIZAR
 1. Para comenzar con la comparación de las ontologías, dar click en el botón **Iniciar** para comenzar con el proceso.
 2. SE MOSTRARA UN MENSAJE DE NOTIFICACIÓN DONDE DIRA SI HA FINALIZADO LA COMPARACIÓN DE LAS ONTOLOGÍAS CON EXITO
 3. Para visualizar los resultados, baje con el scroll
- TERMINAR
 1. Si desea **Terminar** con la verificación del os resultados, dar click en el botón **Terminar**.
- CANCELAR
NOTA:
 1. El botón cancelar estara disponible en todo momento para **Cancelar** cualquier tipo de operación realizada en los puntos anteriores. Si desea cancelar, dar click en el botón **Cancelar**.

Figura 9. Pantalla de ayuda.

En la sección de **acerca de** se presenta una breve explicación del sistema, sus tecnologías usadas y el motivo del proyecto terminal (ver figura 10).

ACERCA DE



El sistema web de **Comparación de ontologías** es un proyecto terminal que se realizó como requisito para la culminación de la carrera de Ingeniería en computación.

Este sistema implementa un método de comparación por similitud que se aplica tanto en sus clases, propiedades de objetos, propiedades de datos, así como en sus individuos que integran la ontología, para como resultado final mostrar un porcentaje del índice de similitud que existe entre una ontología y otra.

La documentación y el desarrollo del método utilizado esta contenido en el reporte final de manera impresa o digital, que se puede consultar en la hemeroteca de la UAM - Azcapotzalco.

El sistema fue hecho con las tecnologías: Apache Tomcat v7 Oracle JDK OWL-API Netbeans IDE PrimeFaces Bootstrap Framework

Nombre del proyecto terminal: Sistema de comparación de ontologías utilizando un enfoque de medición basado en vectores

Año: Julio 2014

Hecho por: El alumno Ildefonso César Aparicio Orozco actualmente matriculado en la UAM - Azcapotzalco.

Asesorado por: La Dra. Marcela Claudia Bravo Contreras

Figura 10. Información del sistema web.

Después se tiene la sección de resultados que es donde se hará todo el proceso visto en la sección de desarrollo.

Para comenzar a comparar las ontologías, tenemos que seleccionar la sección de **iniciar** que se muestra en la parte superior del área de resultados (ver figura 11).

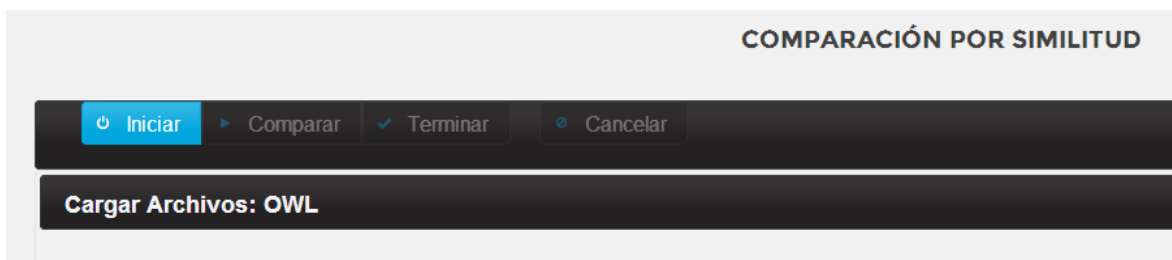


Figura 11. Botón para iniciar el sistema.

Una vez iniciado el sistema, se muestra una nueva sección y unos mensajes que nos sugieren cargar los archivos que necesitamos comparar, para esto se selecciona la opción nueva que dice **Buscar** (ver figura 12).



Figura 12. Botón buscar.

Enseguida aparece el explorador de archivos, por lo que tenemos que navegar al directorio donde se alojan los archivos OWL, se pueden seleccionar simultáneamente las ontologías que se deseen comparar (ver figura 12), y una vez dando click en abrir nos mostrara en el sistema los archivos ya seleccionados por el sistema(ver figura 13).

Nombre	Fecha de modifica...	Tipo	Tamaño
961373447231.owl	10/07/2013 04:07 a...	Archivo OWL	1,134 KB
10761384730117 - copia.owl	17/03/2014 01:54 ...	Archivo OWL	23 KB
10761384730117.owl	17/11/2013 05:15 ...	Archivo OWL	17 KB
13601373446452 - copia.owl	17/03/2014 01:32 ...	Archivo OWL	11 KB
13601373446452.owl	10/07/2013 03:54 a...	Archivo OWL	6 KB
13741373447016.owl	10/07/2013 04:03 a...	Archivo OWL	64 KB
17361373446582.owl	10/07/2013 03:56 a...	Archivo OWL	2 KB
catalog-v001	09/07/2014 04:19 ...	Archivo XML	1 KB

Figura 13. Selección de archivos OWL (ontologías).

Cargar Archivos: OWL

+ Buscar
↶ Subir
⊗ Cancelar

10761384730117 - copia.owl	22.7 KB	<input type="text"/>	<input type="button" value="x"/>
10761384730117.owl	17.0 KB	<input type="text"/>	<input type="button" value="x"/>
13601373446452 - copia.owl	10.5 KB	<input type="text"/>	<input type="button" value="x"/>
13601373446452.owl	5.8 KB	<input type="text"/>	<input type="button" value="x"/>
13741373447016.owl	63.9 KB	<input type="text"/>	<input type="button" value="x"/>
17361373446582.owl	1.8 KB	<input type="text"/>	<input type="button" value="x"/>

Figura 14. Vista de archivos abiertos en el sistema.

Una vez subidos los archivos, se tiene la opción de elegir si se desean eliminar o subir de nuevo más archivos faltantes. En caso contrario se sigue con el proceso para poder subir las ontologías y cargarlas al sistema para su posterior comparación, para esto se da click ahora en el botón subir y nos aparecerán los nombres de los archivos cargados (ver figura 15).

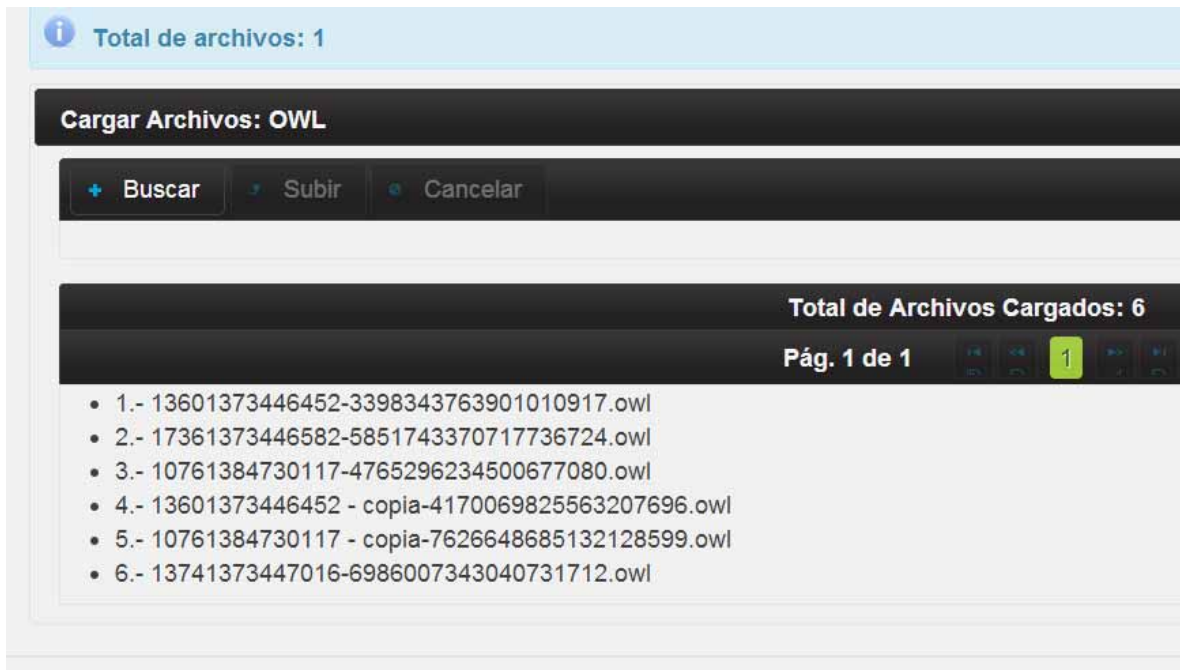


Figura 15. Archivos subidos al sistema.

Como nos percatamos, un anuncio nos dirá cuántas ontologías se han cargado en el sistema.

Como ya se tienen los archivos dentro del sistema web, ahora se seleccionara el botón Comparar, una vez hecho esto, se mostrara un mensaje de erro por sí ocurrió un percance o un mensaje de proceso finalizado en caso de haber terminado con éxito (ver figura 16 y 17).

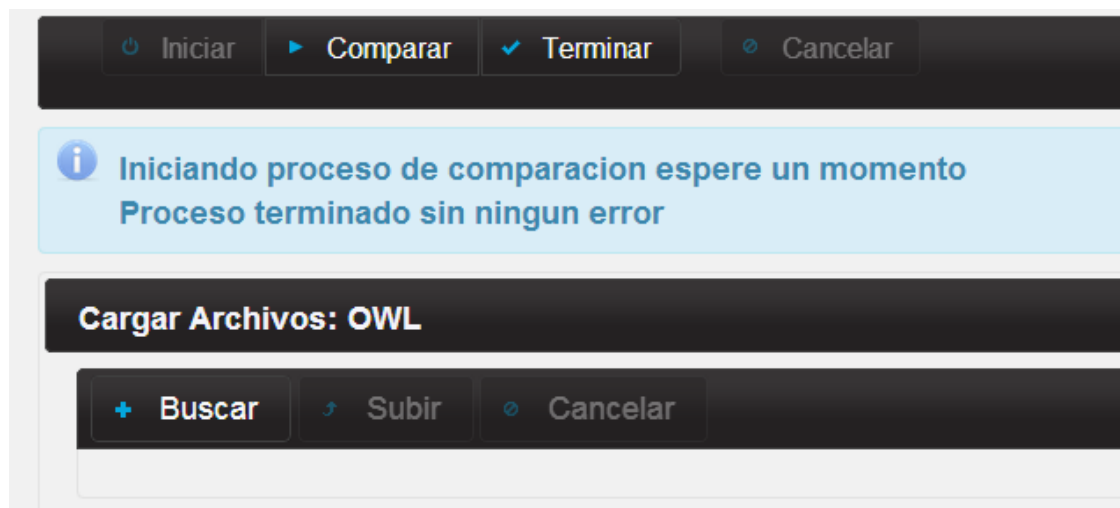


Figura 16. Mensaje de proceso de comparación terminado.

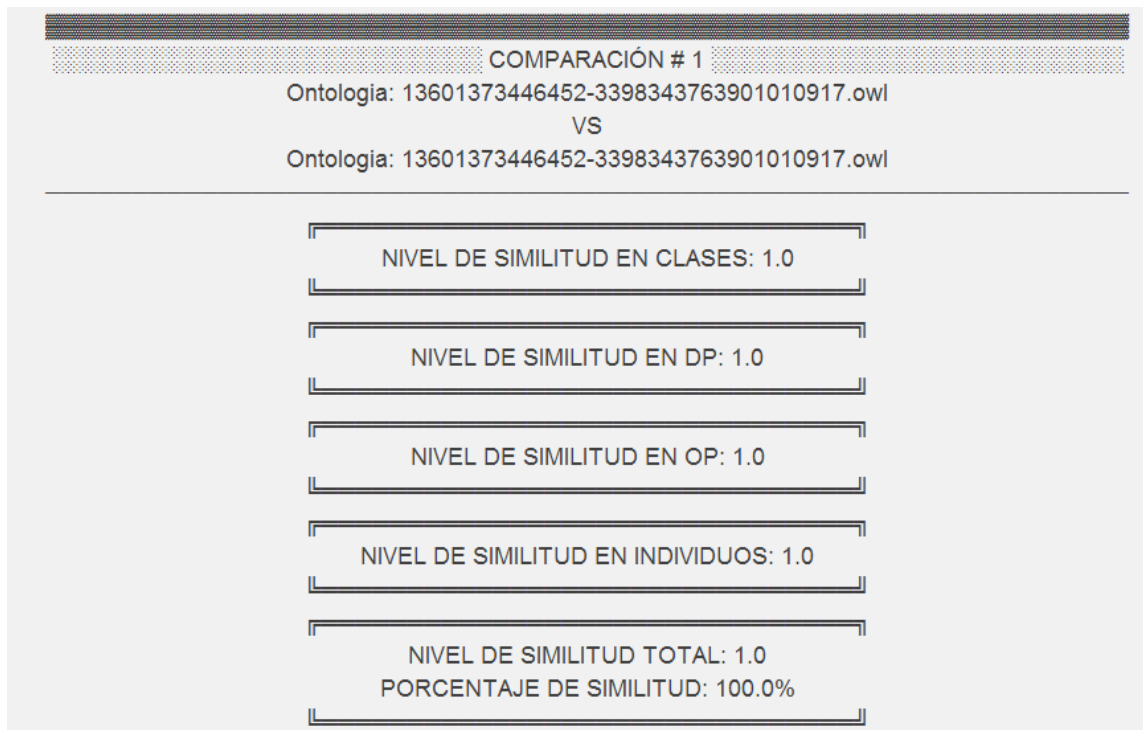


Figura 17. Resultados de la primera comparación.

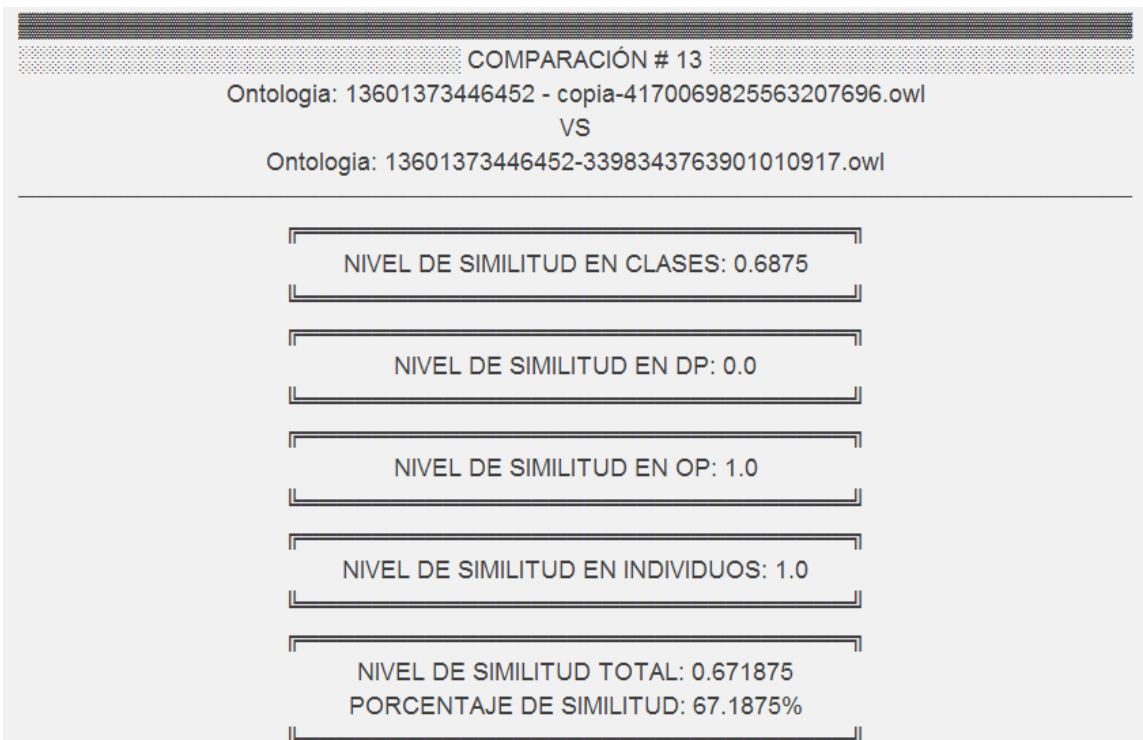


Figura 18. Otra comparación de una iteración n.

Como se pueden apreciar en las figuras 17 y 18, se muestran los resultados arrojados por el sistema web una vez finalizada a comparación.

Para volver a realizar una nueva comparación, se tiene que dar click en el botón de terminar como se muestra en la figura 19.

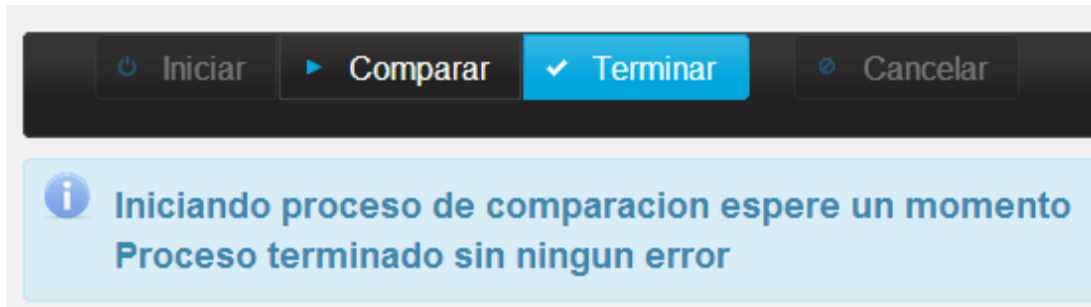


Figura 19. Botón de Terminar.

En resumen es lo que se hace en el sistema comparador de ontologías, el proceso que se lleva a cabo está contenido en la sección de desarrollo.

8 ANÁLISIS DE RESULTADOS

Si se analizan los resultados de las figuras 17 y 18, se puede percatar de que los resultados que salen en pantalla son todas las comparaciones de todas las ontologías que se cargaron en el directorio.

Aparece el nombre de las dos ontologías que se compararon, así como la similitud en las características de las ontologías y por último el promedio total que representa el índice de similitud entre una ontología y otra.

La comparación como ya se mencionó antes consiste en obtener la similitud de cada una de las características que las conforman, para después sacar un promedio el cual nos dice el índice de similitud que existe o no entre las ontologías que se compararon.

Cada ontología fue cargada en memoria una vez subida al sistema, en el momento en el que el usuario inicia la comparación lo primero que pasa es obtener tomar las ontologías y extraer cada uno de los elementos que conforman las características de las ontologías, para después ser almacenadas en listas temporales para su posterior utilización.

El comparador una vez obtenidos los conjuntos de características (Clases, propiedades de objeto, propiedades de datos e individuos) convierte esas listas (vectores) en enteros, por ejemplo:

Al momento en que se tiene una ontología A y una ontología B como se vio en casos anteriores, se extraen el conjunto de clases y antes de aplicar el algoritmo de similitud, se convierten en enteros para poder así hacer la división y obtener así el cociente que nos dirá el índice de similitud de las clases, propiedades de objeto, propiedades de datos e individuos respectivamente.

9 CONCLUSIONES

El mundo de la web está compuesto por demasiada información hoy en día, la cual no siempre tiene un orden o una norma de estructuración para que los motores de búsqueda tengan la gran facilidad de obtener resultados óptimos y precisos. Es por eso que este proyecto refleja una parte de la importancia que puede llegar a tener una web semántica.

El proyecto presente hace una comparación por un método que utiliza conjuntos para así poder manejar con más precisión el conjunto de elementos que conforman una ontología. Este no es el único método pro el cual se pueden obtener este tipo de información, puesto que existen varias formas más, pero como opción más viable se eligió está en el desarrollo del proyecto.

El comparador actual que se tiene implementado como un sistema web puede ser anexado para hacerlo compatible con un motor de búsqueda semántico que utiliza ontologías como fuentes de información, ya que así se hace un mejor análisis se las mismas y pueda ser un poco más preciso.

El proyecto terminal presente destaca la manipulación de cada una de las características que conforman la ontología y el aprovechamiento de su estructura para poder hacer un análisis más exhaustivo y poder así explotar en gran potencial que tiene las ontologías.

El comparador se convierte en un elemento importante para que en un futuro alguien más se interese y pueda hacer la convergencia de este comparador junto con un clasificador de ontologías, para así formar un motor de búsqueda mucho más fiable de los que existen en la actualidad, es así como se concluye este proyecto, con la gran esperanza de que en un futuro de algunos frutos de conocimiento y se espera pueda ser aprovechado por otras mentes interesadas en el mismo para tener una web más organizada y llevar la tecnología a las cosas que aun solo son teorías o existen en la ciencia ficción.

10 BIBLIOGRAFÍA

- [1] E. Ruiz, «Ontologías en Informática,» 2013.
- [2] P. B. E. R. J. Madhavan, «Generic Schema Matching using Cupid VLDB,» de *VLDB conference*, Roma, Italy, 2001.
- [3] M. G. Sierra, «Sistema clasificador de ontologías,» México, D.F., 2013.
- [4] J. P. Martínez, «Extracción automatizada y representación de servicios Web mediante ontologías,» México, D.F., 2011.
- [5] L. Y. G. Jurado, «Sistema de recuperación de información de textos de investigación de la Web,» México, D.F., 2012.
- [6] F. O. Z. E. Nasser Yazdani, «Ontology matching using vector space,» de *European Conference on Information*, Dubai, 2008.
- [7] M. F. P. C. David Vallet, «An Ontology Based Information Retrieval Model,» España, Madrid, 2005.
- [8] H. J. a. L. O. Berners-Lee T., «The Semantic Web,» *Scientific American*, vol. 284, n° 5, pp. 34-43, 2001.
- [9] A. L. Tello, «Ontologías en la Web Semántica,» Jornadas de Ingeniería Web' 01, Cáceres, 2005.