



# Instituto Politécnico Nacional

## Centro de Investigación en Computación

---

**Laboratorio de Procesamiento Inteligente de Información  
Geoespacial**

**Cálculo de la visibilidad de conceptos en ontologías**

Tesis para obtener el grado de  
**Maestro en Ciencias de la Computación**

Presenta:  
**Ing. Henocho Rodríguez Franco**

Directores de Tesis:  
**Dr. Rolando Quintero Téllez**  
**Dr. Marco Antonio Moreno Ibarra**

**MÉXICO, D.F.**

**Diciembre 2011**





# INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 11:00 horas del día 24 del mes de noviembre de 2011 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:  
**Centro de Investigación en Computación**  
para examinar la tesis titulada:

**"CÁLCULO DE LA VISIBILIDAD DE CONCEPTOS EN ONTOLOGÍAS"**

Presentada por el alumno:

**RODRÍGUEZ**

Apellido paterno

**FRANCO**

Apellido materno

**HENOCH**

Nombre(s)

Con registro:

B	0	9	1	6	6	4
---	---	---	---	---	---	---

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA

Directores de Tesis


  
Dr. Rolando Quintero Téllez

  
Dr. Marco Antonio Moreno Ibarra

  
Dr. Grigori Sidorov

  
Dr. Miguel Jesús Torres Ruiz

  
Dr. Miguel Félix Mata Rivera

  
Dr. Gilberto Lorenzo Martínez Luna

PRESIDENTE DEL COLEGIO DE PROFESORES

  
Dr. Luis Alfonso Villa Vargas  
INSTITUTO POLITÉCNICO NACIONAL  
CENTRO DE INVESTIGACION  
EN COMPUTACION  
DIRECCION





**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

*CARTA CESIÓN DE DERECHOS*

En la Ciudad de México, el día 28 de Noviembre de 2011, el que suscribe, Ing. Henoch Rodríguez Franco alumno del Programa de Maestría en Ciencias de la Computación, con número de registro B091664, adscrito al Centro de Investigación en Computación, manifiesta que es único autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Rolando Quintero Téllez y el Dr. Marco Antonio Moreno Ibarra y en este acto cede los derechos del trabajo intitulado “*Cálculo de la visibilidad de conceptos en ontologías*”, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección [hafx1@hotmail.com](mailto:hafx1@hotmail.com). Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

---

Ing. Henoch Rodríguez Franco



## RESUMEN

Las ontologías se han utilizado ampliamente dentro de la computación, ya que representan el conocimiento sobre dominios en específico. De forma general, se conforman de conceptos y relaciones entre conceptos, en ocasiones representados al estilo de un modelo de red. Resulta útil evaluar la distancia conceptual entre los conceptos de una ontología que representa la similitud semántica entre ellos. Esto puede utilizarse para aplicaciones como la recuperación de información, sistemas de recomendación o la integración de datos.

Utilizando como referencia el algoritmo *DIS-C*, el cual evalúa la distancia conceptual entre conceptos de una ontología con base en la ponderación de las relaciones entre los conceptos por parte del usuario. Lo que agrega cierta subjetividad a la evaluación de la similitud. En este trabajo se propone una modificación a *DIS-C*, enfocada en ponderar automáticamente las relaciones conceptuales. Esto se ajusta refinando las ponderaciones hasta alcanzar un estado óptimo definido por un umbral. Para esto se propone una métrica, denominada *visibilidad*, para evaluar la asequibilidad de cada concepto, considerando la ontología como si fuera un grafo fuertemente conexo, esto representa las peculiaridades geométricas de la ontología (basada en un grafo). El enfoque permite utilizarlo en ontologías que no necesariamente posean una estructura jerárquica.

El algoritmo propuesto se implementó utilizando el lenguaje *Java* y *Graphviz* como herramienta de visualización. Se realizaron pruebas que muestran su utilidad en diversos tipos de ontologías, además se realizó una comparación contra el algoritmo de confusión lo que permitió analizar diferencias y similitudes entre los métodos.





## ABSTRACT

Ontologies have been widely used in computer science, they represent knowledge about specific domains. In general, are composed by concepts and relationships between concepts, sometimes are represented as a network model. It is useful to assess the conceptual distance between the concepts of ontology; it represents the semantic similarity between them. This can be used for applications such as information retrieval, recommender systems or data integration.

Using as reference the DIS-C algorithm, which evaluates the conceptual distance between concepts in ontology, the user defines a weight for the relations between the concepts. It adds some subjectivity to the judgment of similarity. In this work, we propose a modification of DIS-C, focused on the conceptual relationships automatically weighed. For this we propose a metric, called visibility, to assess the affordability of each concept, considering the ontology like a strongly connected graph, it represents geometric peculiarities of ontology (graph based representation). The weights are refined in order to reach an optimal state defined by a threshold. The approach allows use our method in ontologies that not necessarily possess a hierarchical structure.

The proposed algorithm was implemented using the Java language and Graphviz as a visualization tool. The method was tested in different ontologies, in addition a comparison was made against confusion algorithm in order to analyze the differences and similarities between methods.



## AGRADECIMIENTOS

Deseo agradecer sinceramente a mis asesores Dr. Rolando Quintero Téllez y Dr. Marco Antonio Moreno Ibarra, por su confianza, apoyo, enseñanzas y por haberme guiado a través de la realización de esta tesis.

Agradezco al IPN, CONACYT y PIFI, por el apoyo económico, mediante el cual ha sido posible la realización de mis estudios de posgrado.

También extiendo mi agradecimiento a todos mis compañeros de estudio y trabajo por su apoyo y los gratos momentos de convivencia.

Finalmente agradezco al Centro de Investigación en Computación (CIC), por los recursos, espacio y años de formación que me ha brindado para lograr mis estudios de posgrado.



## ÍNDICE GENERAL

RESUMEN.....	7
ABSTRACT .....	9
AGRADECIMIENTOS .....	11
ÍNDICE GENERAL.....	13
ÍNDICE DE FIGURAS .....	15
ÍNDICE DE MATRICES.....	18
ÍNDICE DE TABLAS .....	19
ÍNDICE DE CÓDIGOS.....	20
1 CAPÍTULO I.....	23
INTRODUCCIÓN.....	23
1.1 Generalidades .....	23
1.2 Planteamiento del problema.....	24
1.3 Objetivo .....	25
1.4 Metas .....	25
1.5 Organización del documento .....	26
2 CAPÍTULO II.....	27
ESTADO DEL ARTE.....	27
2.1 Mediciones de similitud semántica con modelos de red .....	27
2.1.1 Distancia conceptual en redes semánticas .....	27
2.1.2 Similitud semántica en una taxonomía considerando en el contenido de información .....	29
2.1.3 Confusion .....	32
2.1.4 Algoritmo <i>DIS-C</i> .....	34
2.2 Medidas de similitud.....	35
2.2.1 Similarity fooding .....	36
2.2.2 Proceso de correspondencia basada en grafos (GMO).....	37
2.2.3 Similitud léxica .....	37
2.2.4 Similitud entre propiedades.....	38
2.2.5 Similitud entre superclases .....	38
2.3 Otras Aplicaciones.....	39
2.3.1 PageRank .....	39
2.3.2 La similitud semántica para entidades de diferentes ontologías.....	40
2.3.3 Similitud de escenas espaciales .....	41

2.3.4	TDD – Un modelo integral para la evaluación cualitativa de similitud espacial <sup>43</sup>	
2.3.5	Aplicación en la Web Semántica Geoespacial .....	46
2.3.6	La similitud semántica en la recuperación de la información geográfica	47
2.4	Semántica .....	50
2.5	Ontologías.....	51
2.6	Discusión .....	53
3	CAPÍTULO III .....	55
	METODOLOGÍA.....	55
3.1	<i>Visibilidad</i> .....	55
3.2	Ejemplo de aplicación.....	58
3.2.1	Algoritmo <i>Floyd–Warshall</i> .....	60
3.2.2	Cálculo de la <i>visibilidad y distancia conceptual</i> .....	61
3.3	Análisis de resultados.....	68
4	CAPÍTULO IV .....	71
	PRUEBAS EXPERIMENTALES Y RESULTADOS .....	71
4.1	Resultados y pruebas experimentales del método propuesto .....	71
4.1.1	Análisis de resultados .....	79
4.2	Algoritmo <i>DIS-C</i> vs algoritmo <i>confusion</i> .....	80
5	CAPÍTULO V.....	85
	CONCLUSIONES Y TRABAJO FUTURO .....	85
5.1	Conclusiones .....	85
5.2	Contribuciones .....	86
5.3	Trabajo futuro.....	86
	REFERENCIAS .....	89
	ANEXOS.....	95
5.4	Ontología No. 2 .....	95
5.5	Ontología No. 3 .....	97
5.6	Ontología No. 4 .....	99
5.7	Ontología No. 5 .....	101
5.8	Ontología No. 6 .....	103
5.9	Ontología No. 7 .....	105

## ÍNDICE DE FIGURAS

Figura 1. Fragmento de la taxonomía de Wordnet. Las líneas continuas representan enlaces, las líneas de puntos indican que algunos nodos intermedios se omiten, para ahorrar espacio.....	29
Figura 2. Otro fragmento de la taxonomía de WordNet.....	30
Figura 3. Jerarquía de seres vivos.....	33
Figura 4. Conjunto de medidas de similitud y los elementos de ontologías a los que se enfocan (Zagal, 2008).....	36
Figura 5. A y B son enlaces anteriores de C.....	39
Figura 6. Cálculo simplificado de <i>PageRank</i> .....	40
Figura 7. Conexión de ontologías independientes (a) ontología parcial de WordNet (b) ontología parcial de los datos espaciales normalizados de transferencia.....	41
Figura 8. Dos escenas similares y sus diferencias (Bruns & Egenhofer, 1996).....	42
Figura 9. Entorno de red conceptual de las relaciones topológicas: (a) dos polígonos; (b) una línea y un polígono; (c) un punto y un polígono; (d) dos líneas; (e) dos puntos; (f) un punto y una línea (Li & Fonseca, 2006).....	44
Figura 10. (a) Dirección de la red; (b) ejemplos de patrones; (c) clasificación de similitud de los patrones de (b) (Egenhofer & Franzosa, 1995) y (Li & Fonseca, 2006).....	45
Figura 11. Entorno de la métrica de distancia (Li & Fonseca, 2006).....	45
Figura 12. Cinco escenas serán comparadas usando el modelo TDD (Li & Fonseca, 2006).....	45
Figura 13. Resultados de la clasificación de similitud usando el modelo TDD (Li & Fonseca, 2006).....	46
Figura 14. Recuperación basada en semánticas en una ontología simplificada de figuras geométricas (Janowicz, Raubal, & Kuhn, 2011).....	49
Figura 15. Ejemplo de relaciones semánticas.....	50
Figura 16. Descripción relación semántica.....	51
Figura 17. Representación de una ontología.....	52
Figura 18. <i>aisb</i> , Implica que “ <i>a</i> ” es del mismo tipo que “ <i>b</i> ”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011).....	53
Figura 19. <i>ahashb</i> , Implica que “ <i>b</i> ” es una propiedad de “ <i>a</i> ” o que “ <i>a</i> ” tiene una propiedad llamada “ <i>b</i> ”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011).....	53
Figura 20. <i>adoesb</i> , Implica que “ <i>b</i> ” es una habilidad de “ <i>a</i> ”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011).....	53
Figura 21. conceptualización <i>K</i> .....	57
Figura 22. Grafo <i>GkVγ, Aγ</i> de la conceptualización <i>K</i> .....	59
Figura 23. Grafo con valores de peso conceptual para cada relación en la conceptualización <i>K</i> .....	60
Figura 24. Algoritmo de Floyd–Warshall.....	61
Figura 25. Ontología (Cardona, 2010) visualizada en <i>Graphviz</i> . Utilizando como referencia el Código 1.....	72
Figura 26. <i>aisb</i> , Implica que “ <i>a</i> ” es del mismo tipo que “ <i>b</i> ”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011).....	72

Figura 27. <i>ahashb</i> , Implica que “ <i>b</i> ” es una propiedad de “ <i>a</i> ” o que “ <i>a</i> ” tiene una propiedad llamada “ <i>b</i> ”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011) .....	72
Figura 28. <i>adoesb</i> , Implica que “ <i>b</i> ” es una habilidad de “ <i>a</i> ”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011). .....	72
Figura 29. Ontología de la Figura 25 con valores resultantes de peso conceptual, diseñada en <i>Graphviz</i> . Utilizando como referencia el Código 5 para su diseño.....	76
Figura 30. Resultados de la implementación y análisis del <i>error cuadrático medio</i> después de 55 iteraciones. ....	77
Figura 31. Resultados de la implementación y análisis de <i>visibilidad</i> en la iteración 55, después haberse cumplido la condición del umbral de convergencia en el algoritmo <i>DIS-C</i> . ....	78
Figura 32. Resultados de la implementación y análisis de distancia conceptual para solo un tipo de relación subyacente en la conceptualización de la Figura 25 después de 7 iteraciones.....	79
Figura 33. Jerarquía de seres vivos.....	80
Figura 34. Valores resultantes de <i>visibilidad</i> de la jerarquía de seres vivos.....	82
Figura 35. Ontología (Rasgado, 2006) .....	95
Figura 36. Ontología de la Figura 35 con valores de pesos conceptuales. Utilizando como referencia la Matriz 8. ....	96
Figura 37. Valores resultantes de <i>visibilidad</i> de la Ontología No. 2.....	96
Figura 38. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 4 iteraciones. ....	96
Figura 39. Ontología (Rasgado, 2006).....	97
Figura 40. Ontología de la Figura 39 con valores de pesos conceptuales. Utilizando como referencia la Matriz 9. ....	98
Figura 41. Valores resultantes de <i>visibilidad</i> de la Ontología No. 3.....	98
Figura 42. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 5 iteraciones. ....	98
Figura 43. Ontología (Cimiano, 2006).....	99
Figura 44. Ontología de la Figura 43 con valores de pesos conceptuales. Utilizando como referencia la Matriz 10.....	100
Figura 45. Valores resultantes de <i>visibilidad</i> de la Ontología No. 4.....	100
Figura 46. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 10 iteraciones.....	100
Figura 47. Ontología (Cimiano, 2006).....	101
Figura 48. Ontología de la Figura 47 con valores de pesos conceptuales. Utilizando como referencia la Matriz 11.....	102
Figura 49. Valores resultantes de <i>visibilidad</i> de la Ontología No. 5.....	102
Figura 50. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 6 iteraciones. ....	102
Figura 51. Ontología (Cimiano, 2006).....	103
Figura 52. Ontología de la Figura 51 con valores de pesos conceptuales. Utilizando como referencia la Matriz 12.....	104
Figura 53. Valores resultantes de <i>visibilidad</i> de la Ontología No. 6.....	104



Figura 54. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 6 iteraciones. ....	104
Figura 55. Ontología (Staab & Studer, 2009). ....	105
Figura 56. Ontología de la Figura 55 con valores de pesos conceptuales. Utilizando como referencia la Matriz 13.....	106
Figura 57. Valores resultantes de <i>visibilidad</i> de la Ontología No. 7.....	106
Figura 58. Valores de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 7 iteraciones. ....	106

## ÍNDICE DE MATRICES

Matriz 1. Matriz de adyacencia del grafo <b>GK</b> .....	59
Matriz 2. Matriz de adyacencia con valores de peso conceptual para cada relación en el grafo <b>GK</b> .....	60
Matriz 3. Matriz de pesos conceptuales para la conceptualización K.....	62
Matriz 4. Matriz de adyacencia con valores de peso conceptual después de 29 iteraciones.....	65
Matriz 5. Matriz de adyacencia con valores de peso conceptual después de 59 iteraciones.....	66
Matriz 6. Matriz de adyacencia con valores de distancia conceptual después de 82 iteraciones.....	67
Matriz 7. Matriz de adyacencia resultante una vez implementado el algoritmo Floyd Warshall. Los valores resultantes pertenecen a la ontología de la Figura 25 después de 55 iteraciones.....	75
Matriz 8. Matriz resultante de pesos conceptuales después de 4 iteraciones.....	95
Matriz 9. Matriz resultante de pesos conceptuales después de 5 iteraciones.....	97
Matriz 10. Matriz resultante de pesos conceptuales después de 10 iteraciones.....	99
Matriz 11. Matriz resultante de pesos conceptuales después de 6 iteraciones.....	101
Matriz 12. Matriz resultante de pesos conceptuales después de 6 iteraciones.....	103
Matriz 13. Matriz resultante de pesos conceptuales después de 7 iteraciones.....	105

## ÍNDICE DE TABLAS

Tabla 1. Tabla con valores resultantes de la evaluación de <i>confusion</i> de la jerarquía de seres vivos. ....	34
Tabla 2. Tabla con valores resultantes de <i>visibilidad</i> de la 1ª iteración. ....	62
Tabla 3. Tabla con valores resultantes de distancia conceptual para el tipo de relación “is” de la 1ª iteración. ....	63
Tabla 4. Tabla con valores resultantes de distancia conceptual para el tipo de relación “has” de la 1ª iteración. ....	63
Tabla 5. Tabla con valores resultantes de distancia conceptual para el tipo de relación “does” de la 1ª iteración. ....	63
Tabla 6. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K de la 1ª iteración. ....	63
Tabla 7. Tabla con valores resultantes de <i>visibilidad</i> después de 30 iteraciones. ....	65
Tabla 8. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K después de 30 iteraciones. ....	65
Tabla 9. Tabla con valores resultantes de <i>visibilidad</i> después de 60 iteraciones. ....	66
Tabla 10. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K después de 60 iteraciones. ....	66
Tabla 11. Tabla con valores resultantes de <i>visibilidad</i> después de 83 iteraciones. ....	67
Tabla 12. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K después de 83 iteraciones. ....	67
Tabla 13. Tabla con valores resultantes de la evaluación de <i>confusion</i> de la jerarquía de seres vivos. ....	81
Tabla 14. Tabla con valores resultantes de la evaluación del algoritmo <i>DIS-C</i> de la jerarquía de seres vivos. ....	81

## ÍNDICE DE CÓDIGOS.

Código 1. Código para visualizar un grafo utilizando el programa <i>Graphviz</i> .....	72
Código 2. Código donde se ejemplifica la forma de ingresar los datos de una ontología en el programa.....	73
Código 3. Código donde se presenta la implementación de los costos de “entrar” y “salir” de los conceptos directamente relacionados en la ontología.....	74
Código 4. Código donde se presenta la implementación del algoritmo <i>Floyd-Warshall</i> . .....	74
Código 5. Código para diseñar una ontología resultante utilizando el programa <i>Graphviz</i> . Tomando como referencia los resultados obtenidos en la Matriz 7.....	76
Código 6. Código donde se presenta la implementación del cálculo de <i>visibilidad</i> , <i>error cuadrático medio</i> y el umbral de convergencia.....	77
Código 7. Código donde se presenta la implementación del cálculo de la distancia conceptual.....	78





# 1 CAPÍTULO I.

## INTRODUCCIÓN

### 1.1 Generalidades

Una *distancia o proximidad conceptual* se refiere a la similitud semántica basada en un modelo de red o basado en grafos sobre representaciones conceptuales como las redes semánticas, jerarquías u ontologías (Schwering, 2008). Esta distancia representa que tan parecidos o relacionados semánticamente se encuentran los conceptos (Rodríguez & Egenhofer, 2004). La similitud semántica es un tema central en el área del procesamiento semántico y tiene una larga tradición en las ciencias cognitivas, ya que puede utilizarse con diversos fines. Dentro de las ciencias de la computación se han estudiado en la inteligencia artificial (Rissland, 2006), además encontramos aplicaciones de la similitud semántica en la recuperación de información (Janowicz, Raubal, & Kuhn, 2011), en buscadores de contenidos multimedia (Lew, Sebe, Djeraba, & Jain, 2006) y en la integración semántica de información (Fonseca, 2008) (Shvaiko & Euzenat, 2008). Mientras que dentro de las ciencias cognitivas la similitud se ha aplicado para aprender acerca de la cognición humana, el razonamiento y la categorización de las diferencias y similitudes en las conceptualizaciones (Goldstone, Medin, & Halberstadt, 1997).

En el contexto de las ciencias de la información geográfica se han utilizado mediciones de similitud para aplicaciones como la comparación de escenas espaciales (Nedas & Egenhofer, 2008) y la verificación de consistencia de datos geográficos (Sheeren, Mustière, & Zucker, 2009) (Moreno-Ibarra, 2007).

De acuerdo con la forma de representar conocimiento del dominio, las medidas de similitud se clasifican de la siguiente forma: similitud geométrica, de propiedades, de

red, de alineamiento y modelo transformacional (Schwering, 2008). Cada una de ellas depende en gran parte de la aplicación para la que sean utilizadas, ya que como se ha mencionado, dependen de la representación.

En este caso, se propone utilizar un enfoque basado en el *modelo de red*, donde se utiliza un algoritmo que iterativamente evalúa que tan parecidos son dos conceptos en relación a la semántica que explícitamente expresa una ontología. Como se mencionó anteriormente los modelos de red para similitud son empleados en representaciones del conocimiento en forma de redes semánticas o basadas en grafos. Estas se componen de nodos (conceptos) y aristas (relaciones). Los nodos representan las unidades de conocimiento, por ejemplo, objetos, conceptos o propiedades. Mientras que las aristas enlazan a los nodos entre si y representan relaciones entre ellos de forma explícita. Aunque el modelo de representación siempre tiene la misma estructura, los modelos de red pueden diferir restringiendo la dirección de las relaciones. Esto representa que las mediciones de similitud basada en el modelo de red son dependientes del contexto (Rodríguez & Egenhofer, 2004) y de la semántica que describe la ontología.

## 1.2 Planteamiento del problema

En este trabajo se definen e integran un conjunto de criterios que ayudan a evaluar la proximidad conceptual entre dos conceptos. De acuerdo con (Rada, Mili, Bicknell, & Blettner, 1989), una *distancia conceptual* es la longitud del camino más corto que conecta los conceptos en una conceptualización, la cual representa la similitud semántica. Por lo tanto, el cálculo de la distancia entre conceptos depende directamente del espacio que los separa. De este modo, en este enfoque las mediciones de similitud deben cumplir con la propiedad de *solucionabilidad*, la cual representa que la representación debe ser lo suficientemente rica como para permitir que exista un camino entre todos los conceptos. Esto representa que no se puede calcular la proximidad entre conceptos que no están conectados. Por otra parte, las medidas de similitud en el modelo de red supone que cada relación es relevante para determinar un juicio sobre ellas mismas (Schwering, 2008).



En este caso se considera al algoritmo *DIS-C*, que se usa para calcular la *distancia conceptual* entre conceptos de una ontología, se basa en la analogía de una ontología a un *grafo fuertemente conexo*. El grafo define las relaciones entre conceptos y cada relación tiene asignada una ponderación que se considera para valorar la proximidad entre conceptos. Las ponderaciones las define el usuario de forma arbitraria.

Para resolver eliminar la arbitrariedad en las ponderaciones, se propone la utilización de una medida denominada *visibilidad*, la cual describe que tan asequible o accesible es un concepto desde cualquier parte de la ontología. La *visibilidad* se calcula considerando las relaciones que entran y que salen de un nodo en el grafo conceptual. En cierto sentido esto refleja la configuración geométrica de la ontología, esto permite que la *visibilidad* se incorpore en la ponderación automática de cada una de las relaciones entre conceptos. Con el fin de optimizar los valores para las ponderaciones, se hace un refinamiento iterativo, logrando que el algoritmo *DIS-C* pueda automáticamente valorar la similitud quitando la subjetividad de las ponderaciones que define un usuario.

### 1.3 Objetivo

Diseñar e implementar un método automatizado para evaluar la *distancia conceptual* y *visibilidad* de los conceptos de una ontología, con base en un método automático para calcular las ponderaciones de las relaciones entre conceptos con el fin de eliminar la subjetividad introducida por el usuario.

### 1.4 Metas

- Análisis del estado del arte en modelos de red para evaluar similitud semántica.
- Diseño e implementación de los métodos para transformar la ontología en un grafo doblemente conexo.
- Diseño e implementación de los métodos para ponderar las relaciones de forma automática.

- Implementación de los algoritmos de camino mínimo para evaluar *distancia conceptual* entre todos los conceptos de un grafo.
- Diseño e implementación de la medida de *visibilidad*.
- Diseño e implementación para evaluar la *distancia conceptual* para cada tipo de relación entre conceptos.
- Diseño e implementación de *distancia conceptual* considerando la evaluación de la *visibilidad* y la evaluación de la *distancia conceptual* para cada tipo de relación.
- Implementación del refinamiento de la ponderación de las relaciones.

## 1.5 Organización del documento

El resto del trabajo se organiza de la siguiente manera: el capítulo 2 presenta el estado del arte donde se describen los trabajos relacionados con el tema y donde se presenta la base teórica de este trabajo. El capítulo 3 presenta la metodología describiendo el funcionamiento del algoritmo para el cálculo de *distancia conceptual*. En el capítulo 4 se presentan las pruebas experimentales y resultados, esta parte muestra la implementación del algoritmo, también se muestra una comparación del algoritmo *DIS-C* con respecto al algoritmo *confusión* (Levachkine & Guzmán-Arenas, 2007). En el capítulo 5 se presentan las conclusiones, aportaciones y trabajo futuro de este trabajo. Finalmente, en la parte de anexos se presentan los resultados de otras pruebas experimentales realizadas con diferentes ontologías.

# 2 CAPÍTULO II

## ESTADO DEL ARTE

En este capítulo, se presenta el estado del arte donde se describen trabajos relacionados con medidas de similitud y trabajos relacionados con la recuperación de información. Al final se presenta una breve discusión sobre este capítulo.

### 2.1 Mediciones de similitud semántica con modelos de red

#### 2.1.1 Distancia conceptual en redes semánticas

Este trabajo es uno de los primeros casos publicados de la descripción de una métrica *distancia conceptual*; la cual se aplica a una ontología. En su métrica, la distancia se calcula midiendo la distancia entre nodos, cada nodo representa un concepto, contando el número de nodos intermedios. Los conceptos que son vecinos cercanos el uno del otro la medida de distancia calculada es bastante pequeña, indicando que semántica y conceptualmente son cercanos.

Los autores proponen una métrica, llamada “distancia”. Dando un enfoque basado en las aristas donde es la forma más natural y directa en la evaluación de similitud semántica en una taxonomía. Este enfoque estima la distancia, por ejemplo, la longitud de la arista, entre los nodos que se corresponde con los conceptos de comparación. La *distancia conceptual* se puede medir convenientemente con la distancia geométrica entre los nodos que representan los conceptos cuando están en la taxonomía.

Se señala que para una taxonomía jerárquica, la distancia debe satisfacer las propiedades de una métrica, llamada propiedad cero, propiedad simétrica, propiedad positiva y la desigualdad triangular.

Una función  $f(x, y)$  es una métrica si las siguientes propiedades son satisfechas:

- |                                     |                         |
|-------------------------------------|-------------------------|
| 1. $f(x, x) = 0$                    | propiedad cero.         |
| 2. $f(x, y) = f(y, x)$              | propiedad simétrica.    |
| 3. $f(x, y) \geq 0$                 | propiedad positiva.     |
| 4. $f(x, y) + f(y, z) \geq f(x, z)$ | desigualdad triangular. |

En adición se define a la *distancia conceptual*, como el número mínimo de relaciones semánticas que hay que atravesar para poder conectar los términos. En donde la forma más sencilla de determinar la distancia entre dos nodos (conceptos compuestos o elementales) A y B en una ontología es el camino más corto que une A y B, que es el mínimo número de aristas que separa A y B.

$$\text{Distancia}(A, B) = \text{mínimo número de aristas que separan A y B}$$

Sugiriendo que la “distancia” puede ser utilizada con éxito para evaluar la *distancia conceptual* entre grupos de conceptos cuando es utilizada en una ontología de relaciones jerárquicas. Así mismo se menciona que los juicios de distancia se correlacionan significativamente con los juicios de distancia que la gente hace y ayuda a determinar si una ontología es más adecuada o menos adecuada que otra.

Los autores realizaron experimentos en los que la distancia se aplica a los pares de conceptos y grupos de conceptos en una base jerárquica del conocimiento, mostrando el poder de las relaciones jerárquicas en la representación de información referente a la *distancia conceptual* entre conceptos.

Los autores aplicaron el método de la distancia a un dominio de médico y encontraron que la distancia de la función simula evaluaciones de los humanos en las parte de *distancia conceptual*.

En la discusión se menciona que en el mundo real, las distancias entre dos nodos adyacentes no son necesariamente iguales. Por lo tanto, es necesario tener en cuenta la ponderación de las aristas que conectan los dos nodos. Donde ciertos aspectos deben ser considerados para determinar el peso de la relación de forma automática.

### 2.1.2 Similitud semántica en una taxonomía considerando en el contenido de información

Resnik presenta una medida de similitud semántica en una taxonomía basada en la noción de contenido de la información. Argumentando que uno de los pasos cruciales de una Inteligencia Artificial para ser inteligente es que pueda “entender” el lenguaje natural que hablamos todos los días y pueda ser capaz de extraer palabras importantes como los nombres de una sentencia. Así la Inteligencia Artificial debe primero “entender” las relaciones entre los diferentes nombres de la taxonomía como los nombres son la base para la consolidación de nuestros conocimientos. Una de las relaciones más importantes entre los nombres es los similares que son. Donde la pregunta es ¿cómo se comparan esas relaciones? Resnik menciona que incluso los seres humanos a veces no están de acuerdo que dos palabras son similares (ej. “verde” y la “experiencia”).

A continuación se muestran dos taxonomías extraídas de WordNet (véase Figura 1).

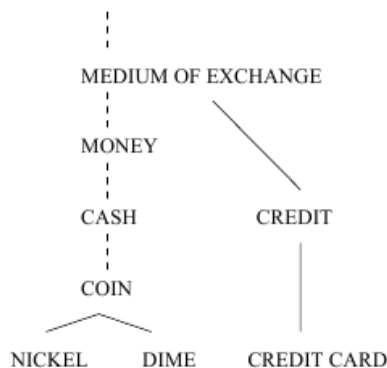


Figura 1. Fragmento de la taxonomía de Wordnet. Las líneas continuas representan enlaces, las líneas de puntos indican que algunos nodos intermedios se omiten, para ahorrar espacio.

Considérese que el concepto “COIN” incluye tanto “NICKEL” y “DIME”, mientras que el concepto “MEDIUM OF EXCHANGE” subsume los dos subárboles “MONEY” y “CREDIT”.

Una manera obvia de medir la similitud entre dos palabras es simplemente contando las aristas de una palabra a otra, así que contar la similitud entre “NICKEL” y “CREDIT CARD” es empezar desde “NICKEL”, siguiendo por la taxonomía del árbol hasta encontrar un concepto que subsume a ambos, “MEDIUM OF EXCHANGE”, siguiendo

hasta "CREDIT CARD" para un total de 6 aristas. Por lo tanto la siguiente fórmula se utiliza para medir la similitud:

$$S_{imedge}(w1, w2) = (2xMAX) - [minlen(c1, c2)]$$

Donde *MAX* es igual a la longitud máxima del posible camino, *c* es igual a todas las gamas de posibles conceptos sobre una palabra. Es necesario distinguir las similitudes entre conceptos distintos y entre palabras, porque una palabra como "Orange" corresponde a dos conceptos distintos, el color y la fruta. (Véase "NICKEL" y "GOLD" y sus 4 conceptos inmediatos en la Figura 2). Por lo tanto, tenemos que dar cuenta de todos esos conceptos y tomar la longitud entre aristas más corta usada para conectar uno de los conceptos de *w1* hacia *w2*. Finalmente, lo maximizamos para restar de  $2xMAX$  la longitud máxima posible.

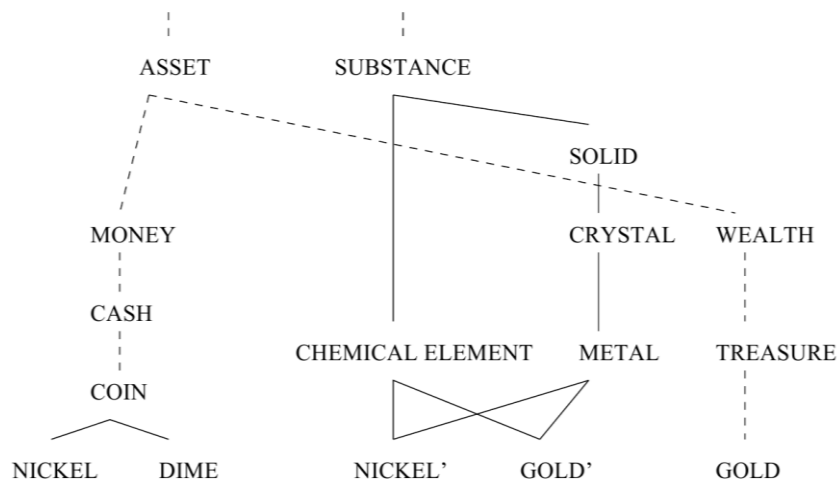


Figura 2. Otro fragmento de la taxonomía de WordNet.

Nótese que "NICKEL", "NICKEL'" y "NICKEL'" son conceptos distintos como sus inmediatos ascendentes son distintos.

Mediante el uso de la argumentación estándar de la teoría de la información, el contenido de la información de un concepto se define como el negativo de la probabilidad de registro,  $-\log p(c)$ , donde  $p(c)$  es la probabilidad de encontrar dicho concepto *c*.

Por ejemplo en la Figura 1 “MONEY” tiene un menor contenido de información que “NICKEL” así que la probabilidad de encontrar el concepto,  $p(MONEY)$  es mucho mayor que la probabilidad de encontrar  $p(NICKEL)$ . En consecuencia “NICKEL” contiene más información, es decir, más específico que “MONEY” ya que es cuantificable. Además, cada vez que el concepto “NICKEL” es utilizado el concepto “MONEY” es también utilizado como el nodo principal de “NICKEL”. Como resultado, Resnik concluyo que cuanto más alto los términos en el árbol de la taxonomía, menor información contienen, pero son encontrados con mayor frecuencia. Por ejemplo “MATTER” tendría un  $p(MATTER)$  cercano a 1 y un contenido de información cercano a 0. Resnik utiliza la siguiente ecuación para definir la similitud entre dos conceptos distintos:

$$Sim(c1, c2) = \max c \text{ in } S(c1, c2) [-\log p(c)]$$

Donde  $c$  es igual al concepto y  $S(c1, c2)$  es igual al conjunto de conceptos  $c$ , que subsume a ambos  $c1$  y  $c2$  y  $-\log p(c)$  es igual a la información la información contenida, que subsume a  $c1$  y  $c2$ . Entonces esta ecuación significa simplemente que queremos un concept, con un contenido máximo de información que incluye tanto  $c1$  y  $c2$ . Por lo tanto, en la Figura 2, “COIN” es utilizado para la similitud entre conceptos “NICKEL” y “DIME” en lugar de “CASH”, ya que es el que tiene mayor contenido de información y también subsume tanto a “NICKEL” y “DIME”.

La ecuación para la comprobación de dos palabras (recordando que una palabra puede tener varios conceptos distintos, por ejemplo “ORANGE”) es necesario tener en cuenta todos los conceptos que van antes y más allá de  $w1$  y  $w2$ .

$$Simword(w1, w2) = \text{take max from all } c1, c2 [sim(c1, c2)]$$

Donde  $c1$  va más allá de  $w1$  y  $c2$  va más allá de  $w2$ . La fórmula significa que queremos comparar todos los conceptos de  $w1$  y  $w2$  y tener el máximo valor similar  $Sim(c1, c2)$  donde  $w1$  y  $w2$  pueden ser una instancia del concepto máximo. Por ejemplo en la Figura 2 para “NICKEL” y “GOLD” la palabra “NICKEL” tiene padres: “COIN”, “CHEMICAL ELEMENT” y “METAL”. El  $c$  máximo de “NICKEL” y “GOLD” podría ser

elegido para ser “ASSET”, “CHEMICAL ELEMENT” y “METAL”. Al igual que en el mundo real, suponemos que  $p(ASSET) > p(METAL) > p(CHEMICAL\ ELEMENT)$ , es decir, más general y menos contenido de información, “CHEMICAL ELEMENT” sería elegido como su información contenida,  $Sim(NICKEL', GOLD')$ , puede ser calculada utilizando la ecuación  $Sim(c1, c2)$  y se asigna a  $Simword(GOLD, NICKEL)$  para las dos palabras “GOLD” y “NICKEL”.

### 2.1.3 Confusion

Presentan una medida denominada *confusion* que se aplica sobre jerarquías de conceptos. Esta medida puede interpretarse como la confusión o el error de tener un valor  $r$ , en vez de un valor  $s$ , es decir cuando un valor simbólico se utiliza en lugar del valor deseado o correcto. Para fines de este trabajo se emplea como una *distancia conceptual*, la cual existe entre dos términos de una jerarquía. Algunos términos relacionados con la confusión y jerarquías (Levachkine & Guzmán-Arenas, 2007)

- Definición *Jerarquía*. Para un elemento del conjunto  $E$ , una jerarquía  $H$  de  $E$  es otro conjunto de elementos donde cada  $e_i$  es un valor simbólico que representa a cualquier elemento de  $E$  o una partición; y  $\cup_i \{r_i | e_i \propto r_i\} = E$  (La unión de todos los conjuntos representados por los  $e_i$  de  $E$ ).
- Definición *Jerarquía Simple*: Una jerarquía simple es un árbol con raíz  $E$  y si un nodo tiene un hijo, éste forma partición del padre.
- Definición *Jerarquía Ordenada*: En una jerarquía ordenada, los nodos de algunas particiones obedecen una relación de orden.
- Definición *Padre\_de (v)*. En un árbol que representa una jerarquía, el nodo *padre\_de* un nodo es aquel del cual desciende.
- Definición *Hijos\_de (v)* son los valores que descienden de  $v$ .
- Definición *Raíz* es el nodo que no tiene padre.
- Definición Si  $r, s \in H$ , entonces la confusión de usar  $r$  en vez de  $s$ , se denota como  $conf(r, s)$ , es:



- $conf(r, r) = conf(r, s) = 0$ , donde  $s$  es cualquier ascendente de  $r$ . (Regla 1)
- $conf(r, s) = 1 + conf(r, padre\_de(s))$ . (Regla 2)
- Para medir  $conf$ , se cuentan los enlaces *descendentes* de  $r$  a  $s$ , al valor reemplazado.
- Definición Para jerarquías simples compuestas por conjuntos ordenados, la confusión de usar  $r$  en vez de  $s$ ,  $conf'(r, s)$ , se define como:
  - $conf'(r, r) = conf(r, cualquier\ ascendente\ de\ r) = 0$ ;
  - Si  $r$  y  $s$  son hermanos  $conf'(r, s) = 1$ ; Si el padre no está en un conjunto ordenado; entonces,  $conf'(r, s)$  es la distancia relativa de  $r$  a  $s$  siendo; el número de pasos requeridos para llegar de  $r$  a  $s$  en el orden definido, dividido entre la cardinalidad - 1 del padre. (Regla 3)
  - $conf(r, s) = 1 + conf(r, padre\_de(s))$

Ejemplo de *confusion* para una jerarquía de seres vivos (véase Figura 3).

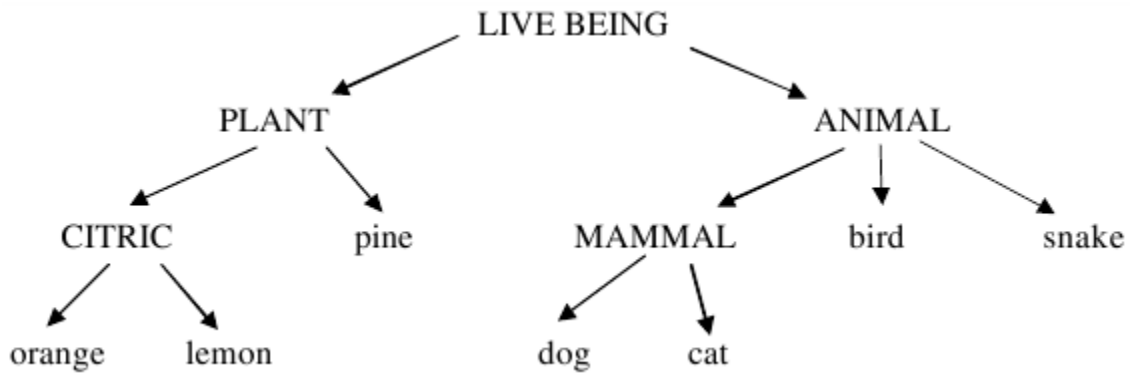


Figura 3. Jerarquía de seres vivos .

En la siguiente tabla se muestra el uso de la *confusion* en la fila  $r$  en lugar de la columna  $s$  para la jerarquía de seres vivos.

↓ r	→ s												
	conf	Live being	Animal	Plant	Mammal	Bird	Snake	Citric	pine	cat	dog	lemon	orange
Live being	0	1	1	2	2	2	2	2	2	3	3	3	3
Animal	0	0	1	1	1	1	2	2	2	2	2	3	3
Plant	0	1	0	2	2	2	1	1	3	3	2	2	2
Mammal	0	0	1	0	1	1	2	2	1	1	3	3	3
Bird	0	0	1	1	0	1	2	2	2	2	3	3	3
Snake	0	0	1	1	1	0	2	2	2	2	3	3	3
citric	0	1	0	2	2	2	0	1	3	3	1	1	1
Pine	0	1	0	2	2	2	1	0	3	3	2	2	2
Cat	0	0	1	0	1	1	2	2	0	1	3	3	3
Dog	0	0	1	0	1	1	2	2	1	0	3	3	3
Lemon	0	1	0	2	2	2	0	1	3	3	0	1	1
Orange	0	1	0	2	2	2	0	1	3	3	1	0	0

Tabla 1. Tabla con valores resultantes de la evaluación de *confusion* de la jerarquía de seres vivos.

#### 2.1.4 Algoritmo *DIS-C*

El algoritmo *DIS-C* toma como referencia los trabajos mencionados en el capítulo 2, donde se explica que estos solo funcionan con ontologías de relaciones taxonómicas lo que limita su aplicación. El algoritmo *DIS-C* no tiene esa limitante por lo que es aplicable a cualquier tipo de ontología. La idea principal de este trabajo de tesis es determinar el valor de *distancia conceptual* a cada uno de los tipos de relaciones en la ontología. Esto se logra al convertir esta última en un grafo dirigido y ponderado en el que cada concepto se convierte en un nodo y cada relación se convierte en un par de aristas. Sobra decir que esto es aplicable de manera directa a las relaciones binarias, pero que es posible aplicarlo a relaciones n-arias. Una vez que se tiene el grafo, es posible aplicar diferentes técnicas de teoría de grafos. El paso natural consiste en calcular las distancias mínimas, con lo que se puede encontrar la distancia entre conceptos que no se encuentran directamente relacionados.

Cabe mencionar, que de ello surge la propuesta de una métrica llamada *visibilidad* que sirve para calcular los valores de *distancia conceptual* que buscamos, pero que además representa la “importancia” de un concepto para una ontología en particular.

Así, los pasos a seguir para calcular la *distancia conceptual* en *DIS-C* son los siguientes:

1. Para cada tipo de relación, se debe asignar la *distancia conceptual* que genera la existencia de dicha relación. Además se debe asignar una ponderación en cada sentido de la relación; por ejemplo, si tenemos la relación “es” y que “gato es animal”, se debe establecer la distancia de “gato” a “animal” y de “animal” a “gato”.

A manera de ilustrar se puede establecer que  $distancia(gato, animal) = 0$  y  $distancia(animal, gato) = 1$ . Más formalmente, sea  $K(C, \mathfrak{R}, R)$  una conceptualización donde  $C$  es el conjunto de conceptos,  $\mathfrak{R}$  es el conjunto de tipos de relaciones y  $R$  es el conjunto de relaciones en la conceptualización. Se puede decir que para cada tipo de relación  $\rho \in \mathfrak{R}$ , se deben establecer los valores de  $\delta^\rho$ , para la relación  $\rho$  en el sentido normal, y de  $\bar{\delta}^\rho$ , para la relación  $\rho$  en el sentido inverso. En el ejemplo anterior  $\delta^{es} = 0$  y  $\bar{\delta}^{es} = 1$ .

2. Se crea el grafo  $G_K(V, A)$  para la conceptualización  $K$ . Primero, se agrega cada concepto  $c \in C$  como un vértice del grafo  $G_K$ , es decir que  $V = C$ . Ahora para cada relación  $a\rho b \in R$ , donde  $a, b \in C$  y  $\rho \in \mathfrak{R}$ ; se añaden las aristas  $(a, b, \delta^\rho)$  y  $(b, a, \bar{\delta}^\rho)$ .
3. Se calculan los caminos mínimos entre cada par de vértices, los cuales están mapeados directamente a los conceptos en la ontología; por lo que el resultado es la *distancia conceptual* diseminada a todos los conceptos en una conceptualización  $K$ .

## 2.2 Medidas de similitud

Existen diferentes tipos de similitud con base en términos: se enfoca en el nombre de las entidades en las ontologías, principalmente en el nombre de las clases y propiedades.

Similitud semántica: Su alcance va más allá de los nombres de las entidades, se enfoca en los componentes que definen la semántica de una clase:

- Similitud entre propiedades de clases: Considera las coincidencias existentes entre las propiedades de dos clases.
- Similitud entre superclases: Se refiere al par de superclases con respecto a dos clases comparadas.

La Figura 4 resume el ámbito de aplicación del conjunto de medidas de similitud. Se utilizan la similitud léxica y similitud entre propiedades como un enfoque local, las

cuales se combinan con la similitud entre clases ascendentes (que considera las estructuras de las ontologías). Una medición global es alcanzada al integrar otras medidas pueden ser agregadas o sustituidas como los algoritmos usados para medir similitud estructural; como “Similarity Flooding (Melnik, Garcia-Molina, & Erhard, 2001)” y “Graph Matching Ontology (Jian, Hu, Cheng, & Qu, 2005)”.

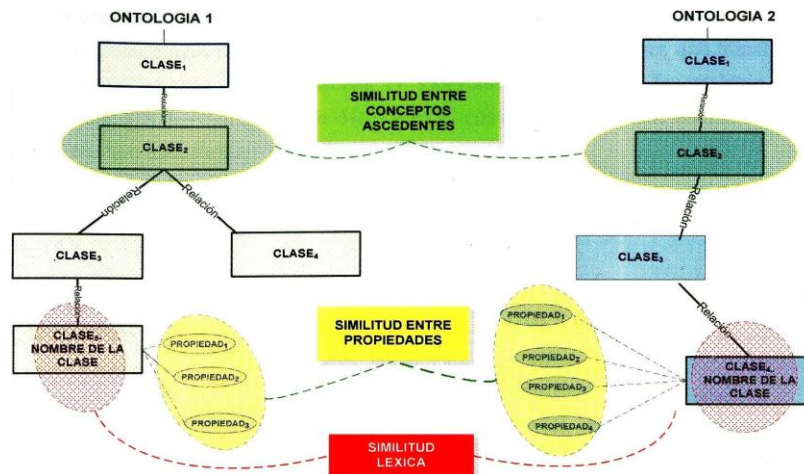


Figura 4. Conjunto de medidas de similitud y los elementos de ontologías a los que se enfocan (Zagal, 2008).

### 2.2.1 Similarity flooding

Es una técnica basada en modelos estructurales, este algoritmo es usado para establecer correspondencias en varios modelos o estructuras de datos (Melnik, Garcia-Molina, & Erhard, 2001). Como entrada recibe dos modelos de datos que después serán convertidos a grafos. *Similarity flooding* funciona bajo el principio: “si dos grafos son similares entonces, también sus conceptos adyacentes (en los grafos) en cierto grado también son similares”.

El algoritmo comienza al obtener las correspondencias iniciales entre los elementos (de los dos grafos creados), utilizando una función de similitud basada en cadenas de texto, después se retornan las similitudes iniciales entre los elementos que corresponden.

Con las correspondencias iniciales, el algoritmo procede a establecer más correspondencias entre otros elementos, basándose en la suposición que si dos

elementos corresponden satisfactoriamente con alguna medida de similitud, también sus elementos vecinos son similares.

Las iteraciones continúan incrementando las similitudes de los conceptos en los grafos, hasta un punto fijo en que todas las medidas de similitud para todos los conceptos se han estabilizado. Este algoritmo corresponde a la categoría de técnica basado en el nivel estructural de los modelos.

### 2.2.2 Proceso de correspondencia basada en grafos (GMO)

La idea principal de este trabajo es medir la similitud entre dos entidades (de dos ontologías), considerando la acumulación de similitudes de sentencias involucradas (tripletas) (Jian, Hu, Cheng, & Qu, 2005). Las entidades a comparar se consideran dentro de la misma estructura: sujeto, predicado y objeto. Por otro lado, GMO también toma en cuenta la acumulación de las similitudes de entidades vecinas respecto a las dos entidades que están siendo comparadas.

Como entrada GMO, acepta un conjunto de pares de entidades relacionadas, que son encontradas previamente por otros algoritmos de similitud (*matching*). A la salida GMO proporciona parejas de entidades adicionales al comparar la similitud estructural.

### 2.2.3 Similitud léxica

La distancia de Levenshtein o distancia de edición (*edit distance*), fue creada en 1965 por el científico ruso Vladimir Levenshtein. La idea consiste en determinar el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra, estas operaciones son: inserción, eliminación o sustitución de un carácter.

Por ejemplo, la distancia de Levenshtein entre los términos “hotel” y “hostal” es de dos porque se necesitan al menos dos ediciones elementales para cambiar un término en el otro término.

1. hotel (?) hostal

2. hostel (?) hostal (inserción de 's' antes de t)
3. hostel (?) hostal (sustitución de 'e' por 'a')
4. hostal (?) hostal

#### 2.2.4 Similitud entre propiedades

Esta medida tiene el propósito de comparar la distancia entre el conjunto de propiedades de dos clases, considerando la simetría de la similitud; es decir que la  $Similitud\_propiedades("hotel", "botel")$  sea la misma que  $Similitud\_propiedades("botel", "hotel")$

El objetivo es identificar si cada propiedad  $p_n$  del conjunto de propiedades  $P$  de una clase  $C_1$ , coincide con otra propiedad  $p'_m$  del conjunto de propiedades de  $P'$  de otra clase  $C_2$ . Para realizar dicha comparación entre cada propiedad, se utilizan las etiquetas de ambas propiedades (o nombres) como entrada a una medida de similitud léxica, en este caso se ocupa la distancia de Levenshtein.

#### 2.2.5 Similitud entre superclases

Esta medida se enfoca en la similitud entre superclases (o similitud entre conceptos ascendentes) a partir de dos clases iniciales. La similitud entre algún par de superclases puede influir para establecer una correspondencia entre dos clases. Para calcular esta medida de similitud es necesario calcular previamente la similitud léxica y de propiedades de todas las parejas de clases.

A continuación se define la similitud entre superclases, dado  $C_1, C_2$  una pareja de subclases:

$$Similitud\_superclases(C_1, C_2) = Similitud\_parcial(SuperClases(C_1), SuperClases(C_2))$$

La siguiente ecuación expresa la similitud parcial: dado  $C_{a1}, C_{a2}$  una pareja de superclases obtenida desde  $C_1, C_2$ :

$$Similitud\_parcial(C_{a1}, C_{a2}) = \frac{Similitud\_lexica(C_{a1}, C_{a2}) + Similitud\_propiedades(C_{a1}, C_{a2})}{2}$$

La idea intuitiva de la similitud parcial es obtener un valor que integre preliminarmente la similitud léxica y entre propiedades de forma equitativa. El objetivo es compara parejas de superclases y seleccionar la pareja de superclases con mayor similitud entre sí, a fin de encontrar de manera aproximada la pareja de superclases en donde coincidan o converjan las clases  $C_1$  y  $C_2$ .

## 2.3 Otras Aplicaciones

### 2.3.1 PageRank

*PageRank* es un conjunto de algoritmos utilizados para asignar de forma numérica la relevancia de los documentos o páginas web indexadas por un motor de búsqueda (Page, Brin, Motwani, & Winograd, 1999).

El *PageRank* de una página ( $A$ ) está definido de manera recursiva por los *PageRanks* de las diferentes páginas que enlazan a la página ( $A$ ).

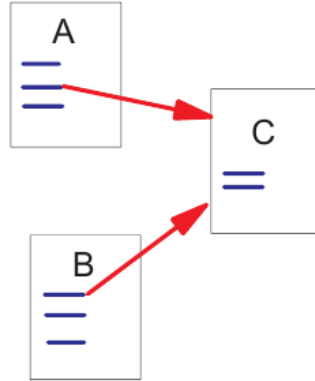


Figura 5. A y B son enlaces anteriores de C.

Donde el algoritmo es:

$$PR(A) = (1 - d) + d * \sum_{i=1}^n \frac{PR(Ti)}{C(Ti)}$$

dónde:

$PR(A)$  es el *PageRank* de la página  $A$ .

$d$  es un factor de amortiguación que tiene un valor entre 0 y 1.

$PR(T_i)$  son los valores de *PageRank* que tienen cada una de las páginas  $i$  que enlazan a  $A$ .

$C(T_i)$  es el número total de enlaces salientes de la página  $i$  (sean o no hacia  $A$ ).

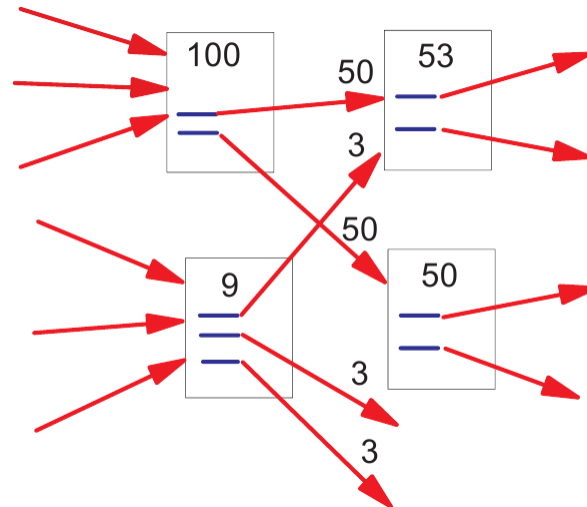


Figura 6. Cálculo simplificado de *PageRank*.

El *PageRank* de las páginas  $T_i$  que enlazan a la página  $A$  no influye en el *PageRank* de  $A$  de manera uniforme. Dentro del algoritmo del *PageRank*, el *PageRank* de la página  $T$  esta siempre valorado por el número de enlaces  $C(T)$  en la página  $T$ . Esto significa que cuantos más enlaces tenga una página  $T$ , menor será el *ranking* de la página  $A$  enlazada en la página  $T$ . Explicado de otra manera que cuantos más enlaces tenga una página menor será el *ranking* de *PageRank* de las páginas enlazadas, motivo por el cual cuando un directorio web apunta a una página esto no repercute demasiado en su *PageRank*.

Finalmente la suma de los *PageRanks* de todas las páginas  $T_i$  es multiplicada por un factor  $d$  que toma un valor intermedio entre 0 y 1, de manera que el *ranking* de que una página haya sido enlazada por otra se reduce o aumenta dependiendo de ese valor.

### 2.3.2 La similitud semántica para entidades de diferentes ontologías

Los enfoques tradicionales de similitud semántica trata de calcular la distancia semántica entre las definiciones dentro de una ontología única. Por lo tanto el



resultado de la evaluación de similitud semántica es independiente del resultado de otras ontologías.

Por lo que, en (Rodríguez & Egenhofer, 2004), se propone un método para calcular la similitud semántica entre diferentes tipos de ontologías. Esta función de similitud determina la similitud entre clases mediante un proceso de comparación sobre conjuntos de sinónimos, vecindades semánticas y características distintivas como lo son las funciones y atributos. Permitiendo conectar ontologías independientes con otras ontologías (véase Figura 7).

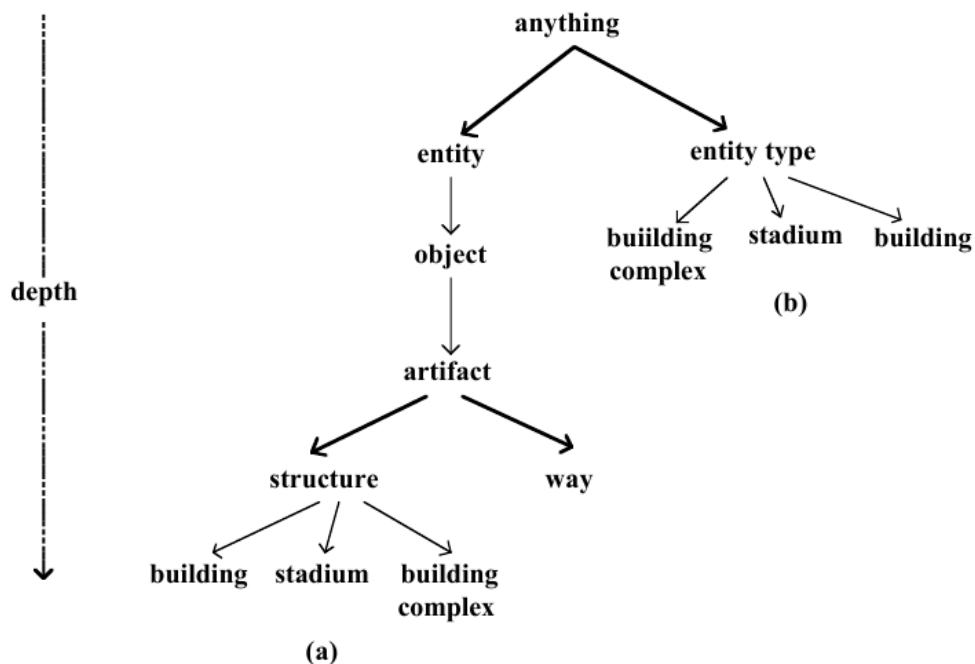


Figura 7. Conexión de ontologías independientes (a) ontología parcial de WordNet (b) ontología parcial de los datos espaciales normalizados de transferencia.

Los resultados experimentales de similitud entre ontologías diferentes indican que el modelo funciona bien cuando las ontologías tienen representaciones completas y detalladas en la entidad de clases.

### 2.3.3 Similitud de escenas espaciales

(Bruns & Egenhofer, 1996) presentan un modelo para la evaluación de la similitud espacial, basada en las relaciones topológicas y relaciones de distancia, y sus vecinos

conceptuales correspondientes. El modelo utiliza el diseño de tres tipos de vecinos conceptuales para tres tipos de relaciones espaciales, las cuales son:

- Relaciones topológicas.
- Relaciones de distancia.
- Relaciones de dirección.

Dadas dos escenas con igual número de objetos espaciales, diferentes relaciones espaciales y considerando los esquemas de vecinos conceptuales, existe un número mínimo de cambios graduales, que transforman una escena en otra, el número de cambios graduales es considerado proporcional a la similitud entre las dos escenas. Por lo que entre más cambios se tengan, las escenas son menos similares.

La similitud está determinada por un proceso que, gradualmente reemplaza las relaciones espaciales por sus vecinos conceptuales, donde dicho procesos intenta construir una escena a partir de otra. Las escenas derivadas se clasifican con respecto al número de cambios, por lo que aquellas escenas que tienen menos cambios, son más similares que aquellas con más cambios.

En la Figura 8 se muestra un análisis de una escena, donde la única diferencia entre la Figura 8a y la Figura 8b, es que el objeto C, ha sido movido; para determinar el grado de similitud entre ambas escenas, donde se evalúan las relaciones espaciales y se cuentan sus diferencias.

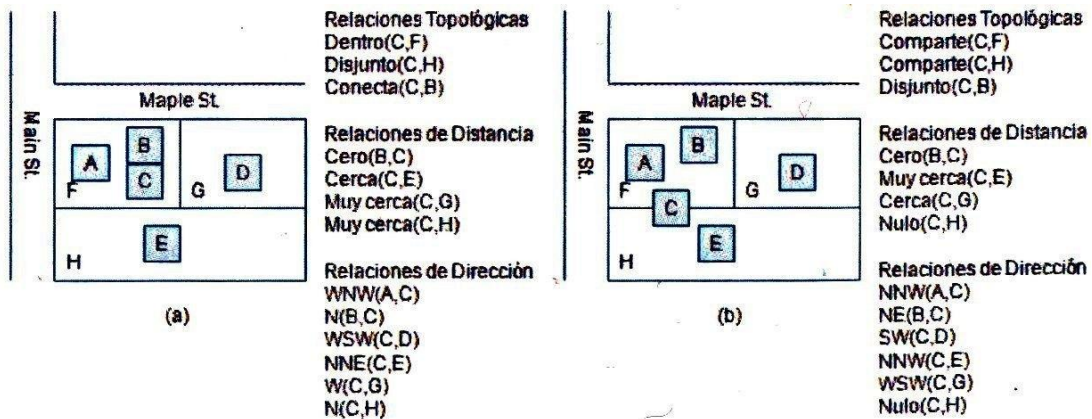


Figura 8. Dos escenas similares y sus diferencias (Bruns & Egenhofer, 1996).

Así, finalmente se menciona que la *web semántica geoespacial* permitirá a los usuarios recuperar de forma más precisa los datos que ellos necesitan, basándose en la semántica asociada a dichos datos.

#### 2.3.4 TDD – Un modelo integral para la evaluación cualitativa de similitud espacial

(Li & Fonseca, 2006) proponen el modelo TDD (Topología – Dirección - distancia) para medir similitud entre escenarios espaciales.

El modelo TDD se enfoca a escenas con dos objetos para realizar el análisis de similitud. Este modelo mide las coincidencias ( $C$ ) y diferencias ( $D$ ) entre escenas espaciales. La medida de similitud final ( $S$ ) es una combinación de ambas ( $S = C - D$ ).

El modelo TDD se basa en el modelo de alineamiento de estructura que es considerado en la comparación de un par de estímulos, donde las partes de un objeto deben de ser alineadas o colocadas con las partes de otro.

El modelo TDD se enfoca en la similitud de relación y la similitud de atributos. Considerando que la similitud de relación y la similitud de atributos tienen diferentes impactos sobre el juicio de las coincidencias y diferencias (Tversky, 1977). Dependiendo del contexto; las diferencias de atributos y de relación serán más o menos relevantes en la evaluación de la similitud. Enfocado en este problema la solución propuesta en el modelo TDD es tener diferentes pesos asociados a cada atributo y a cada relación. De este modo los usuarios del modelo TDD colocan los pesos de forma interactiva así ellos reflejan la situación particular en el cual la medida de similitud está siendo hecha.

Otra contribución del modelo TDD incluye un orden de prioridad en la aplicación (Topología → Dirección → distancia) dentro de la valoración espacial de similitud. La estrategia adoptada para ordenar la prioridad en la aplicación del modelo TDD es colocando los costos de los pesos de transformación sobre los tres tipos diferentes de relación espacial.

$$peso(topología) > peso(dirección) > peso(distancia)$$

La relación topológica en el modelo TDD considera la red conceptual entre dos polígonos (véase Figura 9). Los seis nodos pueden ser divididos en tres grupos: disjuntos, sobrepuesto (en diferentes niveles) e iguales. Los nodos encontrados, sobre puestos, contenidos y contenidos-encontrados pueden ser vistos como como sobre puestos en diferentes niveles.

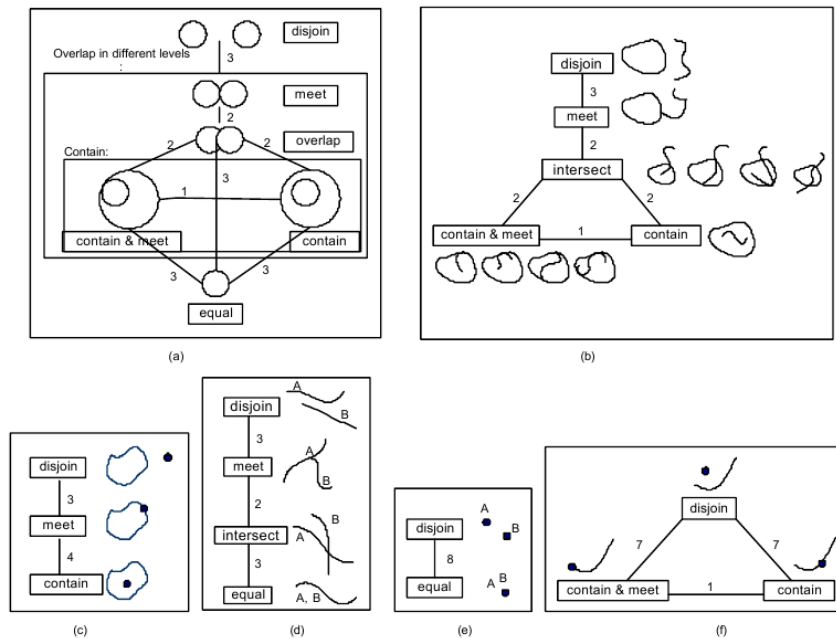


Figura 9. Entorno de red conceptual de las relaciones topológicas: (a) dos polígonos; (b) una línea y un polígono; (c) un punto y un polígono; (d) dos líneas; (e) dos puntos; (f) un punto y una línea (Li & Fonseca, 2006).

La relación de dirección en el modelo TDD considera que  $\frac{p}{2}$  es la mayor distancia direccional entre las escenas espaciales de dos objetos. Hay 5 nodos en la direccional  $\frac{p}{2}$  (Egenhofer & Franzosa, 1995):  $\{east, west\}$ ,  $\{northeast, southwest\}$ ,  $\{north, south\}$ ,  $\{northwest, southeast\}$  y  $\{same\}$  (véase Figura 10).

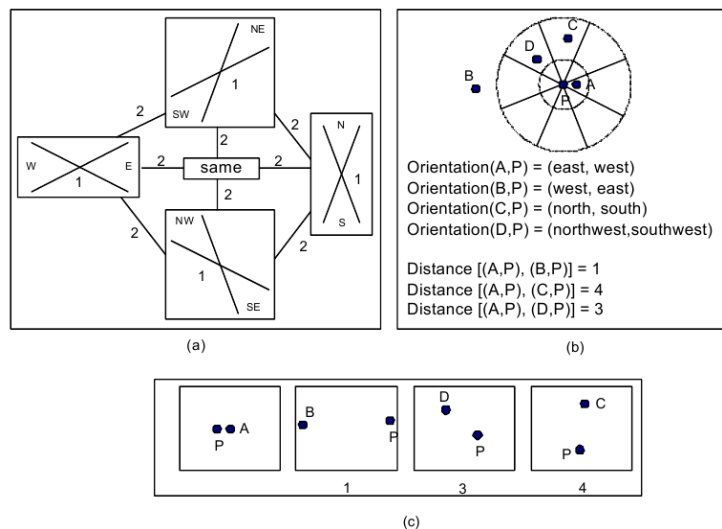


Figura 10. (a) Dirección de la red; (b) ejemplos de patrones; (c) clasificación de similitud de los patrones de (b) (Egenhofer & Franzosa, 1995) y (Li & Fonseca, 2006).

La relación métrica de distancia es adoptada la tradicional métrica de gradualidad 4  $\{equal, near, medium, far\}$  (véase Figura 11) para medir la similitud en el entorno de la relación métrica de distancia entre un par de escenas espaciales.

Ejemplo del uso del modelo de similitud TDD. Donde la similitud es medida como el número de transformaciones necesarias para ir de un nodo en la red a otro nodo. Las menos transformaciones entre los dos nodos, indica una mayor similitud que comparten esos dos nodos.

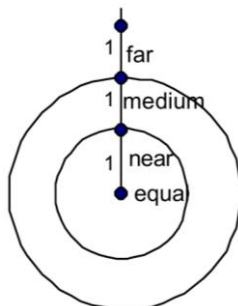


Figura 11. Entorno de la métrica de distancia (Li & Fonseca, 2006).

Considerando las siguientes 5 escenas, (A,L), (B,L), (C,L), (D,L) y (E,L) en la Figura 12.

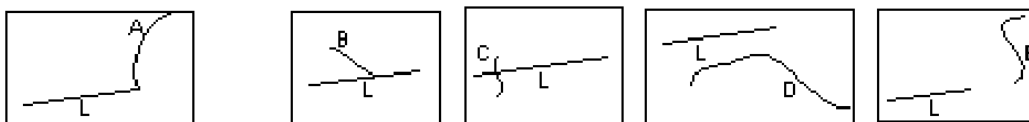


Figura 12. Cinco escenas serán comparas usando el modelo TDD (Li & Fonseca, 2006).

Sus relaciones espaciales son:

$$Scene(A, L) = \{meet, (northeast, southwest), equal\}$$

$$Scene(B, L) = \{meet, (northwest, southeast), equal\}$$

$$Scene(C, L) = \{intersect, (west, east), equal\}$$

$$Scene(D, L) = \{disjoin, (southeast, northwest), near\}$$

$$Scene(E, L) = \{disjoin, (northeast, southwest), medium\}$$

Donde se obtuvo la siguiente clasificación de similitud del más similar al menos similar con respecto a  $(A, L)$  (véase Figura 13):



$$(A, L), (B, L) > (A, L), (E, L) > (A, L), (C, L) > (A, L), (D, L)$$

Figura 13. Resultados de la clasificación de similitud usando el modelo TDD (Li & Fonseca, 2006).

En su definición más amplia, la recuperación de información está enfocada con encontrar información relevante basada en la consulta del usuario.

La similitud ha sido aplicada en varias tareas y en varios dominios. Esto hace que la selección de una apropiada medida para una aplicación en particular sea una tarea desafiante.

### 2.3.5 Aplicación en la Web Semántica Geoespacial

(Egenhofer, 2002) describe los elementos necesarios para construir lo que él denomina como la *web semántica geoespacial*, haciendo énfasis en aspectos para la construcción de la semántica formal, así como el desarrollo de múltiples ontologías espaciales y de términos. Donde una ontología espacial es aquella que contiene conceptos y relaciones de un dominio geográfico. Mientras que en las ontologías basadas en términos se enfocan en la naturaleza del texto.

Indicando, además que se requieren mecanismos para que la semántica se represente de tal forma que sea procesable tanto para personas como para computadoras, así como el procesamiento de consultas espaciales apoyadas por ontologías. Además, se

mencionan aspectos tales como la medición de resultados recuperados, basados en la concordancia entre la necesidad de información expresada y la semántica disponible en las fuentes de información y sistemas de búsquedas.

Egenhofer detalla y describe un nuevo marco de trabajo para la recuperación de información geográfica basada en la semántica de ontologías espaciales y de términos. Adicionalmente, el artículo visualiza y describe la representación de la semántica en diferentes partes del proceso de recuperación (personas, interfaces, sistemas de búsquedas y fuentes de información). Se enfatiza en la necesidad de la creación de servicios, en la carencia donde hay una carencia del procesamiento semántico de datos geográficos.

### 2.3.6 La similitud semántica en la recuperación de la información geográfica

(Janowicz, Raubal, & Kuhn, 2011) examinaron varias medidas de similitud desde diferentes dominios donde encontraron patrones genéricos los cuales conjuntamente forman un marco de trabajo para describir como la similitud computada.

Definiendo la recuperación de información por el grado de relevancia sin declarar la forma de medir esta relevancia. Basándose en la siguiente definición, se explica el rol de la similitud restringiendo la definición de tal manera.

$$m[\mathfrak{R}(O, (Q, \langle I, \mapsto \rangle))]$$

$O$  es el conjunto de conceptos objetivo ( $C_t$ ) en una ontología.

$Q$  es un concepto en particular o la selección de una concepto en una enunciado para la búsqueda ( $C_s$ ).

$I$  y  $\mapsto$  son información contextual adicional en el tiempo de ejecución ( $C_c$ ).

$\mathfrak{R}$  es la relación de similitud entre un par de conceptos.

$m$  es el grado de similitud entre un par de conceptos.

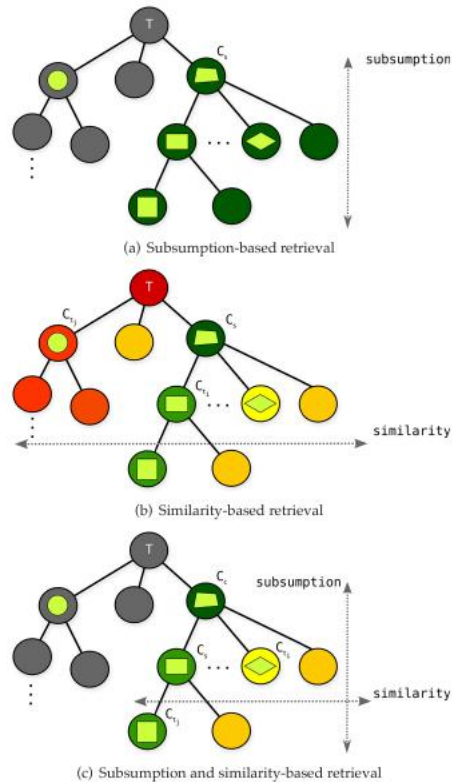
Formalmente, el conjunto de resultados de una recuperación basada en la subsunción se define como  $\mathbb{RS} = \{C \mid C \in O \wedge C \sqsubseteq Q\}$  así cada concepto en  $\mathbb{RS}$  es una subsunción de la búsqueda o en la consulta del concepto, esto cumple con los criterios de

búsqueda del usuario (véase Figura 14). En consecuencia no hay grado de relevancia  $m$ , en otras palabras este siempre será 1. La falta de relevancia en la información y la rigidez de la subsunción hace que la selección de un concepto de búsqueda sea adecuada, pero es el gran reto en la recuperación basada en la subsunción.

Para la recuperación basada en la similitud como se muestra en la Figura 14 el conjunto de resultados está definido como  $\mathbb{RS} = \{C \mid C \in O \wedge sim(Q, C) > t\}$ ; donde  $t$  es el umbral definido por el usuario o la aplicación. En la recuperación basada en la similitud, el concepto de búsqueda es el concepto que el usuario está realmente buscando, no importa si es parte de la ontología de consultar o no. Es más flexible que el enfoque basado en la subsunción. Además, los resultados se clasifican como una lista ordenada de valores de similitud descendente en representación de la relevancia  $m$ . Esto hace que sea más fácil para el usuario, donde solo selecciona un concepto apropiado de los resultados. Sin embargo, no se garantiza que los conceptos devueltos coinciden con todos los criterios de búsqueda del usuario.

Para superar estas deficiencias, se combinaron la subsunción y el razonamiento de similitud con la introducción de contextos para reducir el conjunto de los potenciales conceptos de objetivo. Como se muestra en la Figura 14, solo esos conceptos son comparados por similitud que son los subconceptos del concepto contexto  $C_c$ . De esta manera el usuario debe de seleccionar un mínimo de características de todos los conceptos objetivo. Lo anterior es definido como  $\mathbb{RS} = \{C \mid C \in O \wedge C \sqsubseteq C_c \wedge sim(Q, C) > t\}$ .





**Figura 14. Recuperación basada en semánticas en una ontología simplificada de figuras geométricas (Janowicz, Raubal, & Kuhn, 2011).**

En la Figura 14 se muestra una ontología de figuras geométricas. Donde se asume que el usuario está buscando por cuadriláteros con características específicas. En el caso de la subsunción, el conjunto de resultados contiene tipos como: *rectángulo*, *rombo*, *cuadrado* y así sucesivamente, sin información adicional acerca de su grado de relevancia. En el caso de la similitud el conjunto de resultados contiene información adicional de relevancia para esos tipos pero también figuras geométricas como un *círculo* el cual no satisfacía todos los requerimientos especificados por el usuario. Teniendo en cuenta sin embargo de que el círculo se presentara al final de la lista de importancia debido a su baja similitud. Para el caso de la combinación de subsunción y similitud un usuario puede preferir *cuadriláteros* con ángulos rectos como el *rectángulo* especificado en el concepto de búsqueda y *cuadrilátero* como concepto del contexto. En contraste con el caso de recuperación basada en la similitud, el conjunto de resultados no contiene *círculo*, pero aún ofrece información sobre el grado de pertinencia.

## 2.4 Semántica.

El término semántica, se refiere a los aspectos del significado (el contenido mental asignado a un objeto) o la interpretación de un determinado código simbólico, lenguaje o representación formal (Liddell & Scott, 1996). En principio, cualquier medio de expresión (código, lenguas, etc.) admite una correspondencia entre: palabras o expresiones simbólicas, situaciones y entidades (ya sean físicos o abstractos), esta correspondencia permite describir a dicho medio de expresión. El propósito de la semántica, es precisamente entender un concepto en un sentido. En este aspecto, es entender lo que se quiere decir.

Actualmente la definición de los sentidos de las palabras existe en forma de diccionarios. Es importante mencionar que las ontologías definen un concepto con base en su dominio, la semántica radica en las relaciones en los conceptos (véase Figura 15).

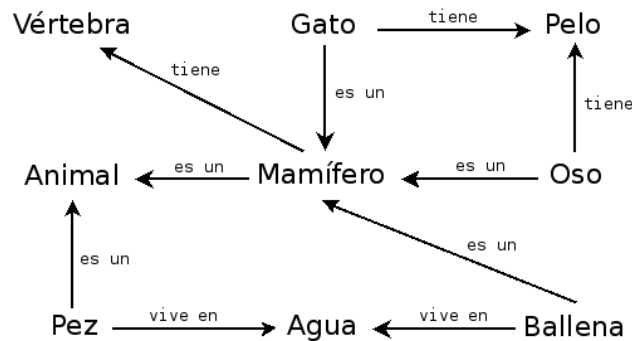


Figura 15. Ejemplo de relaciones semánticas.

Así, un esquema de representación en red es una forma de representación de conocimiento en la que los conceptos y sus interrelaciones se representan mediante un grafo. En caso de que no existan ciclos, estas redes pueden ser visualizadas como árboles.

Dado un conjunto de conceptos, elementos semánticos o términos relacionados semánticamente mediante alguna relación, una red semántica representa estas relaciones en forma de grafo. Explícitamente, dado un conjunto de términos

$\{t_1, t_2, \dots, t_n\}$  y cierta relación semántica entre ellos se construye un grafo  $G = (V, A)$  cumpliendo las siguientes condiciones:

El conjunto  $V$  es el conjunto de vértices o nodos del grafo. Este conjunto estará formado por  $n$  elementos (tantos vértices como términos). Cada uno de los vértices del grafo representará uno de los términos, por tanto los vértices del grafo se llamarán:  $t_1, t_2, \dots, t_n$ .

El conjunto  $A$  es el conjunto de aristas o líneas del grafo. Dados dos vértices (términos) del grafo  $t_i$  y  $t_j$  existirá una línea  $a_{ij}$  que une los vértices  $t_i$  y  $t_j$  si y sólo si los términos  $t_i$  y  $t_j$  están relacionados (véase Figura 16).

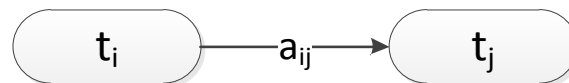


Figura 16. Descripción relación semántica.

## 2.5 Ontologías.

Thomas Gruber (Gruber, 1993) define una ontología, como una especificación explícita de una conceptualización y a su vez, una conceptualización, como una visión abstracta y simplificada de la realidad (objetos y sus relaciones) que se desea representar con un propósito particular. Cada base del conocimiento, sistema basado de conocimiento o agentes a nivel de conocimiento están comprendidos con alguna conceptualización de forma explícita o implícita. El término de conceptualización es definido como una estructura  $\langle D, R \rangle$ , donde  $D$  es un dominio y  $R$  es un conjunto de relaciones relevantes sobre  $D$ . (Gruber, 1993).

En el contexto de los sistemas información, las ontologías son un mecanismo utilizado originalmente en el campo de la Inteligencia Artificial que permiten describir con algún detalle la semántica de un dominio particular, sea un área de conocimiento (como la física o la medicina) o un área de interés específica (como el estudio de una bacteria o de un fenómeno espacial) a partir de los conceptos y relaciones entre ellos.

De esta manera, es posible enriquecer el conocimiento del dominio e inferir o construir nuevo conocimiento (Moreno, 2007).

Los beneficios que ofrecen las ontologías en el campo de los sistemas de información se traducen en la interoperabilidad semántica de los sistemas. Esto permite al usuario comunicarse con los sistemas de una forma más natural, dándole la posibilidad de extraer contenidos con mayor orientación al objeto que persigue.

En su concepción, las ontologías se componen de clases, conceptos, instancias y atributos o propiedades. Una instancia es un caso concreto de una clase y las clases definen un concepto. Por ejemplo "Juan" es una instancia de la clase "Persona" así como "México" es una instancia de la clase "País". Las propiedades definen una relación binaria entre dos instancias. Por ejemplo la propiedad "vive en" relaciona la instancia "Juan" con la instancia "México" (Horridge, Holger, Rector, Stevens, & Wroe, 2009).

El enfoque GEONTO-MET se describe en detalle en (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011) donde se utiliza un conjunto de relaciones axiomáticas "is", "has", "does" que permite una interpretación directa de las relaciones entre los conceptos dentro de la conceptualización (véase Figura 17, Figura 18, Figura 19 y Figura 20.)

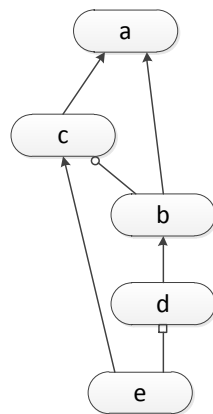


Figura 17. Representación de una ontología

Dónde:



Figura 18.  $a(is)b$ , Implica que “a” es del mismo tipo que “b”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011)



Figura 19.  $a(has)b$ , Implica que “b” es una propiedad de “a” o que “a” tiene una propiedad llamada “b”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011)



Figura 20.  $a(does)b$ , Implica que “b” es una habilidad de “a”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011).

## 2.6 Discusión

Independientemente de las ventajas y desventajas de cada modelo, estos tratan de medir la similitud semántica de diferentes maneras. En el caso del modelo basado en grafos se usa la teoría de grafos realizando el análisis del camino más corto para calcular el valor de similitud semántica. (Rada, Mili, Bicknell, & Blettner, 1989) Proponen una medida para calcular la *distancia conceptual* que mide la *distancia conceptual* entre dos conceptos dentro de una ontología. Esta medida fue diseñada principalmente para las ontologías con relaciones taxonómicas. Midiendo entre dos nodos o un conjunto de nodos como el promedio de las distancia mínima entre cada par de nodos relacionados. (Resnik, 1999) Describe una medida de similitud semántica basada en la noción del contenido de la información. Suponiendo que la distancia representada por una relación deben de ser conceptos que están relacionados.

De lo anterior es necesario reflexionar sobre si la distancia semántica es adecuada para medir la *distancia conceptual*. La propuesta que proponemos en este trabajo, no trata de medir la similitud semántica, sino que consiste en encontrar la *distancia conceptual*, teniendo en cuenta las ideas presentadas en (Rada, Mili, Bicknell, & Blettner, 1989) y (Resnik, 1999).

Por ejemplo, si se pregunta entre un “pájaro” y un “gato” ¿Cuál de ambos es más parecido a un “murciélago”? Si se utiliza el criterio de similitud de semántica, para cualquier persona el “pájaro” se parece más a un “murciélago”, mientras que para un ejemplo considerando la *distancia conceptual* se preguntaría ¿qué tan relacionados están una “tarjeta de crédito” y “comida”? en principio la similitud semántica indicaría que ambos conceptos no tienen parecido, pero conceptualmente se podría

indicar que se puede comprar “comida” mediante el uso de la “tarjeta de crédito”. De lo anterior debemos tener en consideración que la similitud semántica es diferente de la *distancia conceptual*, ya que la *distancia conceptual* es una medida que nos dice que fuertemente están relacionados dos conceptos, mientras que la similitud semántica nos indica que tan parecidos son. Por otra parte la distancia conceptual depende de la estructura de la propia ontología, donde si existe una relación directa entre dos conceptos la relación entre estos será más proximidad y donde, si no existe una relación directa la relación será menos próxima.

También cabe mencionar que todos los métodos mencionados funcionan solo con ontologías de relaciones taxonómicas lo que limita su aplicación. El algoritmo *DIS-C* no tiene esa limitante por lo que es aplicable a cualquier tipo de ontología. Por otra parte se pretende que el algoritmo funcione sin la necesidad de que algún individuo asigne un valor a cada relación; pues tal asignación, el algoritmo *DIS-C* la realizará de forma automática.

# 3 CAPÍTULO III

## METODOLOGÍA

En este capítulo se presenta la metodología describiendo el funcionamiento del algoritmo *DIS-C*, presentando aportaciones para este algoritmo para el cálculo de *distancia conceptual* de forma automática. Tomando como referencia los trabajos de (Rada, Mili, Bicknell, & Blettner, 1989), (Resnik, 1999), (Levachkine & Guzmán-Arenas, 2007) y propio algoritmo *DIS-C*.

### 3.1 Visibilidad

La *visibilidad* describe que tan asequible o accesible es un concepto desde cualquier parte de la ontología.

Sea  $K(C, \mathfrak{R}, R)$  una conceptualización  $K$ , donde  $C$  es el conjunto de conceptos,  $\mathfrak{R}$  es el conjunto de tipos de relaciones y  $R$  es el conjunto de relaciones.

Para calcular la *visibilidad* de un concepto  $a$  dentro de la conceptualización  $K$ , es decir  $v_a$ , se debe hacer lo siguiente: se crea el grafo dirigido  $G_K(V_G, A_G)$  para la conceptualización  $K(C, \mathfrak{R}, R)$ , haciendo que cada concepto  $c \in C$  sea un vértice del grafo  $G_K$ , es decir que  $V_G = C$ . Ahora para cada relación  $apb \in R$ , donde  $a, b \in C$  y  $\rho \in \mathfrak{R}$ , se añade la arista  $(a, b, \rho)$  a  $A_G$ .

El siguiente paso consiste en generar iterativamente, a partir de  $G_K$ , el grafo dirigido y ponderado  $\Gamma_K^j(V_\gamma^j, A_\gamma^j)$ . Para ello, en la  $j$ -ésima iteración, se hace que  $V_\gamma^j = V_G$ ,  $A_\gamma^j = A_G$  y para cada arista  $(a, b, \rho) \in A_G$ , se añaden las aristas  $(a, b, \omega_{ab}^j)$  y  $(a, b, \omega_{ba}^j)$ , donde  $\omega_{ab}^j$  es el peso de la arista que va del nodo  $a$  al nodo  $b$  en la  $j$ -ésima iteración. Estos se calculan sumando tres términos: el “costo de salir” del nodo de origen multiplicado

por la *visibilidad* de dicho nodo, la *distancia conceptual* de la relación  $\rho$  que conecta los nodos, el “costo de llegar” al nodo destino por el ocultamiento<sup>1</sup> del mismo, esto se expresa de la siguiente manera:

$$\omega_{ab}^j = v_a^{j-1} \omega_a^o + [\delta^\rho]^{j-1} + \frac{1}{v_b^{j-1}} \omega_b^i$$

$$\omega_{ba}^j = v_b^{j-1} \omega_b^o + [\bar{\delta}^\rho]^{j-1} + \frac{1}{v_a^{j-1}} \omega_a^i$$

Siendo precisamente  $v_x^j$  la *visibilidad* del vértice  $x \in V_G$  en la  $j$ -ésima iteración, y que con el grafo  $\Gamma_K^j$  se calcula su valor de como la razón de la distancia media desde dicho nodo hacia todos los demás nodos del grafo, entre la distancia media desde todos los nodos del grafo hacia el nodo, es decir:

$$v_x^j = \frac{\frac{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(x, y)}{\text{card}(V_\gamma^j)}}{\frac{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(y, x)}{\text{card}(V_\gamma^j)}} = \frac{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(x, y)}{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(y, x)}$$

donde  $\text{dist}_{\Gamma_K^j}$  es la distancia mínima desde el nodo  $x$  hacia el nodo  $y$  según en grafo  $\Gamma_K^j$ . Además,  $\forall x \in V_G, v_x^0 = 1$ , es decir que el valor inicial de *visibilidad* para todos los nodos es igual a 1. Además, se involucran los términos  $[\delta^\rho]^j$  y  $[\bar{\delta}^\rho]^j$  que son los valores de *distancia conceptual* de la relación que hay entre  $a$  y  $b$  (directa e inversa, respectivamente) cuyos valores son los que se buscan. Esta distancia se calcula como el promedio de las distancias  $\omega^j$  entre las aristas que comparten el mismo tipo de relación, es decir:

$$[\delta^\rho]^j = \frac{\sum_{(a,b) \in \rho^*} \omega_{ab}^j}{\text{card}(\rho^*)}$$

---

<sup>1</sup> El ocultamiento es lo inverso a la visibilidad.



$$[\bar{\delta}\rho]^j = \frac{\sum_{(a,b) \in \rho^*} \omega_{ba}^j}{\text{card}(\rho^*)}$$

donde  $\rho^* = \{(a, b, \rho) \in A_G\}$  es el conjunto de aristas que representan una relación  $\rho$ .

Inicialmente, estas distancias valen 0, es decir que  $\forall \rho \in \mathfrak{R}, [\delta\rho]^0 = 0, [\bar{\delta}\rho]^0 = 0$ .

Por otro lado,  $\omega_x^i$  es el costo de “llegar” al vértice  $x \in V_G$ , definido como  $\omega_x^i = \frac{o_x}{i_x + o_x}$  y

$\omega_x^o$  es el costo de “salir” del concepto  $x \in V_G$ , definido como  $\omega_x^o = \frac{i_x}{i_x + o_x}$ , donde  $i_x$  es el número de aristas dirigidas al nodo  $x$  ( $i_x = \text{card}(\cup\{y\}, (y, x, -) \in A_G)$ ) y  $o_x$  es el número de aristas que salen del nodo  $x$  ( $o_x = \text{card}(\cup\{y\}, (x, y, -) \in A_G)$ ).

Para ejemplificar esto, consideremos la ontología de la Figura 21. El concepto  $a$  tiene 2 conceptos que se relacionan con él ( $b$  y  $c$ ), por lo que el valor de  $i_a = 2$  (es decir, el número de relaciones que “llegan” al concepto  $a$ ). Por otro lado, el concepto  $a$  no se relaciona con ninguno otro concepto en la ontología, por lo que  $o_a = 0$  (es decir, no “sale” ninguna relación desde el concepto  $a$ ).

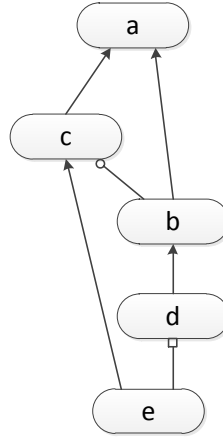


Figura 21. conceptualización K.

Entonces podemos establecer el “costo de llegar” al concepto  $a$  como  $\omega_a^i = \frac{o_a}{i_a + o_a} = \frac{0}{2+0} = 0$ , y el “costo de salir” de  $a$  como  $\omega_a^o = \frac{i_a}{i_a + o_a} = \frac{2}{2+0} = 1$ . De manera análoga, para el concepto  $b$ :  $i_b = 1$  (una relación “entra” a  $b$ ),  $o_b = 2$  (dos relaciones “salen” de  $b$ ); entonces,  $\omega_b^i = \frac{o_b}{i_b + o_b} = \frac{2}{1+2} = \frac{2}{3}$  y  $\omega_b^o = \frac{i_b}{i_b + o_b} = \frac{1}{1+2} = \frac{1}{3}$ .

Ahora supongamos que se realiza el cálculo para la primera iteración, es decir  $j = 1$ , entonces el valor de la arista que va de  $a$  hacia  $b$  será  $\omega_{ab}^1 = v_a^0 \omega_a^0 + [\delta^\rho]^0 + \frac{1}{v_b^0} \omega_b^i = \omega_a^0 + \omega_b^i = 1 + \frac{2}{3} = \frac{5}{3}$ . De igual forma,  $\omega_{ba}^1 = \frac{1}{3}$ .

Este proceso se realiza comenzando con  $j = 1$ , y aumentando el valor de  $j$  en una unidad, hasta que  $\frac{\sum_{x \in V_\gamma} (v_x^j - v_x^{j-1})^2}{\text{card}(V_\gamma)} \leq \epsilon_K$ ; donde  $\frac{\sum_{x \in V_\gamma} (v_x^j - v_x^{j-1})^2}{\text{card}(V_\gamma)}$  se define como el *error cuadrático medio* de la diferencia de los valores de *visibilidad* de la iteración  $j$  y los valores de *visibilidad* de la iteración  $j - 1$ ,  $\epsilon_K$  es el umbral de convergencia. Una vez cumplida la condición, el algoritmo termina.

Así finalmente este análisis considera cada uno de los tipos de relaciones que subyacen entre los conceptos en una ontología, como resultado se obtiene una métrica que pondera los conceptos y su tipo de relación.

### 3.2 Ejemplo de aplicación

Por otra parte en GEONTO-MET se ha propuesto una metodología para la conceptualización del dominio geográfico, la cual será utilizada para ejemplificar la aplicación del algoritmo DIS-C propuesto. Se propone el uso de las relaciones definidas en GEONTO-MET (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011) para aplicar el algoritmo *DIS-C*.

Como se mencionó anteriormente, para aplicar *DIS-C* se debe establecer la ponderación o peso conceptual para cada tipo relación y su inversa. En este trabajo, presentamos una aproximación al cálculo automático de los pesos conceptuales en una ontología. Teniendo una conceptualización  $K$ . Como la que se muestra en la Figura 21.

Se define como un grafo  $G_k(V_G, A_G)$  para la conceptualización  $K$ . Dónde:  $V_G = C$ ,  $A_G = (a, b, \omega_{ab}), (b, a, \omega_{ba})$ , tal y como se describe en el algoritmo *DIS-C*. Obteniéndose un grafo como el que se presenta en la Figura 22.

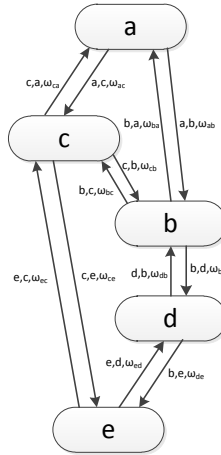


Figura 22. Grafo  $G_K(V_\gamma, A_\gamma)$  de la conceptualización K.

El grafo  $G_K$  lo podemos representar con la matriz de adyacencia, siguiente:

$$G_K = \begin{bmatrix} 0 & \omega_{ab} & \omega_{ac} & \infty & \infty \\ \omega_{ba} & 0 & \omega_{bc} & \omega_{bd} & \infty \\ \omega_{ca} & \omega_{cb} & 0 & \infty & \omega_{ce} \\ \infty & \omega_{db} & \infty & 0 & \omega_{de} \\ \infty & \infty & \omega_{ec} & \omega_{ed} & 0 \end{bmatrix}$$

Matriz 1. Matriz de adyacencia del grafo  $G_K$ .

dónde,  $\omega_{xy}$  representa el peso en el grafo de una relación entre dos conceptos,  $\infty$  representa la inexistencia de una relación.

Haciendo uso de la ecuación  $\omega_x^i = \frac{o_x}{i_x + o_x}$  del costo de “llegar” y de la ecuación

$\omega_x^o = \frac{i_x}{i_x + o_x}$  del costo de “salir” se obtiene lo siguiente:

$$\omega_{ab}^1 = 1.67$$

$$\omega_{ac}^1 = 1.33$$

$$\omega_{ad}^1 = \infty$$

$$\omega_{ae}^1 = \infty$$

$$\omega_{ba}^1 = 0.33$$

$$\omega_{bc}^1 = 0.67$$

$$\omega_{bd}^1 = 0.83$$

$$\omega_{be}^1 = \infty$$

$$\omega_{ca}^1 = 0.67$$

$$\omega_{cb}^1 = 1.33$$

$$\omega_{cd}^1 = \infty$$

$$\omega_{ce}^1 = 1.67$$

$$\omega_{da}^1 = \infty$$

$$\omega_{db}^1 = 1.17$$

$$\omega_{dc}^1 = \infty$$

$$\omega_{de}^1 = 1.50$$

$$\omega_{ea}^1 = \infty$$

$$\omega_{eb}^1 = \infty$$

$$\omega_{ec}^1 = 0.33$$

$$\omega_{ed}^1 = 0.50$$

Así obtenemos un primer grafo y su matriz de adyacencia correspondientes a la conceptualización  $K$ .

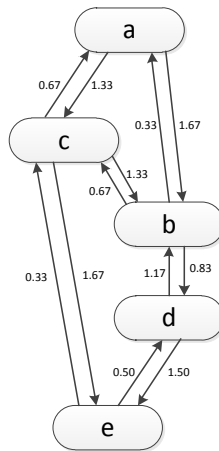


Figura 23. Grafo con valores de peso conceptual para cada relación en la conceptualización  $K$ .

$$K = \begin{bmatrix} 0 & 1.67 & 1.33 & \infty & \infty \\ 0.33 & 0 & 0.67 & 0.83 & \infty \\ 0.67 & 1.33 & 0 & \infty & 1.67 \\ \infty & 1.17 & \infty & 0 & 1.50 \\ \infty & \infty & 0.33 & 0.50 & 0 \end{bmatrix}$$

Matriz 2. Matriz de adyacencia con valores de peso conceptual para cada relación en el grafo  $G_K$ .

### 3.2.1 Algoritmo Floyd–Warshall

Con la matriz de adyacencia se realiza un análisis con algoritmo *Floyd-Warshall* (Floyd, 1962) (Warshall, 1962) sobre el primer grafo resultante para encontrar el camino mínimo entre cada par de nodos del grafo. El algoritmo compara todos los posibles caminos a través del grafo entre cada par de vértices, mejorando paulatinamente la estimación del camino, hasta que se llega a la estimación más óptima.

De lo anterior tenemos que el camino mínimo  $P_{ab}^{(k)}$ , desde el vértice  $a$  al vértice  $b$ , atravesando  $k$  vértices en el conjunto  $\{1,2,3, \dots, k-1\}$ , se define de la siguiente manera.

$$P_{ab}^{(k)} = \min\{P_{ab}^{(k-1)}, P_{ak}^{(k-1)} + C_{kb}^{(k-1)}\}$$

Para  $k = 0$  tenemos que:

$$P_{ab}^{(0)} = \omega_{ab}$$

Dónde  $\omega_{ab}$  es el peso de la relación  $a,b$ .

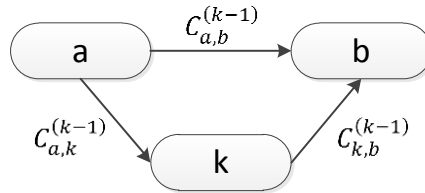


Figura 24. Algoritmo de Floyd-Warshall

$P_{ab}^{(k)} = \min\{P_{ab}^{(k-1)}, P_{ak}^{(k-1)} + C_{kb}^{(k-1)}\}$  y  $P_{ab}^{(0)} = \omega_{ab}$  son las definiciones base del algoritmo *Floyd-Warshall*. Funciona ejecutando primero  $P_{ab}^{(1)}$  para todos los pares de  $(a,b)$ , usándolos para después hallar  $P_{ab}^{(2)}$  para todos los pares  $(a,b)$ . Este proceso continúa hasta que  $k = n$ , y habremos encontrado el camino más corto para todos los pares de vértices  $(a,b)$  usando algún vértice intermedio.

### 3.2.2 Cálculo de la visibilidad y distancia conceptual

Basándose en los valores de la Matriz 2, realizamos el análisis con el Algoritmo Floyd-Warshall obteniendo lo siguiente:

$\omega_{ab}^1 = 1.67$	$\omega_{cd}^1 = 2.17$
$\omega_{ac}^1 = 1.33$	$\omega_{ce}^1 = 1.67$
$\omega_{ad}^1 = 2.50$	
$\omega_{ae}^1 = 3.00$	$\omega_{da}^1 = 1.50$
	$\omega_{db}^1 = 1.17$
$\omega_{ba}^1 = 0.33$	$\omega_{dc}^1 = 1.83$
$\omega_{bc}^1 = 0.67$	$\omega_{de}^1 = 1.50$
$\omega_{bd}^1 = 0.83$	
$\omega_{be}^1 = 2.33$	$\omega_{ea}^1 = 1.00$
	$\omega_{eb}^1 = 1.67$
$\omega_{ca}^1 = 0.67$	$\omega_{ec}^1 = 0.33$
$\omega_{cb}^1 = 1.33$	$\omega_{ed}^1 = 0.50$

Con los resultados anteriores se obtiene una matriz con valores de peso conceptual en la primera iteración. Donde el camino mínimo entre 2 vértices cualesquiera del grafo Figura 21 será el obtenido en la Matriz 3.

$$K = \begin{bmatrix} 0 & 1.67 & 1.33 & 2.50 & 3.00 \\ 0.33 & 0 & 0.67 & 0.83 & 2.33 \\ 0.67 & 1.33 & 0 & 2.17 & 1.67 \\ 1.50 & 1.17 & 1.83 & 0 & 1.50 \\ 1.00 & 1.67 & 0.33 & 0.50 & 0 \end{bmatrix}$$

**Matriz 3. Matriz de pesos conceptuales para la conceptualización K.**

Al obtener la matriz resultante de la primera iteración se realiza el análisis para cada uno de los nodos con la ecuación de *visibilidad*:

$$v_x^j = \frac{\frac{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(x, y)}{\text{card}(V_\gamma^j)}}{\frac{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(y, x)}{\text{card}(V_\gamma^j)}} = \frac{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(x, y)}{\sum_{y \in V_\gamma^j} \text{dist}_{\Gamma_K^j}(y, x)}$$

Ejemplo:

$$v_a^1 = \frac{0 + 1.67 + 1.33 + 2.50 + 3.00}{0 + 0.33 + 0.67 + 1.50 + 1.00} = \frac{2.13}{0.88} = 2.43$$

Así, se obtienen valores para cada uno de los nodos, como se muestra a continuación:

Nodo	<i>visibilidad</i>
<b>c</b>	<b>1.40</b>
<b>d</b>	<b>1.00</b>
<b>b</b>	<b>0.71</b>
<b>a</b>	<b>0.41</b>
<b>e</b>	<b>0.41</b>

**Tabla 2. Tabla con valores resultantes de *visibilidad* de la 1ª iteración.**

De igual forma al obtener la matriz resultante se puede obtener la *distancia conceptual* para cada uno de los tipos de relación en la conceptualización. Como se definió anteriormente, esta distancia se calcula como el promedio de los pesos  $\omega$  entre las

aristas que comparten el mismo tipo de relación, esto es  $[\delta^\rho]^j = \frac{\sum_{(a,b) \in \rho^*} \omega_{ab}^j}{\text{card}(\rho^*)}$  y

$[\bar{\delta}^\rho]^j = \frac{\sum_{(a,b) \in \rho^*} \omega_{ba}^j}{\text{card}(\rho^*)}$ , donde  $\rho^* = \{(a, b, \rho) \in A_G\}$  es el conjunto de aristas que representan una relación  $\rho$ .

Ejemplo:

$$[\delta^{is}]^1 = \frac{0.67 + 0.33 + 1.17 + 0.33}{4} = 0.63$$

$$[\bar{\delta}^{is}]^1 = \frac{1.33 + 1.67 + 0.83 + 1.67}{4} = 1.38$$

De esta manera se obtienen valores de *distancia conceptual* para cada tipo de relación en este ejemplo de forma directa e inversa, esto se muestra en las Tablas 2, 3, 4.

Tipo "is"			
$\delta$		$\bar{\delta}$	
c->a	0.67	a->c	1.33
b->a	0.33	a->b	1.67
d->b	1.17	b->d	0.83
e->c	0.33	c->e	1.67
DIS-C	0.63	DIS-C	1.38

Tabla 3. Tabla con valores resultantes de distancia conceptual para el tipo de relación "is" de la 1ª iteración.

Tipo "has"			
$\delta$		$\bar{\delta}$	
b->c	0.67	c->b	1.33
DIS-C	0.67	DIS-C	1.33

Tabla 4. Tabla con valores resultantes de distancia conceptual para el tipo de relación "has" de la 1ª iteración.

Tipo 2 "does"			
$\delta$		$\bar{\delta}$	
e->d	0.50	d->e	1.50
DIS-C	0.50	DIS-C	1.50

Tabla 5. Tabla con valores resultantes de distancia conceptual para el tipo de relación "does" de la 1ª iteración.

De los resultados anteriores en resumen se obtiene la siguiente:

Tipo	$\delta$	$\bar{\delta}$
"is"	0.63	1.38
"has"	0.67	1.33
"does"	0.5	1.5

Tabla 6. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K de la 1ª iteración.

Con el fin de optimizar los valores para las ponderaciones, se hace un refinamiento iterativo, logrando que el algoritmo *DIS-C* pueda automáticamente valorar la similitud quitando la subjetividad de las ponderaciones que define un usuario. Así el siguiente paso consiste en calcular el *error cuadrático medio*, que sirve para indicarle al algoritmo *DIS-C* en que momento debe terminar el proceso de refinamiento.

Para calcular el *error cuadrático medio* se definió la siguiente expresión:

$\frac{\sum_{x \in V_\gamma} (v_x^j - v_x^{j-1})^2}{\text{card}(V_\gamma)} \leq \epsilon_K$ ; donde  $\epsilon_K$  es un valor preestablecido (para este análisis se define con el valor de 0.0001) que indica el umbral de convergencia y es comparado con el valor del *error cuadrático medio*. Así también recordemos que el valor de  $v_x^0$  es 1. De este modo obtenemos lo siguiente:

$$\frac{\sum_{x \in V_\gamma} (v_x^j - v_x^{j-1})^2}{\text{card}(V_\gamma)} \leq \epsilon_K$$

$$\frac{(v_a^1 - v_a^{1-1})^2 + (v_b^1 - v_b^{1-1})^2 + (v_c^1 - v_c^{1-1})^2 + (v_d^1 - v_d^{1-1})^2 + (v_e^1 - v_e^{1-1})^2}{\text{card}(V_\gamma)} \leq \epsilon_K$$

$$\frac{(2.44 - 1)^2 + (0.72 - 1)^2 + (1.4 - 1)^2 + (1 - 1)^2 + (0.41 - 1)^2}{5} \leq 0.001$$

$$\mathbf{0.5257 \leq 0.001}$$

El resultado del cálculo del *error cuadrático medio* de la 1ª iteración es: 0.5257; este

resultado no cumple con la condición  $\frac{\sum_{x \in V_\gamma} (v_x^j - v_x^{j-1})^2}{\text{card}(V_\gamma)} \leq \epsilon_K$  por lo tanto los resultados

de *visibilidad* de la 1ª iteración serán utilizados nuevamente en las ecuaciones

$$\omega_{ab}^j = v_a^{j-1} \omega_a^o + [\delta^\rho]^{j-1} + \frac{1}{v_b^{j-1}} \omega_b^i \quad \text{y} \quad \omega_{ba}^j = v_b^{j-1} \omega_b^o + [\bar{\delta}^\rho]^{j-1} + \frac{1}{v_a^{j-1}} \omega_a^i \quad \text{.para} \quad \text{así}$$

recalcular la *distancia conceptual* entre conceptos y así obtener también nuevos

valores de *visibilidad*. Este procedimiento se realizara continuamente hasta llegar a

cumplir la condición de convergencia.



A continuación se muestran los resultados de las iteraciones 29, 59 y 82.

**Iteración 29**

$$\begin{array}{ll}
 \omega_{ab}^{29} = 144.46 & \omega_{cd}^{29} = 175.88 \\
 \omega_{ac}^{29} = 144.11 & \omega_{ce}^{29} = 142.89 \\
 \omega_{ad}^{29} = 282.83 & \\
 \omega_{ae}^{29} = 287.00 & \omega_{da}^{29} = 38.56 \\
 & \omega_{db}^{29} = 19.44 \\
 \omega_{ba}^{29} = 19.12 & \omega_{dc}^{29} = 36.45 \\
 \omega_{bc}^{29} = 17.01 & \omega_{de}^{29} = 139.02 \\
 \omega_{bd}^{29} = 138.37 & \\
 \omega_{be}^{29} = 159.91 & \omega_{ea}^{29} = 38.81 \\
 & \omega_{eb}^{29} = 52.42 \\
 \omega_{ca}^{29} = 19.92 & \omega_{ec}^{29} = 18.90 \\
 \omega_{cb}^{29} = 53.01 & \omega_{ed}^{29} = 32.99
 \end{array}$$

$$K = \begin{bmatrix}
 0 & 144.46 & 144.11 & 282.83 & 287.00 \\
 19.12 & 0 & 17.01 & 138.37 & 159.91 \\
 19.92 & 53.01 & 0 & 175.88 & 142.89 \\
 38.56 & 19.44 & 36.45 & 0 & 139.02 \\
 38.81 & 52.42 & 18.90 & 32.99 & 0
 \end{bmatrix}$$

Matriz 4. Matriz de adyacencia con valores de peso conceptual después de 29 iteraciones.

Nodo	visibilidad
c	1.81
b	1.24
d	0.37
e	0.20
a	0.14

Tabla 7. Tabla con valores resultantes de visibilidad después de 30 iteraciones.

Tipo	$\delta$	$\bar{\delta}$
"is"	19.34	142.46
"has"	17.01	53.01
"does"	32.99	139.02

Tabla 8. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K después de 30 iteraciones.

Resultado del cálculo del *error cuadrático medio* después de 30 iteraciones:

$$0.0008 \leq 0.0001$$

**Iteración 59**

$$\begin{array}{ll}
 \omega_{ab}^{59} = 334.22 & \omega_{ad}^{59} = 661.20 \\
 \omega_{ac}^{59} = 333.88 & \omega_{ae}^{59} = 665.63
 \end{array}$$

$$\begin{aligned}\omega_{ba}^{59} &= 38.06 \\ \omega_{bc}^{59} &= 35.21 \\ \omega_{bd}^{59} &= 326.98 \\ \omega_{be}^{59} &= 366.96\end{aligned}$$

$$\begin{aligned}\omega_{da}^{59} &= 76.38 \\ \omega_{db}^{59} &= 38.32 \\ \omega_{dc}^{59} &= 73.53 \\ \omega_{de}^{59} &= 303.31\end{aligned}$$

$$\begin{aligned}\omega_{ca}^{59} &= 38.86 \\ \omega_{cb}^{59} &= 105.25 \\ \omega_{cd}^{59} &= 407.65 \\ \omega_{ce}^{59} &= 331.75\end{aligned}$$

$$\begin{aligned}\omega_{ea}^{59} &= 76.67 \\ \omega_{eb}^{59} &= 114.21 \\ \omega_{ec}^{59} &= 37.81 \\ \omega_{ed}^{59} &= 75.90\end{aligned}$$

$$K = \begin{bmatrix} 0 & 334.22 & 333.88 & 661.20 & 665.63 \\ 38.06 & 0 & 35.21 & 326.98 & 366.96 \\ 38.86 & 105.25 & 0 & 407.65 & 331.75 \\ 76.38 & 38.32 & 73.53 & 0 & 303.31 \\ 76.67 & 114.21 & 37.81 & 75.90 & 0 \end{bmatrix}$$

Matriz 5. Matriz de adyacencia con valores de peso conceptual después de 59 iteraciones.

Nodo	visibilidad
c	1.84
b	1.30
d	0.33
e	0.18
a	0.12

Tabla 9. Tabla con valores resultantes de visibilidad después de 60 iteraciones.

Tipo	$\delta$	$\bar{\delta}$
"is"	38.26	331.71
"has"	35.21	105.25
"does"	75.90	303.31

Tabla 10. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K después de 60 iteraciones.

Resultado del cálculo del *error cuadrático medio* después de 60 iteraciones:

$$0.0002 < 0.0001$$

### Iteración 82

$$\begin{aligned}\omega_{ab}^{82} &= 492.06 \\ \omega_{ac}^{82} &= 491.73 \\ \omega_{ad}^{82} &= 976.35 \\ \omega_{ae}^{82} &= 980.89\end{aligned}$$

$$\omega_{be}^{82} = 538.55$$

$$\omega_{ca}^{82} = 53.36$$

$$\omega_{cb}^{82} = 145.28$$

$$\omega_{cd}^{82} = 600.25$$

$$\omega_{ce}^{82} = 489.16$$

$$\omega_{ba}^{82} = 52.56$$

$$\omega_{bc}^{82} = 49.39$$

$$\omega_{bd}^{82} = 484.29$$

$$\omega_{da}^{82} = 105.35$$

$$\begin{aligned}\omega_{db}^{82} &= 52.79 \\ \omega_{dc}^{82} &= 102.18 \\ \omega_{de}^{82} &= 434.97\end{aligned}$$

$$\begin{aligned}\omega_{eb}^{82} &= 163.88 \\ \omega_{ec}^{82} &= 52.30 \\ \omega_{ed}^{82} &= 111.09\end{aligned}$$

$$\omega_{ea}^{82} = 105.66$$

$$K = \begin{bmatrix} 0 & 492.06 & 491.73 & 976.35 & 980.89 \\ 52.56 & 0 & 49.39 & 484.29 & 538.55 \\ 53.36 & 145.28 & 0 & 600.25 & 489.16 \\ 105.35 & 52.79 & 102.18 & 0 & 434.97 \\ 105.66 & 163.88 & 52.30 & 111.09 & 0 \end{bmatrix}$$

Matriz 6. Matriz de adyacencia con valores de distancia conceptual después de 82 iteraciones.

Nodo	visibilidad
c	1.85
b	1.32
d	0.32
e	0.18
a	0.11

Tabla 11. Tabla con valores resultantes de visibilidad después de 83 iteraciones.

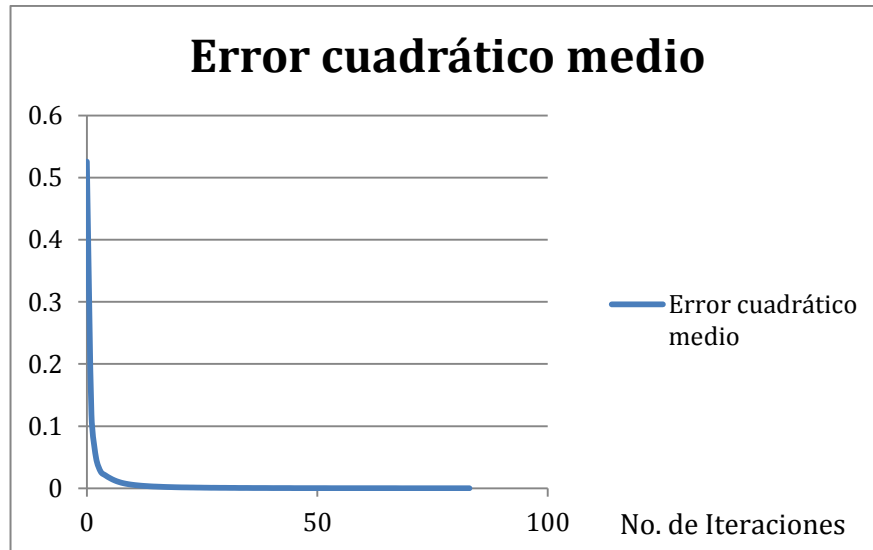
Tipo	$\delta$	$\bar{\delta}$
"is"	52.75	489.31
"has"	49.39	145.28
"does"	111.09	434.97

Tabla 12. Tabla con valores resultantes de distancia conceptual para cada uno de los tipos de relaciones subyacentes en la conceptualización K después de 83 iteraciones.

Resultado del cálculo del *error cuadrático medio* después de 83 iteraciones:

$$0.0001 \leq 0.0001$$

La Gráfica 1 indica cómo al paso del número de iteraciones, el *error cuadrático medio* disminuye, indicando que la estimación del valor de la *visibilidad* y las distancias conceptuales converge a un valor estable.



Gráfica 1. Cálculo del *error cuadrático medio*.

### 3.3 Análisis de resultados

Como se mencionó, la métrica de *visibilidad* permite ordenar los conceptos de acuerdo a su asequibilidad y accesibilidad en la ontología. Tomando como referencia la Tabla 11, se presentan cada uno de los conceptos ordenados de mayor a menor basándose en los valores de *visibilidad* obtenidos.

Nodo	<i>visibilidad</i>
<b>c</b>	<b>1.85</b>
<b>b</b>	<b>1.32</b>
<b>d</b>	<b>0.32</b>
<b>e</b>	<b>0.18</b>
<b>a</b>	<b>0.11</b>

Así, se puede apreciar que el concepto *c* resulta ser el concepto principal con respecto a los demás, siendo este el punto de referencia principal para poder vincular a los demás conceptos.

Para el caso del análisis de *distancia conceptual*, se obtienen los valores de cada uno de los tipos de relación que subyacen dentro una conceptualización en particular, interpretándose como la proximidad conceptualmente hablando que tiene un concepto que está directamente relacionado con otro concepto, implicando que no necesariamente ambos conceptos tengan que ser similares uno con respecto a otro.

En el siguiente capítulo se presentan resultados con ontologías que permiten apreciar estas propiedades.



# 4 CAPÍTULO IV

## PRUEBAS

### EXPERIMENTALES Y

### RESULTADOS

En este capítulo se muestran las pruebas experimentales, forma de implementación del algoritmo, resultados obtenidos, análisis de resultados y una comparación de resultados entre los algoritmos *DIS-C* y *Confusion*.

#### 4.1 Resultados y pruebas experimentales del método propuesto

El algoritmo fue implementado en lenguaje de Java donde se realizaron pruebas experimentales con 7 ontologías de las cuales 6 se presentan en los anexos. También se utilizó un programa llamado *Graphviz* ([www.Graphviz.org](http://www.Graphviz.org)) para poder visualizar las ontologías y los grafos resultantes. A continuación se presenta el código (véase Código 1) fuente de la forma que se ingresan los datos y las características para visualizar un grafo en el programa *Graphviz*.

```
digraph abstract {
node[style=filled, color=blue4,fillcolor=beige];
edge[color=black, arrowhead=normal]; egresado -> cosa
edge[color=blue, arrowhead=odot]; egresado -> dato_laboral
edge[color=blue, arrowhead=odot]; egresado -> dato_personal
edge[color=blue, arrowhead=odot]; egresado -> dato_académico
edge[color=black, arrowhead=normal]; dato_laboral -> cosa
edge[color=black, arrowhead=normal]; dato_personal -> cosa
edge[color=black, arrowhead=normal]; dato_académico -> cosa
edge[color=black, arrowhead=normal]; escuela -> dato_académico
edge[color=black, arrowhead=normal]; cédula_profesional -> dato_académico
edge[color=black, arrowhead=normal]; grado -> dato_académico
edge[color=black, arrowhead=normal]; facultad -> dato_académico
edge[color=black, arrowhead=normal]; universidad -> dato_académico
```

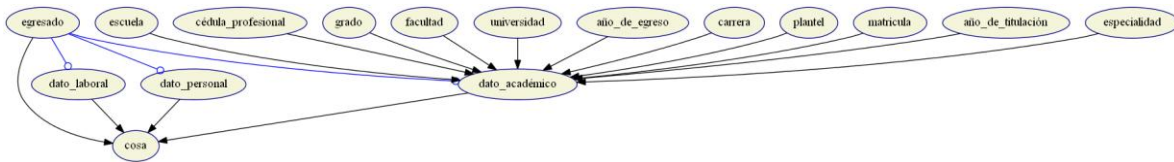
```

edge[color=black, arrowhead=normal]; año_de_egreso -> dato_académico
edge[color=black, arrowhead=normal]; carrera -> dato_académico
edge[color=black, arrowhead=normal]; plantel -> dato_académico
edge[color=black, arrowhead=normal]; matricula -> dato_académico
edge[color=black, arrowhead=normal]; año_de_titulación -> dato_académico
edge[color=black, arrowhead=normal]; especialidad -> dato_académico
}

```

**Código 1.** Código para visualizar un grafo utilizando el programa *Graphviz*.

Una vez ingresados los datos se ejecuta el código fuente y el programa *Graphviz* muestra la ontología o grafo (véase Figura 25).



**Figura 25.** Ontología (Cardona, 2010) visualizada en *Graphviz*. Utilizando como referencia el Código 1.

Ahora utilizando la Figura 25 como referencia se pueden ingresar los datos en el programa *DIS-C* donde la función *g.addEdge()* tiene como entrada cuatro parámetros. Para ejemplificar este procedimiento tomaremos como referencia las Figuras Figura 26, Figura 27, Figura 28 mencionadas en el capítulo 3.



**Figura 26.** *a(is)b*, Implica que “a” es del mismo tipo que “b”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011)



**Figura 27.** *a(has)b*, Implica que “b” es una propiedad de “a” o que “a” tiene una propiedad llamada “b”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011)



**Figura 28.** *a(does)b*, Implica que “b” es una habilidad de “a”. (Torres, Quintero, Moreno-Ibarra, Guzmán, & Menchaca-Méndez, 2011).

De las figuras anteriores se utilizará primero como ejemplo la Figura 26 para ingresar los datos de dos conceptos con su relación directa y su tipo de relación.

```

g.addEdge("a","b","1.0","0.0")

```



Donde "a" representa el nodo origen y "b" representa el nodo destino, el siguiente parámetro es "1.0", que indica la relación entre estos dos nodos y por último la tenemos el parámetro "0.0", que indica el tipo de relación entre los nodos "a" y "b".

Para las Figura 27, Figura 28 tenemos lo siguiente: `g.addEdge("a","b","1.0","1.0")` y `g.addEdge("a","b","1.0","2.0")`, se puede apreciar que el último parámetro cambia de valor; pues "0.0" corresponde al tipo de relación "is", "1.0" corresponde al tipo de relación "has" y por último "2.0" corresponde al tipo de relación "does".

En el siguiente código (véase Código 2) se muestra el ingreso de la ontología de la Figura 25 en programa *DIS-C*

```
g.addEdge("egresado", "cosa", 1.0, 0);
g.addEdge("egresado", "dato_laboral", 1.0, 1);
g.addEdge("egresado", "dato_personal", 1.0, 1);
g.addEdge("egresado", "dato_académico", 1.0, 1);
g.addEdge("dato_laboral", "cosa", 1.0, 0);
g.addEdge("dato_personal", "cosa", 1.0, 0);
g.addEdge("dato_académico", "cosa", 1.0, 0);
g.addEdge("escuela", "dato_académico", 1.0, 0);
g.addEdge("cédula_profesional", "dato_académico", 1.0, 0);
g.addEdge("grado", "dato_académico", 1.0, 0);
g.addEdge("facultad", "dato_académico", 1.0, 0);
g.addEdge("universidad", "dato_académico", 1.0, 0);
g.addEdge("año_de_egreso", "dato_académico", 1.0, 0);
g.addEdge("carrera", "dato_académico", 1.0, 0);
g.addEdge("plantel", "dato_académico", 1.0, 0);
g.addEdge("matricula", "dato_académico", 1.0, 0);
g.addEdge("año_de_titulación", "dato_académico", 1.0, 0);
g.addEdge("especialidad", "dato_académico", 1.0, 0);
```

**Código 2. Código donde se ejemplifica la forma de ingresar los datos de una ontología en el programa.**

En la siguiente parte se presenta la implementación de los pesos  $\omega_x^i$  y  $\omega_x^o$  que representan los costos de "entrar" y "salir" de los conceptos en la ontología (Figura 25). Cabe mencionar que estas ecuaciones es invariante entre todas las iteraciones, por lo que sólo se calcula al inicio.

En el Código 3 se muestra un fragmento de la implementación de los pesos  $\omega_x^i$  y  $\omega_x^o$  (en código "w\_ia" y "w\_oa") de los conceptos directamente relacionados.

```
double i_a = g.getInputs(name),
       o_a = g.getOutputs(name),
```

```

w_ia = o_a / (i_a + o_a),
w_oa = i_a / (i_a + o_a);

```

**Código 3. Código donde se presenta la implementación de los costos de “entrar” y “salir” de los conceptos directamente relacionados en la ontología**

A continuación se muestra la implementación del algoritmo *Floyd-Warshall*. En la Figura 25 se observa que “universidad” está relacionado con “dato\_académico” por lo tanto su valor de peso conceptual de “entrar” y “salir” es calculado con las ecuaciones  $\omega_x^i$  y  $\omega_x^o$ . En el Código 4 se muestra la implementación de este algoritmo en *DIS-C*. Obteniéndose como resultado la matriz de distancia mostrada en la Matriz 7.

```

public double[][] cm_FloydWarshal() {
    int size = nodos.size();
    double [][] dd = new double[size][size];
    double [][] d = new double[size][size], tmp;

    for(int i=0; i<size; i++) {
        for(int j=0; j<size; j++) {
            d[i][j] = i==j?0:Double.POSITIVE_INFINITY;
        }
    }

    for(int i=0; i<aristas.size(); i++) {
        Edge a = aristas.elementAt(i);
        d[a.a.id][a.b.id] = a.w;
    }

    for(int k=0; k<size; k++) {
        for(int i=0; i<size; i++) {
            for(int j=0; j<size; j++) {
                dd[i][j] = MIN(d[i][j], d[i][k]+d[k][j]);
            }
        }
        tmp = d; d = dd; dd = tmp;
    }

    return d;
}

```

**Código 4. Código donde se presenta la implementación del algoritmo *Floyd-Warshall*.**

Como se mencionó anteriormente, el algoritmo *Floyd-Warshall* nos ayuda a determinar los pesos conceptuales de aquellos conceptos que no están directamente relacionas. Estos pesos obtenidos son tanto para en un sentido como en su sentido inverso respecto al par de nodos que se está analizando, por ejemplo, “escuela” y “egresado”, su peso conceptual en un sentido  $DIS - C(escuela, egresado) = 190.27$  y

para su sentido inverso  $DIS - C(egresado, escuela) = 442.84$ . Esto se puede observar en la Matriz 7.

	egresado	cosa	dato laboral	dato personal	dato académico	escuela	cédula profesional	grado	facultad	universidad	año de egreso	carrera	plantel	matrícula	año de titulación	especialidad
egresado	0.00	24.67	9.47	9.47	9.27	442.84	442.84	442.84	442.84	442.84	442.84	442.84	442.84	442.84	442.84	442.84
cosa	444.46	0.00	444.05	444.05	443.86	877.42	877.42	877.42	877.42	877.42	877.42	877.42	877.42	877.42	877.42	877.42
dato laboral	162.19	25.87	0.00	171.66	171.46	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03
dato personal	162.19	25.87	171.66	0.00	171.46	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03	605.03
dato académico	165.58	29.26	175.05	175.05	0.00	433.56	433.56	433.56	433.56	433.56	433.56	433.56	433.56	433.56	433.56	433.56
escuela	190.27	53.95	199.74	199.74	24.69	0.00	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25
cédula profesional	190.27	53.95	199.74	199.74	24.69	458.25	0.00	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25
grado	190.27	53.95	199.74	199.74	24.69	458.25	458.25	0.00	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25
facultad	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	0.00	458.25	458.25	458.25	458.25	458.25	458.25	458.25
universidad	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	0.00	458.25	458.25	458.25	458.25	458.25	458.25
año de egreso	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	458.25	0.00	458.25	458.25	458.25	458.25	458.25
carrera	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	458.25	458.25	0.00	458.25	458.25	458.25	458.25
plantel	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	458.25	458.25	458.25	0.00	458.25	458.25	458.25
matrícula	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	0.00	458.25	458.25
año de titulación	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	0.00	458.25
especialidad	190.27	53.95	199.74	199.74	24.69	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	458.25	0.00

**Matriz 7. Matriz de adyacencia resultante una vez implementado el algoritmo Floyd Warshall. Los valores resultantes pertenecen a la ontología de la Figura 25 después de 55 iteraciones.**

Ahora bien, una vez obtenidos los resultados de la matriz de adyacencia (véase Matriz 7) correspondientes de la Figura 25, el programa *DIS-C* muestra en pantalla el código fuente (véase Código 5) para poder presentar mediante el uso del programa *Graphviz* los resultados de los pesos conceptuales de los conceptos relacionados y los pesos conceptuales de su relación inversa respectivamente en un grafo (Figura 29)

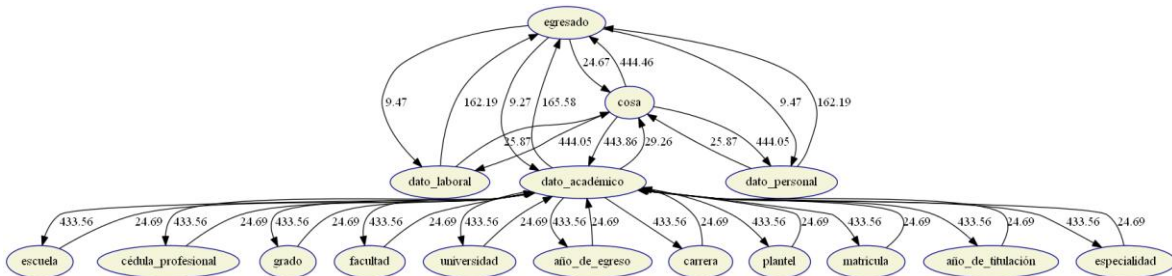
```
digraph abstract {
node[style=filled, color=blue4,fillcolor=beige];
edge[color=black, arrowhead=normal]; egresado -> cosa
edge[color=black, arrowhead=normal]; cosa -> egresado
edge[color=blue, arrowhead=odot]; egresado -> dato_laboral
edge[color=blue, arrowhead=odot]; dato_laboral -> egresado
edge[color=blue, arrowhead=odot]; egresado -> dato_personal
edge[color=blue, arrowhead=odot]; dato_personal -> egresado
edge[color=blue, arrowhead=odot]; egresado -> dato_académico
edge[color=blue, arrowhead=odot]; dato_académico -> egresado
edge[color=black, arrowhead=normal]; dato_laboral -> cosa
edge[color=black, arrowhead=normal]; cosa -> dato_laboral
edge[color=black, arrowhead=normal]; dato_personal -> cosa
edge[color=black, arrowhead=normal]; cosa -> dato_personal
edge[color=black, arrowhead=normal]; dato_académico -> cosa
edge[color=black, arrowhead=normal]; cosa -> dato_académico
edge[color=black, arrowhead=normal]; escuela -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> escuela
edge[color=black, arrowhead=normal]; cédula_profesional -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> cédula_profesional
edge[color=black, arrowhead=normal]; grado -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> grado
```

```

edge[color=black, arrowhead=normal]; facultad -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> facultad
edge[color=black, arrowhead=normal]; universidad -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> universidad
edge[color=black, arrowhead=normal]; año_de_egreso -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> año_de_egreso
edge[color=black, arrowhead=normal]; carrera -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> carrera
edge[color=black, arrowhead=normal]; plantel -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> plantel
edge[color=black, arrowhead=normal]; matricula -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> matricula
edge[color=black, arrowhead=normal]; año_de_titulación -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> año_de_titulación
edge[color=black, arrowhead=normal]; especialidad -> dato_académico
edge[color=black, arrowhead=normal]; dato_académico -> especialidad
}

```

**Código 5. Código para diseñar una ontología resultante utilizando el programa *Graphviz*. Tomando como referencia los resultados obtenidos en la Matriz 7.**



**Figura 29. Ontología de la Figura 25 con valores resultantes de peso conceptual, diseñada en *Graphviz*. Utilizando como referencia el Código 5 para su diseño.**

En la siguiente parte se muestra la implementación del *error cuadrático medio* y la ecuación de *visibilidad*  $v_x^j$  y su inverso que es el *ocultamiento*  $\frac{1}{v_x^j}$ .

El cálculo del valor de *visibilidad* de un concepto, se analiza una vez efectuado el análisis en el algoritmo *Floyd-Warshall* en donde se obtiene la matriz resultante con los valores de peso conceptual. Estos valores de la matriz son utilizados calculando el promedio de los pesos conceptuales de “salida” concepto en donde se está efectuando el análisis entre el promedio de los pesos conceptuales de “entrada” del mismo nodo o concepto.

Una vez obtenidos los valores de *visibilidad* para todos los nodos o conceptos se puede obtener el *error cuadrático medio* entre los valores de la *visibilidad* de una iteración y los valores de *visibilidad* de la iteración anterior. El *error cuadrático medio* compara

estos valores de *visibilidad* de tal forma que si ambos valores son casi iguales con una diferencia mínima (esta diferencia es indicada con el umbral de convergencia, 0.0001 en este ejemplo) significa que estos valores son estables. Por lo tanto se puede interpretar que estos valores son los valores definitivos para la conceptualización que se está analizando.

Esta implementación se encuentra en el código siguiente.

```
double eps = 1000;
    double eps_K = 1e-04;

int i = 0;
    for (; i<1000 && eps>eps_K; i++ ) {

while(it.hasNext()) {
    String name = it.next();
    Node n = nodos.get(name);
    double from = 0.0, to = 0.0;
    for( int i=0; i<size; i++) {
        from += table[n.id][i];
        to += table[i][n.id];
    }
    double v_a = ((Double)n.data.get("v_a")).doubleValue();
    eps += sqr(v_a - (from / to));
    n.data.put("v_a", new Double(from/to));
    n.data.put("h_a", new Double(to/from));
}
    System.out.printf("Error cuadrático medio: %10.4f\n", eps/size);
    return eps / size;
}
}
```

**Código 6. Código donde se presenta la implementación del cálculo de *visibilidad*, *error cuadrático medio* y el umbral de convergencia.**

En la Figura 30 y la Figura 31 se muestran los resultados de obtenidos después de efectuado el análisis del *error cuadrático medio* y *visibilidad*.

```
Error cuadrático medio:      0.0001
Iteración: 55
```

**Figura 30. Resultados de la implementación y análisis del *error cuadrático medio* después de 55 iteraciones.**

```
v_año_de_egreso: 0.6959475176760447
v_año_de_titulación: 0.6959475176760447
v_carrera: 0.6959475176760447
v_cosa: 0.061318510514272465
v_cédula_profesional: 0.6959475176760447
v_dato_académico: 0.20109540552175367
v_dato_laboral: 0.41768377331251005
```

```

v_dato_personal: 0.41768377331251005
v_egresado: 1.623281738887124
v_escuela: 0.6959475176760447
v_especialidad: 0.6959475176760447
v_facultad: 0.6959475176760447
v_grado: 0.6959475176760447
v_matricula: 0.6959475176760447
v_plantel: 0.6959475176760447
v_universidad: 0.6959475176760447

```

**Figura 31. Resultados de la implementación y análisis de *visibilidad* en la iteración 55, después haberse cumplido la condición del umbral de convergencia en el algoritmo *DIS-C*.**

En la siguiente parte se muestra la implementación de las ecuaciones  $[\delta^\rho]^j = \frac{\sum_{(a,b) \in \rho^*} \omega_{ab}^j}{\text{card}(\rho^*)}$  y  $[\bar{\delta}^\rho]^j = \frac{\sum_{(a,b) \in \rho^*} \omega_{ba}^j}{\text{card}(\rho^*)}$  correspondientes al análisis de la *distancia conceptual*, el uso de estas ecuaciones parte de los resultados obtenidos por el algoritmo *Floyd-Warshall* que como se mencionó estos resultados son representados en una matriz de pesos (Matriz 7). Este análisis finalmente otorga un valor que describe el *distancia conceptual* para cada tipo de relación (directa e inversa, respectivamente) en una conceptualización en particular. Esto, finalmente, se calcula obteniendo el promedio de los pesos conceptuales entre las aristas que comparten el mismo tipo de relación en un sentido y por otro lado obteniendo el promedio de los pesos conceptuales entre las aristas que comparten el mismo tipo de relación en su sentido inverso, respectivamente. Lo anterior mencionado se representa en el código que se presenta a continuación.

```

for(Edge ont:g.aristas) {
    tipo_hist[ont.tipo]++;
    for(Edge gra:this.aristas) {
        if( ont.a.id==gra.a.id && ont.b.id==gra.b.id ) {
            tipo_sum_ida[ont.tipo]+=gra.w;
        } else if( ont.a.id==gra.b.id && ont.b.id==gra.a.id) {
            tipo_sum_vuelta[ont.tipo]+=gra.w;
        }
    }
}

for(int i=0; i<max_tipo; i++) {
    tipo_sum_ida[i] = tipo_sum_ida[i] / tipo_hist[i];
    tipo_sum_vuelta[i]= tipo_sum_vuelta[i] / tipo_hist[i];
}

```

**Código 7. Código donde se presenta la implementación del cálculo de la distancia conceptual.**

Finalmente, en la Figura 32 se muestran los resultados obtenidos de la implementación de *distancia conceptual*. Se puede observar que para el caso de la ontología representada en la Figura 25 existen dos tipos de relaciones la cual es una relación de tipo “0” y la otra de tipo “1”, de cual la *distancia conceptual* resultante en la relación tipo “0” es de 24.67 para su forma directa y 427.54 en su forma inversa y para la relación tipo “1” es de 9.26 para su forma directa y 160.38 en su forma inversa.

Tipo: 0, Directa:	24.67,	Inversa:	427.54
Tipo: 1, Directa:	9.26,	Inversa:	160.38

**Figura 32. Resultados de la implementación y análisis de distancia conceptual para solo un tipo de relación subyacente en la conceptualización de la Figura 25 después de 7 iteraciones.**

#### 4.1.1 Análisis de resultados

Para definir la proximidad entre conceptos de la ontología tomamos como referencia la *distancia conceptual* donde el valor 9.47 (véase Matriz 7) es el valor mínimo de *distancia conceptual* en la ontología (véase Figura 25). Este valor indica que conceptualmente estos conceptos están muy relacionados, ejemplo:

$$DIS - C(\text{egresado}, \text{dato laboral}) = 9.47$$

$$DIS - C(\text{egresado}, \text{dato personal}) = 9.47$$

$$DIS - C(\text{egresado}, \text{dato académico}) = 9.47$$

por lo tanto los demás valores de *distancia conceptual* son mayores a este valor. Ahora tenemos que el valor de *distancia conceptual* máximo en la ontología (véase Figura 25) encontramos 877.42 (véase Matriz 7). Este valor indica que conceptualmente estos conceptos están muy pocos relacionados, ejemplo:

$$DIS - C(\text{cosa}, \text{escuela}) = 877.42$$

$$DIS - C(\text{cosa}, \text{cédula profesional}) = 877.42$$

$$DIS - C(\text{cosa}, \text{especialidad}) = 877.42$$

Para el análisis de los resultados de *visibilidad* encontramos que *egresados* (véase Figura 31) es el concepto más asequible con respecto a los demás conceptos, donde este se puede tomar como referencia para poder llegar a los demás conceptos. Por ejemplo: si se quiere referir “*dato académico*” resulta más accesible, primero tomar como referencia “*egresados*” por lo que “*egresados*” otorga una mayor información concreta para llegar al concepto “*dato académico*” Ahora para el caso del concepto con

el menor valor de visibilidad se encuentra “cosa” (véase Figura 31) indicando que este concepto es que menor información concreta otorga, para llegar a los demás conceptos de la ontología.

Ahora para el análisis de los resultados de *distancia conceptual* para cada tipo de relación (véase Figura 32) se puede interpretar que la proximidad entre dos conceptos es mayor si pertenecen al tipo de relación “1” que al tipo de relación “0”. Por otra parte se observa que el tipo de relaciones en su forma directa, el valor numérico es menor que para su forma inversa. Indicando que el costo de salida y entrada en forma directa de un concepto a otro es menor, con respecto al costo de salida y entrada en su forma inversa.

#### 4.2 Algoritmo DIS-C vs algoritmo confusion

Cómo se ha mencionado el algoritmo DIS-C calcula la *distancia conceptual* entre conceptos utilizando un modelo de red. Por lo que para comparar sus resultados se debe comparar con un método de similares características, es decir un método que utilice el modelo de red, como es el caso de *confusion* (Levachkine & Guzmán-Arenas, 2007). En esta subsección se muestra una comparación de resultados del algoritmo DIS-C con respecto al algoritmo *confusion*. La comparación se realiza tomando como referencia los resultados presentados en (Levachkine & Guzmán-Arenas, 2007), en donde se utiliza una jerarquía de seres vivos (véase Figura 33). Y se obtienen los resultados de la tabla (véase Tabla 13).

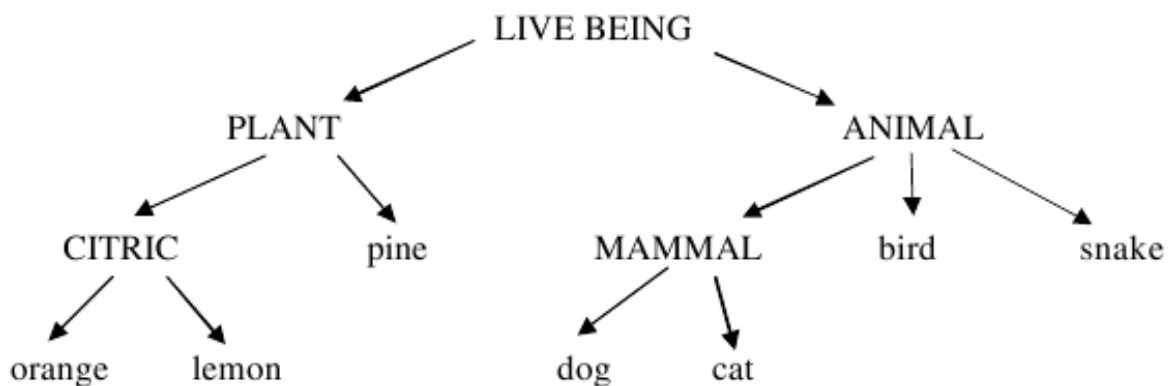


Figura 33. Jerarquía de seres vivos.



↓	→	s											
r	<i>confusion</i>	live being	animal	plant	mammal	bird	snake	citric	pine	cat	dog	lemon	orange
	live being	0	1	1	2	2	2	2	2	3	3	3	3
	animal	0	0	1	1	1	1	2	2	2	2	3	3
	plant	0	1	0	2	2	2	1	1	3	3	2	2
	mammal	0	0	1	0	1	1	2	2	1	1	3	3
	bird	0	0	1	1	0	1	2	2	2	2	3	3
	snake	0	0	1	1	1	0	2	2	2	2	3	3
	citric	0	1	0	2	2	2	0	1	3	3	1	1
	pine	0	1	0	2	2	2	1	0	3	3	2	2
	cat	0	0	1	0	1	1	2	2	0	1	3	3
	dog	0	0	1	0	1	1	2	2	1	0	3	3
	lemon	0	1	0	2	2	2	0	1	3	3	0	1
	orange	0	1	0	2	2	2	0	1	3	3	1	0

Tabla 13. Tabla con valores resultantes de la evaluación de *confusion* de la jerarquía de seres vivos.

La jerarquía de seres vivos (véase Figura 33) es ingresada al algoritmo DIS-C para así evaluarla, y comparar ambos métodos. En la tabla (véase Tabla 14) se muestran los resultados obtenidos al utilizar el algoritmo DIS-C.

	live being	plant	citric	orange	lemon	pine	animal	mammal	dog	cat	bird	snake
live being	0.00	4.25	8.29	11.77	11.77	7.61	4.46	8.45	11.93	11.93	7.76	7.76
plant	16.46	0.00	4.04	7.52	7.52	3.36	20.92	24.91	28.39	28.39	24.22	24.22
citric	30.39	13.93	0.00	3.48	3.48	17.29	34.86	38.84	42.32	42.32	38.16	38.16
orange	44.97	28.51	14.57	0.00	18.05	31.86	49.43	53.41	56.89	56.89	52.73	52.73
lemon	44.97	28.51	14.57	18.05	0.00	31.86	49.43	53.41	56.89	56.89	52.73	52.73
pine	31.11	14.64	18.69	22.17	22.17	0.00	35.57	39.55	43.03	43.03	38.87	38.87
animal	16.40	20.65	24.69	28.17	28.17	24.00	0.00	3.98	7.46	7.46	3.30	3.30
mammal	30.55	34.80	38.84	42.32	42.32	38.16	14.15	0.00	3.48	3.48	17.45	17.45
dog	45.15	49.40	53.44	56.92	56.92	52.75	28.75	14.60	0.00	18.08	32.05	32.05
cat	45.15	49.40	53.44	56.92	56.92	52.75	28.75	14.60	18.08	0.00	32.05	32.05
bird	31.27	35.51	39.55	43.03	43.03	38.87	14.87	18.85	22.33	22.33	0.00	18.16
snake	31.27	35.51	39.55	43.03	43.03	38.87	14.87	18.85	22.33	22.33	18.16	0.00

Tabla 14. Tabla con valores resultantes de la evaluación del algoritmo *DIS-C* de la jerarquía de seres vivos.

```

v_animal: 1.7459220475766823
v_bird: 0.9629540684601188
v_cat: 0.7165725591098432
v_citric: 1.025957900705567
v_dog: 0.7165725591098432
v_lemon: 0.7297199488298107
v_live_being: 0.26746474568368833
v_mammal: 1.0279613714148201
v_orange: 0.7297199488298107
v_pine: 0.9651967666093292
v_plant: 1.6362701162608917
v_snake: 0.9629540684601188

```

Figura 34. Valores resultantes de *visibilidad* de la jerarquía de seres vivos.

Al analizar las comparaciones encontramos lo siguiente:

- El algoritmo *DIS-C*, agrega mayor detalle a la *distancia conceptual*, al utilizar valores del conjunto de los números reales. Por otra parte, *confusion* en este caso presenta valores en el conjunto de los números enteros.
- El algoritmo *DIS-C* otorga una ponderación para cada tipo de relación, mientras que *confusion* en el caso de jerarquías simples utiliza exclusivamente un tipo de relación
- El algoritmo *DIS-C* no se limita al análisis de estructuras jerárquicas a diferencia *confusion*.
- El algoritmo *DIS-C* otorga valores de *distancia conceptual* que no son simétricos, ejemplo:

$$DIS - C(\text{live\_being}, \text{plant}) = 4.25; DIS - C(\text{plant}, \text{live\_being}) = 16.46.$$

A diferencia del algoritmo *confusion* que en ciertos casos otorga valores simétricos, ejemplo:

$$\text{conf}(\text{orange}, \text{cat}) = 3; \text{conf}(\text{cat}, \text{orange}) = 3.$$

- El algoritmo *DIS-C* presenta un valor de visibilidad donde indica que el concepto más asequible de la jerarquía.
- Ambos algoritmos dependen de la propia estructura de la jerarquía para realizar su análisis, por lo que son *dependientes del contexto*.

- Encontramos que para *confusion* las relaciones  $conf(planta, animal) = 1$  y  $conf(animal, planta) = 1$ , para ambos casos es igual a 1. Mientras que para *DIS-C* la relación  $DIS - C(planta, animal) = 20.92$  y para la relación  $DIS - C(animal, planta) = 20.65$ . En el último se puede apreciar que si redondeamos los valores de 20.90 y 20.65 obtendríamos valores similares, de la misma forma que en *confusion* ambos valores de las relaciones son similares. Cabe mencionar que en la diferencia en la cifras decimales para la relación  $DIS - C(planta, animal)$  es mayor con respecto a  $DIS - C(animal, planta)$ , por lo que  $DIS - C(planta, animal)$ , el concepto *animal* tiene mayor información heredada; *mammal, bird, snake, dog, cat*.
- Ahora encontramos que para *confusion* la relaciones en la parte baja de la jerarquía  $conf(orange, lemon)$  y  $conf(lemon, orange)$  tienen el valor simétrico para ambos casos. Y encontramos que para *DIS-C* en la misma combinación de relaciones también tienen un valor simétrico esto por que como ambos conceptos forman la parte final de jerarquía no hay más información que se derive de estos. El mismo caso ocurre para las relaciones  $(dog, cat)$  y  $(cat, dog)$ .
- Para ambos casos los valores obtenidos de *distancia conceptual* su interpretación radica en tomar como referencia el valor mínimo de entre todas las distancias conceptuales de la matriz de resultante, considerando ese valor mínimo como la distancia más corta conceptualmente hablando y el valor máximo para la distancia más lejana.
- Encontramos que las relaciones  $(animal, bird)$  y  $(animal, snake)$  tanto para *confusión* como para *DIS-C* obtienen el valor mínimo de *distancia conceptual* en toda la jerarquía. Y para el caso de  $(dog, orange)$ ,  $(dog, lemon)$ ,  $(cat, orange)$ ,  $(cat, lemon)$  obtenemos los valores más altos tanto para *confusion* como para *DIS-C* de *distancia conceptual* en toda la jerarquía. Indicando que el algoritmo *DIS-C* es consistente con sus resultados.



# 5 CAPÍTULO V

## CONCLUSIONES Y

## TRABAJO FUTURO

### 5.1 Conclusiones

En este trabajo de tesis utilizó como referencia el algoritmo *DIS-C*, el cual evalúa la *distancia conceptual* entre conceptos de una ontología con base en la ponderación de las relaciones entre los conceptos por parte del usuario. Lo que agrega cierta subjetividad a la evaluación.

Como parte de la propuesta se realizó una modificación al algoritmo *DIS-C*, enfocado en ponderar automáticamente las relaciones conceptuales, permitiendo eliminar la subjetividad del criterio del usuario para este tipo de análisis, que a su vez, esta misma automatización ajusta las ponderaciones hasta alcanzar un estado óptimo definido por un umbral. Dentro de este análisis se logró el objetivo de diseñar e implementar una métrica, denominada *visibilidad*, para evaluar la asequibilidad de cada concepto en la ontología. Este enfoque permitió utilizarlo en ontologías que no necesariamente posean una estructura jerárquica, realizándose pruebas en diversos tipos de ontologías. Además se realizó una comparación contra el algoritmo de *confusion* lo que permitió analizar diferencias y similitudes entre los métodos. Por otra parte se encontró que este tipo de algoritmo puede ser utilizado en aplicaciones de la recuperación de información, sistemas de recomendación o la integración de datos.

Como algo importante de mencionar, los resultados obtenidos en el algoritmo *DIS-C*, no se pueden definir si son buenos o son malos o si definen la proximidad de un concepto con respecto a otro de forma absoluta. Puesto que en un principio no se

puede definir si una ontología es correcta o incorrecta. Se puede decir que solo son una referencia que se fundamenta en el uso diversos métodos aplicados sobre la propia estructura de la ontología.

## 5.2 Contribuciones

Si bien en el estado del arte solo se reporta a (Rada, Mili, Bicknell, & Blettner, 1989) donde se propone una métrica para calcular la *distancia conceptual* entre dos conceptos dentro de una ontología. Esta medida fue diseñada principalmente para las ontologías con relaciones taxonómicas. Midiendo entre dos nodos o un conjunto de nodos las distancia mínima entre cada par de nodos relacionados. Por otra parte el algoritmo *DIS-C*, es un algoritmo que evalúa la *distancia conceptual* entre conceptos de una ontología con base en la ponderación de las relaciones entre los conceptos por parte del usuario. Lo que agrega cierta subjetividad a la evaluación.

De lo anterior la contribución de este trabajo reside en la obtención de una métrica de evaluación para la *distancia conceptual* y una métrica llamada *visibilidad* esta última permite evaluar la asequibilidad de los conceptos de la ontología; ambas métricas trabajan de forma automática dentro un algoritmo llamado *DIS-C*, el cual se automatizó en este trabajo. Este algoritmo funciona utilizando la información de la propia estructura de la ontología, lo que permite utilizarlo en ontologías que no necesariamente posean una estructura taxonómica.

## 5.3 Trabajo futuro

Se pretende explorar la opción de permitir que el usuario pueda predefinir la ponderación del peso inicial de las relaciones; es decir, que en lugar de que el valor inicial de  $\delta_p^0$  sea 0, que este sea predefinido, permitiendo así involucrar la experiencia del usuario. De esta manera es posible combinar la subjetividad del criterio del usuario con la objetividad del algoritmo *DIS-C*.







# REFERENCIAS

**Bruns, H. and Egenhofer, M.** (1996) Similarity of spatial scenes, 7<sup>th</sup> International symposium on spatial data handling, Delf.

**Cardona H., A.** (2010) Integración de fuentes de datos espaciales con base en ontologías. M.S.C. Thesis. pp. 117.

**Cimiano, P.** (2006) Ontology learning and population from text. *Algorithms, evaluation and applications*. Springer. pp. 98, 133, 270. ISBN 978-0-387-30632-2.

**Egenhofer, M.** (2002) Toward the Semantic Geospatial Web, *ACM-GIS 2002*, A. Voisard and S. C. Chen (eds.), november.

**Egenhofer, M. and Franzosa, R.** (1995) On the equivalence of topological relations. *International journal of geographical information systems*.9(2): pp. 133-152.

**Floyd, R. W.** (1962). "Algorithm 97: Shortest Path". *Communications of the ACM* 5 (6): 345. doi:10.1145/367766.368168

**Fonseca, F. T.** (2008) Ontology-Based Geospatial Data Integration. *Encyclopedia of GIS*, 2008: 812~815

**Gentner, D., Markman A. B., Rattermann M.J., and Kotovsky L.** (1995) Similarity is like analogy. *Similarity*, pp. 111-148.

**Goldstone, R. L., Medin, D. L., and Halberstadt, J.** (1997) Similarity in context. *Memory and Cognition* 25, 237-255.

**Gruber, T.** (1993) A translation approach to portable ontology specifications. *Knowl. Acquis.* London, UK.

**Hahn, U., Chater, N., and Richardson, L.B.** (2003) Similarity as transformation. *Cognition*, 87(1):1-32.

**Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.** (2009) A practical guide to building OWL ontologies using the protégé-OWL plugin and CO-ODE tolos, Edition 1.2, March 13, 2009.

**Janowicz, K., Raubal, M. and Kuhn, W.** (2011) The semantics of similarity in geographic information retrieval. *Journal of spatial information science* Num. 2, pp. 29–57.

**Levachkine, S., Guzmán-Arenas, A.** (2007) Hierarchy as a new data type for qualitative variables. *Expert Systems with Applications*. 32(3): 899-910.

**Lewis, W. D.** (2001) Measuring conceptual distance using WordNet: the design of a metric for measuring semantic similarity. *Coyote papers: Working papers in linguistics, language in cognitive science*. Ed. Circle University of Arizona Linguistics. Vol. 12.

**Lew,M., Sebe,N., Djeraba,C., and Jain, R.** (2006) Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications* 2, 1 (2006), 1–19. doi:10.1145/1126004.1126005.

**Li, B. and Fonseca, F. T.** (2006) TDD – A comprehensive model for qualitative spatial similarity assesment. *Spatial cognition and computation*. 6(1). Pp. 31-62.

**Liddell, H. G. and Scott, R.** (1996) *Semantikos. A Greek – English Lexicon*. 1996.

**Luger, G. F.** (2002) *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 5/e. Pearson Education India.

**Moreno-Ibarra, M.** (2007) Semantic Similarity Applied to Generalization of Geospatial Data. *Lecture Notes in Computer Science*. Vol. 4853. 2007. pp. 247-255.

**Nedas, K. and Egenhofer, M.** (2008) Spatial-Scene Similarity Queries, *Transactions in GIS* 12 (6): 661-681, 2008.

**Page, L., Brin, S., Motwani, R., Winograd, T.** (1999) The *PageRank* Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab.

**Rada, R.; Mili, H.; Bicknell, E.; Blettner, M.** (1989) Development and application of a metric on semantic nets. *IEEE transactions on systems, man and cybernetics*, 19(1):17–30, 1989.

**Rasgado C., A. D.** (2006) Unión de ontologías usando propiedades semánticas. Ph.D. Thesis. pp. 83, 86.

**Resnik, P.** (1999) Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research* 11: 95–130.

**Rissland, E. L.** (2006) AI and similarity. *IEEE Intelligent Systems* 21, 3 (2006), 39–49. doi:10.1109/MIS.2006.38.

**Rodríguez, A. and Egenhofer, M. J.** (2004) Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science* 18: 229–56

**Shvaiko, P., and Euzenat, J.** (2008) Ten challenges for ontology matching. In Proc. On the Move to Meaningful Internet Systems (OTM) (2008), R. Meersman and Z. Tari, Eds., vol. 5332 of Lecture Notes in Computer Science, Springer, pp. 1164–1182.

**Schwering, A.** (2008) Approaches to semantic similarity measurement for geospatial data - a survey. *Transactions in GIS*, 12(1):5–29.

**Sheeren, D., Mustière, S., Zucker, J. D.** (2009) “A data mining approach for assessing consistency between multiple representations in spatial databases”. *Int. J. of Geographical Information Science*. Vol. 23. 2009. pp. 961- 992.

**Staab, S. and Studer, R.** (2009) Handbook on ontologies. Berlin: Springer-Verlag. 2<sup>a</sup> pp. 170. ISBN 978-3-540-70999-2.

**Tversky, A.** (1977) Features of similarity. *Psychological Review* 84(4): pp. 327-352.

**Torgerson, W. S.** (1958) Theory and methods of scaling. Wiley New York.

**Torgerson, W. S.** (1965) Multidimensional scaling of similarity. *Psychometrika*, 30(4):379-393.

**Torres, M., Quintero, R., Moreno-Ibarra, Guzmán, G. y Menchaca-Méndez, R.** (2011) GEONTO-MET: An Approach to Conceptualizing the Geographic Domain”, accepted to *the International Journal of Geographical Information Science*, Taylor & Francis, Por aparecer.

**Warshall, S.** (1962). "A theorem on Boolean matrices". *Journal of the ACM* 9 (1): 11-12. doi:10.1145/321105.321107

**www.Graphviz.org** “Graph visualization software”

**Zagal, R. E.** (2008) Alineamiento de ontologías usando el método Boosting. M.S.C. Thesis. pp. 71 -75.





# ANEXOS

En esta parte de anexos se presentan los demás resultados de las pruebas experimentales realizadas con otras ontologías utilizando el algoritmo *DIS-C*.

## 5.4 Ontología No. 2

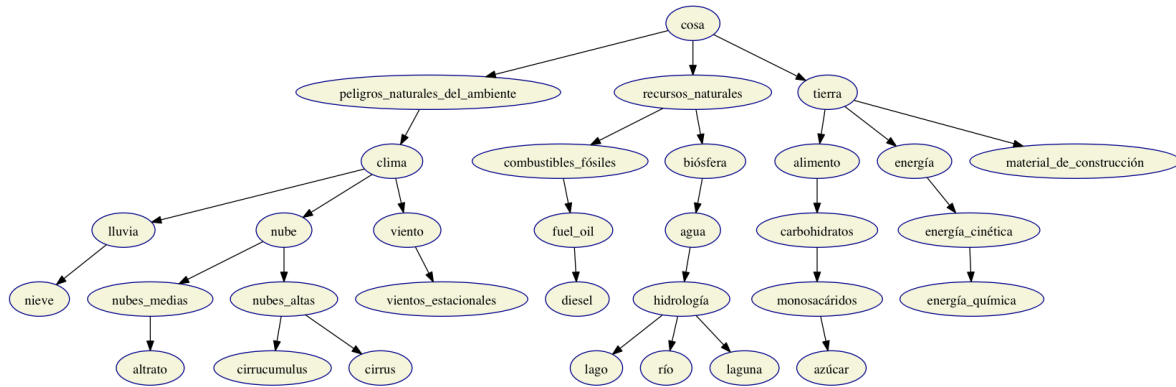


Figura 35. Ontología (Rasgado, 2006)

	cosa	peligros naturales del ambiente	clima	lluvia	nieve	nube	nubes medias	altrato	nubes altas	cirrocumulus	cirrus	viento	vientos estacionales	recursos naturales	combustibles fósiles	fuel oil	diesel	biósfera	agua	hidrología	lago	río	laguna	tierra	alimento	carbohidratos	monosacáridos	azúcar	energía	energía cinética	energía química	material de construcción
cosa	0.00	3.51	7.59	11.10	14.38	11.29	14.87	18.20	15.03	18.18	18.18	11.10	14.38	3.76	7.37	11.11	14.39	7.37	11.11	15.11	18.16	18.16	18.16	3.88	7.43	11.17	14.93	18.25	7.43	11.17	14.45	6.85
peligros naturales del ambiente	8.06	0.00	4.08	7.59	10.87	7.78	11.36	14.69	11.52	14.67	14.67	7.59	10.87	11.82	15.43	19.17	22.45	15.43	19.17	23.17	26.22	26.22	26.22	11.94	15.49	19.23	22.99	26.31	15.49	19.23	22.51	14.91
clima	14.73	6.67	0.00	3.52	6.79	3.70	7.28	10.61	7.45	10.59	10.59	3.52	6.79	18.49	22.10	25.84	29.12	22.10	25.84	29.84	32.90	32.90	32.90	18.61	22.17	25.91	29.66	32.98	22.16	25.91	29.19	21.58
lluvia	21.91	13.85	7.17	0.00	3.28	10.88	14.45	17.78	14.62	17.77	17.77	10.69	13.97	25.66	29.27	33.01	36.29	29.28	33.01	37.01	40.07	40.07	40.07	25.79	29.34	33.08	36.83	40.16	29.34	33.08	36.36	28.76
nieve	29.30	21.23	14.56	7.39	0.00	18.26	21.84	25.17	22.01	25.15	25.15	18.08	21.36	33.05	36.66	40.40	43.68	36.67	40.40	44.40	47.46	47.46	47.46	33.18	36.73	40.47	44.22	47.54	36.73	40.47	43.75	36.14
nube	21.74	13.68	7.01	10.52	13.80	0.00	3.58	6.91	3.74	6.89	6.89	10.52	13.80	25.49	29.11	32.84	36.12	29.11	32.85	36.85	39.90	39.90	39.90	25.62	29.17	32.91	36.66	39.99	29.17	32.91	36.19	28.59
nubes medias	28.74	20.68	14.00	17.52	20.80	7.00	0.00	3.33	10.74	13.89	13.89	17.52	20.80	32.49	36.10	39.84	43.12	36.11	39.85	43.84	46.90	46.90	46.90	32.62	36.17	39.91	43.66	46.99	36.17	39.91	43.19	35.59
altrato	36.16	28.10	21.43	24.94	28.22	14.42	7.42	0.00	18.17	21.31	21.31	24.94	28.22	39.92	43.53	47.27	50.55	43.53	47.27	51.27	54.32	54.32	54.32	40.04	43.59	47.33	51.09	54.41	43.59	47.33	50.61	43.01
nubes altas	28.55	20.49	13.82	17.34	20.61	6.82	10.39	13.72	0.00	3.15	3.15	17.34	20.61	32.31	35.92	39.66	42.94	35.92	39.66	43.66	46.72	46.72	46.72	32.44	35.99	39.73	43.48	46.80	35.99	39.73	43.01	35.40
cirrocumulus	36.14	28.08	21.41	24.93	28.20	14.41	17.98	21.31	7.59	0.00	10.74	24.93	28.20	39.90	43.51	47.25	50.53	43.52	47.25	51.25	54.31	54.31	54.31	40.03	43.58	47.32	51.07	54.39	43.58	47.32	50.60	42.99
cirrus	36.14	28.08	21.41	24.93	28.20	14.41	17.98	21.31	7.59	10.74	0.00	24.93	28.20	39.90	43.51	47.25	50.53	43.52	47.25	51.25	54.31	54.31	54.31	40.03	43.58	47.32	51.07	54.39	43.58	47.32	50.60	42.99
viento	21.91	13.85	7.17	10.69	13.97	10.88	14.45	17.78	14.62	17.77	17.77	0.00	3.28	25.66	29.27	33.01	36.29	29.28	33.01	37.01	40.07	40.07	40.07	25.79	29.34	33.08	36.83	40.16	29.34	33.08	36.36	28.76
vientos estacionales	29.30	21.23	14.56	18.08	21.36	18.26	21.84	25.17	22.01	25.15	25.15	7.39	0.00	33.05	36.66	40.40	43.68	36.67	40.40	44.40	47.46	47.46	47.46	33.18	36.73	40.47	44.22	47.54	36.73	40.47	43.75	36.14
recursos naturales	7.94	11.45	15.53	19.04	22.32	19.23	22.81	26.14	22.97	26.12	26.12	19.04	22.32	0.00	3.61	7.35	10.63	3.62	7.35	11.35	14.41	14.41	14.41	11.82	15.37	19.11	22.87	26.19	15.37	19.11	22.39	14.79
combustibles fósiles	15.09	18.60	22.67	26.19	29.47	26.38	29.95	33.28	30.12	33.27	33.27	26.19	29.47	7.15	0.00	3.74	7.02	10.76	14.50	18.50	21.55	21.55	21.55	18.97	22.52	26.26	30.01	33.34	22.52	26.26	29.54	21.94
fuel oil	21.92	25.42	29.50	33.02	36.29	33.20	36.78	40.11	36.95	40.09	40.09	33.02	36.29	13.97	6.83	0.00	3.28	17.59	21.33	25.33	28.38	28.38	28.38	25.80	29.35	33.09	36.84	40.16	29.35	33.09	36.37	28.76
diesel	29.29	32.80	36.88	40.39	43.67	40.58	44.16	47.49	44.32	47.47	47.47	40.39	43.67	21.35	14.20	7.37	0.00	24.96	28.70	32.70	35.76	35.76	35.76	33.17	36.72	40.46	44.21	47.54	36.72	40.46	43.74	36.14
biósfera	15.08	18.59	22.67	26.19	29.46	26.37	29.95	33.28	30.12	33.26	33.26	26.19	29.46	7.14	10.75	14.49	17.77	0.00	3.74	7.74	10.79	10.79	10.79	18.97	22.52	26.26	30.01	33.33	22.52	26.26	29.54	21.93
agua	21.91	25.42	29.50	33.02	36.29	33.20	36.78	40.11	36.95	40.09	40.09	33.02	36.29	13.97	17.59	21.32	24.60	6.83	0.00	4.00	7.05	7.05	7.05	25.80	29.35	33.09	36.84	40.16	29.35	33.09	36.37	28.76
hidrología	28.45	31.96	36.04	39.55	42.83	39.74	43.32	46.65	43.48	46.63	46.63	39.55	42.83	20.51	24.12	27.86	31.14	13.37	6.54	0.00	3.06	3.06	3.06	32.33	35.88	39.62	43.38	46.70	35.88	39.62	42.90	35.30
lago	36.11	39.62	43.70	47.22	50.49	47.40	50.98	54.31	51.15	54.29	54.29	47.22	50.49	28.17	31.78	35.52	38.80	21.03	14.20	7.66	0.00	10.72	10.72	40.00	43.55	47.29	51.04	54.36	43.55	47.29	50.57	42.96
río	36.11	39.62	43.70	47.22	50.49	47.40	50.98	54.31	51.15	54.29	54.29	47.22	50.49	28.17	31.78	35.52	38.80	21.03	14.20	7.66	10.72	10.72	40.00	43.55	47.29	51.04	54.36	43.55	47.29	50.57	42.96	
laguna	36.11	39.62	43.70	47.22	50.49	47.40	50.98	54.31	51.15	54.29	54.29	47.22	50.49	28.17	31.78	35.52	38.80	21.03	14.20	7.66	10.72	10.72	0.00	40.00	43.55	47.29	51.04	54.36	43.55	47.29	50.57	42.96
tierra	7.88	11.39	15.47	18.98	22.26	19.17	22.75	26.08	22.91	26.06	26.06	18.98	22.26	11.64	15.25	18.99	22.27	15.25	18.99	22.99	26.05	26.05	26.05	0.00	3.95	7.29	11.04	14.37	3.95	7.29	10.57	2.97
alimento	15.45	18.66	22.74	26.26	29.53	26.44	30.02	33.35	30.19	33.33	33.33	26.26	29.53	18.91	22.52	26.26	29.54	22.53	26.26	30.26	33.32	33.32	33.32	7.27	10.83	14.57	18.32	21.64	10.82	14.56	17.84	10.24
carbohidratos	21.98	25.49	29.57	33.08	36.36	33.27	36.85	40.18	37.01	40.16	40.16	33.08	36.36	25.74	29.35	33.09	36.37	29.35	33.09	37.09	40.14	40.14	40.14	14.10	16.82	20.60	24.45	28.30	16.82	20.60	24.45	
monosacáridos	28.79	32.29	36.37	39.89	43.16	40.07	43.65	46.98	43.82	46.96	46.96	39.89	43.16	32.54	36.15	39.89	43.17	36.16	39.89	43.89	46.95	46.95	46.95	20.90	13.63	6.80	0.00	3.32	24.45	28.19	31.47	23.87
azúcar	36.19	39.70	43.77	47.29	50.57	47.48	51.05	54.38	51.22	54.37	54.37	47.29	50.57	39.94	43.55	47.29	50.57	43.56	47.30	51.30	54.35	54.35	54.35	28.31	21.03	14.21	7.40	0.00	31.86	35.60	38.88	31.27
energía	15.16	18.66	22.74	26.26	29.53	26.44	30.02	33.35	30.19	33.33	33.33	26.26	29.53	18.91	22.52	26.26	29.54	22.53	26.26	30.26	33.32	33.32	33.32	7.27	10.83	14.57	18.32	21.64	0.00	3.74	7.02	10.24
energía cinética	21.98	25.49	29.57	33.08	36.36	33.27	36.85	40.18	37.01	40.16	40.16	33.08	36.36	25.74	29.35	33.09	36.37	29.35	33.09	37.09	40.15	40.15	40.15	14.10	17.65	21.39	25.14	28.47	6.83	0.00	3.28	17.07
energía química	29.35	32.86	36.94	40.46	43.73	40.64	44.22	47.55	44.39	47.53	47.53	40.46	43.73	33.11	36.72	40.46	43.74	36.73	40.46	44.46	47.52	47.52	47.52	21.47	25.02	28.76	32.52	35.84	14.20	7.37	0.00	24.44
material de construcción	15.64	19.14	23.22	26.74	30.01	26.92	30.50	33.83	30.67	33.81	33.81	26.74	30.01	19.39	23.00	26.74	30.02	23.01	26.74	30.74	33.80	33.80	33.80	7.75	11.31	15.05	18.80					

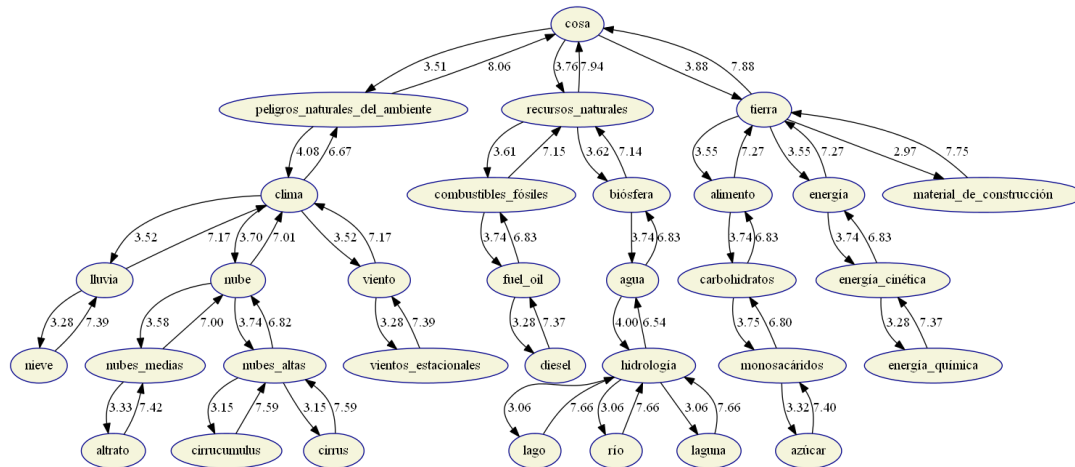


Figura 36. Ontología de la Figura 35 con valores de pesos conceptuales. Utilizando como referencia la Matriz 8.

```

v_agua: 1.0350081586348034
v_alimento: 1.1680655763790893
v_altrato: 0.8357653248670553
v_azúcar: 0.8540534137517567
v_biósfera: 1.1770867983449373
v_carbohidratos: 1.03272952753475
v_cirrucumulus: 0.8341743421875214
v_cirrus: 0.8341743421875214
v_clima: 1.2430510687274285
v_combustibles_fósiles: 1.168733784031014
v_cosa: 0.5125498111792594
v_diesel: 0.9090020034837232
v_energía: 1.1654544785197865
v_energía_cinética: 1.0319369073110933
v_energía_química: 0.9101170083557397
v_fuel_oil: 1.0324348048685446
v_hidrología: 0.9607971188007967
v_lago: 0.8446555972854223
v_laguna: 0.8446555972854223
v_lluvia: 1.0370483508947461
v_material_de_construcción: 1.100024621850524
v_monosacáridos: 0.9445363842742468
v_nieve: 0.8989900015043989
v_nube: 1.0589438202773593
v_nubes_altas: 0.9500588776504645
v_nubes_medias: 0.9359780655190516
v_peligros_naturales_del_ambiente: 1.4252689774649765
v_recursos_naturales: 1.440631466315617
v_río: 0.8446555972854223
v_tierra: 1.4478283861480412
v_viento: 1.0370483508947461
v_vientos_estacionales: 0.8989900015043989

```

Figura 37. Valores resultantes de visibilidad de la Ontología No. 2.

Tipo: 0, Directa: 2.80, Inversa: 5.76

Figura 38. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 4 iteraciones.



## 5.5 Ontología No. 3

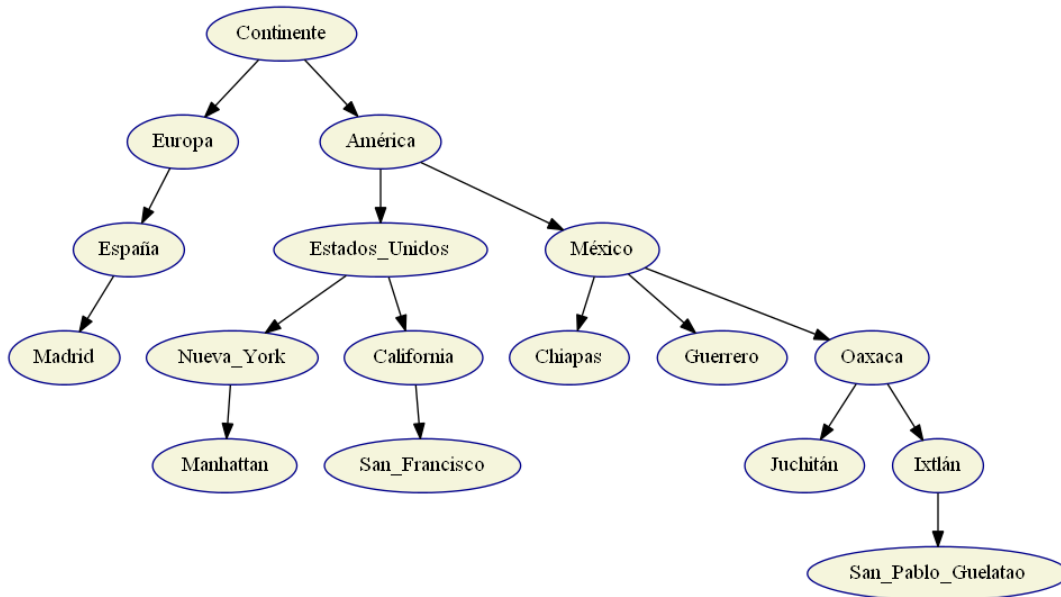


Figura 39. Ontología (Rasgado, 2006).

	Continente	Europa	España	Madrid	América	Estados Unidos	Nueva York	Manhattan	California	San Francisco	México	Chiapas	Guerrero	Oaxaca	Juchitán	Ixtlán	San Pablo Guelatao
Continente	0.00	4.07	8.34	12.14	4.58	8.97	13.07	16.95	13.07	16.95	9.11	12.66	12.66	13.31	17.01	17.41	21.39
Europa	10.59	0.00	4.28	8.08	15.18	19.56	23.66	27.54	23.66	27.54	19.70	23.25	23.25	23.91	27.60	28.01	31.98
España	19.74	9.14	0.00	3.80	24.32	28.70	32.81	36.69	32.81	36.69	28.85	32.40	32.40	33.05	36.75	37.15	41.13
Madrid	29.36	18.77	9.63	0.00	33.95	38.33	42.43	46.31	42.43	46.31	38.47	42.02	42.02	42.68	46.37	46.78	50.75
América	10.42	14.49	18.77	22.56	0.00	4.38	8.49	12.36	8.49	12.36	4.53	8.07	8.07	8.73	12.43	12.83	16.81
Estados Unidos	19.91	23.97	28.25	32.05	9.49	0.00	4.10	7.98	4.10	7.98	14.01	17.56	17.56	18.21	21.91	22.32	26.29
Nueva York	29.27	33.33	37.61	41.41	18.84	9.36	0.00	3.88	13.46	17.34	23.37	26.92	26.92	27.57	31.27	31.67	35.65
Manhattan	39.00	43.07	47.35	51.14	28.58	19.09	9.74	0.00	23.20	27.07	33.10	36.65	36.65	37.31	41.00	41.41	45.39
California	29.27	33.33	37.61	41.41	18.84	9.36	13.46	17.34	0.00	3.88	23.37	26.92	26.92	27.57	31.27	31.67	35.65
San Francisco	39.00	43.07	47.35	51.14	28.58	19.09	23.20	27.07	9.74	0.00	33.10	36.65	36.65	37.31	41.00	41.41	45.39
México	19.83	23.90	28.18	31.98	9.41	13.80	17.90	21.78	17.90	21.78	0.00	3.55	3.55	4.20	7.90	8.30	12.28
Chiapas	29.92	33.98	38.26	42.06	19.49	23.88	27.98	31.86	27.98	31.86	10.08	0.00	13.63	14.28	17.98	18.39	22.36
Guerrero	29.92	33.98	38.26	42.06	19.49	23.88	27.98	31.86	27.98	31.86	10.08	13.63	0.00	14.28	17.98	18.39	22.36
Oaxaca	29.16	33.22	37.50	41.30	18.74	23.12	27.22	31.10	27.22	31.10	9.32	12.87	12.87	0.00	3.70	4.10	8.08
Juchitán	39.10	43.17	47.45	51.24	28.68	33.06	37.17	41.04	37.17	41.04	19.27	22.81	22.81	9.94	0.00	14.05	18.02
Ixtlán	38.43	42.49	46.77	50.57	28.01	32.39	36.49	40.37	36.49	40.37	18.59	22.14	22.14	9.27	12.97	0.00	3.98
San Pablo Guelatao	48.26	52.33	56.61	60.41	37.84	42.22	46.33	50.20	46.33	50.20	28.43	31.98	31.98	19.11	22.80	9.84	0.00

Matriz 9. Matriz resultante de pesos conceptuales después de 5 iteraciones.

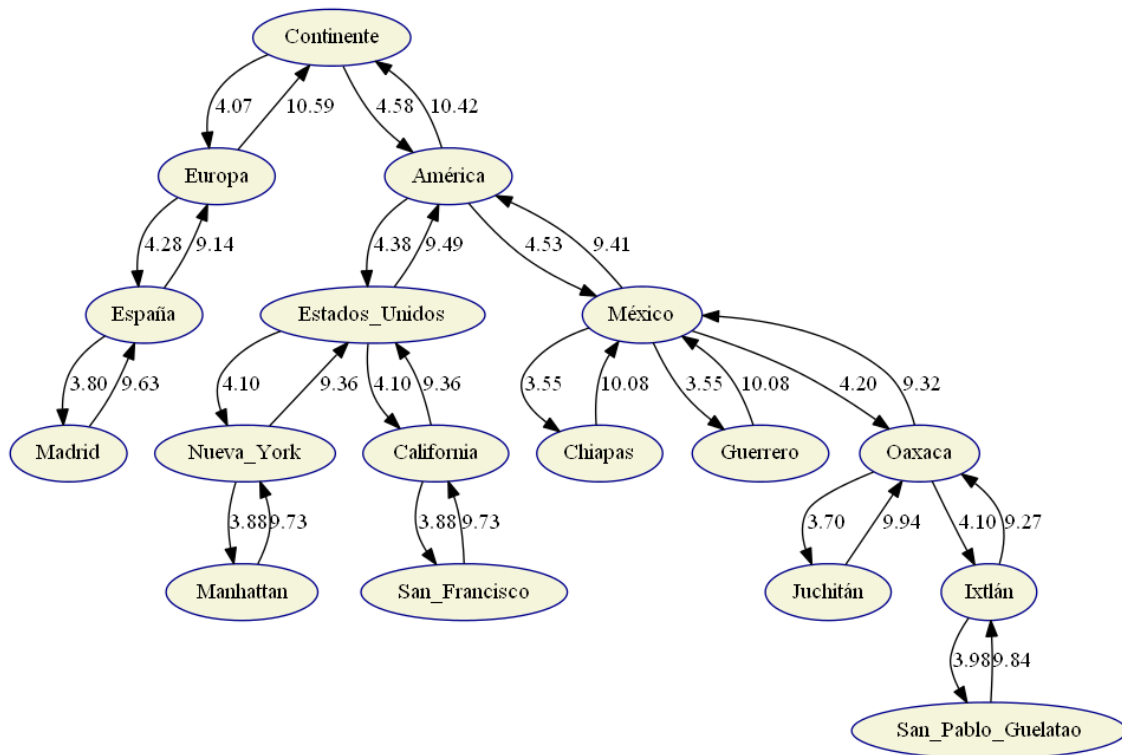


Figura 40. Ontología de la Figura 39 con valores de pesos conceptuales. Utilizando como referencia la Matriz 9.

```

v_América: 1.8399025775905753
v_California: 0.9656191228972795
v_Chiapas: 0.9105895060898983
v_Continente: 0.44584322300531515
v_España: 1.1355034512698627
v_Estados_Unidos: 1.2639273863118594
v_Europa: 1.4152431318758507
v_Guerrero: 0.9105895060898983
v_Ixtlán: 0.8079305800391288
v_Juchitán: 0.7736896045968767
v_Madrid: 0.9428003269267664
v_Manhattan: 0.7966418723457839
v_México: 1.3145924177109831
v_Nueva_York: 0.9656191228972795
v_Oaxaca: 0.9809954964576764
v_San_Francisco: 0.7966418723457839
v_San_Pablo_Guelatao: 0.696962672554963
  
```

Figura 41. Valores resultantes de visibilidad de la Ontología No. 3.

Tipo: 0, Directa: 3.36, Inversa: 8.00

Figura 42. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 5 iteraciones.

## 5.6 Ontología No. 4

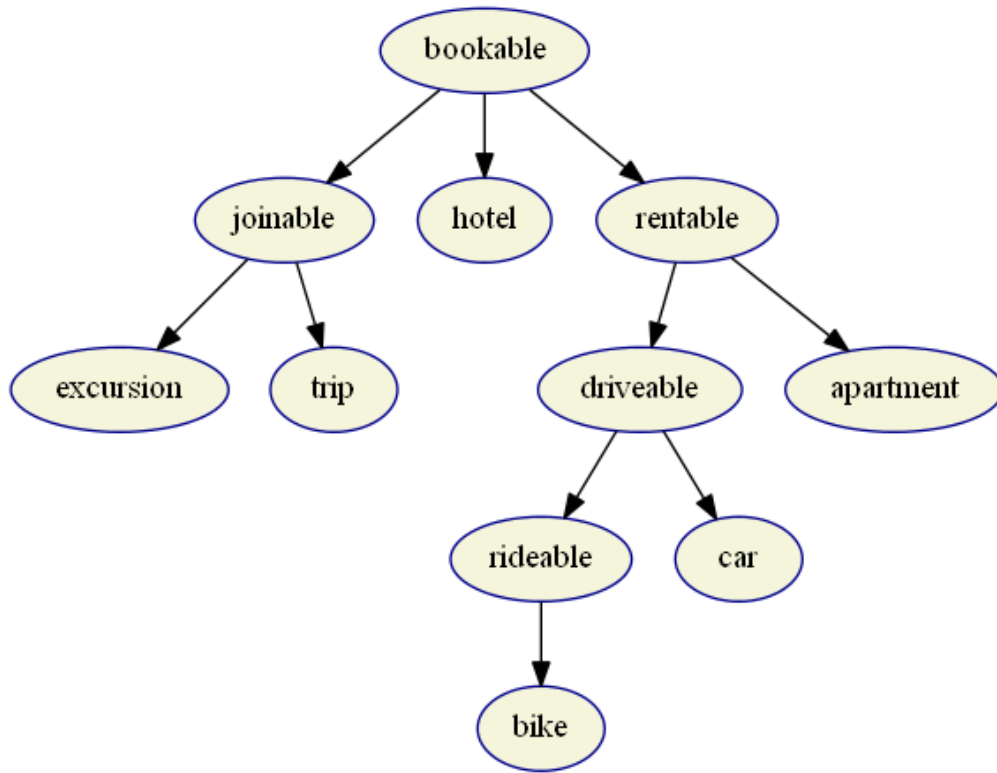


Figura 43. Ontología (Cimiano, 2006).

	bookable	joinable	excursion	trip	hotel	rentable	driveable	rideable	bike	car	apartment
bookable	0.00	6.39	11.95	11.95	5.34	6.58	12.74	18.74	24.88	18.43	12.10
joinable	29.45	0.00	5.55	5.55	34.79	36.03	42.19	48.19	54.33	47.88	41.55
excursion	56.31	26.86	0.00	32.41	61.65	62.90	69.06	75.05	81.20	74.74	68.41
trip	56.31	26.86	32.41	0.00	61.65	62.90	69.06	75.05	81.20	74.74	68.41
hotel	29.94	36.34	41.89	41.89	0.00	36.52	42.68	48.68	54.82	48.37	42.04
rentable	29.42	35.82	41.37	41.37	34.76	0.00	6.16	12.16	18.30	11.85	5.52
driveable	55.76	62.16	67.71	67.71	61.10	26.34	0.00	6.00	12.14	5.69	31.86
rideable	81.96	88.36	93.91	93.91	87.30	52.54	26.20	0.00	6.14	31.89	58.06
bike	109.13	115.53	121.08	121.08	114.47	79.71	53.37	27.17	0.00	59.05	85.23
car	82.76	89.15	94.70	94.70	88.10	53.34	27.00	32.99	39.14	0.00	58.86
apartment	56.48	62.88	68.43	68.43	61.82	27.06	33.22	39.22	45.36	38.91	0.00

Matriz 10. Matriz resultante de pesos conceptuales después de 10 iteraciones.

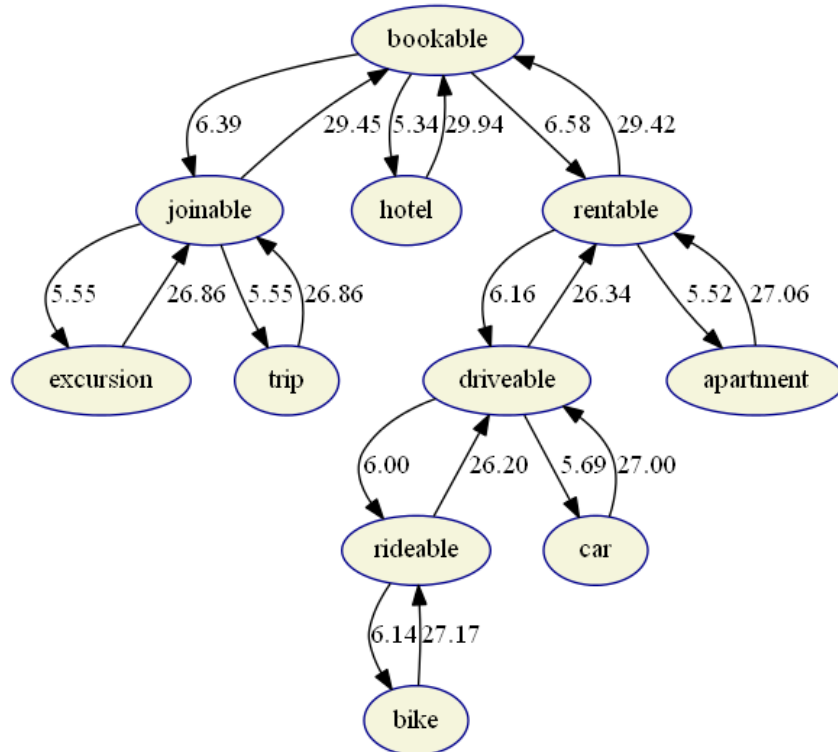


Figura 44. Ontología de la Figura 43 con valores de pesos conceptuales. Utilizando como referencia la Matriz 10.

```

v_apartment: 0.9391810499024981
v_bike: 0.47609463714978917
v_bookable: 0.2229174401041888
v_car: 0.6259198142754456
v_driveable: 0.9652169329531485
v_excursion: 0.9501475566159622
v_hotel: 1.4304963823710097
v_joinable: 1.5823323127508817
v_rentable: 1.8612206188373228
v_rideable: 0.6229108716795607
v_trip: 0.9501475566159622

```

Figura 45. Valores resultantes de visibilidad de la Ontología No. 4.

Tipo: 0, Directa: 5.34, Inversa: 24.76

Figura 46. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 10 iteraciones.

## 5.7 Ontología No. 5

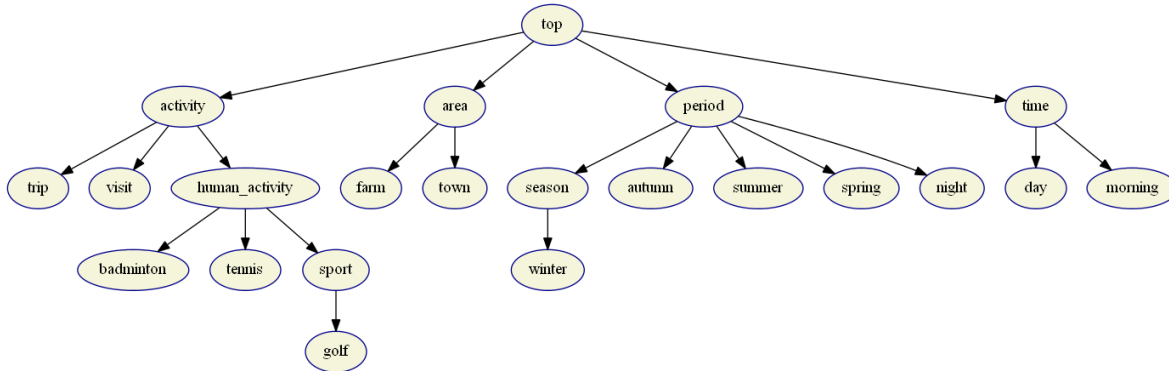


Figura 47. Ontología (Cimiano, 2006).

	top	activity	trip	visit	human activity	badminton	tennis	sport	golf	area	farm	town	period	season	winter	autumn	summer	spring	night	time	day	morning
top	0.00	4.06	7.01	7.01	7.79	10.84	10.84	11.22	14.68	3.81	6.83	6.83	4.18	7.59	10.89	7.09	7.09	7.09	7.09	3.81	6.83	6.83
activity	17.83	0.00	2.95	2.95	3.73	6.78	6.78	7.16	10.62	21.63	24.66	24.66	22.01	25.42	28.72	24.91	24.91	24.91	24.91	21.63	24.66	24.66
trip	33.38	15.55	0.00	18.50	19.28	22.33	22.33	22.70	26.17	37.18	40.21	40.21	37.56	40.97	44.27	40.46	40.46	40.46	40.46	37.18	40.21	40.21
visit	33.38	15.55	18.50	0.00	19.28	22.33	22.33	22.70	26.17	37.18	40.21	40.21	37.56	40.97	44.27	40.46	40.46	40.46	40.46	37.18	40.21	40.21
human activity	32.59	14.76	17.72	17.72	0.00	3.05	3.05	3.42	6.89	36.40	39.42	39.42	36.77	40.18	43.48	39.68	39.68	39.68	39.68	36.40	39.42	39.42
badminton	48.00	30.17	33.12	33.12	15.41	0.00	18.45	18.83	22.30	51.80	54.83	54.83	52.18	55.59	58.89	55.08	55.08	55.08	55.08	51.80	54.83	54.83
tennis	48.00	30.17	33.12	33.12	15.41	18.45	0.00	18.83	22.30	51.80	54.83	54.83	52.18	55.59	58.89	55.08	55.08	55.08	55.08	51.80	54.83	54.83
sport	47.30	29.47	32.42	32.42	14.71	17.75	17.75	0.00	3.47	51.10	54.13	54.13	51.48	54.89	58.19	54.38	54.38	54.38	54.38	51.10	54.13	54.13
golf	62.59	44.76	47.71	47.71	30.00	33.04	33.04	15.29	0.00	66.39	69.42	69.42	66.77	70.18	73.48	69.67	69.67	69.67	69.67	66.39	69.42	69.42
area	17.90	21.96	24.91	24.91	25.69	28.74	28.74	29.12	32.58	0.00	3.03	3.03	22.08	25.49	28.79	24.99	24.99	24.99	24.99	21.71	24.73	24.73
farm	33.19	37.25	40.21	40.21	40.98	44.03	44.03	44.41	47.87	15.29	0.00	18.32	37.37	40.78	44.08	40.28	40.28	40.28	40.28	37.00	40.02	40.02
town	33.19	37.25	40.21	40.21	40.98	44.03	44.03	44.41	47.87	15.29	18.32	0.00	37.37	40.78	44.08	40.28	40.28	40.28	40.28	37.00	40.02	40.02
period	17.78	21.84	24.79	24.79	25.57	28.62	28.62	29.00	32.46	21.59	24.61	24.61	0.00	3.41	6.71	2.91	2.91	2.91	2.91	21.59	24.61	24.61
season	32.92	36.98	39.93	39.93	40.71	43.75	43.75	44.13	47.60	36.72	39.75	39.75	15.14	0.00	3.30	18.04	18.04	18.04	18.04	36.72	39.75	39.75
winter	48.01	52.07	55.03	55.03	55.80	58.85	58.85	59.23	62.69	51.82	54.84	54.84	30.23	15.10	0.00	33.14	33.14	33.14	33.14	51.82	54.84	54.84
autumn	33.45	37.51	40.46	40.46	41.24	44.29	44.29	44.66	48.13	37.25	40.28	40.28	15.67	19.08	22.38	0.00	18.57	18.57	18.57	37.25	40.28	40.28
summer	33.45	37.51	40.46	40.46	41.24	44.29	44.29	44.66	48.13	37.25	40.28	40.28	15.67	19.08	22.38	18.57	0.00	18.57	18.57	37.25	40.28	40.28
spring	33.45	37.51	40.46	40.46	41.24	44.29	44.29	44.66	48.13	37.25	40.28	40.28	15.67	19.08	22.38	18.57	18.57	0.00	18.57	37.25	40.28	40.28
night	33.45	37.51	40.46	40.46	41.24	44.29	44.29	44.66	48.13	37.25	40.28	40.28	15.67	19.08	22.38	18.57	18.57	18.57	0.00	37.25	40.28	40.28
time	17.90	21.96	24.91	24.91	25.69	28.74	28.74	29.12	32.58	21.71	24.73	24.73	22.08	25.49	28.79	24.99	24.99	24.99	24.99	0.00	3.03	3.03
day	33.19	37.25	40.21	40.21	40.98	44.03	44.03	44.41	47.87	37.00	40.02	40.02	37.37	40.78	44.08	40.28	40.28	40.28	40.28	15.29	0.00	18.32
morning	33.19	37.25	40.21	40.21	40.98	44.03	44.03	44.41	47.87	37.00	40.02	40.02	37.37	40.78	44.08	40.28	40.28	40.28	40.28	15.29	18.32	0.00

Matriz 11. Matriz resultante de pesos conceptuales después de 6 iteraciones.

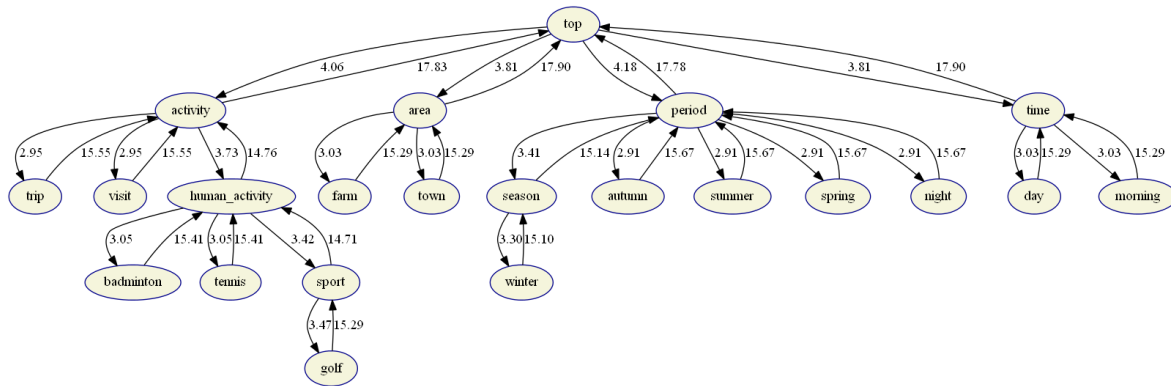


Figura 48. Ontología de la Figura 47 con valores de pesos conceptuales. Utilizando como referencia la Matriz 11.

```

v_activity: 1.6738586148162111
v_area: 1.5022494461023965
v_autumn: 0.9752241338305835
v_badminton: 0.7344335073784025
v_day: 0.9778955456196529
v_farm: 0.9778955456196531
v_golf: 0.6075956063738256
v_human_activity: 1.0380804149690215
v_morning: 0.9778955456196529
v_night: 0.9752241338305835
v_period: 1.6509616758835113
v_season: 1.0134664880882962
v_sport: 0.7554723600782356
v_spring: 0.9752241338305835
v_summer: 0.9752241338305835
v_tennis: 0.7344335073784025
v_time: 1.5022494461023967
v_top: 0.22663070047821438
v_town: 0.9778955456196531
v_trip: 0.9743572594535721
v_visit: 0.9743572594535721
v_winter: 0.7557573401907278

```

Figura 49. Valores resultantes de visibilidad de la Ontología No. 5.

Tipo: 0, Directa: 2.80, Inversa: 13.27

Figura 50. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 6 iteraciones.

## 5.8 Ontología No. 6

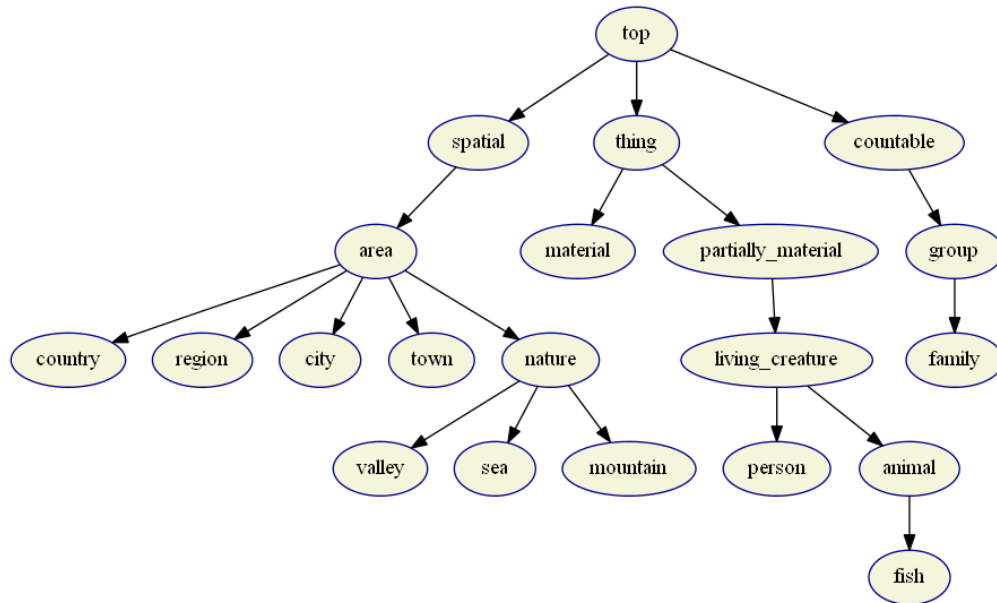


Figura 51. Ontología (Cimiano, 2006).

	top	spatial	area	country	region	city	town	nature	valley	sea	mountain	thing	material	partially material	living creature	person	animal	fish	countable	group	family
top	0.00	4.24	8.96	12.44	12.44	12.44	12.44	13.16	16.77	16.77	16.77	4.49	8.03	8.64	13.04	16.74	17.14	21.12	4.14	8.39	12.17
spatial	14.20	0.00	4.72	8.20	8.20	8.20	8.20	8.92	12.53	12.53	12.53	18.69	22.24	22.84	27.25	30.94	31.35	35.32	18.34	22.59	26.37
area	26.13	11.93	0.00	3.48	3.48	3.48	3.48	4.20	7.81	7.81	7.81	30.62	34.16	34.77	39.17	42.87	43.27	47.25	30.26	34.52	38.30
country	39.22	25.02	13.09	0.00	16.57	16.57	16.57	17.29	20.90	20.90	20.90	43.71	47.26	47.86	52.27	55.96	56.36	60.34	43.36	47.61	51.39
region	39.22	25.02	13.09	16.57	0.00	16.57	16.57	17.29	20.90	20.90	20.90	43.71	47.26	47.86	52.27	55.96	56.36	60.34	43.36	47.61	51.39
city	39.22	25.02	13.09	16.57	16.57	0.00	16.57	17.29	20.90	20.90	20.90	43.71	47.26	47.86	52.27	55.96	56.36	60.34	43.36	47.61	51.39
town	39.22	25.02	13.09	16.57	16.57	16.57	0.00	17.29	20.90	20.90	20.90	43.71	47.26	47.86	52.27	55.96	56.36	60.34	43.36	47.61	51.39
nature	38.38	24.18	12.25	15.73	15.73	15.73	15.73	0.00	3.61	3.61	3.61	42.87	46.42	47.02	51.43	55.13	55.53	59.50	42.52	46.77	50.55
valley	51.34	37.13	25.21	28.68	28.68	28.68	28.68	12.95	0.00	16.56	16.56	55.82	59.37	59.97	64.38	68.08	68.48	72.45	55.47	59.72	63.50
sea	51.34	37.13	25.21	28.68	28.68	28.68	28.68	12.95	16.56	0.00	16.56	55.82	59.37	59.97	64.38	68.08	68.48	72.45	55.47	59.72	63.50
mountain	51.34	37.13	25.21	28.68	28.68	28.68	28.68	12.95	16.56	16.56	0.00	55.82	59.37	59.97	64.38	68.08	68.48	72.45	55.47	59.72	63.50
thing	14.12	18.36	23.08	26.56	26.56	26.56	26.56	27.28	30.89	30.89	30.89	0.00	3.55	4.15	8.56	12.25	12.65	16.63	18.25	22.51	26.29
material	27.04	31.28	36.00	39.48	39.48	39.48	39.48	40.20	43.81	43.81	43.81	12.92	0.00	17.07	21.47	25.17	25.57	29.55	31.17	35.43	39.21
partially material	26.58	30.82	35.54	39.02	39.02	39.02	39.02	39.74	43.35	43.35	43.35	12.46	16.00	0.00	4.41	8.10	8.51	12.48	30.71	34.97	38.74
living creature	38.43	42.68	47.39	50.87	50.87	50.87	50.87	51.59	55.20	55.20	55.20	24.31	27.86	11.85	0.00	3.70	4.10	8.07	42.57	46.82	50.60
person	51.25	55.50	60.22	63.69	63.69	63.69	63.69	64.42	68.03	68.03	68.03	37.13	40.68	24.68	12.82	0.00	16.92	20.89	55.39	59.64	63.42
animal	50.60	54.84	59.56	63.04	63.04	63.04	63.04	63.76	67.37	67.37	67.37	36.48	40.03	24.02	12.17	15.87	0.00	3.97	54.73	58.99	62.77
fish	63.34	67.59	72.31	75.78	75.78	75.78	75.78	76.51	80.12	80.12	80.12	49.22	52.77	36.77	24.91	28.61	12.74	0.00	67.48	71.73	75.51
countable	14.24	18.49	23.21	26.68	26.68	26.68	26.68	27.41	31.02	31.02	31.02	18.73	22.28	22.88	27.29	30.98	31.39	35.36	0.00	4.25	8.03
group	26.36	30.60	35.32	38.80	38.80	38.80	38.80	39.52	43.13	43.13	43.13	30.85	34.39	35.00	39.40	43.10	43.50	47.48	12.12	0.00	3.78
family	38.92	43.16	47.88	51.36	51.36	51.36	51.36	52.08	55.69	55.69	55.69	43.41	46.95	47.56	51.96	55.66	56.06	60.03	24.68	12.56	0.00

Matriz 12. Matriz resultante de pesos conceptuales después de 6 iteraciones.

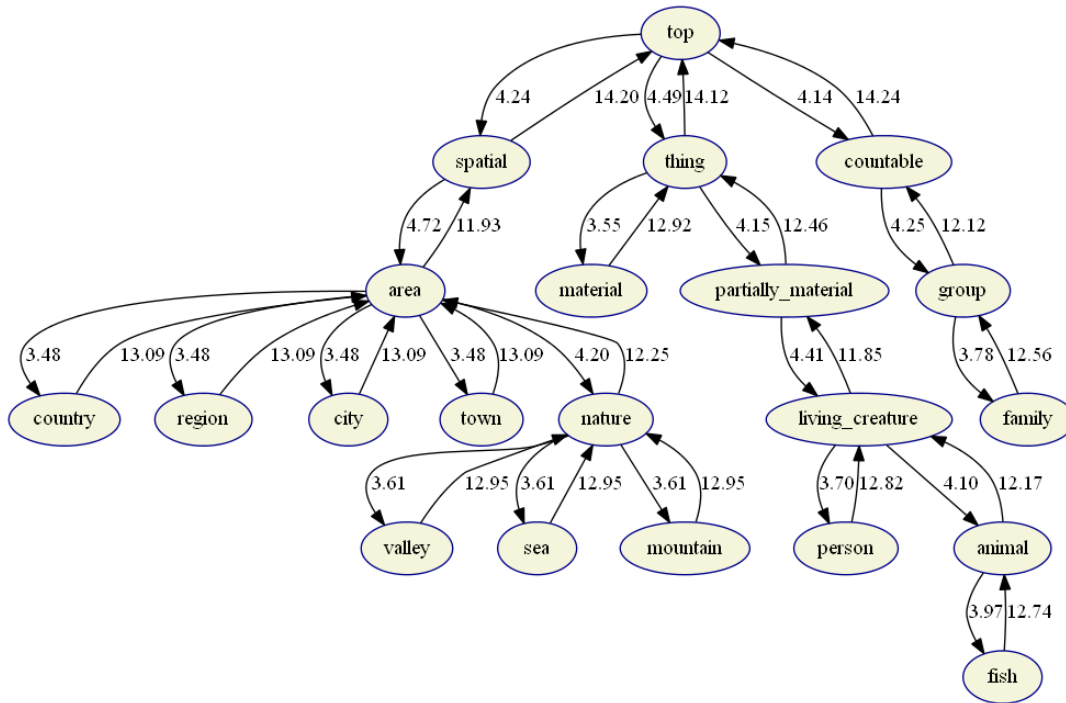


Figura 52. Ontología de la Figura 51 con valores de pesos conceptuales. Utilizando como referencia la Matriz 12.

```

v_animal: 0.803882159272237
v_area: 1.3074392816720755
v_city: 0.9112443132759976
v_countable: 1.569848250729914
v_country: 0.9112443132759976
v_family: 0.9343175305186565
v_fish: 0.6960721892820413
v_group: 1.1692253248303655
v_living_creature: 0.9626263930751096
v_material: 1.1415491687624102
v_mountain: 0.7542058919691877
v_nature: 0.9619743131889597
v_partially_material: 1.2053347308618028
v_person: 0.7843358428665969
v_region: 0.9112443132759976
v_sea: 0.7542058919691877
v_spatial: 1.7884270877157389
v_thing: 1.7079946319686796
v_top: 0.33128218949006283
v_town: 0.9112443132759976
v_valley: 0.7542058919691877
  
```

Figura 53. Valores resultantes de visibilidad de la Ontología No. 6.

Tipo: 0, Directa: 3.35, Inversa: 10.90

Figura 54. Valores resultantes de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 6 iteraciones.



## 5.9 Ontología No. 7

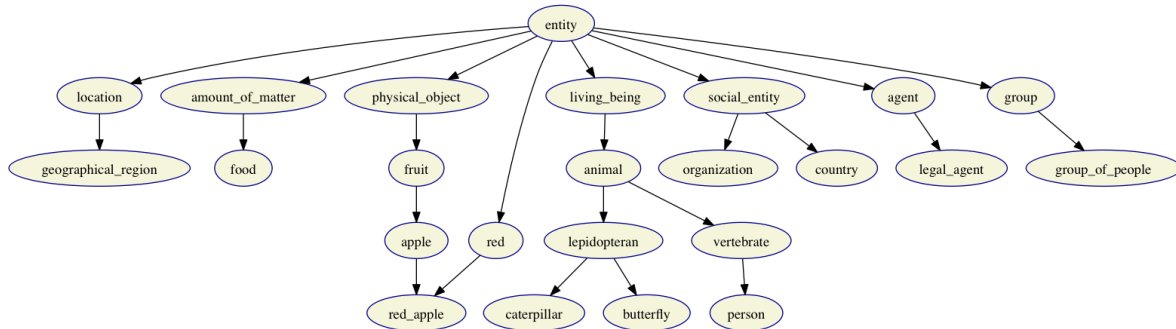


Figura 55. Ontología (Staab & Studer, 2009).

	entity	location	geographical region	amount of matter	food	physical object	fruit	apple	red apple	red	living being	animal	lepidopteran	caterpillar	butterfly	vertebrate	person	social entity	organization	country	agent	legal agent	group	group of people
entity	0.00	4.85	9.38	4.85	9.38	4.88	9.91	15.05	9.37	4.83	4.89	10.06	15.08	19.67	19.67	14.95	19.77	5.10	9.50	9.50	4.85	9.38	4.85	9.38
location	17.97	0.00	4.53	22.82	27.35	22.85	27.88	33.02	27.34	22.80	22.86	28.03	33.04	37.64	37.64	32.91	37.73	23.07	27.47	27.47	22.82	27.35	22.82	27.35
geographical region	33.95	15.99	0.00	38.80	43.33	38.83	43.86	49.01	43.33	38.79	38.85	44.02	49.03	53.63	53.63	48.90	53.72	39.05	43.46	43.46	38.80	43.33	38.80	43.33
amount of matter	17.97	22.82	27.35	0.00	4.53	22.85	27.88	33.02	27.34	22.80	22.86	28.03	33.04	37.64	37.64	32.91	37.73	23.07	27.47	27.47	22.82	27.35	22.82	27.35
food	33.95	38.80	43.33	15.99	0.00	38.83	43.86	49.01	43.33	38.79	38.85	44.02	49.03	53.63	53.63	48.90	53.72	39.05	43.46	43.46	38.80	43.33	38.80	43.33
physical object	17.95	22.80	27.33	22.80	27.33	0.00	5.03	10.17	14.84	22.79	22.85	28.02	33.03	37.63	37.63	32.90	37.72	23.05	27.45	27.45	22.80	27.33	22.80	27.33
fruit	33.41	38.26	42.79	38.26	42.79	15.46	0.00	5.14	9.81	25.84	38.30	43.47	48.48	53.08	53.08	48.35	53.17	38.51	42.91	42.91	38.26	42.79	38.26	42.79
apple	38.67	43.52	48.05	43.52	48.05	30.74	15.28	0.00	4.67	20.70	43.57	48.74	53.75	58.35	58.35	53.62	58.44	43.78	48.18	48.18	43.52	48.05	43.52	48.05
red apple	34.00	38.85	43.38	38.85	43.38	38.88	31.13	15.85	0.00	16.03	38.90	44.07	49.08	53.68	53.68	48.95	53.77	39.10	43.50	43.50	38.85	43.38	38.85	43.38
red	17.98	22.83	27.36	22.83	27.36	22.86	27.89	20.39	4.54	0.00	22.87	28.04	33.05	37.65	37.65	32.92	37.74	23.08	27.48	27.48	22.83	27.36	22.83	27.36
living being	17.95	22.80	27.33	22.80	27.33	22.83	27.86	33.00	27.32	22.78	0.00	5.17	10.18	14.78	14.78	10.05	14.87	23.05	27.45	27.45	22.80	27.33	22.80	27.33
animal	33.27	38.11	42.64	38.11	42.64	38.15	43.18	48.32	42.64	38.10	15.32	0.00	5.01	9.61	9.61	4.88	9.70	38.37	42.77	42.77	38.11	42.64	38.11	42.64
lepidopteran	48.62	53.47	58.00	53.47	58.00	53.50	58.53	63.67	57.99	53.45	30.67	15.35	0.00	4.60	4.60	20.24	25.05	53.72	58.12	58.12	53.47	58.00	53.47	58.00
caterpillar	64.97	69.82	74.35	69.82	74.35	69.85	74.88	80.02	74.34	69.81	47.03	31.71	16.35	0.00	20.95	36.59	41.41	70.07	74.48	74.48	69.82	74.35	69.82	74.35
butterfly	64.97	69.82	74.35	69.82	74.35	69.85	74.88	80.02	74.34	69.81	47.03	31.71	16.35	20.95	0.00	36.59	41.41	70.07	74.48	74.48	69.82	74.35	69.82	74.35
vertebrate	48.84	53.69	58.22	53.69	58.22	53.72	58.75	63.89	58.21	53.67	30.89	15.57	20.58	25.18	25.18	0.00	4.82	53.94	58.34	58.34	53.69	58.22	53.69	58.22
person	65.05	69.90	74.43	69.90	74.43	69.93	74.96	80.10	74.42	69.88	47.10	31.78	36.79	41.39	41.39	16.21	0.00	70.15	74.55	74.55	69.90	74.43	69.90	74.43
social entity	17.84	22.69	27.22	22.69	27.22	22.72	27.75	32.89	27.21	22.67	22.73	27.91	32.92	37.52	37.52	32.79	37.61	0.00	4.40	4.40	22.69	27.22	22.69	27.22
organization	34.08	38.93	43.46	38.93	43.46	38.96	43.99	49.13	43.45	38.91	38.97	44.14	49.15	53.75	53.75	49.02	53.84	16.24	0.00	20.64	38.93	43.46	38.93	43.46
country	34.08	38.93	43.46	38.93	43.46	38.96	43.99	49.13	43.45	38.91	38.97	44.14	49.15	53.75	53.75	49.02	53.84	16.24	20.64	0.00	38.93	43.46	38.93	43.46
agent	17.97	22.82	27.35	22.82	27.35	22.85	27.88	33.02	27.34	22.80	22.86	28.03	33.04	37.64	37.64	32.91	37.73	23.07	27.47	27.47	0.00	4.53	22.82	27.35
legal agent	33.95	38.80	43.33	38.80	43.33	38.83	43.86	49.01	43.33	38.79	38.85	44.02	49.03	53.63	53.63	48.90	53.72	39.05	43.46	43.46	15.99	0.00	38.80	43.33
group	17.97	22.82	27.35	22.82	27.35	22.85	27.88	33.02	27.34	22.80	22.86	28.03	33.04	37.64	37.64	32.91	37.73	23.07	27.47	27.47	22.82	27.35	0.00	4.53
group of people	33.95	38.80	43.33	38.80	43.33	38.83	43.86	49.01	43.33	38.79	38.85	44.02	49.03	53.63	53.63	48.90	53.72	39.05	43.46	43.46	38.80	43.33	15.99	0.00

Matriz 13. Matriz resultante de pesos conceptuales después de 7 iteraciones.

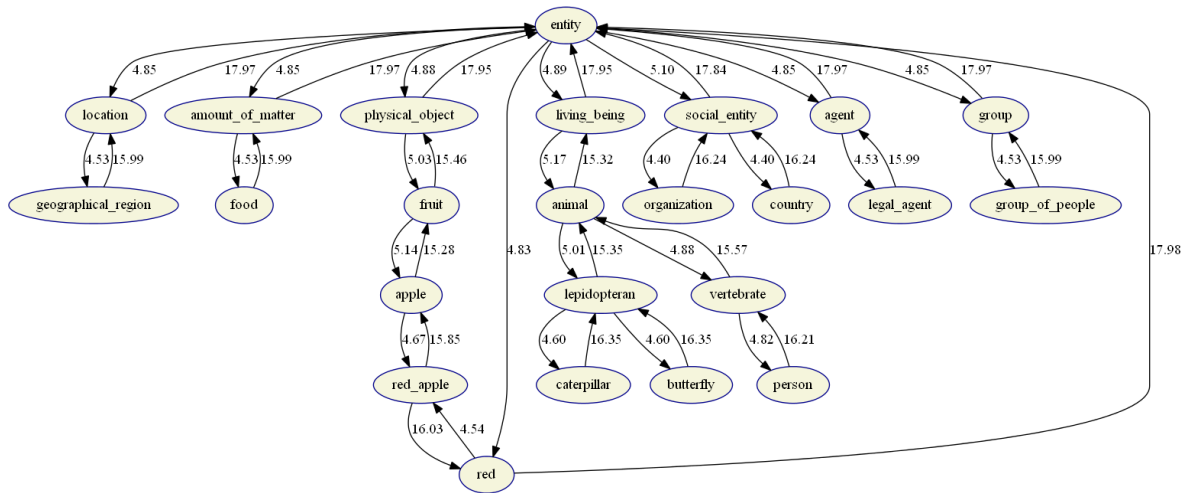


Figura 56. Ontología de la Figura 55 con valores de pesos conceptuales. Utilizando como referencia la Matriz 13.

```

v_agent: 1.3695707789980087
v_amount_of_matter: 1.3695707789980083
v_animal: 0.9952932401842076
v_apple: 0.9848510266940615
v_butterfly: 0.6304159273523497
v_caterpillar: 0.6304159273523497
v_country: 0.9563005888210426
v_entity: 0.29874936030472604
v_food: 0.9572303631434982
v_fruit: 1.0337381674099542
v_geographical_region: 0.9572303631434982
v_group: 1.3695707789980087
v_group_of_people: 0.9572303631434982
v_legal_agent: 0.9572303631434982
v_lepidopteran: 0.7670284986279513
v_living_being: 1.4578442709368449
v_location: 1.3695707789980083
v_organization: 0.9563005888210426
v_person: 0.63632781574046
v_physical_object: 1.432182980849812
v_red: 1.3399536887661319
v_red_apple: 0.9089098434636521
v_social_entity: 1.4042152521838098
v_vertibrate: 0.7636069158646502
    
```

Figura 57. Valores resultantes de visibilidad de la Ontología No. 7.

Tipo: 0, Directa: 4.16, Inversa: 14.26

Figura 58. Valores de distancia conceptual para cada tipo de relación subyacente en la conceptualización después de 7 iteraciones.