



INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN



Recuperación e integración de datos geoespaciales
mediante alineación de ontologías

TESIS

Que para obtener el grado de:
Doctorado en Ciencias de la Computación

P r e s e n t a :

M en C. JULIO CESAR VIZCARRA ROMERO

Directores de tesis:

DR. MIGUEL JESÚS TORRES RUIZ

DR. ROLANDO QUINTERO TÉLLEZ

ENERO, 2016



INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 10:00 horas del día 01 del mes de diciembre de 2015 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:

"Recuperación e integración de datos geoespaciales mediante alineación de ontologías"

Presentada por el alumno:

VIZCARRA

RÓMERO

JULIO CÉSAR

Apellido paterno

Apellido materno

Nombre(s)

Con registro:

A	1	2	0	4	4	0
---	---	---	---	---	---	---

aspirante de: **DOCTORADO 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. Miguel Jesús Torres Ruiz



Dr. Rolando Quintero Téllez



Dr. Grigori Sidorov



Dr. Francisco Hiram Calvo Castro



Dr. Miguel Félix Mata Rivera



Dr. Marco Antonio Moreno Ibarra

PRESIDENTE DEL COLEGIO DE PROFESORES





Dr. Alfonso Villa Vargas



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México D.F. el día 11 del mes de diciembre del año 2015, el que suscribe Vizcarra Romero Julio Cesar alumno del Programa de Doctorado en Ciencias de la Computación con número de registro A120440 , adscrito al Centro de Investigación en Computación , manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Miguel Jesús Torres Ruiz y del Dr. Rolando Quintero Téllez y cede los derechos del trabajo intitulado "recuperación e integración de datos espaciales mediante alineación de 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 kaivzro@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Vizcarra Romero Julio Cesar

RESUMEN

Una manera de realizar búsquedas, es el escenario donde un individuo genera representaciones conceptuales de forma automática por medio de conceptos y relaciones, limitando el espacio de conocimiento dónde se realizan dichas búsquedas. En la actualidad, existe un creciente manejo de conceptos especializados en diversos ámbitos que han surgido con la inclusión de la tecnología y la ciencia en el procesamiento de las consultas para la recuperación de las fuentes de datos geoespaciales. Con la finalidad de representar explícitamente la semántica de un conjunto de términos, las ontologías han sido ampliamente utilizadas, dichas ontologías son útiles en los sistemas de gestión del conocimiento y en bases de conocimiento.

El principal objetivo de esta tesis es proveer una metodología de recuperación de datos geoespaciales con un enfoque basado en conocimiento, relacionado a una consulta con las fuentes de información, usando conceptos de uso común y en dominios especializados. Como objetivos particulares se tienen: 1) la construcción automática de ontologías multi-dominio, las cuales son generadas en forma no ambigua y con una dimensión mínima considerando textos no estructurados; 2) crear un mecanismo de colaboración entre ontologías en una base de conocimiento para enriquecer descripciones de consulta y fuentes de datos.

La metodología propuesta se compone, en general, de tres etapas, a saber: la conceptualización, la síntesis y el análisis. La etapa de conceptualización se encarga de generar ontologías multi-dominio automáticamente, tomando un texto no estructurado, en esta etapa son involucrados procedimientos de expansión multi-dominio, métricas de distancia conceptual y procesamiento de árboles Steiner para la creación de la ontología. En la etapa de síntesis se concentra cada representación (ontología) de cada objeto geoespacial, mediante un repositorio de ontologías, además de la descripción y el acceso del objeto; así mismo se provee un mecanismo de control y gestión para su recuperación. Finalmente, la etapa de análisis se encarga de usar la conceptualización para construir una ontología que representa a un conjunto de conceptos de búsqueda; posteriormente esta ontología de consulta se compara con las ontologías de la etapa de síntesis usando tres tipos de similitud por alineación (léxica, estructural y ontológica). Como resultado final se recupera una lista ordenada decrecientemente, de acuerdo con su similitud con las fuentes de datos geoespaciales relacionados a la consulta.

ABSTRACT

One form for searching is the scenario where a person generates conceptual representations automatically by means of concepts and relationships limiting the knowledge space where conduct such searches. In addition, there is a growing specialized management of concepts in various areas that have emerged with the inclusion of technology and science in the query processing for recovery of geospatial data sources. In order to explicitly represent the semantics of a set of terms, ontologies have been widely used, such ontologies are used in knowledge management systems and knowledge base.

The main objective of this thesis is to provide a method for recovery of geospatial data with a knowledge-based approach, relating to consultation with the sources of information, using concepts and commonly used in specialized domains. As particular objectives we have: 1) the automatic construction of multi-domain ontologies which are generated unambiguously and with a minimum dimension considering unstructured texts; 2) create a mechanism for collaboration between ontologies in a knowledge base to enrich descriptions of consultation and data sources.

The proposed methodology consists generally of three stages, namely: conceptualization, synthesis and analysis. Conceptualization stage is responsible for generating multi-domain ontologies automatically taking an unstructured text, at this stage are involved procedures like multi-domain expansion, conceptual distance metrics and processing Steiner trees for the creation of ontology. In the synthesis stage each representation (ontology) of each geospatial object is concentrated by a repository of ontologies, along with the description of the object and access; also a control mechanism for recovery and management is provided. Finally, the analysis stage uses the conceptualization to build an ontology representing a set of search terms; this ontology query subsequently compared ontologies synthesis step using three types of similarity by alignment (lexical, structural and ontological). As a final result an ordered list decreasingly according to their similarity to spatial data sources related to the query retrieves.

AGRADECIMIENTOS

Este trabajo se lo dedico con todo mi amor a mis padres.

Gracias mama por todo el apoyo moral, buenos consejos, palabras de aliento que he tenido siempre de tu parte. Gracias por nunca dejarme solo y darme la vida, así como haberme inculcado los valores que me convierten en una persona de bien.

Gracias papa † por haberme guiado a un mejor sendero, por ser un ejemplo a seguir, inspirarme a continua en la vida y ayudarme hasta los últimos momentos. Tu recuerdo vivirá por siempre.

Papas cada logro mío es un logro de ustedes.

Ab Imo Pectore

Gracias a mis asesores Dr. Rolando Quintero y Dr. Miguel Torres por guiarme al término de la tesis, mejorar mis ideas y la escritura de tesis y artículos.

Agradezco a todos los miembros del jurado que con sus comentarios, sugerencias y opiniones enriquecieron el desarrollo de la tesis.

Et monere et moreni propium est verae amicitiae

Gracias a mi amigo Víctor por hacer mi estancia en el centro más amena con su amistad.

Sine Amicitia, vita esse nullam

Benignus omnibus gratias

Resumen	iv
Abstract.....	v
Agradecimientos	vi
Índice	vii
Índice de figuras	xi
Índice de tablas	x
Índice de algoritmos	x

ÍNDICE

1. ANTECEDENTES	1
1.1. INTRODUCCIÓN	1
1.2. DESCRIPCIÓN DEL PROBLEMA.....	4
1.3. OBJETIVOS	6
1.3.1. <i>Objetivo general</i>	6
1.3.2. <i>Objetivos particulares</i>	6
1.4. HIPÓTESIS DE LA PROPUESTA.....	7
1.5. ORGANIZACIÓN DE LA TESIS	8
2. TRABAJOS RELACIONADOS.....	10
2.1. RECUPERACIÓN DE INFORMACIÓN.....	10
2.1.1. <i>Enfoques basados en ontologías</i>	10
2.1.2. <i>Enfoques basados en anotaciones semánticas</i>	14
2.1.3. <i>Enfoques basados en modelos de similitud semántica</i>	18
2.1.4. <i>Enfoque basado en minería de datos</i>	18
2.2. INTEGRACIÓN DE INFORMACIÓN	19
2.2.1. <i>Integración basada en ontologías</i>	19
2.2.2. <i>Integración basada en descripciones semánticas</i>	21
2.3. INTEROPERABILIDAD E INTERCAMBIO	23
2.3.1. <i>Interoperabilidad basada en ontologías</i>	23
2.3.2. <i>Interoperabilidad basada en anotaciones</i>	25
2.4. CONSTRUCCIÓN DE ONTOLOGÍAS	26
2.4.1. <i>Construcción automática de ontologías</i>	26
2.4.2. <i>Construcción semi-automática de ontologías</i>	28
2.4.3. <i>Construcción manual de ontologías</i>	29
3. MARCO TEÓRICO.....	32
3.1. INTRODUCCIÓN	32
3.2. ONTOLOGÍAS	32
3.2.1. <i>El enfoque de Gruber</i>	34

3.2.2.	<i>El enfoque de Guarino</i>	36
3.3.	KROVETZ STEMMER (KSTEM).....	40
3.4.	WORDNET.....	41
3.4.1.	<i>Diseño y contenido</i>	41
3.4.2.	<i>Cobertura</i>	42
3.4.3.	<i>Relaciones</i>	42
3.4.4.	<i>Apuntadores</i>	43
3.5.	ÁRBOL STEINER.....	44
3.5.1.	<i>Versión no dirigida</i>	45
3.5.2.	<i>Versión dirigida</i>	45
3.5.3.	<i>Problema generalizado de Steiner dirigido de red</i>	45
3.5.4.	<i>Problema del árbol Steiner de grupo</i>	46
3.5.5.	<i>Algoritmos de aproximación</i>	46
3.6.	ALGORITMO DE APROXIMACIÓN SKETCHLS.....	48
3.7.	CÁLCULO DE LA VISIBILIDAD DE CONCEPTOS EN ONTOLOGÍAS.....	50
4.	METODOLOGÍA	53
4.1.	MARCO DE TRABAJO GENERAL.....	53
4.2.	ETAPA DE CONCEPTUALIZACIÓN.....	55
4.2.1.	<i>Preprocesamiento léxico</i>	56
4.2.2.	<i>Expansión sobre la base de conocimiento multi-dominio</i>	57
4.2.2.1.	<i>EXPANSIÓN ONTOLÓGICA EN LA ONTOLOGÍA BASE</i>	58
4.2.2.2.	<i>EXPANSIÓN ONTOLÓGICA EN ONTOLOGÍAS EXTERNAS</i>	59
4.2.3.	<i>Medida de similitud</i>	60
4.2.4.	<i>Desambiguación y reducción de la dimensión de la ontología base</i>	61
4.2.5.	<i>Fusión de la ontología base y ontologías externas</i>	63
4.3.	ETAPA DE SÍNTESIS.....	66
4.4.	ETAPA DE ANÁLISIS.....	69
5.	RESULTADOS EXPERIMENTALES	73
5.1.	INTRODUCCIÓN.....	73
5.2.	BASES DE CONOCIMIENTO MULTI-DOMINIO.....	75
5.2.1	<i>Ontología base</i>	75
5.2.2	<i>Ontologías externas</i>	75
5.3.	RESULTADOS ETAPA DE CONCEPTUALIZACIÓN.....	76
5.3.1.	<i>Procesamiento léxico</i>	76
5.3.2.	<i>Expansión de la ontología</i>	76
5.3.3.	<i>Reducción de dimensión y desambiguación</i>	79
5.3.4.	<i>Fusión de la ontología base y ontologías externas</i>	80
5.4.	RESULTADOS ETAPA DE SÍNTESIS.....	84
5.5.	RESULTADOS ETAPA DE ANÁLISIS.....	86
5.6.	DIAGRAMA DEL SISTEMA.....	95
5.7.	CONSIDERACIONES GENERALES.....	96
6.	CONCLUSIONES	98

6.1.	APORTACIONES	99
6.2.	LIMITACIONES.....	100
6.3.	TRABAJO A FUTURO	100
REFERENCIAS		102
ANEXOS		109
ANEXO 1.	FRAGMENTO DE NOTICIAS	109
ANEXO 2.	LISTADO DE PALABRAS USADAS COMO STOP WORDS EN INGLÉS	112
ANEXO 3.	CONSTRUCCIÓN DE LA ONTOLOGÍA DE CONSULTA EN EL SISTEMA DE RECUPERACIÓN	114
ANEXO 4.	COMPARACIÓN POR ALINEACIÓN DE LA ONTOLOGÍA DE CONSULTA EN EL SISTEMA DE RECUPERACIÓN	122
ANEXO 5.	OBJETOS GEOESPACIALES RECUPERADOS	124

ÍNDICE DE FIGURAS

Figura 1.1	Descripción general del problema. _____	5
Figura 1.2	Esquema general de integración semántica. _____	8
Figura 3.1	Tipos de ontologías de acuerdo con su nivel de dependencia. _____	39
Figura 3.2	Fragmento de la ontología con relaciones axiomáticas entre conceptos _____	43
Figura 3.3	Fragmento de la ontología con relaciones axiomáticas entre conceptos _____	43
Figura 3.4	Fragmento de la ontología con relaciones axiomáticas entre conceptos _____	45
Figura 3.5	Sketch de q_1, q_2, q_3 . _____	49
Figura 3.6	Una arista entre v y l_2 es descubierta, q_1 está conectado a q_3 . _____	49
Figura 3.7	BFS1 y BFS2 también se reúnen (a través de l_3). _____	50
Figura 3.8	El grafo Steiner resultante es resaltado. _____	50
Figura 3.9	Ontología ejemplo. _____	52
Figura 4.1	Estructura conceptual de la metodología. _____	53
Figura 4.2	Esquema general de la metodología. _____	55
Figura 4.3	Esquema general de la fase de conceptualización. _____	56
Figura 4.4	Preprocesamiento léxico y expansión sobre la base de conocimiento. _____	57
Figura 4.5	Esquema general de la Fase de conceptualización. _____	58
Figura 4.6	Algoritmo DIS-C. _____	61
Figura 4.7	Algoritmo. SketchLs para Steiner Tree. _____	61
Figura 4.8	Procesamiento de D-Stainer a Group D-Stainer. _____	62
Figura 4.9	Ontología final multi-dominio. _____	64
Figura 4.10	Fase de Síntesis. _____	67
Figura 4.11	Fase de Análisis. _____	71
Figura 5.1	. Caso de estudio General. _____	73
Figura 5.2	Caso de estudio general para metadatos. _____	74
Figura 5.3	Caso de estudio particular delimitado a un dominio. _____	75
Figura 5.4	Expansión en la ontología. _____	77

Figura 5.5 Grafo DIS-C para “mouse - computer”.	78
Figura 5.6 Agrupación de terminales del grafo DIS-C .	79
Figura 5.7 Ontología de “mouse - computer”.	80
Figura 5.8 El concepto “lixisol” en la ontología ENVO.	81
Figura 5.9 Grafo desambiguado-reducido y ontología para “tree-lixisol”.	83
Figura 5.10 Ontología multi-dominio para “tree-lixisol”.	84
Figura 5.11 Fase de Síntesis.	85
Figura 5.12 Fase de Análisis.	87
Figura 5.13 Esquema de consultas y noticia-ontología.	87
Figura 5.14 Entrada al sistema.	88
Figura 5.15 . Consulta al sistema.	88
Figura 5.16 Primeros 10 resultados para la consulta por comparación léxica.	89
Figura 5.17 Primeros 10 resultados para la consulta por similitud estructural.	90
Figura 5.18 Primeros 10 resultados para la consulta por similitud léxica.	90
Figura 5.19 Primeros 10 resultados para la consulta por similitud ontológica.	91
Figura 5.20 Precisión para las 3 medidas de similitud	92
Figura 5.21 Recall para las 3 medidas de similitud	92
Figura 5.22. F-measure para las 3 medidas de similitud	93
Figura 5.23 Relación Precision/Recall para la de similitud estructural.	93
Figura 5.24 Relación Precision/Recall para la de similitud léxica.	93
Figura 5.25 Relación Precision/Recall para la de similitud ontológica.	94
Figura 5.26 Localización geográfica de la noticia	94
Figura 5.27 Noticias cercanas espacialmente a la noticia	95
Figura 5.28 Ver canal emitido	95
Figura 5.29 Implementación del sistema.	96

ÍNDICE DE TABLAS

Tabla 5.1 Resumen expansión ejemplo 1, computer y mouse	76
Tabla 5.2 Sentidos de mouse	80
Tabla 5.3 Sentidos de computadora	80
Tabla 5.4 Procesamiento léxico y alineación de tree y lixisol	82
Tabla 5.5 Sentidos de soil	82
Tabla 5.6 Resumen expansión ejemplo 2, tree y lixisol	82
Tabla 5.7 Sentidos para tree	84
Tabla 5.8 Esquema de la base de datos donde se guardan noticias/canales	85
Tabla 5.9 Noticia/Id. Relación	86

ÍNDICE DE ALGORITMOS

Algoritmo 1 Expansión para ontología base (AOEB)	59
Algoritmo 2 Expansión sobre la ontología externa (AOEE)	60

<i>Algoritmo 3 Group D-Steiner (AGST)</i>	63
<i>Algoritmo 4. Fusión de las ontologías</i>	66
<i>Algoritmo 5 Síntesis petición de ontologías</i>	68
<i>Algoritmo 6 Síntesis petición de objetos espaciales</i>	69
<i>Algoritmo 7 Análisis</i>	72

1. Antecedentes

1.1. Introducción

En los años recientes un gran volumen de fuentes de información geográfica han llegado a estar disponibles en Internet, estas son tanto diversas, distribuidas y heterogéneas (como por ejemplo, bases de datos, bases de conocimiento, colección de documentos, entre otras), por lo que para habilitar la interoperabilidad, los sistemas remotos deben, no solo poder localizar y acceder a estas fuentes de información, sino también interpretar y procesar los datos recuperados [1].

En el contexto de los Sistemas de Información Geográfica (SIG), la complejidad y riqueza de los datos geográficos y la dificultad de su representación producen una serie de problemas muy específicos, para lograr la interoperabilidad con este tipo de datos [2]. Algunos trabajos han sugerido que estructuras sofisticadas como las ontologías son buenas candidatas para representar y resumir los datos geográficos, a la par de lograr que los mismos puedan ser compartidos mediante una conceptualización [3] [4].

Así, la *integración semántica* de los datos geoespaciales se puede describir como el proceso de integración de los atributos de diferentes fuentes de datos geoespaciales. Esta tarea puede ser un método complicado y amplio, con una solución no sencilla o una solución correcta. Uno de los principales retos de la *integración semántica* es la gestión de las diferencias semánticas o la heterogeneidad semántica entre los datos geoespaciales. Por tanto, con la integración se busca evitar la heterogeneidad, mediante la aplicación de un conjunto de métodos y técnicas.

El propósito de la integración semántica es compartir información entre diferentes usuarios o dominios de aplicación. Hoy en día, existen grandes cantidades de datos geoespaciales disponibles, pero el mayor problema se presenta en combinar estos datos para que puedan funcionar de manera adecuada en un SIG.

El problema de la integración de fuentes de datos heterogéneas y en algunos casos no estructurados, conlleva a problemas relacionados directamente con la *recuperación de la información geográfica*. No obstante diversos métodos han sido propuestos utilizando como medio fundamental las ontologías [5] [6] [7]. En los últimos años, el problema de la precisión en la recuperación de información puede ser visto como consecuencia de la falta de significado o de semántica [8], tanto en la *web* como en las bases de datos y repositorios heterogéneos no estructurados de información.

Los datos y la integración de la información, así como la interoperabilidad han ganado cada vez más atención por varias razones, a saber, una de ellas es el aumento de la interconexión que ofrece la *web* y todos los datos estructurados y semi-estructurados que son accesibles desde la misma. Esto conlleva a un acceso fácil a un gran número de fuentes autónomas, pero heterogéneas de gran variedad. Además, los progresos en el trabajo especializado y la consiguiente necesidad de aumentar la reutilización y el análisis de los datos conducen a la creación de información y conocimiento [9].

Con respecto a la distribución de los datos, los sistemas de bases de datos múltiples fueron el principal tema de investigación asociado con la interoperabilidad. Estos sistemas involucraban sólo algunas bases de datos y nodos de computación, ya sea conectado punto a punto o en una red de área local. Con los avances en la tecnología de las comunicaciones y la llegada de la *web*, la perspectiva de la interoperabilidad se ha ampliado a un ámbito más global dejando atrás este tipo de desarrollos [10].

Los desarrolladores de nuevos sistemas hacen frente a varios retos, ya que normalmente todos los procesos en el sistema de distribución están ocultos a los usuarios finales. Ahora, existe un incremento en la cantidad y el número de categorías de los datos, porque se tiene que manejar información multimedia, tal como del tipo visual, sonido, gráficos, geográficos, entre otros. El número de nodos involucrados aumenta y la tecnología aplicada en los sistemas utiliza un ancho de banda mucho más amplio que antes; por lo cual, es necesario aumentar el nivel de conocimiento de la información, así como lograr que los datos se conviertan en la información que se trata de integrar [11].

En los últimos años, la investigación se ha enfocado en la solución de la heterogeneidad. Uno de los problemas de esto y quizás el más difícil de resolver es la *heterogeneidad*

semántica. Con la aparición de la tecnología de *middleware*, la aplicación de las normas y la computación distribuida han permitido a los desarrolladores hacer frente a los problemas a nivel de información y no sólo a nivel sintáctico y estructural. Sin embargo, investigaciones relacionadas con la semántica apuntan a comprender, descubrir y extraer la información esencial de acuerdo con un contexto o dominio de aplicación, como una técnica para la integración.

Por otra parte, la amplia variedad de paquetes de software y modelos de datos, han creado una demanda de mecanismos que permitan el intercambio de información almacenada en diferentes fuentes. Por tanto, en un intento de lograr la interoperabilidad y heterogeneidad semántica se han propuesto diversas propuestas con diferentes técnicas de integración, tales como [12] [13] [14].

La integración de datos es una de las principales áreas, tanto de investigación como de negocio, que se encarga de permitir a los usuarios acceder a datos almacenados en fuentes de datos heterogéneas, presentando una única vista unificada de esos datos, de forma que el usuario no llegue a percibir esa heterogeneidad [15].

Existen múltiples ejemplos de la importancia que tiene la integración de datos en la vida diaria en todos los ámbitos. A continuación se muestran dos de ellos:

1. La cantidad de fuentes de información que encontramos en la *web* es cada vez más extensa y más heterogénea. Aquí el problema de la integración de la información está presente en cada consulta que realiza el usuario [16].
2. La cantidad de grandes empresas que nacen como unión de empresas más pequeñas, tienen la necesidad de integrar y compartir una gran cantidad de información procedente de numerosas fuentes de datos diferentes.

Las fuentes de datos heterogéneas, se incrementan a diario, al igual que los datos pobres, inexactos, incompletos, inconsistentes, desactualizados, duplicados, etc. En resumen, datos carentes de calidad. Por eso, sin técnicas adecuadas para integrar y fusionar tales datos, la calidad de los datos en la información final resulta baja [17].

Aunque hasta ahora no se haya referenciado, la heterogeneidad semántica es uno de los retos clave en la integración de datos procedentes de distintas fuentes. La semántica trata el significado de los datos, en contraste con la sintaxis, que sólo tiene en cuenta la adecuación de los datos a la estructura en la que se encuentran guardados. Según [18], semántica es la interpretación que las personas atribuyen a los datos, de acuerdo con su entendimiento del mundo real. En el área de las bases de datos, la semántica se refiere a la interpretación que las personas dan a los datos y a los atributos que referencian esos datos dependiendo del contexto.

La integración semántica es el área de investigación que se centra en contrastar y decidir sobre la similitud de los datos provenientes de distintas fuentes, usando herramientas basadas en la semántica [19].

El objetivo de este trabajo es crear una metodología de integración semántica para resolver los problemas de heterogeneidad semántica en fuentes de datos heterogéneos distribuidos en la *web*, facilitando la recuperación con una mejora calidad de datos recuperados cercanos a lo deseado con mejor precisión.

1.2. Descripción del problema

Hoy en día, en diferentes áreas del conocimiento, en diversos desarrollos tecnológicos y más aún en la *web*, se presenta el problema de la *heterogeneidad e interoperabilidad semántica* de la información; aunado a que las conceptualizaciones son diferentes, debido al conocimiento particular sobre un dominio (expertos o usuarios comunes). La problemática a resolver radica en la diversificación de los datos geoespaciales en diferentes conceptualizaciones, modos de representación con diversos significados, lo cual incrementa en tiempo, esfuerzo y complejidad la búsqueda y recuperación. De acuerdo con lo anterior, el problema de la heterogeneidad semántica origina una ineficiente y poco confiable recuperación de información geoespacial.

Con base en lo anterior, en la Figura 1.1 se muestra de manera esquemática el problema a resolver, se puede apreciar cómo un usuario desea recuperar una serie objetos geoespaciales

localizados dentro de *web*. El usuario inicialmente conceptualiza un conjunto de conceptos relevantes, así como un conjunto de relaciones entre ellos y a otros conceptos, esto con el fin de establecer un contexto y formular una consulta, dentro de este procesamiento se genera un extracto de conocimiento que a su vez es un subconjunto del conocimiento general el cual posee el individuo, es decir, la forma y modo en cómo en que se defina la consulta depende directamente del dominio del tema y cuanto se conoce.



Figura 1.1 Descripción general del problema.

Una vez construida la consulta del usuario, ésta se envía a la *web*. Las distintas fuentes de información son también construidas bajo un propósito y son descritas por una persona con conocimiento y *expertise* específico en el tema, estas descripciones pueden ser definidas por medio de metadatos o simplemente textos planos. Las fuentes de datos descritas por metadatos pueden ser: una base de datos geoespacial, un documento *web*, descripciones servicios web geoespaciales, un servidor, noticias Geo RSS, conjunto de gráficas o datos analizados, etc.

Cada fuente de información fue creada para un fin dentro del contexto y es descrita generalmente por un especialista en el dominio e igualmente interviene el conocimiento o *expertise* de quien describe los objetos geoespaciales.

Por lo anterior, es necesario crear una metodología para la recuperación de conjuntos de datos geoespaciales, esto con un enfoque basado en conocimiento para incrementar o definir mejor el conocimiento relacionado con cada objeto geoespacial y la consulta.

Con este trabajo de tesis se pretende recuperar información espacial mediante un enfoque semántico, basado en conocimiento de distintas fuentes de datos heterogéneas sobre la *web*, los resultados pueden ser jerarquizados de acuerdo con la consulta del usuario. Para este propósito, se propone el diseño de un método que cuantifique la *similitud conceptual* por medio de métricas de *alineación* entre las diferentes representaciones (ontologías) de los datos geoespaciales, además se puedan establecer los criterios de relevancia para recuperar la información geoespacial.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar una metodología para la búsqueda y recuperación de datos geoespaciales, basada en métodos de integración semántica (alineación) y utilizando un conjunto de diversas representaciones conceptuales (ontologías).

1.3.2. Objetivos particulares

Los objetivos de esta propuesta de tesis se enfocan principalmente en:

- Proponer una metodología de recuperación de datos geoespaciales con un enfoque de representación basado en conocimiento.

- Proponer un algoritmo para construir ontologías automáticamente, basado en una base de conocimiento multi-dominio, así como métricas de comparación entre ontologías basadas en alineación.
- Implementar un caso de estudio para la recuperación de datos geoespaciales, a partir de objetos diversos y heterogéneos, para devolver a éstos en un orden de relevancia de acuerdo con una consulta.

1.4. Hipótesis de la propuesta

De acuerdo con la descripción del problema, se ha considerado como hipótesis que la recuperación de datos geoespaciales se puede realizar de forma inteligente y basada en conocimiento, considerando un método de integración que considere medir la similitud, por alineación de diversas representaciones conceptuales que refieren a objetos geográficos. Todo esto con base en diversos dominios, en donde la información puede variar en su significado. Considerando además, que la información es heterogénea, en diversos formatos y formas de representación, sin importar las características implícitas de los objetos geográficos, tales como la escala, el sistema de referencia, la proyección, entre otras.

Para la recuperación de información espacial se necesitan procesos de *descubrimiento*, *integración* y *recuperación* de datos geoespaciales en la *web*, además de la construcción de ontologías de manera automatizada sobre una base de conocimiento multi-dominio.

En la Figura 1.2 se muestra el esquema general del objetivo de la tesis, donde el conocimiento relacionado a las fuentes de información y la consulta están descritos por medio de una ontología. La metodología debe ser capaz de interpretar y comparar estos extractos de conocimiento, con la finalidad de generar una lista ordenada con base en la similitud de la consulta y los objetos espaciales, la cual será devuelta al usuario.

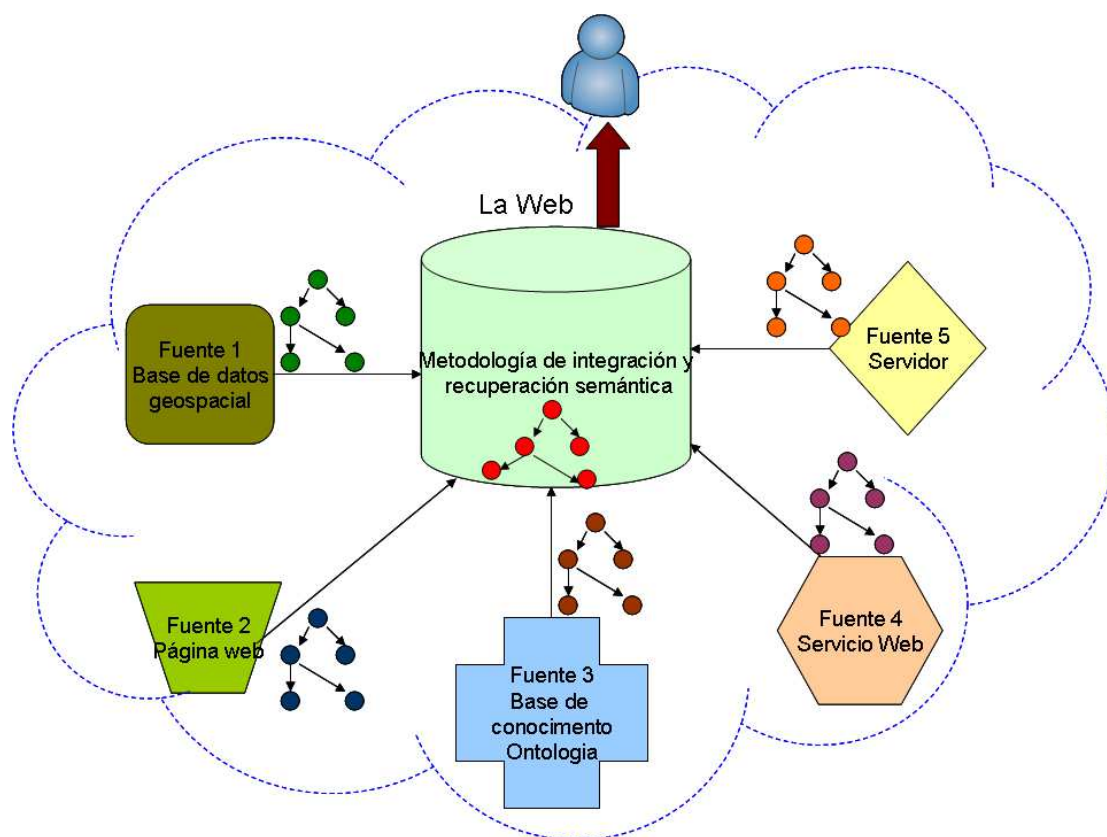


Figura 1.2 Esquema general de integración semántica.

1.5. Organización de la tesis

El resto del trabajo se ha organizado de la siguiente manera: en el Capítulo 2 se presenta un análisis relacionado con el estado del arte, es decir, la relación con algunos trabajos relacionados a la metodología propuesta.

En el Capítulo 3 se describe el marco teórico de la tesis, considerando particularmente las metodologías y herramientas necesarias y utilizadas como parte de la metodología propuesta, entre las cuales están tópicos enfocados en ontologías, WordNet, métodos como el árbol de Steiner, distancia conceptual DIS-C y la visibilidad de conceptos en ontologías.

El Capítulo 4 presenta una descripción formal de la metodología propuesta para la recuperación semántica de información geoespacial con base en extractos de conocimiento, éstos representados por ontologías sobre diversas fuentes de datos heterogéneas.

En el Capítulo 5 se presentan los resultados obtenidos implementando la metodología, se presentan pruebas de escritorio y salidas gráficas de las etapas que comprenden la metodología. Además, se presenta un análisis con gráficas que representan a las métricas *precision*, *recall* y *F-measure*; así como resultados del *performance* en tiempo de ejecución.

En el Capítulo 6 se presentan las conclusiones, limitaciones, aportaciones y trabajos futuros que se pueden desarrollar.

Finalmente en la sección de anexos se puede encontrar la lista más amplia de los resultados, se presenta la lista de *stop-words* utilizados, las salidas del sistema en el procesamiento de construcción de ontologías y resultados de la recuperación de objetos geoespaciales.

2. Trabajos Relacionados

2.1. Recuperación de información

2.1.1. Enfoques basados en ontologías

El trabajo de Zhan, Q. [20] está enfocado en el descubrimiento y recuperación de datos espaciales en entornos distribuidos sobre infraestructuras de datos espaciales. Para resolver la problemática se propuso un modelo de descripción semántica basado en una ontología, la cual representa explícitamente la información geográfica en un nivel abstracto y concreto. El modelo consiste en tres partes: perfil del dato, contenido del dato y enlace del dato. La primera parte describe de manera abstracta al usuario, el contenido del dato describe en un nivel concreto el contenido del dato y finalmente el enlace del dato indica al usuario como acceder al dato.

En la metodología para el descubrimiento y recuperación se utilizó una ontología del tipo híbrida [10] de alto nivel o general que comparte descripciones de otras ontologías. La implementación se efectúa por medio de clases, mediante un algoritmo compuesto por tres pasos: en el primero los términos de búsqueda del usuario se asignan a los conceptos en la ontología de dominio, basado en una ontología híbrida. En el segundo paso, los conceptos se amplían sobre la base de la jerarquía de conceptos en la ontología de dominio. Entonces, si los conceptos tienen subclases, a continuación se añaden las subclases para ampliar la consulta; de lo contrario, la superclase se marca. Si hay una superclase, ésta se añade al ampliar la consulta. En el tercer paso se ejecuta la expansión de la consulta, se buscan las descripciones de información geográfica y éstas son devueltas a los usuarios. Si los resultados son adecuados, la búsqueda ha terminado. De lo contrario, se repite la segunda etapa. Como resumen del trabajo se presenta una solución para el descubrimiento y recuperación de datos por medio de su semántica y no por palabras clave y como resultado la consulta del usuario es procesada con mayor precisión con respecto al significado.

En [21] se presenta una nueva métrica para medir la relación semántica entre las palabras. Su enfoque está basado en ontologías para representar el conocimiento general en la construcción de una red semántica. A su vez, se basa en las propiedades lingüísticas y métricas para establecer una medida de relación semántica. Se propone un sistema de

recuperación de información web llamado DySe para ponderar documentos digitales en relación a los intereses del usuario. Este sistema implementa un enfoque de contexto, las palabras clave son procesadas según su contexto. En la búsqueda, el usuario debe especificar en la consulta un conjunto de palabras clave a ser buscadas y se especifica el dominio al que pertenecen y sobre estos criterios se recuperan páginas web, todo esto es llevado a cabo con una base de conocimiento desde el punto de vista lingüístico la cual es WordNet.

El sistema realiza cuatro funciones: *obtención* donde se recuperan documentos web relacionados con las palabras clave de la búsqueda, en el *pre-procesado* se descartan todos los elementos que no representan información útil, se realiza *minería* con un análisis semántico, asignando una calificación con respecto a la consulta y por último la *devolución* de resultados ponderados para establecer el orden de relevancia con respecto a la consulta. Para la creación de las similitudes en la red, se realiza con una ontología WordNet, la cual se crea dinámicamente, entonces la red semántica toma en consideración los hipónimos, para establecer valores de peso para cada relación en WordNet en la construcción de la red semántica dinámica. Para el proceso de establecer pesos se calcula la relevancia de un término considerando la propiedad de polisemia; es decir, el número de posibles significados para ese término. Para establecer el peso de relevancia en la red semántica, éste se determina mediante la fórmula $w(i) = 1/Polisemia(i)$. Después de establecer la relevancia con respecto a un dominio, la similitud de la consulta con la red semántica se determina vía una medición híbrida entre probabilidad y semántica. Esto se realiza de dos formas; a nivel de documento y de palabra, en la primera se establecen clases para los términos mediante sus frecuencias en el documento, la segunda a nivel de palabra se mide la longitud de dos términos dentro de la red semántica y se escoge la mejor ruta. Estos dos métodos anteriores consideran el componente sintáctico por frecuencia de los términos dentro del documento y el dominio de interés, además se considera el componente semántico por cada pareja de términos en el documento. En resumen, el trabajo presenta un sistema de recuperación web basado en ontologías, una red semántica dinámica, cadenas léxicas, y la definición de una estrategia para el ordenamiento de información y la clasificación de los resultados por medio de una métrica que pondera la relación semántica entre las palabras.

Otro trabajo relacionado es el de Kishor [22], el cual presenta un enfoque para la medición de similitud semántica entre palabras y en una estructura jerárquica. Esta métrica es usada

para representar el contenido de la información, además se presenta un mecanismo de búsqueda usando la API del Google, el cual expande la consulta del usuario por cada término de la consulta. Las palabras del usuario son remplazadas por sinónimos descubiertos por medio de una medida de similitud y éstas entran a la API de búsqueda de Google. El trabajo utiliza la base de datos léxica WordNet, se implementan las relaciones semánticas de hipónimos entre conceptos como relaciones de especialización de categorías conceptuales. El algoritmo implementado para la búsqueda es el siguiente: ingresar un par de palabras para medir su similitud semántica, se calcula la similitud semántica entre dos palabras con enfoque basado en contenido. Mientras más alto sea el valor numérico, mayor será la similitud entre ellos. La siguiente fase en el algoritmo la consulta del usuario es expandida con base en términos similares para la tarea de recuperación semántica, la expansión no solo toma sinónimos sino hiperónimos e hipónimos. Finalmente en el sistema, el mecanismo de búsqueda *web* acepta el conjunto de sinónimos, el cual es generado por el módulo de expansión como palabras clave extra que se incluyen a la consulta, además de las especificadas por el usuario, se usa el mecanismo de búsqueda de Google para suplementar la consulta con los términos analizados. En resumen, se presenta un método basado en el contenido de información para agregación de conceptos usando la jerarquía de WordNet, ya realizado el reemplazo de una consulta por un conjunto de sinónimos con respecto a una medida de similitud para la recuperación de información. La metodología sirve de apoyo para los usuarios para complementar sus consultas.

En el trabajo de Barathi [23] se introduce un enfoque de captura del contexto específico del usuario usando WordNet, basado en una medición de relación semántica de conjunto de términos con el número de ocurrencias de las palabras clave en la página *web*. Esto da como resultado la identificación del contexto de la consulta del usuario, una vez que la consulta identificada y formulada, es enriquecida para obtener más páginas *web* relevantes a las necesidades del usuario. Para establecer la relación entre un par de conceptos se realiza un recuento de bordes simples en la jerarquía IS-A de WordNet, se utiliza una función logarítmica para la medición de distancia de ruta de un concepto al otro. Se presenta un sistema de desambiguación llamado SWSIR, el cual cuenta con cuatro fases: la primera es indizar un conjunto de páginas *web*, las palabras vacías son removidas, las palabras clave son extraídas y guardadas con el fin de hacer una eficiente recuperación de páginas. El segundo paso es utilizar el algoritmo Leacock-Chodorow para el cálculo de

frecuencias y medir de similitud entre conceptos. El tercer paso es, mediante el algoritmo K-core, generar K-valores de relevancia. Finalmente la consulta del usuario es desambiguada y el contexto específico del usuario es identificado por medio del uso de una ontología sobre WordNet. En el algoritmo K-core está basado en clústeres semánticos de palabras, este clúster de términos interrelacionados en el sentido que ellos aparecen con un número de páginas *web*. La frecuencia de cada palabra es calculada por la generación de k-valores. En el sistema SWSIR se crea un archivo índice junto con una tabla de frecuencia de palabras clave, entonces el clúster semánticamente relacionado es generado, las medidas de similitud son procesadas para cada consulta de usuario, con base en la lista de conceptos resultantes se identifica el contexto y se refina la consulta. En resumen el SWSIR introduce un enfoque donde, tomando una consulta de usuario, se captura el contexto específico usando WordNet, basado en una medición de la relación y las frecuencia de apariciones de palabras clave en las páginas *web*. La consulta es enriquecida para obtener más páginas relevantes.

En el trabajo de Akmal [24] se crea una metodología para la recuperación de información de productos y la reutilización basado en una ontología. En dicho trabajo se presenta un enfoque basado en ontologías que pueden determinar la similitud entre dos clases, con base en sus características. Para la evaluación del enfoque propuesto se establece el estudio de caso sobre el servicio del sistema producto de problemas de diseño. La metodología se basa en el uso de ontologías obtenidas con FCA (Formal Concept Analysis) [53]. En el trabajo se desarrolla una medida de similitud sobre ontologías enfocada en la recuperación de información. En los experimentos se hizo comparación con otras medidas de similitud del estado del arte para medir su eficiencia en diferentes escenarios y ontologías de distintos tipos y dominios. Además de la comparación con otras similitudes, se propone un caso de estudio de recuperación sobre la *web* para evaluar todas las similitudes, esta evaluación es llevada a cabo con los motores de búsqueda proporcionados por Google Scholar y Scirus de Elsevier. El alcance de este experimento se limita a los procesos de eliminación de materiales mecánicos. Se construye una ontología con base en libros de texto y fuentes de información, después se aplican las medidas de similitud del estado del arte y la propuesta.

Para el caso de Lipani [25] se presenta una metodología de recuperación para nanopublicaciones con la finalidad de difundir información sobre los experimentos y los datos experimentales. Toda la información de las publicaciones se guarda por medio de tripletas

donde se describe el experimento, de donde proviene y la información de la publicación. Se diseña una ontología de recuperación de información con la finalidad de que el contenido de una nano-publicación sea expresado utilizando la ontología y se asegure un entendimiento común sobre las publicadas, declaraciones o afirmaciones. La ontología de dominio consiste en un vocabulario de conceptos específicos para dicho dominio y las relaciones entre ellos. En general, la ontología es usada como una capa intermedia que contribuye al módulo de comprensión del lenguaje natural; es decir, sirve de base para otros procesos y solo se puede aplicar sistemas razonadores para inferir información almacenada de publicaciones, la ontología genera una integración semántica de información de publicaciones para ser consultada y esta sea compartida.

2.1.2. Enfoques basados en anotaciones semánticas

Sanjay [26] introduce un marco de trabajo de anotaciones semánticas con el objetivo de integrar y recuperar información sobre recursos heterogéneos distribuidos sobre la *web*. Este proceso es enfocado sobre la semántica y los requerimientos del usuario. Para la construcción de las anotaciones se realizó una herramienta llamada KIM [27] como ambiente de desarrollo de lenguaje de anotaciones. Las anotaciones semánticas definen la base del conocimiento además de los procesos de descubrimiento de conocimiento y control del mismo, en este caso los conceptos son llamados entidades las cuales pueden ser descritas e interconectadas, en general se define un proceso de descubrimiento de piezas de conocimiento dentro de los datos sin o con poca ayuda humana. El control de la base de conocimiento es responsable de descubrir o manipular el conocimiento obtenido en las fuentes de datos heterogéneas. Las anotaciones semánticas proveen una más precisa descripción del conocimiento en un documento y en su semántica en el dominio; son realizadas sobre la asignación de las entidades en los enlaces de texto a sus descripciones semánticas. La anotación semántica automática permite: destacar, la indexación, la extracción de información y recuperación, el patrón de la entidad y el patrón de búsqueda inteligente de una plataforma *web* inteligente. Las anotaciones permiten agregar semántica a documentos no estructurados o semi-estructurados en la *web*, el proceso de tipificación de modelos semánticos y lenguaje natural se hace por medio de una vinculación dinámica por ontologías.

Los metadatos en los documentos *web* proporcionan información, tanto de clase como ejemplo, de las entidades/relaciones en la ontología. La anotación semántica automática permite a muchas aplicaciones como: relieve, categorización, búsqueda semántica, la generación de metadatos más avanzados, sin problemas de recorrido entre el texto no estructurado y el conocimiento formal [28]. Por otro lado, la plataforma de KIM proporciona la infraestructura y servicios para la anotación semántica automática, indexación y recuperación de documentos y permite un sistema escalable y personalizable basado en ontologías de extracción de información, así como la gestión de anotaciones y documentos. En resumen, se estudia la gestión del conocimiento y anotación semántica como una tecnología clave en un entorno de *web* inteligente, con la finalidad de presentar un marco de gestión del conocimiento, un marco para la anotación semántica y búsqueda semántica.

El trabajo de Sarah [29] se enfoca hacia la recuperación de documentos mediante una combinación de métodos lingüísticos estadísticos y conceptuales para el establecimiento de relevancia de la consulta. Su técnica de procesamiento de lenguaje natural es utilizada para extraer términos relevantes dentro de los documentos por medio de lematización de palabras. Entonces la ontología, vía un proceso conceptual, anota documentos y expande la consulta con respecto a una medida de similitud basada en métodos probabilísticos, este proceso de expansión está basado en el algoritmo de propagación por activación. Como resultado la metodología desarrolla un sistema llamado O-Rank. Comparado con otras variantes de modelos de *rankeo*, el enfoque combina las características de los documentos como su lingüística, estadística y similitud conceptual, mejora el algoritmo de propagación de la activación, incrementando distancias con respecto a pesos de diferentes relaciones conceptuales, se expande la consulta con respecto a los conceptos de la consulta, considera frases en vez de sus palabras. En la etapa de procesamiento de los documentos HTML se remueven las etiquetas HTML para poder analizar partes del texto y sobre éste se reemplaza por sus raíces posteriormente, para extraer la frases apropiadas; se cuenta su frecuencia en el documento y como resultado se obtienen vectores, los cuales por una fórmula estadística se calculan sus pesos. El procesador de ontología tiene la labor de asignar pesos a los enlaces de la ontología y es básicamente la multiplicación de la medida de similitud y una medida de especificidad; la primera se refiere a la similitud de conceptos y la segunda medición es el tipo y número de relaciones. Como resumen se ha presentado un modelo basado en ontología para recuperar documentos con un grado de relevancia. Se

implementan un sistema de anotaciones en los documentos para la expansión mediante una ontología de aplicación.

El trabajo de anotaciones de Sridevi [30] presenta una metodología de análisis de correlación basado en ontología y en términos, ejecuta un procesamiento para una ponderación y posterior anotación semántica. El análisis de correlación basado en la ontología se lleva a cabo para encontrar las relaciones entre los términos, con la combinación de este análisis de correlación y el modelo de espacio vectorial tradicional se calcula la similitud del documento. Los documentos de ciencias ambientales desde el directorio de ciencia USGS son anotados con la ontología y se representa en formato RDF usando la API de la herramienta GATE. Usando el modelo de consulta Jena SPARQL, los documentos se pueden recuperar, no sólo en función de su contenido textual, sino también en función de sus características o anotaciones. Los documentos que están anotados con el concepto se recuperan con mayor rango. En la anotación semántica se les asignan a las entidades del texto enlaces hacia su descripción semántica y proveen información sobre el contenido *web*, de este modo una mejor decisión sobre el contenido puede ser hecha. Como pruebas experimentales se hace una comparativa sobre la arquitectura KIM [27]. La precisión con anotación y sin anotación muestra una mejora para un caso específico del 11% con el incremento de una medición de relevancia. El trabajo establece un esquema de anotación que combina el análisis de correlación basado en la ontología y la frecuencia de conceptos en el documento. El enfoque puede ser visto como evoluciones de índices basados en palabras clave y se sustituye por conceptos de una ontología, en la recuperación de información la correlación mejora la clasificación del documento.

En la metodología de Dhingra [31] se combina un sistema de anotaciones basados en una ontología para la recuperación inteligente de información llamado SemCrawler. Se propone un marco de trabajo para la búsqueda de las ontologías/documentos sobre la *web* semántica. El marco propuesto es implementado y evaluado con diferentes colecciones de páginas *web*. Este sistema tiene características de extracción de documentos en formato HTML anotados con una ontología RDF / OWL. Se diseña un *crawler* que obtiene páginas *web*, las cuales son filtradas comprobando si tienen alguna anotación de una ontología, si se cumple esta condición se extraen los *links* y se guarda el contenido de la página en un repositorio semántico en formato *triple store*. El dominio de la búsqueda y de la ontología fue limitada a las computadoras portátiles. El sistema de anotaciones basado en una

ontología ayuda en la búsqueda de relaciones entre las entidades de la página *web*. En este trabajo la ontología y el sistema de anotaciones establecieron un carácter semántico a la captura y la recuperación de páginas *web*, pero limitado a una ontología con un dominio específico y sin establecer una medida de similitud entre los contenidos de las páginas *web* para una posible recuperación ordenada. El repositorio semántico de contenidos solo capturó las oraciones sujeto-predicado-objeto de las páginas con base en anotaciones de una ontología identificando entidades y relaciones para la recuperación de información.

En este trabajo de Rodriguez-Garcia [32] se propone una plataforma semántica para la anotación de servicios en la nube y la recuperación con base en sus descripciones. Se realiza la anotación automáticamente sobre diferentes servicios en la nube de su descripción considerando el lenguaje natural de su descripción, la cual puede estar disponible en una variedad de formatos de documentos como XML, HTML. La plataforma propuesta, se ha implementado teniendo en cuenta los entornos multi-ontología (OWL2 ontologías) con el fin de ser capaz de hacer frente a varios dominios. El sistema hace posible la compatibilidad con la evolución de los documentos originales, manteniendo así la coherencia entre las descripciones en lenguaje natural y las anotaciones, que se almacenan utilizando un modelo basado en la *web* semántica. Describiendo la metodología, una vez que los servicios fueron provistos de información semántica por medio de anotaciones, se crean índices semánticos. Un modelo de espacio vectorial clásico es utilizado con el fin de construir estos índices. La última fase de la metodología se refiere a la búsqueda semántica; las consultas basadas en palabras clave se modelan en forma de un vector y se comparan con los vectores de servicios mediante el uso de similitud coseno. Las ventajas de la metodología es la búsqueda multi-dominio considerando ontologías de varios dominios así como la posibilidad de actualización de anotaciones y descripciones de servicios. Como una desventaja se presenta la carencia de medidas de similitud entre servicios por lo que el indexado se basó en la medición de la frecuencia de palabras dentro de las descripciones.

2.1.3. Enfoques basados en modelos de similitud semántica

El enfoque de Janowicz [33] orienta su trabajo a la *web* semántica geoespacial con dos enfoques: de razonamiento y de similitud basados en mediciones semánticas. Se presenta una teoría de similitudes concernientes al contexto sobre el lenguaje modelado de *web services* (WMSL) y por ello se puede realizar la integración y recuperación dentro arquitecturas orientadas a servicios. En la implementación se desarrolla un marco de trabajo para WMSL, con base en la intersección de descriptores lógicos. En general la medición de similitudes consiste en las siguientes fases:

- Definir el concepto de búsqueda y el contexto.
- Generar la forma canónica normal para comparar conceptos.
- Alinear las piezas para la comparación.
- Aplicar las funciones de similitud con respecto a las partes.
- Derivar las similitudes normalizadas.

Para el proceso de establecer la medida de similitud de un concepto, a éste se le extraen los atributos o propiedades de su especificación en el WMSL para formar una matriz, esta matriz se compara con otros conceptos y sus respectivas matrices. Así, a los pares idénticos se les asigna un peso, similares o diferentes según sea el caso. La similitud espacial se puede determinar por medio de un grafo de vecinos conceptuales, se utiliza la topología para establecer la distancia relativa de cada concepto dentro del vecindario. Finalmente, se realiza la suma normalizada derivada de la comparación de atributos por cada concepto para la asignación de las similitudes.

2.1.4. Enfoque basado en minería de datos

En el trabajo de Xiaohui [34] se propone un nuevo modelo de recuperación de conocimiento para resolver los problemas de recuperación mediante la construcción y la minería con base en una ontología personalizada. El modelo toma la consulta del usuario, extrae el conjunto de dominios relevantes que son potenciales de la base de conocimiento, se muestra los dominios al usuario, el usuario identifica el conocimiento relacionado incluyendo dominios positivos y negativos, y finalmente el conocimiento identificado por el usuario se construye una ontología parcial del dominio. Una vez que la ontología del dominio del usuario es construida, el conocimiento del usuario necesita ser minado desde el repositorio de información local y la ontología construida. Como resultado del proceso

anterior se produce un conjunto de dominios relacionados a los intereses del usuario y ayuda a interpretar las necesidades del mismo. El modelo de recuperación de conocimiento usa ontologías para especificar el perfil de conocimiento del usuario y capturar la información de usuario que necesita. En la implementación del modelo se crea el sistema KRM que significa “modelos de recuperación de información”. En él se entrena con un conjunto de documentos positivos y negativos para la generación del perfil de usuario, el usuario puede ingresar desde cualquier dominio, posteriormente con el repositorio local de información que compone una colección de documentos visitados por el usuario. El perfil de conocimiento es relacionado con el dominio de entrada en el sistema y el perfil de usuario minado del repositorio de información local describiendo finalmente conocimiento de fondo. En los experimentos se ingresaron 394,070 temas en un sistema de clasificación, para el repositorio de información local se utiliza un catálogo para buscar temas guardado en una librería, se realizan comparaciones mediante la medida F, relacionada con precisión y *recall*. En resumen, dos fuentes de conocimiento son usados; en del repositorio local de información que contiene documentos visitados por el usuario y la ontología de temas construida a partir de la consulta del usuario, con el fin de analizar el conocimiento descubierto. Con el fin de analizar el conocimiento descubierto, un esquema de dos dimensiones de la especificidad y la exhaustividad se presenta para evaluar las unidades de conocimiento y las relaciones semánticas relacionadas en una ontología. Un perfil de usuario es finalmente generado a partir de LIR de un usuario, que es un conjunto de entrenamiento que consta de un subconjunto de positivo y negativo subconjunto de documentos.

2.2. Integración de información

2.2.1. Integración basada en ontologías

En el trabajo de Tanasescu [35] se realiza la integración de datos geoespaciales en la *web* semántica 2.0, por medio de la construcción de un modelo cognitivo entre servicios *web* semánticos. Esta investigación se le conoce como la aproximación *Emerges*, donde se realiza una integración por medio de una representación y manipulación unificada de distintas fuentes de datos geoespaciales. Este proceso se lleva cabo utilizando ontologías espaciales de alto nivel. Los aspectos generales a los cuales la integración fue orientada son los siguientes:

- **Objeto/Campo.** Corresponde al problema distinción del campo y del objeto, por ejemplo distinguir cuando una montaña es el objeto o el campo.
- **Cognitivo.** Este aspecto es la relación de la representación espacial y la cognición humana, es decir, diferentes formas de mostrar un objeto espacial y como los seres humanos lo interpretan y entienden.
- **Multi-representaciones.** Un objeto espacial puede tener varios significados dependiendo del contexto y propósito.

Existen dos fases en la metodología la primera es la descripción de los objetos espaciales, para definir lo anterior se utilizó la teoría de Galton para definir la pertenencia a un tipo por sus características. Para el elemento de generalización en las representaciones de los datos conceptuales, éste es analizado por una ontología de arquetipos previendo abstracciones genéricas, las cuales serán representadas (i.e., casa, hospital, contenedor, etc.) posteriormente. Para la segunda fase, se determina la descripción de contexto espacial con la finalidad de obtener el foco de interés del objeto por su rol, tarea, localización; especificado en el servicio web de la aplicación. Finalmente los resultados son mostrados en una interfaz *web* por medio de Google Maps, mostrando los objetos espaciales por polígono o puntos dependiendo del tipo de dato espacial que sea representado.

En [36], el objetivo fue implementar una base de conocimiento para realizar la integración semántica en la *web*, se discute el significado de los posibles retos en la integración de los datos, el caso de estudio es integrar datos para el gobierno electrónico en los servicios electrónicos. Existieron dos tipos de integración la horizontal (entre agencias) y la vertical (intergubernamental). Se muestra como este proceso puede ser realizado por medio del uso de ontologías sobre los servicios *web* con base en la arquitectura orientada a servicios (SOA), utilizando minería de datos y explotando tecnologías *web* semánticas. Las ontologías utilizadas son descritas mediante el lenguaje de modelado de ontologías (OWL) debido a la manera de describir e inferir y establecer relaciones semánticas como un esquema común global entre las fuentes. En este artículo se muestran las herramientas y modelos que se pueden emplear para la integración de distintas fuentes y organismos para el comercio electrónico, se presenta el uso de ontologías como una de las mejores opciones el manejo semántico mediante la estructura conceptual de las ontologías para llevar a cabo el objetivo.

En el trabajo de Haav [37] se implementa un meta-modelo de ontologías con lenguajes específicos de dominio (DSL) para la integración de artefactos de software. Los lenguajes DSL se han utilizado durante mucho tiempo con el fin de acortar el ciclo de vida del desarrollo de software y que sea rentable en un dominio particular de interés. Con base en esto se creó un enfoque para la integración de software para el desarrollo de DSL utilizando la representación semántica en forma de ontologías formales descritas en OWL. El papel de OWL es servir como un lenguaje común para la representación de la semántica de los artefactos de software. La novedad del trabajo es la integración semántica de software distribuido en un meta-modelo DSL coherente, así como la simplificación del ciclo de vida del desarrollo y evolución de una conexión DSL. Así, este trabajo está enfocado en crear un marco de integración de módulos de software por medio de una ontología y generar un conjunto de descripciones comunes para la integración de estos componentes de software.

2.2.2. Integración basada en descripciones semánticas

Vaccari [38] se enfoca en resolver los problemas de la heterogeneidad semántica en las infraestructuras de datos espaciales, mediante un geo-servicio de integración semántico para coordinar diversos servicios geográficos. En este enfoque, los proveedores de servicios comparten el conocimiento explícito de las interacciones en las que sus servicios sean contratados, estos modelos de interacción son utilizados operacionalmente como ancla para describir la semántica de la interacción. Por tanto, se logra el descubrimiento de servicios *web* y la integración mediante el uso de *coincidencia semántica* entre interacciones particulares y descripciones de los servicios *web*. Para ello se introduce una solución específica llamada estructura de la preservación de la unión semántica. Para la formalización de la metodología se utilizó el lenguaje *Lightweight Coordination Calculus (LCC)*. Como caso de estudio se implementó en un modelo organizacional de la Agencia de Infraestructura Distribuida GIS en Trentino, Italia; donde este marco de trabajo se representa por un gran número de agencias especializadas en GIS. Sobre este escenario se efectúa la integración bajo el proceso más común que es el servicio de solicitud de mapa, donde éste debe ser geo-referenciado bajo la información del usuario. Una vez formalizado el servicio *web*, se procede a un mapeo mediante una ontología para realizar las relaciones semánticas entre los modelos de rol de interacción y la descripción del punto del servicio *web*, este proceso es hecho automáticamente. Para la integración de los servicios *web* se ejecuta el descubrimiento y posteriormente la coordinación. Con respecto al

emparejamiento de la ontología entre las operaciones del servicio, éste se realiza semánticamente a través de la solución SPSM en Java, la cual hace la correspondencia entre cada elemento del servicio y cada elemento de la interacción. Este trabajo se enfoca en la integración de geo-servicios *web* con el propósito de combatir la heterogeneidad semántica mediante la representación conceptual de la ontología específicamente, bajo el proceso de solicitud de mapa entre un geo-servicio *web*.

Por otro lado, Liping [39] presenta una aproximación para resolver automáticamente los problemas en el entorno de la *web* semántica geoespacial. El enfoque simula el proceso usado por expertos en el ámbito geoespacial donde se ejecutan funciones secuenciales comenzado con datos crudos para derivar el conocimiento deseado creando una ruta entre ellos. Con enriquecimiento de la semántica en las descripciones de los servicios y el soporte de la ontología, la ruta puede ser formada automáticamente a través del razonamiento de reversa desde el resultado deseado a los datos crudos geoespaciales utilizando servicios *web* semánticos. En la automatización de los servicios *web* hay 2 etapas: la primera es encadenar servicios *web* mediante el emparejamiento de entradas y salidas del servicio y, la segunda, es el procesamiento de pre-condiciones y post-condiciones entre servicios intermedios cuando no se emparejan adecuadamente. Sobre estas etapas se puede crear un servicio *web* de composición que mediante reglas dependiendo las entradas y sus salidas el servicio de razonamiento hacia atrás puede ser utilizado para encadenar e integrar los servicios. El razonamiento hacia atrás se entiende de arriba hacia abajo donde empieza por el objetivo y se aproxima a las reglas para generar las salidas requeridas. Este razonamiento es usado para la automatización de un servicio *web* geoespacial mediante un límite de reglas para hacerlo eficiente. Las entradas y salidas pueden ser integradas en varios niveles, éstas pueden ser emparejadas en el nivel de instancia con todas las propiedades concretas y condiciones, además se puede emparejar a nivel de dato ignorando los detalles asociados con cada instancia, este tipo de emparejamiento es considerado más flexible durante la primera etapa, mediante una ontología apropiada, más significado y mayor flexibilidad puede ser realizada. La integración en el proceso de tipificación realizado mediante una ontología geoespacial, que es la encargada de dar significado e integrar servicios *web* geoespaciales y datos, además de emparejar entradas y salidas mediante inferencias en las relaciones de la ontología. Finalmente en la etapa de razonamiento de reversa cada servicio *web* geoespacial es

semánticamente descrito y definido en un lenguaje de ontologías, es posible la recolección y registro de servicios *web*, estas expresiones en el lenguaje pueden ser usadas en un razonador semántico. En esta etapa el razonamiento hacia atrás es derivado para la abstracción y encadenamiento de los servicios *web*. Este trabajo se enfoca hacia la composición e integración automática mediante la tecnología de ontología con un sistema basado en reglas, el razonamiento de reversa fue adoptado en la automatización de servicios *web* geoespaciales con tecnología *web* semántica.

2.3. Interoperabilidad e intercambio

2.3.1. Interoperabilidad basada en ontologías

La metodología presentada en [40] describe un lenguaje llamado AERIA, el cual fue creado con el propósito de soportar un esquema semántico de representación y de integración aplicado para los repositorios multicapas, los cuales representan un amplio conjunto de fuentes de datos. De este modo, se puede proveer una vista integrada total para el control de la información. El lenguaje AERIA tiene un buen balance entre expresividad y fácil uso que define sintaxis conceptual y semántica. Este lenguaje es una extensión de SKOS[41] que define esquemas de peso ligero para soportar esquemas *web* de integración y abstracción con la capacidad de representar arbitrariamente relaciones entre conceptos, es decir, generalización del inter-esquema. Este lenguaje fue desarrollado presentando algunas ventajas de los lenguajes de ontologías *web* como lo son OWL, RDF, etc. AERIA [42] tiene la capacidad de representar arbitrariamente relaciones entre conceptos mapeados en SKOS mediante relaciones semánticas, además de la capacidad de representar con base en abstracción mapas entre conceptos mediante la generalización entre esquemas, el olvido de conceptos y el colapso de conceptos. El caso de estudio se basó en comercio electrónico donde se asume que una compañía tiene distintas fuentes de datos conteniendo información acerca de artículos, clientes, ventas, el proceso de producción, esquema de la organización, etc. Estos esquemas conceptuales de las fuentes son representados por una notación gráfica simplificada que es el modelo entidad relación extendido [43]. Para la integración se propuso un método de integración progresiva por abstracción, donde el primer paso es por cada esquema donde no se representa ningún concepto en uno solo se da una abstracción de nivel más alto, esto se realiza con la finalidad de integrar esquemas en un esquema global. El segundo paso es para las relaciones que tienen la misma semántica de manera intuitiva y

por lo tanto identificarlas por el mismo nombre. Por último el tercer paso es relacionar conceptos que son similares entre los esquemas no ligados directamente mediante procedimientos basados en la subsunción de las relaciones. Los repositorios de esquemas multicapas basados en *web* descritos por medio de representación e integración de modelos conceptuales provee a las organizaciones el trato con amplios conjuntos de fuentes de datos. El lenguaje AERIA tiene una sintaxis basada sobre lenguajes *web* semánticos para la representación de esquemas, además este lenguaje es enfocado para usuarios sin un perfil alto sobre lenguajes *web* semánticos. Como la sintaxis de AERIA está basada sobre una ontología OWL, técnicas de razonamiento automático suportan la inferencia de nuevas relaciones y modelos.

Manoj [44] define un servicio para la integración de datos espaciales en la *web* por medio de una ontología para distintos proveedores de datos espaciales como servicios. Mediante la arquitectura SOA para *web* services, el servicio proporciona una descripción de servicio a ser descubiertos. Se propone un método de evaluación de similitud de servicios geoespaciales OGC para resolver la heterogeneidad semántica. Se presenta una conceptualización basada en una ontología de localización geoespacial que es usada para la evaluación de similitud del servicio *web* geoespacial, con base en una petición se establecen un conjunto de servicios candidatos con el fin de establecer servicios relevantes, la similitud se calcula con métricas de medición euclidiana y distancia jerárquica. La recuperación se compone de dos fases: la *adquisición*, prepara la información necesaria para el catálogo de servicio y la fase de *consulta* que provee un servicio de descubrimiento basado en ontología para los servicios relacionados. Esta fase se compone por un módulo procesador de consultas donde la ontología describe el proceso de obtención de características geográficas provenientes del servicio *web* geoespacial. En este módulo la consulta del usuario se analiza en sub-consultas y se integran antes de ser enviado al cliente. Los componentes principales del sistema de integración y recuperación son los servicios geoespaciales, los repositorios datos geográficos y el servicio de metadatos del servicio de descripciones. El usuario se comunica con el catálogo de servicios para descubrir servicios relevantes dependiendo de las necesidades, y para el acceso el servicio de procesamiento de consultas divide la consulta en sub consultas para distintos servidores. El núcleo de la interoperabilidad se encuentra en el servicio de integración de repositorios, el cual interactúa con el catálogo de servicio, envía la consulta al servicio de procesamiento de consulta. En el catálogo de servicios se realiza el servicio de análisis espacial y el ajuste de

formatos como los son GML entre servicios geoespaciales. Para la similitud mediante la ontología es determinada mediante la cercanía de las propiedades: lugares que son referidos mediante diferentes nombres, lugares que pueden estar en diferentes niveles de jerarquía topológica o administrativa, y lugares que tienen una cierta medida de proximidad o conectividad.

2.3.2. Interoperabilidad basada en anotaciones

En [45] se propone un marco de trabajo que permite una gestión de anotaciones semánticas de datos geoespaciales en la *web*, como un estándar de intercambio y que habilita la interoperabilidad sintáctica. El servicio de anotaciones es responsable de obtener información de diversas fuentes de datos y en combinación se produce un resultado con más significado. La anotación de un objeto es útil, tanto para las características espaciales como para las características derivadas del contenido; de este modo se define un proceso de anotación específico para cada uno. La metodología utiliza flujos que describen cada proceso de anotación, cada flujo contiene información del esquema de anotación de los datos que serán usados durante el proceso. Además, las ontologías que describen esos datos semánticamente, definen que operaciones que deben ser mejoradas y la forma de guardarse las anotaciones. Cabe señalar que para la creación del esquema de anotaciones se usan los metadatos del dato geoespacial y el esquema es llenado con términos de la ontología, además de anotaciones adicionales que pueden ser llenadas manualmente. El propósito de la creación de anotaciones semánticas es la utilidad para el descubrimiento, integración y adición de significado a los datos espaciales en la *web*. El trabajo se enfoca en proporcionar un mecanismo de anotaciones de datos geoespaciales, distribuidos y heterogéneos en la *web* y como caso de estudio para la generación de estrategias de información en la agricultura. Este sistema es implementado como un servicio *web*, considerando la automatización del proceso de anotación, la integración de los datos y la generación de anotaciones sobre las consultas.

En el trabajo de Harold [46] se estudia el diseño, sintaxis, implementación de la semántica de negocios electrónicos (e-negocios) y su conocimiento posicional integrado y por medio de un sistema *slots* (ranuras). Se entiende por posicional el lugar donde se encuentra un concepto en el campo de conocimiento y sobre esta posición se obtiene su significado, y por la ranura como un solo lugar donde pueden corresponder varios conceptos y están

ligados al nombre del lugar, se relacionan estos por pares. Para ejemplificar lo anterior se define el ejemplo envío (pc, 45.7, mexico, colombia), cada posición define su significado, en este caso sería el primero artículo, precio, origen u destino respectivamente. Por otro lado por ranura el ejemplo sería envío (carga->Pc; precio->45.7; origen->México; destino-> Colombia). De este modo se crea el lenguaje posicional-ranura (POSL) que es la unión de los dos para poder describir el conocimiento y la semántica en la integración en la Web semántica. Por medio de este lenguaje se integra la representación de conocimiento el cual es frecuentemente necesitado particularmente en la *web* semántica, esta integración es hecha reconstruyendo la noción relación objeto con sus componentes compartidos.

2.4. Construcción de ontologías

2.4.1. Construcción automática de ontologías

El trabajo de Jike [47] se centra en la construcción de ontologías para textos escritos en chino utilizando técnicas de reconocimiento de patrones tales como *fuzzy clustering* [48] con el fin de obtener los contextos. Se limita al ámbito de la exploración petrolera en China. Se utilizó la base de conocimientos HowNet [49] con el fin de obtener el contexto, se aplicó la técnica de la identificación de posición y el tipo de término (nombre, adjetivo, etc.) mediante el uso de frecuencia, este sistema utiliza un tamaño de ventana manteniendo pista de su posición en la oración. La ontología se construye por *clustering* asignando el nombre del concepto y atributos asociaciones con base en la información morfológica del término chino y su sintaxis. Este estudio propone una ontología del dominio chino de exploración de petróleo. Se emplean tecnologías tales como mecanismos de contexto, procesamiento de lenguaje natural, segmentación de palabras para el chino y la agrupación para desarrollar un mecanismo de construcción de ontologías a partir de documentos de texto chinos no estructurados. La calidad de la ontología derivada del método de construcción de ontologías propuesto puede no ser tan buena como lo que puede ser construido por los expertos. Además para algunos casos especiales, como un dominio con términos y conceptos cambiantes rápidamente o con la semántica compleja, es muy difícil construir la ontología de dominio adecuado usando esta metodología.

En el trabajo de Azevedo [50] se implementan técnicas de procesamiento de lenguaje natural [51], análisis sintáctico, análisis semántico y un módulo de axiomas OWL. Para este

propósito se realiza un procesamiento sintáctico utilizando el software *Stanford Analyzer 1*, para el procesamiento semántico se identificaron los términos y relaciones usando patrones, finalmente, los axiomas OWL evitan interpretaciones ambiguas. Este trabajo no incluye una base de conocimiento pre-definida en absoluto y es básicamente procesando los términos en su contexto dentro de las oraciones. Se describe en el trabajo un enfoque para el desarrollo automático de ontologías expresivas tomando las definiciones proporcionadas por el usuario. Los resultados obtenidos a través de los experimentos evidencian la necesidad de la creación automática de axiomas expresivos, suficiente para la creación de ontologías. La viabilidad de su enfoque se demuestra a través de la generación de descripciones de axiomas complejos a partir de los conceptos definidos por los usuarios y glosarios que se encuentran en Wikipedia. Se evalúa la metodología en un experimento con frases de entrada enriquecidos con axiomas jerarquía, disyunción, conjunción, negación, así como la cuantificación existencial y universal de imponer la restricción de propiedades para la representación del conocimiento y la construcción automática de ontologías expresivas.

La investigación de Touzi [52] introduce el análisis de conceptos formales (FCA) [53], la agrupación y la lógica difusa [54] en la creación de ontologías para el dominio de minería de datos. La metodología se divide en organización de datos con un procesamiento del *clustering* de conceptos y generación de ontologías difusas con relaciones jerárquicas. Esta metodología fue enfocada para la generación automática de FODM (*Fuzzy Ontology of Data Mining*). El enfoque FODM comienza por la organización de los datos en grupos homogéneos que tienen propiedades comunes que permite deducir semántica de los datos. Se hace la mención que el número de clases es siempre menor que el número de datos de partida, se demuestra que esta solución reduce considerablemente la definición de la ontología, que ofrece una mejor interpretación de los datos y tanto la memoria del espacio optimizado y el tiempo de ejecución.

El mecanismo de construcción de ontologías de Dixit [55] implementa un análisis semántico de los documentos, comenzó el procesamiento léxico de un texto sin formato y el conjunto resultante de conceptos entran a un tratamiento de población semántica automática de dominio a fin de definir la estructura y mostrar las relaciones entre los conceptos. La ontología ya definida y la ontología poblada puede considerarse como un subconjunto de la ontología principal. En la metodología se identificaron las cinco fases

para la extracción automática de la información semántica de los documentos de texto y poblar en la ontología los cuales son:

1. En la primera fase de procesamiento léxico de un texto plano es realizado.
2. En la segunda fase las reglas gramaticales del inglés son aplicadas a la entrada recibida de fase de procesamiento léxico para llevar a cabo el procesamiento sintáctico.
3. El proceso de resolución de la anáfora es el problema a resolver, se define que es un pronombre o una frase.
4. En la fase de procesamiento semántico se identifican las principales partes de un objeto, es decir, frase sujetos, acciones, atributos.
5. Por último, en la fase de población de la ontología, la información se inserta en la base de conocimiento tras la representación en el dominio de la ontología.

En [56] una ontología de gran escala es construida con base en el dominio genómico mediante *clustering*. La ontología que se crea resulta carente de semántica conceptual y, básicamente centrado en una estructura taxonómica. La forma de construcción se puede resumir como procesos de agrupado y *clustering* para conceptos especializados de los datos genómicos. La similitud entre conceptos se calcula utilizando el perfil vecino. El clúster ortogonal es construido mediante la aplicación de un nuevo método de agrupación de todos los posibles genes de proteína en todos los genomas completos de codificación con base en sus similitudes de secuencia de aminoácidos. La originalidad del algoritmo de agrupamiento se encuentra en el uso de una búsqueda cuasi-cíclica (una variante de la búsqueda cíclica) y la incorporación de información filogenética en el agrupamiento. Es computacionalmente eficiente y es posible actualizar periódicamente los contenidos.

2.4.2. Construcción semi-automática de ontologías

En la obra de Küçük [57] los conceptos se toman de artículos de Wikipedia. Esta metodología no utiliza una base de conocimiento u ontologías creadas a-priori, en lugar de esto, los conceptos en el texto se procesan por medio de lenguaje natural con el fin de obtener conceptos propiedades y relaciones que componen la ontología final. Se extraen automáticamente las palabras muy frecuentes y frases (n-gramas) del texto en los artículos, el siguiente paso fue la construcción de la ontología mediante la organización de estos

componentes y propiedades que utilizando otras fuentes de información escrita tales como las normas internacionales y libros de texto. Se propone una ontología de cobertura amplia para el dominio de la energía eólica y la ontología se construye a través de un proceso semi-automático del uso de los recursos *web* relacionados, reduciendo así el coste global del proceso de construcción de la ontología. El proceso de construcción de la ontología es un procedimiento semi-automático basado en el aprendizaje, donde en la primera fase palabras clave plausibles son utilizadas durante la determinación de los conceptos y sus propiedades desde la *web*. Los artículos de Wikipedia relacionados son procesados para determinar la relevancia. Cabe señalar que la ontología propuesta se puede ampliar o personalizar para satisfacer mejor las necesidades de aplicaciones potenciales mediante la intervención de un experto.

Xiong [58] creó una metodología para la construcción de ontologías para organismos marinos. Se implementó una teoría de roles (una entidad bajo un cierto ambiente) [59] para describir las relaciones entre los organismos marinos. La adquisición de conceptos del texto se realiza automáticamente por un proceso de extracción de conceptos que involucra técnicas de agrupamiento, clasificación, fusión, etc. Para la construcción de ontologías se hace necesaria la participación de expertos de varios campos utilizando una herramienta de edición dentro de una interfaz para la perfección de la ontología.

2.4.3. Construcción manual de ontologías

El trabajo de Fernández-López [60] se enfoca en desarrollar un método sistemático para la conceptualización del dominio llamado METHONTOLOGY, este método proporciona un conjunto de tareas para asistir al modelador de la ontología en capturar y estructurar la información requerida, utilizando una serie de tablas, un diccionario de datos y una serie de árboles de conceptos, se incluyen actividades para la planificación del proyecto, la calidad de las ontologías, la documentación, para la implementación de la ontología se describen actividades como evaluación, conceptualización, configuración, integración, implementación, etc. Estas representaciones son especificadas para ciertos dominios. Este trabajo se centra en reducir la brecha existente entre el arte y la ingeniería ontológica por medio de la identificación de un conjunto de actividades a realizar durante el proceso de desarrollo de la ontología. Ellos son: planificar, especificar, adquirir conocimientos, conceptualizar, formalizar, integrar, implementar, evaluar, documentar y mantener.

De acuerdo con la descripción del problema, se consideró como hipótesis que la recuperación de datos geoespaciales se puede realizar de forma inteligente, considerando un proceso de síntesis semántica basada en un método de integración que considere medir la similitud conceptual y la alineación de diversas representaciones conceptuales sobre los objetos geográficos, con base en diversos contextos, en donde esta información puede variar su significado. Considerando además que esta información es heterogénea, en diversos formatos y formas de representación sin importar las características implícitas de los objetos geográficos, tales como la escala, sistema de referencia, proyección, entre otras. En este contexto, podría tratarse del mismo objeto; por lo tanto, se necesitarán procesos de *descubrimiento, integración y recuperación* de datos geoespaciales en la *web*, además de la construcción de una base de conocimiento para construir un motor de búsqueda, basado en medidas de similitud semántica y alineación para contar con un método de recuperación basado en un conjunto de representaciones conceptuales. En la Figura 2.1 se esquematiza de manera general el objetivo fundamental en el desarrollo de la tesis.

La metodología de Uschold & King [61] presenta una serie de pasos para construir ontologías, se compone de cuatro fases: capturar conceptos y relaciones en el dominio de interés, codificación que significa la representación explícita de la conceptualización capturando los conceptos y relaciones en algún lenguaje formal, en esta fase se incluye la integración de ontologías existentes, la evaluación para hacer un juicio técnico de las ontologías creadas, el software asociado y su documentación. Finalmente la documentación donde se establecen guías para la documentación dependiendo el tipo y propósito de la ontología.

En el trabajo de Suarez-Figueroa [62] se presenta la metodología llamada Neon para la construcción de ontologías y redes de ontologías. Esta metodología basada en escenarios soporta diferentes aspectos del proceso de desarrollo de la ontología, así como la reutilización y la evolución dinámica de las ontologías en red en entornos distribuidos, donde el conocimiento se introduce por diferentes personas (expertos de dominio, los profesionales de la ontología) en las diferentes etapas del proceso de desarrollo de ontologías. Esta metodología incluye los siguientes componentes:

- El Glosario NeOn, el cual que identifica y define los procesos y actividades potencialmente involucrados en la construcción de la red de ontologías.

- Un conjunto de nueve escenarios para la construcción de ontologías y redes de ontologías. Cada escenario se descompone en diferentes procesos y actividades tomadas del glosario NeOn.
- Dos modelos de ciclo de vida de la red la ontologías que especifican la forma de organizar los procesos y actividades del glosario NeOn por medio de fases.
- Un conjunto de guías metodológicas prescriptivas para los procesos y actividades.

3. Marco Teórico

3.1. Introducción

En este capítulo se describe el marco teórico del trabajo de tesis, el cual se concentra en proporcionar las definiciones y teorías apropiadas que se utilizan como herramientas para el desarrollo de la tesis, tales como ontología, instancia, concepto, conceptualización, relación, clase, restricción, entre otras.

De igual forma, se describen un conjunto de definiciones y conceptos nuevos que han sido generados, como parte del marco conceptual del trabajo de tesis para la creación de ontologías de dominio. Asimismo, se puntualiza en un modelo formal para el desarrollo de estas ontologías; así como la definición del término esquemas conceptuales, con lo cual se sientan las bases necesarias para utilizar todas estas herramientas en la parte metodológica.

3.2. Ontologías

Una ontología define los términos que se usan para describir y representar un cierto dominio, entendido éste como un área específica de interés o un área de conocimiento. Para que la *web* semántica sea una realidad, la tecnología a utilizar son las ontologías. Por tanto, si la información es catalogada con las ontologías, al hacer las búsquedas de los recursos en la *web*, tales como páginas, documentos, sonido o imagen, se utiliza el significado de las palabras como *entes*, en lugar de palabras clave.

Con la información organizada por medio de ontologías, serían mucho mejores las búsquedas y la integración de datos, debido al recurso de la interpretación del *significado*, que correspondería a los agentes de software.

Las ontologías son vocabularios comunes para los usuarios y las aplicaciones que pertenecen al campo de la Inteligencia Artificial (IA). Agregan el conocimiento que cada persona utiliza mediante estas representaciones con las que “se representa y se entiende el mundo que nos rodea”. Las ontologías se consideran explícitas porque los significados son

comúnmente conocidos y por lo tanto no es necesario tenerlas plasmadas en un documento, además de que no están organizadas jerárquica o matemáticamente en el cerebro de las personas.

Al igual que los seres humanos, las máquinas carecen de ontologías explícitas, pero en su caso no comprenden el entorno y no pueden comunicarse entre sí. Si se pretende que las palabras sean procesadas por las máquinas, es necesario manejar las ontologías en forma explícita; es decir, desarrollarlas formalmente, mediante un vocabulario específico o en “una forma que sea legible para las máquinas”. En el desarrollo de ontologías explícitas debe contemplarse como mínimo un listado de términos, con el significado de cada uno. De esta manera dos sistemas de información podrían interactuar, gracias a la ausencia de problemas semánticos.

Debido a que las ontologías almacenan conocimiento, en una *web* semántica sería posible extraer información automáticamente y procesarla, como lo haría un agente de software cuando un usuario tiene el propósito de buscar, por ejemplo, una impresora en razón de capacidad y calidad. Con las ontologías de carácter comercial, posteriormente, se tendría que escoger un establecimiento para su adquisición y elegir el precio más conveniente. En resumen, las ontologías:

- Favorecen la comunicación entre personas, organizaciones y aplicaciones.
- Permiten la interoperabilidad entre sistemas.
- Facilitan el razonamiento automático.
- Contribuyen a especificar los sistemas de software.

El área de la investigación referente a las ontologías ha tenido un incremento en estos últimos años, debido a que son muchas las áreas en donde es de gran utilidad el contar con una de ellas, y como muestra de ello se encuentra el área de la Inteligencia Artificial, la Lingüística Computacional y la Teoría de Base de Datos y las Ciencias de la Información Geográfica.

Para el desarrollo de una ontología, se parte teniendo como punto de inicio la base del conocimiento que el ser humano posee y con base en ésta se realiza un análisis más

minucioso para tener una conceptualización aceptable de las cosas. Razón por la cual a continuación se hace referencia a algunos trabajos relacionados con ontologías.

3.2.1. El enfoque de Gruber

En el trabajo de Gruber [63], se expone que el conocimiento representado formalmente está basado sobre una conceptualización: los objetos, conceptos, y otras entidades que se asumen existen en alguna área de interés y las relaciones que mantienen entre ellas.

Una conceptualización de manera resumida, es la vista simplificada del mundo que se desea representar para algún propósito. Entonces a partir de esta aseveración se puede definir a una ontología como una *especificación explícita de una conceptualización*. El término es tomado de la Filosofía, donde una ontología es una *cuenta sistemática de la existencia*. Para el área de la Inteligencia Artificial, lo que “existe” es lo que puede ser representado. Entonces se reafirma la definición mencionada anteriormente donde se dice: *una ontología es una descripción formal de los conceptos y las relaciones entre estos*.

En resumen, una ontología es un sistema de representación del conocimiento que resulta de seleccionar un dominio o ámbito del conocimiento, y aplicar sobre él un método con el fin de obtener una representación formal de los conceptos que contiene y de las relaciones que existen entre dichos conceptos. Además, una ontología se construye con relación a un contexto de uso. En otras palabras, una ontología especifica una conceptualización o una forma de ver el mundo, por lo que cada ontología incorpora un punto de vista. Además, una ontología contiene definiciones que proveen del vocabulario para referirse a un dominio. Estas definiciones dependen del lenguaje que se utilice para describirlas. Todas las conceptualizaciones (definiciones, categorizaciones, jerarquías, propiedades, herencia, etc.) de una ontología pueden ser procesadas por una computadora. Los beneficios de utilizar ontologías se pueden resumir de la siguiente forma:

- Proporcionan una forma de representar y compartir el conocimiento utilizando un vocabulario común.
- Permiten usar un formato de intercambio de conocimiento.
- Proporcionan un protocolo específico de comunicación.
- Permiten una reutilización del conocimiento.

Para evaluar el diseño de una ontología se necesitan criterios objetivos que estén fundamentados en el propósito del resultado esperado y basado en notaciones *a priori* de naturaleza o verdad de los objetos implicados en una ontología. Un ejemplo de estos criterios dados por Gruber son los siguientes:

1. **Claridad.** Una ontología debe de tener el significado de los términos bien definidos, es decir, las definiciones deben ser objetivas. Las ontologías se diseñan para requerimientos computacionales o sociales, la definición debe ser independiente del contexto.
2. **Coherencia.** Una ontología debe de ser coherente, esto es que las inferencias sean consistentes con las definiciones. Al menos, los axiomas que se definen deben ser lógicamente consistentes. La coherencia también debe aplicarse a los conceptos que están definidos de manera informal. Por ejemplo, los descritos con lenguaje natural. Si una oración que se puede inferir a partir de los axiomas contradice una definición o un ejemplo dado informalmente, entonces la ontología es incoherente.
3. **Extensión.** Una ontología se debe diseñar anticipándose en el uso del vocabulario que pueda llegar a ocupar; es decir, debe definir términos nuevos para aplicaciones especiales basadas en el vocabulario existente, de manera que no se requiera revisar las definiciones existentes.
4. **Reducir las tendencias en la codificación.** La conceptualización se debe especificar al nivel del conocimiento sin depender de una codificación a nivel de símbolo en particular. Las tendencias de codificación deben ser reducidas al mínimo, ya que los agentes que comparten conocimiento pueden ser implantados en diferentes sistemas y estilos de representación.
5. **Reducir el compromiso ontológico.** Una ontología puede requerir un mínimo de compromiso ontológico, el cual sea suficiente para soportar las actividades relacionadas con compartir el conocimiento. Debido a que el compromiso ontológico está basado en el uso constante de vocabulario, se puede reducir al mínimo especificando una teoría más débil (que se permite en la mayoría de los modelos) y definiendo solamente esos términos que sean esenciales para la comunicación de conocimiento consistente.

Las posibles aplicaciones y usos de las ontologías pueden ser las siguientes, no siendo éstas las únicas:

- Repositorios para la organización del conocimiento.
- Servir de herramienta para la adquisición de información.
- Servir de herramientas de referencia en la construcción de sistemas de bases de conocimiento que aporten consistencia, fiabilidad y falta de ambigüedad a la hora de recuperar información.
- Permitir compartir conocimiento.
- Posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre organizaciones, comunidades científicas, etc.
- Hacer la integración de diferentes perspectivas de usuarios.
- Permitir el tratamiento ponderado del conocimiento para recuperar información de forma automatizada.
- Posibilitar la construcción automatizada de mapas conceptuales y mapas temáticos.
- Permitir la reutilización del conocimiento existente en nuevos sistemas.
- Permitir la interoperabilidad entre sistemas distintos.
- Establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para extenderlo y transformarlo entre diferentes contextos.
- Servir de base para la construcción de lenguajes de representación del conocimiento.

3.2.2. El enfoque de Guarino

Se definen algunos conceptos relacionados con ontología, compromiso ontológico y conceptualización. En algunos casos, el término “ontología” [64] es el nombre extravagante que denota el resultado de actividades familiares como el análisis conceptual y el modelado del dominio. En el sentido filosófico, podemos referir a una ontología como un sistema particular de categorías relacionadas para una cierta visión del mundo. Como tal, este sistema no depende de un lenguaje en particular.

La ontología de Aristóteles es siempre la misma, independientemente del lenguaje utilizado para describirla. Por otra parte, en su uso más frecuente en Inteligencia Artificial, una ontología se refiere a un artefacto de la ingeniería, constituido por un vocabulario específico usado para describir cierta realidad, además de un conjunto de suposiciones

explícitas con respecto al significado previsto de las palabras del vocabulario. Este sistema de suposiciones tiene generalmente la forma de una teoría lógica de primer orden, donde las palabras del vocabulario aparecen como nombres de predicados unarios o binarios, respectivamente, llamados conceptos y relaciones. En el caso más simple, una ontología describe una jerarquía de conceptos conectados por relaciones de pertenencia; en casos más sofisticados, se agregan los axiomas convenientes para expresar otras relaciones entre los conceptos y restringir la interpretación prevista.

Con el fin de solucionar el problema de la interpretación de la ontología, se utiliza una conceptualización, una palabra muy usada en el ámbito de la IA. Dos ontologías pueden usar diferente vocabulario (por ejemplo, español e inglés) y compartir la misma conceptualización.

Una conceptualización se ha definido como una estructura $\langle D, R \rangle$, donde D es un dominio y R es un conjunto de relaciones relevantes en D . Esta definición fue utilizada por [38], quién definió a una ontología como "una especificación de una conceptualización".

Resaltando estas aclaraciones, se puede proceder a refinar la definición de Gruber para hacer más clara la diferencia entre una ontología y una conceptualización: Una *ontología* es una teoría lógica que explica el significado previsto de un vocabulario formal; esto es, su *compromiso ontológico* para una *conceptualización* en particular del mundo. Los modelos previstos del lenguaje lógico utilizados como vocabulario son restringidos por su compromiso ontológico. Una ontología indirectamente refleja su compromiso y el de la conceptualización subyacente, mediante la aproximación de los modelos previstos.

Es importante enfatizar que una ontología es dependiente del lenguaje, mientras que una conceptualización es independiente del lenguaje. De hecho en la IA, el término "ontología" se colapsa en esos dos aspectos, pero una separación clara entre ellos llega a ser esencial para tratar los asuntos relacionados con compartir, fusionar, y traducir ontologías, que en general implican múltiples vocabularios y conceptualizaciones.

Se puede definir una ontología de manera que exista una clara distinción entre los términos ontología y conceptualización. Así podemos clasificar a las ontologías según su exactitud para caracterizar la conceptualización. Hay dos maneras posibles en que una ontología

puede estar más cercana a una conceptualización: desarrollando una axiomatización más rica, y adoptando un dominio más rico y/o un sistema más rico de relaciones conceptuales relevantes. En el primer caso, la distancia entre el conjunto de modelos de la ontología y el conjunto de modelos previstos se reduce. En el segundo caso, es posible (al menos en principio) incluir en el conjunto de relaciones conceptuales relevantes aquellas relaciones que caractericen un estado del mundo, ampliando en el mismo tiempo el dominio para incluir las entidades involucradas con tales relaciones.

Otra posibilidad de incrementar la aproximación de una ontología a la conceptualización consiste en adoptar un modelo lógico, con lo cual se permite expresar coacciones a través de mundos, o solamente refinando mundos como objetos ordinarios del dominio. Una ontología fina logra una especificación más cercana del significado previsto de un vocabulario (y por lo tanto puede ser utilizado para establecer consenso en relación a compartir dicho vocabulario, o a una base de conocimiento que utilice ese vocabulario); pero puede ser difícil de desarrollar, debido al número de axiomas y la expresividad del lenguaje adoptado. Lo anterior se debe a la representación que cuenta con información muy detallada y se utiliza también el término ontología de bajo nivel (*low-level ontology*).

Una ontología tosca, puede consistir de un conjunto mínimo de axiomas escritos en un lenguaje de expresividad mínima, para soportar solamente un conjunto limitado de servicios específicos, previstos para ser compartidos entre los usuarios que están de acuerdo con la conceptualización subyacente. En este caso, la representación de información es general y se conoce también como ontología de alto nivel (*high-level ontology*).

Las ontologías también son clasificadas de acuerdo con su dependencia a una tarea específica o a un punto de vista y a su nivel de generalización, como se muestra en la Figura 3.1.

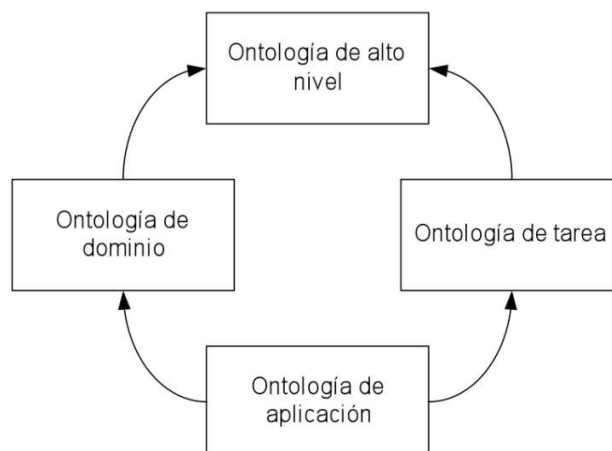


Figura 3.1 Tipos de ontologías de acuerdo con su nivel de dependencia.

Ontología de alto nivel (Top-level Ontology). Describen conceptos muy generales como espacio, tiempo, materia, objeto, eventos, acciones, etc., que son independientes de un dominio o problema en particular; por lo tanto, parece razonable, por lo menos en teoría, tener ontologías a nivel superior unificadas para grandes comunidades de usuarios.

Ontología de Dominio (Domain ontology). Describe el vocabulario relacionado a un dominio genérico como medicina o automóviles.

Ontología de Tarea (Task ontology). Describe el vocabulario relacionado con una tarea en específico como diagnóstico o ventas, cada una especializando los términos introducidos en la ontología de nivel superior.

Ontología de Aplicación (Application ontology). Describe conceptos tanto de un dominio como de una tarea particular que frecuentemente son especializaciones de ambos. Estos conceptos corresponden a los roles desempeñados por las entidades del dominio mientras realizan cierta actividad, como unidad reemplazable o componente de repuesto. Asimismo, representan las necesidades del usuario con respecto a un uso específico, tal como una valoración de la contaminación en la Cd. de México.

3.3. Krovetz Stemmer (KSTEM)

El Krovetz *stemmer* [65] fue presentado en 1993 por Robert Krovetz, es un método de validación-*stemming* de forma lingüística-léxica puesto que se basa en la propiedad flexiva de las palabras y la sintaxis del lenguaje. Posee una buena eficacia y precisión eliminando sufijos flexionales en tres pasos:

1. Transformando los plurales de una palabra a su forma singular.
2. Conversión de pasado de una palabra a su tiempo presente.
3. Eliminar el sufijo 'ing'.

El proceso de conversión primero quita el sufijo y luego entra a un proceso de verificación en un diccionario para cualquier recodificación, posteriormente devuelve la raíz de una palabra. La búsqueda en el diccionario también realiza cualquier transformación que se requieren debido a una excepción de ortografía y también convierte cualquier raíz de en una palabra, cuyo significado puede ser entendido.

La fuerza del análisis derivativo y flexivo está en su capacidad para producir raíces morfológicamente correctas, lidiar con excepciones, procesamiento de prefijos como sufijos Este *stemmer* puede no encontrar las raíces para todas las variantes de la palabra, puede utilizarse como una *pre-stemming* antes de aplicar un algoritmo de derivación. Esto aumentaría la velocidad y la eficacia del proceso *stemming* principal. Tiene una superioridad frente a los algoritmos de Porter y Paice / Husk debido que este es un *stemmer* muy ligero. El *stemmer* de Krovetz intenta aumentar la robustez y precisión por el tratamiento de errores de ortografía y raíces sin sentido.

Si el tamaño del documento de entrada es grande, este *stemmer* se vuelve débil y no funciona de forma muy eficaz. El fallo importante y evidente de algoritmos basados en un diccionario es su incapacidad para lidiar con las palabras, que no están en el léxico. Además, un léxico debe manualmente crearse por adelantado, que requiere esfuerzos significativos. Este *stemmer* no produce constantemente un buen *recall* y precisión.

3.4. WordNet

WordNet es una gran base de datos léxica del inglés. Sustantivos, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos (*synsets*), cada uno expresa un concepto distinto. Los *synsets* están interconectados por medio de relaciones conceptuales-semánticas y léxicas. [66][67]. Esta base se originó en 1986 en la Universidad de Princeton, donde sigue siendo desarrollado y mantenido. George A. Miller, se inspiró en experimentos en IA que trataron de entender semántica humana la memoria [68]. Sólo el conocimiento de conceptos más específicos debe almacenarse con tales conceptos. Así, por ejemplo el enunciado "los canarios tienen plumas" y la declaración "las aves tienen plumas", presumiblemente, la propiedad "tiene plumas" se almacena con el concepto de ave y no se vuelve redundante para cada concepto de clase de ave.

3.4.1. Diseño y contenido

WordNet es una gran red semántica (un grafo) en el cual las palabras están interconectadas por medio de arcos etiquetados que representan las relaciones de significado. Las relaciones léxicas solo conectan simples palabras mientras las relaciones semánticas-conceptuales enlazan conceptos que pueden ser expresadas por más de una palabra.

Un *synset* expresa léxicamente un concepto, un ejemplo de *synset* es $M=\{\text{suelo, piso, territorio}\}$. Los *synsets* en WordNet contienen una breve definición, parafraseando el significado del *synset* y la mayoría de ellos incluyen una o más frases que ilustran el uso de los sinónimos. Los *synsets* están divididos en cinco categorías:

- nombres
- verbos
- adjetivos
- adverbios
- partículas

3.4.2. Cobertura

WordNet en realidad consta de cuatro partes separadas, cada uno conteniendo *synsets* con palabras de las principales categorías sintácticas: sustantivos, verbos, adjetivos y adverbios. La versión actual de WordNet (3.0) contiene más de 117,000 *synsets*, compuesto por más de 81,000 *synsets* de sustantivo, 13,600 *synsets* de verbos, 19,000 *synsets* de adjetivos y 3,600 *synsets* de adverbios. La separación de palabras y *synsets* por diferentes partes de la oración deduce la naturaleza de las relaciones léxicas y semánticas.

3.4.3. Relaciones

Se reconocen dos tipos de relaciones: léxicas y semánticas. Las relaciones léxicas se sostienen entre las formas de la palabra; relaciones semánticas se mantienen entre los significados de la palabra. Los archivos lexicógrafos corresponden a las categorías sintácticas implementadas en WordNet: sustantivo, verbo, adjetivo y adverbio. Todos los *synsets* en un archivo lexicógrafo están en la misma categoría sintáctica. Cada *synset* consiste en una lista de palabras sinónimas o colocaciones como por ejemplo ("medida", "cantidad", "suma") e indicadores que describen las relaciones entre este *synset* y otros *synsets*. Estas relaciones incluyen (pero no están limitadas a) hiperonimia/hiponimia, antonimia, vinculación y meronimia/holonimia. Una palabra o colocación puede aparecer en más de un *synset* y en más de una categoría gramatical. Cada uso de una palabra en un *synset* representa un sentido de la palabra en la categoría gramatical correspondiente al *synset*. Algunos ejemplos de las relaciones incluidas en wordnet son:

Sinonimia. Es la relación más importante de WordNet. Dos expresiones son sinónimas en un contexto C si la sustitución de una por otra en dicho contexto no altera el significado.

Antonimia. Es una relación entre lexemas que muestra una oposición directa entre *synsets*. Así, con los *synsets* {blanco,claro} y {negro,oscuro}, [blanco/negro] son antónimos, [claro/oscuro] son antónimos.

Hiponimia/hiperonimia. Es una relación transitiva y asimétrica. Es fundamental en la definición de nombres en WordNet. Un ejemplo se ilustra en la figura 3.2 donde el concepto pino pertenece a la clase de árbol.

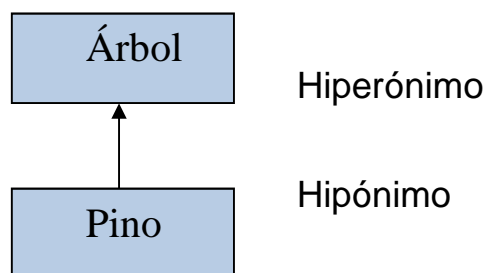


Figura 3.2 Fragmento de la ontología con relaciones axiomáticas entre conceptos

Holonimia/meronimia. Es una relación transitiva y asimétrica. Se asume “el concepto de una parte de un todo” puede ser “una parte de un concepto del todo”. Un ejemplo se ilustra en la figura 3.3 donde una rueda compone es un componente de un coche.

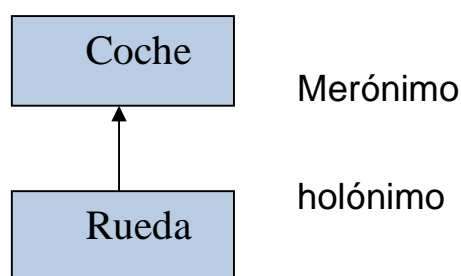


Figura 3.3 Fragmento de la ontología con relaciones axiomáticas entre conceptos

3.4.4. Apuntadores

Los apuntadores se utilizan para representar las relaciones entre las palabras de un *synset* a otro. Los apuntadores semánticos representan las relaciones entre los significados de la palabra y por lo tanto pertenecen a todas las palabras en los *synsets* origen y destino.

Algunos símbolos de apuntadores para los sustantivos son:

- ‘!’ Antónimo
- ‘@’ Hyperónimo
- ‘~’ Hipónimo
- ‘#m’ Miembro holónimo

- '#s' Substancia holónimo
- '#p' Parte holónimo
- '%m' Miembro merónimo
- '%s' Substancia merónimo
- '%p' Parte merónimo

3.5. Árbol Steiner

El problema del árbol Steiner [69][70] ha sido atribuido al matemático del siglo XIX Jakob Steiner, quien estudió el problema de la búsqueda de un único punto cuyas conexiones a los terminales tuvieran la menor longitud total posible. El problema general se denota como el problema del árbol de Steiner euclidiano y un conjunto soluciones óptimas que deben ser los árboles mínimos Steiner (SMTs), como ejemplo se presenta la distancia mínima euclidiana en 3 puntos en la figura 3.4. El problema del árbol de Steiner fue inicialmente de importancia sólo en el contexto de la agrimensura. Sin embargo, la reciente aplicación del árbol de Steiner es en diversas aplicaciones tales como enrutamiento VLSI, estimación de longitudes de cableado, construcción de árbol filogenético en biología y enrutamiento de red.

Utilizando propiedades estructurales fundamentales, Melzak [71] presentó el primer algoritmo para solucionar el problema. Hanan [72] presentó el primer estudio exhaustivo del problema del árbol Steiner rectilíneo en el que se midieron distancias entre puntos utilizando la métrica de distancia rectilínea (o manhattan. Los problemas de árbol Steiner euclidiano y rectilíneos pueden considerarse como casos especiales del problema árbol Steiner en grafos. Dado un grafo no dirigido $G = (V, E)$ con aristas positivas y un conjunto vacío no $Z \subseteq V$ de terminales, el problema es encontrar el árbol más corto de interconexión Z .

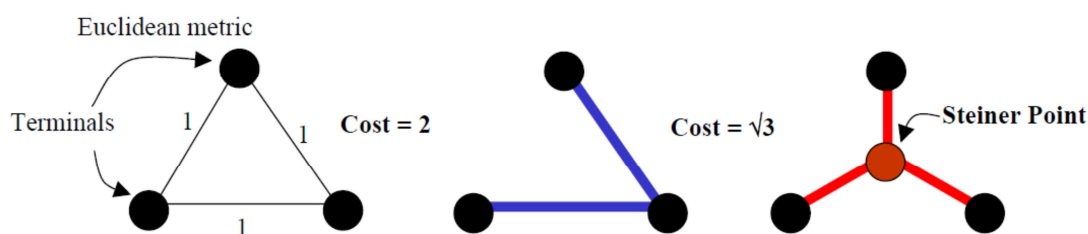


Figura 3.4 Fragmento de la ontología con relaciones axiomáticas entre conceptos

3.5.1. Versión no dirigida

El problema del árbol de Steiner se define de la siguiente manera: dado un grafo $G(V,E)$ con una función de coste c en los bordes y un subconjunto de vértices $X \subseteq V$ (llamados terminales), el objetivo es encontrar un árbol de costo mínimo que incluya todos los vértices en X . El árbol puede incluir vértices que no están en X (también conocidos como vértices de Steiner). El problema del árbol de Steiner es NP-completo incluso cuando el grafo es inducido por puntos en el plano.

3.5.2. Versión dirigida

La versión dirigida del problema del árbol de Steiner es una extensión de la versión no dirigida y se define como sigue. Dado un grafo dirigido $G = (V, A)$, una raíz especificada $r \in V$, y un conjunto de terminales $X \subseteq V$, el objetivo es encontrar la arborescencia de costo mínimo con raíz en r y que abarca todos los vértices en X (en otras palabras r debe tener un camino para cada vértice en X).

3.5.3. Problema generalizado de Steiner dirigido de red

Dado un grafo dirigido $G(V, E)$ y un conjunto $X = \{ \langle u_i, v_i \rangle \}$ de k pares de nodos encontrar el subgrafo de costo mínimo H de G tales que para cada par de nodos $\langle u_i, v_i \rangle \in X$ existe un camino dirigido desde u_i a v_i en H . El costo de la subgrafo H es la suma de los costos de todas las aristas en H . Aquí consideramos un problema un poco más general llamado como DG-Steiner(k, X) cual es el problema de encontrar un subgrafo de costo mínimo H de G que satisface al menos un par de nodo k del conjunto de X . Se asume sin pérdida de

generalidad que entre cada par de vértices $\langle u, v \rangle$ existe una ventaja de costo igual a la distancia más corta de la trayectoria de u a v en G .

3.5.4. Problema del árbol Steiner de grupo

Definición. Dado un grafo no dirigido $G(V,E)$, una raíz r y una colección de grupos $X=\{g_1,g_2\dots g_k\}$, encontrar el árbol de costo mínimo que conecta al menos un vértice en cada uno de los g_i a la raíz. Una aproximación poli-logarítmica fue propuesta por Garg [66] para este problema el conjunto TSP de grafos no dirigidos puede reducirse de manera similar a este problema. Hay una aproximación simple conservando la reducción entre variantes del nodo ponderado cuando el grafo está dirigido.

3.5.5. Algoritmos de aproximación

Se presentan los primeros algoritmos de aproximación no triviales para el problema del árbol de Steiner y el generalizado problema de red Steiner en grafos generales dirigidos. Se presentan unas definiciones básicas y lemas.

Definición 1 Dado una raíz $r \in V(G)$, un entero k y un conjunto $X \subseteq V$ de terminales con $|X| \geq k$, el problema D -STEINER(k,r,X) es construir un árbol con raíz en r , que abarca cualquier terminales k en X y da un mínimo costo posible.

Definición 2: Define $d(T)$, la densidad del árbol T , es el radio de costo del árbol al número de terminales en T ; en otras palabras $d(T)=c(T)/k(T)$.

Existen los siguientes puntos clave:

- Existe un camino de aristas más corto entre cada par de vértices de G .
- Todos los terminales están en las hojas del árbol de Steiner: esto puede arreglarse mediante la conexión de un terminal tonto al terminal real con un borde cero costo.

El algoritmo se basa en encontrar árboles con buena densidad repetitivamente y cada una que abarca sólo un subconjunto de los terminales. La solución final es una Unión de todos estos árboles.

Algoritmos para el cálculo del árbol Steiner [73]: los enfoques existentes pueden ser clasificados según sus estrategias:

- 1) Heurístico de distancia de la red (DNH)
- 2) Expansion y limpieza.
- 3) Programación dinámica
- 4) Búsqueda local.

DNH: Esta heurística [74], [75] construye un grafo completo en los terminales. Los pesos de las aristas reflejan la distancia más corta entre dos terminales en el grafo subyacente. Esta heurística es aplicable a los grafos de tamaño moderado, que pueden caber en la memoria principal. Ha sido emulado por otros enfoques para el cómputo de árbol de top-k en grupo Steiner [76], [77]. Los últimos dos enfoques, sin embargo, resultan tener un ratio de aproximación de $O(n)$, donde n es el número de términos de consulta [78].

Expansión y limpieza: Esta heurística [79] pretende construir el MST en los terminales a partir de un terminal de arbitrario y que abarca el árbol de manera gradual hasta que cubre todas las terminales. Se eliminan los nodos redundantes en una fase de limpieza. [80] explotando este método heurístico mediante dos estrategias diferentes de expansión. En contraste con la heurística original, cada terminal es un punto de partida de un árbol de rendimiento una un posible MST. Mientras que la primera estrategia de expansión elige la arista con un peso mínimo para abarcar un árbol (mínima expansión basada en la arista), la segunda estrategia elige el árbol de expansión que se traduce en un árbol de coste mínimo (expansión equilibrada de MST). Ambos métodos resultan tener un ratio de aproximación de $O(n)$ [81].

Programación dinámica: la primera aproximación dinámica al problema de árbol Steiner fue introducido por Dreyfus y Wagner [82]. Procede al calcular los resultados óptimos para todos los subconjuntos de terminales. Luego se calcula el resultado óptimo para todos los terminales.

Búsqueda local: este heurístico se ha utilizado en el dominio del problema del árbol de Steiner euclidiano y el Steiner árbol computación paralela [83], [84]. En la primera fase que se construye un árbol de interconexión basado en la heurística de distancia de red

introducida por [65]. En la segunda fase, el árbol actual es mejorado iterativamente considerando diferentes nodos del grafo subyacentes que pueden mejorar el costo del árbol actual.

3.6. Algoritmo de aproximación SketchLS

El algoritmo SketchLS [85] presenta una solución al problema de Steiner tree para grafos con millones de nodos y proporciona un error de aproximación de menos de 5% en promedio.

El algoritmo de aproximación tiene dos etapas: la indexación del grafo fuera de línea, y la aproximación rápida de distancias utilizando estos índices en la parte en línea o tiempo real.

Para la primera etapa indexación se ejecutan en los siguientes pasos:

1. Muestreo de semillas. Muestrear uniformemente conjuntos de nodos de semillas S_1, S_2, \dots, S_m de tamaño $1, 2, \dots, 2^{m-1}$ donde $m = \log |V|$, V es el número de nodos del grafo.
2. La ruta más corta en el árbol. Para cada conjunto de semillas se ejecuta el algoritmo de Dijkstra para calcular el árbol de ruta más corta SPTI.
3. Bosquejo. Para cada nodo $v \in V$ y para cada conjunto de semillas S_i obtenemos el nodo $l \in S_i$ el cual es el más cercano a v , y la ruta correspondiente entre v y S_i . Toda esta información se extrae de SPTI calculado en el paso anterior. El nodo $l \in S_i$ se llama el punto de referencia para v en S_i .

Después de cargar los bosquejos (Sketch) y la búsqueda de puntos de referencia l comunes entre ellos, se ejecuta la segunda etapa en tiempo real donde se construye el árbol de aproximación Steiner mediante la fusión de las rutas de los nodos de entrada a los puntos de referencia l .

Se comienza con la carga de sketches y la inicialización k instancias de búsqueda primero por amplitud (BF) sobre los sketches. Para cada proceso BFS se mantiene un conjunto de nodos de frontera, es decir, los nodos que se encuentran actualmente a la máxima distancia desde el nodo de origen.

El nodo v actualmente visitado puede estar conectado en el grafo G original con los nodos visitados por otros procesos. Se comprueba si es el caso, si el nodo v es de hecho un vecino de cualquier nodo previamente visitado mirando los vecinos de v en el grafo original G , y si este es el caso, se construye el camino entre q_i y q_j . Esta es la parte "búsqueda local" en un sentido que se trata de navegar por los vecinos de un nodo en el gráfico original. Se lleva un registro de los pares de nodos conectados en S_{cover} , y tan pronto como todos los nodos de entrada están cubiertos o conectados, se detiene el procedimiento. El conjunto S_{cover} se puede ver como una lista de aristas del grafo con nodos q_1, \dots, q_k , debido que a cada paso se añade una arista de forma (q_i, q_j) . El proceso se debe asegurar de que este grafo no tiene ciclos y que el resultado T es por lo tanto un árbol.

El algoritmo SketchLS en un principio, se cargan los sketches (Figura 3.5) y se inicializan tres procesos BFS. Después de unas pocas iteraciones, las aristas negras (Figura 3.6) denotan las aristas ya visitadas por iteraciones de Primero en Amplitud (BFS). El nodo actual v está en la frontera del primer proceso, y está conectado con el nodo l_2 de la frontera del tercer proceso. Inmediatamente se obtiene el camino (q_1, v, l_2, q_3) . En el siguiente paso (figura 3.7), los restantes procesos BFS1 y BFS2 se encuentran a través de la arista (v, l_3) . En este caso, también se descubre que v_1 y l_4 son vecinos, pero crearía un ciclo en S_{cover} : (q_1, q_2) y (q_2, q_1) , así que se omite este paso y se concluye ahora S_{cover} cubre los tres nodos de entrada (figura 3.8).

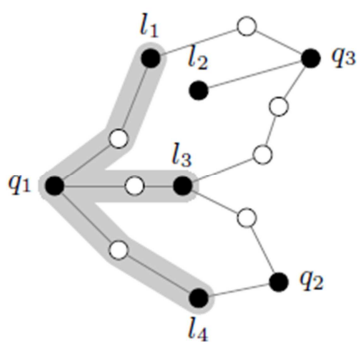


Figura 3.5 Sketch de q_1, q_2, q_3 .

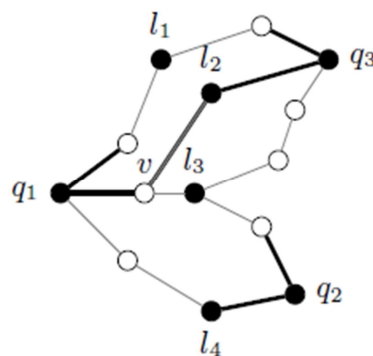


Figura 3.6 Una arista entre v y l_2 es descubierta, q_1 está conectado a q_3 .

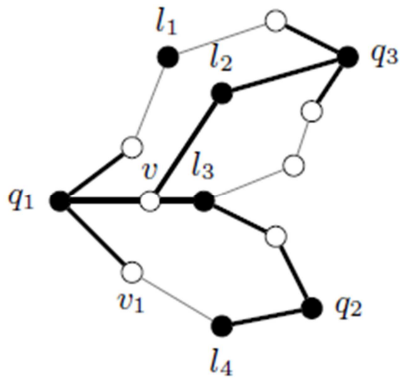


Figura 3.7 BFS1 y BFS2 también se reúnen (a través de l3).

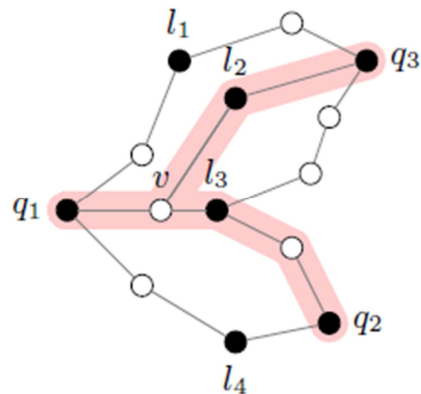


Figura 3.8 El grafo Steiner resultante es resaltado.

3.7. Cálculo de la visibilidad de conceptos en ontologías

El algoritmo DIS-C [86] se basa en el término visibilidad que es la “facilidad” que hay para llegar a un nodo, en otras palabras y desde el punto de vista de la ontología, es la “disponibilidad” del concepto en la conceptualización

Sea $K(C,R,R)$ una conceptualización donde C es el conjunto de conceptos, 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, es decir V_a , se debe hacer lo siguiente: se crea el grafo dirigido $G_k(V_g, A_g)$ para la conceptualización $K(C,R,R)$, haciendo que cada concepto $c \in C$ sea un vértice del grafo G_k , es decir $V_g = C$. Ahora para cada relación $a\rho b \in R$, donde $a, b \in C$ y $\rho \in R$, se añade la arista (a, ρ, b) a A_g .

El siguiente paso consiste en generar iterativamente, a partir de G_k , el grafo dirigido y ponderado $\Gamma_{kj}(V_{yk}, A_{yj})$. Para ello, en la j -ésima iteración, se hace que $V_{yk} = V_g$, y $A_{yk} = A_g$ para cada arista (a, ρ, b) , se añaden las aristas (a, b, w_{jab}) y (a, b, w_{jba}) , donde w_{jab} es el peso de la arista que va del nodo al nodo 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 ρ que conecta los nodos y el “costo de llegar” al nodo destino por el ocultamiento del mismo, esto se expresa de la siguiente manera:

$$w_{ab}^j = V^{j-1}_a w^0_{a+} [\delta^\rho]^{j-1} + \frac{1}{V^{j-1}_b} w^j_b$$

$$w_{ba}^j = V^{j-1}_b w^0_{b+} [\bar{\delta}^\rho]^{j-1} + \frac{1}{V^{j-1}_a} w^j_a$$

Siendo precisamente la visibilidad del vértice x en la j -ésima iteración, se calcula su valor 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:

$$V_x^j = \frac{\frac{\sum_{y \in v_y^j} \text{dist}_{\Gamma_k^j}(x,y)}{\text{card}(v_y^j)}}{\frac{\sum_{y \in v_y^j} \text{dist}_{\Gamma_k^j}(y,x)}{\text{card}(v_y^j)}} = \frac{\sum_{y \in v_y^j} \text{dist}_{\Gamma_k^j}(x,y)}{\sum_{y \in v_y^j} \text{dist}_{\Gamma_k^j}(y,x)}$$

Donde $\text{dist}_{\Gamma_k^j}$ es la distancia mínima desde el nodo hacia el nodo según en grafo Γ_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 $[\delta^\rho]$ y $[\bar{\delta}^\rho]$ 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 w_j entre las aristas que comparten el mismo tipo de relación, es decir:

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

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

Donde $\rho^* \{(a,\rho,b) \in \text{Ag}\}$ es el conjunto de aristas que representan una relación ρ . Inicialmente, estas distancias valen 0, es decir que $\forall \rho \in R, [\delta^\rho]_0=0$ $[\bar{\delta}^\rho]_0=0$.

Por otro lado, w_x^i es el costo de “llegar” al vértice $x \in V_g$, definido como $w_x^i = \frac{O_x}{i_x + O_x}$ y es el costo de “salir” del concepto, definido como o el cual 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 ($O_x = \text{card}(\cup \{y\} (x,y,-))$).

Para ejemplificar esto, consideremos la ontología de la Figura 3.9. 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).

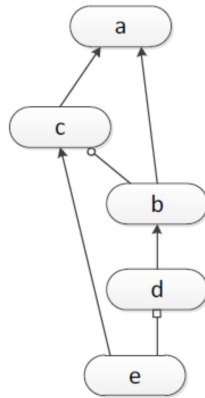


Figura 3.9 Ontología ejemplo.

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

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 $w_{ab}^i = v_a^o w_a^o + [\delta\rho]^0 + \frac{1}{v_b^o} w_b^i = w_a^o + w_b^i = 1 + \frac{2}{3} = \frac{5}{3}$. De igual forma, $w_{ba}^1 = \frac{1}{3}$.

Este proceso se realiza comenzando con 1, y aumentando el valor en una unidad, hasta que $\frac{\sum_{x \in v_y} (v_x^j - v_x^{j-1})^2}{\text{card}(v_y)} \leq E_k$; donde $\frac{\sum_{x \in v_y} (v_x^j - v_x^{j-1})^2}{\text{card}(v_y)}$ 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, E_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.

4. Metodología

El presente trabajo proporciona una metodología que permite integrar y recuperar semánticamente conjuntos de datos geospaciales heterogéneos sobre la *web* mediante la extracción de información cognitiva modelada sobre una ontología.

Los datos geospaciales encontrados sobre páginas *web*, bases de datos, bases de conocimiento, etc., pueden ser descritos por medio de una ontología para su procesamiento semántico con la finalidad de hacer una mejor búsqueda mejor relacionada con las especificaciones del usuario.

4.1. Marco de trabajo general

Con esta metodología se busca proporcionar un mecanismo para la recuperación de objetos espaciales con un enfoque basado en conocimiento descrito explícitamente por medio de ontologías. En la Figura 4.1 se muestra la estructura conceptual del método de recuperación de datos geospaciales sobre la *web* mediante integración de conocimiento, el cual está compuesto de tres etapas fundamentales para llevar a cabo la integración y recuperación de objetos geospaciales.

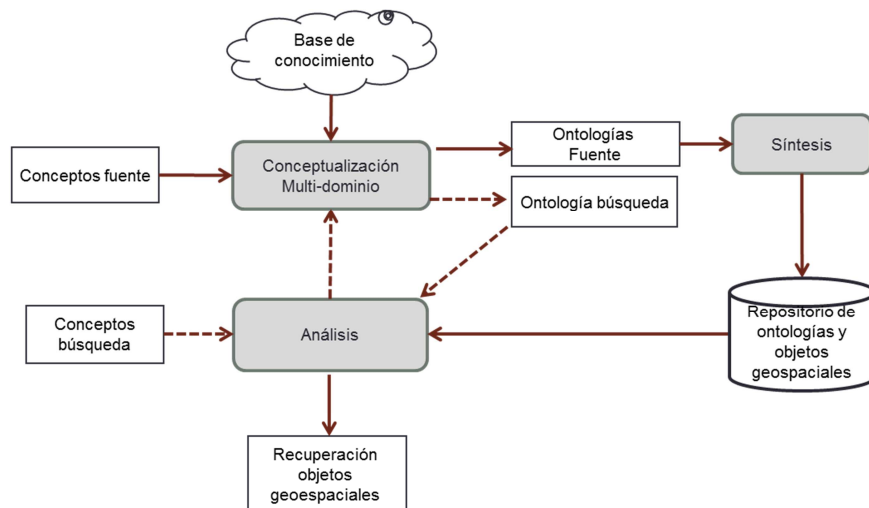


Figura 4.1 Estructura conceptual de la metodología.

Las etapas que componen a la metodología son las siguientes:

- **Conceptualización.** En esta etapa se lleva a cabo el proceso de la construcción de ontologías para un grupo de términos de entrada que describen a un objeto espacial o consulta. Este proceso se compone de subprocesos de tokenizer, eliminación de *Stop Words*, lematización, establecimiento de medidas similitud entre los conceptos, desambiguación y reducción, unión de extractos de ontologías especializadas para formar una ontología multi-dominio.
- **Síntesis.** En esta etapa se crean en un repositorio donde se concentra un conjunto de ontologías que representan cada una de las fuentes correspondientes y las descripciones objetos geoespaciales sobre la *web*, este repositorio facilita la gestión y recuperación mediante técnicas de indexado.
- **Análisis.** En esta etapa se obtiene la consulta de parte del usuario, esta fase utiliza la fase de conceptualización para construir una ontología a partir de un conjunto de términos derivados de la consulta, se generan medidas de similitud entre ontologías por medio de alineación para recuperar objetos espaciales de la fase de síntesis y que están relacionadas semánticamente con la consulta.

En figura 4.2 se presenta el esquema general de la metodología. En la etapa de conceptualización, con base en un conjunto de términos que describen a un objeto geoespacial se genera automáticamente una ontología que describe la porción de conocimiento referida, como resultado de esta etapa se obtiene una ontología por cada objeto geoespacial.

Con respecto a la etapa de síntesis, se concentra cada conceptualización (ontología) y la meta-información de cada objeto geoespacial mediante un repositorio de ontologías y objetos espaciales, se provee un mecanismo de control, gestión y acceso para las siguiente fase de análisis, esta última fase se encarga de usar la fase de conceptualización para construir una ontología que representa a la búsqueda, de este modo esta ontología de consulta se comparara con otras ontologías de la fase de síntesis mediante una métrica de similitud por alineación, el resultado de la metodología es un conjunto de objetos

geográficos ponderados en relación a un nivel de similitud entre la ontología de consulta y de los objeto geoespaciales, se provee métodos de acceso a cualquier objeto.

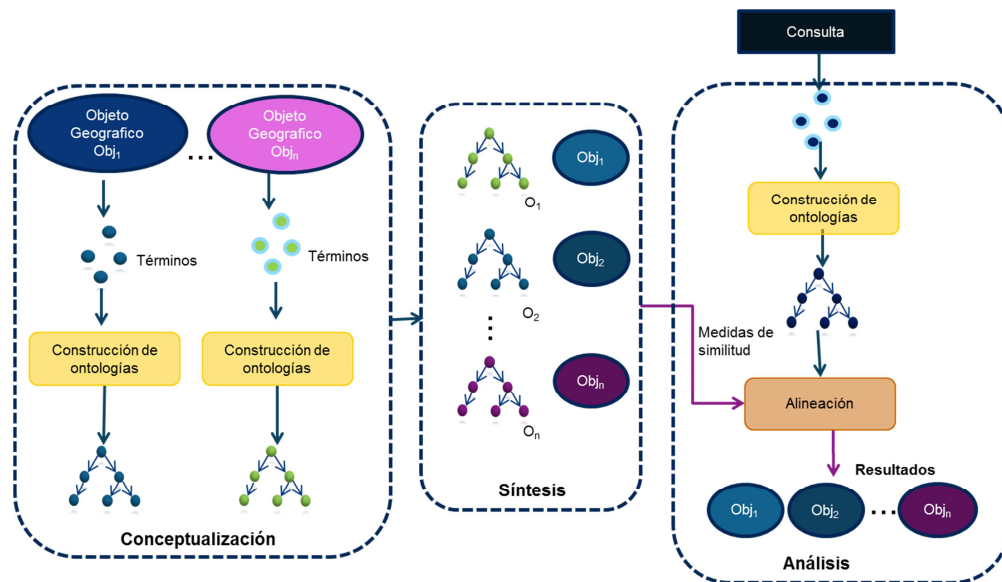


Figura 4.2 Esquema general de la metodología.

4.2. Etapa de conceptualización

En esta sección se describe la etapa de conceptualización constituida por cinco etapas principales. Estas etapas se denominan "pre-procesamiento léxico", "expansión sobre la base de conocimiento multi-dominio", "medida de similitud", "desambiguación y la reducción de la dimensión de la ontología base" y "la fusión de la ontología base y ontologías externas".

En la etapa de "pre-procesamiento léxico", el texto no estructurado es procesado con el fin de ajustar léxicamente el concepto del texto a la forma en la cual está descrito en la base de conocimiento. En la "expansión sobre la base de conocimiento multi-dominio" los conceptos que se encontraron en la base de conocimientos son expandidos, si se encuentran en la ontología base se expanden hasta encontrar una raíz común dentro de la misma ontología, si es el caso y se encuentran en ontologías externas (especializadas) se expanden para crear un extracto de la ontología. La etapa de "medida de similitud" establece medidas de similitud entre los conceptos de una ontología creada por medio de la expansión. La

etapa "desambiguación y la reducción de la dimensión de la ontología base " define dentro de la ontología el sentido exacto de cada uno de los conceptos considerando el resto de ellos, esta etapa además reduce el número de conceptos y relaciones de la ontología de expansión. Finalmente "la fusión de la ontología base y ontologías externas" combina la ontología base con la(s) ontología(s) externa(s) de los conceptos que no fueron descritos en la ontología base, de este modo se crea la ontología final multi-dominio.

La figura 4.3 resume las cinco etapas de la fase de conceptualización las cuales serán descritas en las siguientes secciones.

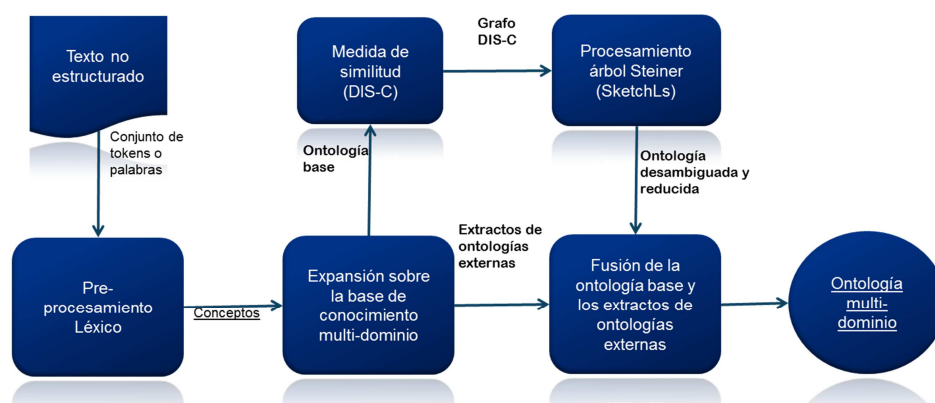


Figura 4.3 Esquema general de la fase de conceptualización.

4.2.1. Preprocesamiento léxico

Para cada término obtenido del texto no estructurado se ejecuta un pre-procesamiento léxico con el fin de proporcionar a las próximas etapas de la metodología un término adecuado tal y como esta descrito en la base de conocimiento (figura 4.4).

Los procesos que componen el procesamiento léxico se enumeran a continuación:

1. **Tokenizer.** En este procesamiento el texto se descompone en términos llamados *tokens*.
2. **Eliminación de stop Words.** Si un término es un stop Word, este se elimina.
3. **Stemming.** Cada concepto es reducido a su raíz (lexema) ejecutando el algoritmo Krovetz.
4. **Eliminación de conceptos desconocidos.** Esto se ejecuta con la finalidad de realizar una optimización y de reducir el procesamiento futuro etapas de la conceptualización.

Cuando un término no se encuentra en la base de conocimiento (KB), entonces se descarta.

Los términos resultantes de esta etapa son entregados al proceso de expansión sobre la base de conocimiento. En caso de que el concepto no se encuentre en la ontología base, se pretende ubicarlo en el resto de la base de conocimiento por medio de ontologías externas. Estas expansiones van a ser descritas en la siguiente sección.

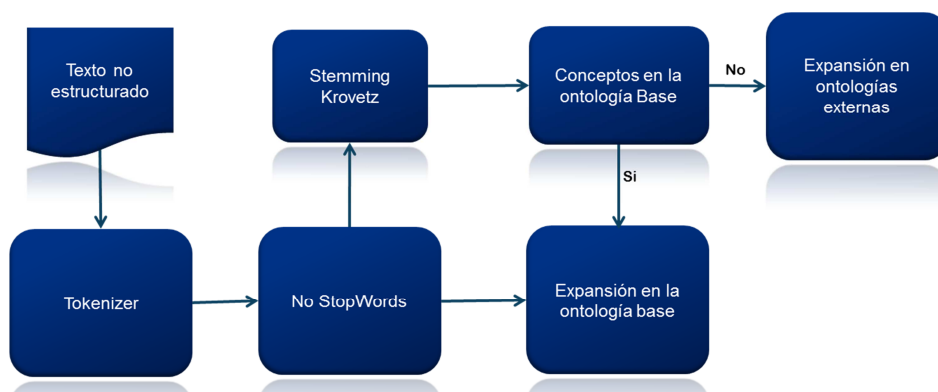


Figura 4.4 Preprocesamiento léxico y expansión sobre la base de conocimiento.

4.2.2. Expansión sobre la base de conocimiento multi-dominio

Definimos la expansión sobre una ontología (O_e) para un concepto C como el proceso iterativo (α iteración) de descubrir nuevos conceptos dentro de una ontología O mediante relaciones semánticas (ρ) que conectan un concepto origen C_x al otro concepto de destino C_y .

$$Oe_{\alpha}^{\rho}(c, O(C, R)) = O_{\alpha}(C_{\alpha}^{\rho}, R_{\alpha}^{\rho})$$

$$C_{\alpha}^{\rho} = \begin{cases} \alpha = 0 & \{c\} \\ \alpha > 0 & C_{\alpha-1} \cup \{y \in C : x \in C_{\alpha-1}, \rho(x, y) \in R\} \end{cases}$$

$$R_{\alpha}^{\rho} = \begin{cases} \alpha = 0 & \emptyset \\ \alpha > 0 & \{\rho(x, y) \in R : x, y \in C, x \in C_{\alpha-1}\} \end{cases}$$

4.2.2.1. EXPANSIÓN ONTOLÓGICA EN LA ONTOLOGÍA BASE

La expansión sobre la ontología base considera dos direcciones de expansión: como nivel superior e inferior (figura 4.5).

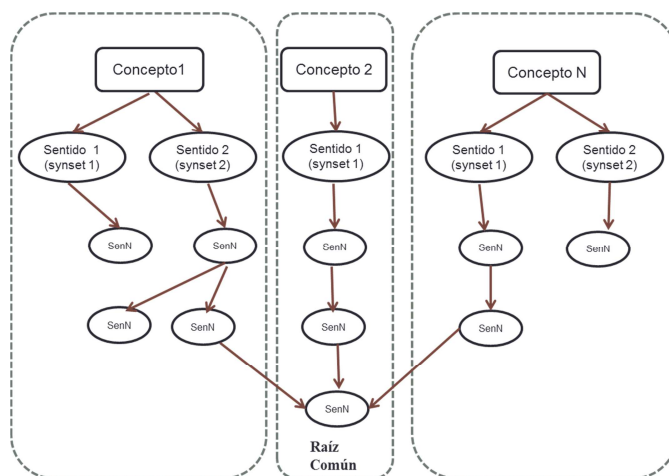


Figura 4.5 Esquema general de la Fase de conceptualización.

Definimos la expansión a un nivel superior (UeW) para una ontología base O_x , como el incremento del radio de relaciones (número de conceptos) para una iteración α como el procedimiento de obtención de nuevos conceptos mediante las relaciones “hiperónimo (@), Instancia de hiperónimo (@i), holónimo miembro (#m), holónimo sustancia (#s), holónimo parte (#p)” .

$$UeW(C_x, O_x, \alpha) = \{C_n \in Oe_\alpha^\rho(C_x, O_x): \rho \in \{ @, @i, \#m, \#s, \#p \} \}$$

En la expansión a un nivel inferior (LeW) los conceptos se obtienen exactamente del mismo modo como la expansión a un nivel superior pero teniendo en cuenta las relaciones: hyponym (\sim), Instancia hyponym ($\sim i$), merónimo miembro (% m), merónimo sustancia (% s), parte merónimo (% p).

$$LeW(C_x, O_x, \alpha) = \{C_n \in Oe_\alpha^\rho(C_x, O_x): \rho \in \{ \sim, \sim i, \%m, \%s, \%p \} \}$$

La expansión sobre la ontología base se ejecuta para cada termino tratado en la etapa anterior de pre-procesamiento léxico, lo cuales nombraremos como sub-expansiones. Con

el fin de garantizar una ontología completa y un grafo fuertemente conexo, la expansión se procesa de forma iterativa hasta que en una α iteración todas las sub-expansiones comparten un concepto común al nombraremos raíz.

La expansión para ontología base (AOEB) se presenta en el Algoritmo 1.

Algorithm 1 Expansion sobre la ontologia base (AOEB(S_c, O, ρ))

Input:

S_c : conjunto de conceptos a expandir

O : ontologia a ser expandida

ρ : relaciones a ser consideradas

Output: la ontologia expandida O_e

$\alpha \leftarrow 0$

repeat

$\alpha \leftarrow \alpha + 1$

$c_i \leftarrow$ cualquier concepto en S_c

$r \leftarrow$ conjunto de conceptos de $Oe_\alpha^\rho(c_i, O)$

foreach concepto $c \in S_c$ distinto a c_i **do**

$r_c \leftarrow$ conjunto de conceptos de $Oe_\alpha^\rho(c, O)$

$r \leftarrow r \cap r_c$

end

until $r \neq \emptyset$;

$O_e \leftarrow \bigcup_{c \in S_c} Oe_\alpha^\rho(c, O)$

return O_e

Algoritmo 1 Expansión para ontología base (AOEB)

4.2.2.2. EXPANSIÓN ONTOLÓGICA EN ONTOLOGÍAS EXTERNAS

Para los conceptos que no pudieron ser expandidos dentro de la ontología base, estos se buscan dentro de ontologías externas, si son localizados se genera su expansión sobre dicha ontología. Para la expansión sobre ontologías externas se considera únicamente la expansión nivel superior ya que en general, la mayor parte de las ontologías externas son del tipo jerárquico. De este modo en la expansión de nivel superior para una ontología externa (UEX), los conceptos son obtenidos mediante el aumento de radio de relaciones una iteración α usando exclusivamente la reacción “subclassoff” (#) es decir, la expansión se realiza a los conceptos superclase.

$$\text{UeX}(C_x, O_x, \alpha) = \{C_n \in Oe_\alpha^\rho(C_x, O_x): \rho \in \{\#\}\}$$

La expansión sobre la ontología externa (AOEE) se describe en el algoritmo 2.

Algorithm 2 Expansion sobre una ontología externa (AOEE(c, O_x, O, ρ))

Input:

c : El concepto a ser expandido

$O_x(C_x, R_x)$: Ontología externa donde la expansion es realizada

$O(C, R)$: La ontología base

ρ : Las relaciones a ser consideradas

Output: La ontología expandida O_e

$\alpha \leftarrow 0$

repeat

$\alpha \leftarrow \alpha + 1$

$Cx_\alpha^\rho \leftarrow$ El conjunto de conceptos de $Oe_\alpha^\rho(c, O_x)$

$r \leftarrow \Psi(C, Cx_\alpha^\rho)$

until $r \neq \emptyset$ or $Cx_\alpha^\rho = Cx_{\alpha-1}^\rho$;

$O_e \leftarrow Oe_\alpha^\rho(c, O_x)$

return O_e

Algoritmo 2 Expansión sobre la ontología externa (AOEE).

4.2.3. Medida de similitud

Una vez que la ontología fue construida en la etapa anterior el siguiente paso es establecer medidas de similitud entre los conceptos que componen la ontología. Para ello se implementó la medida de similitud de distancia conceptual DIS-C (figura 4.6). Como resultado de este procesamiento se obtiene un grafo dirigido ponderado que será entregado a la siguiente etapa de reducción de dimensión y desambiguación. El grafo DIS-C mantiene como información extra el tipo de relación semántica de sus aristas respecto a la ontología de entrada, esta información será útil para la última etapa de conceptualización “fusión de la ontología base y ontologías externas”.



Figura 4.6 Algoritmo DIS-C.

4.2.4. Desambiguación y reducción de la dimensión de la ontología base

En esta etapa el grafo DIS-C es desambiguado y su dimensión es reducida (menor número de nodos y relaciones) hasta obtener la representación mínima de conocimiento para los terminos de entrada a la etapa conceptualización. Como resultado es obtenida la ontología base final (figura 4.7).

Con base en el grafo DIS-C obtenido de la etapa anterior, es aplicado el algoritmo SketchLs [78] el cual produce su árbol de Steiner, este algoritmo fue seleccionado este debido que es una aproximación rápida para grandes grafos con una solución heurística factible en relación tiempo, ejecución y exactitud con una baja complejidad.

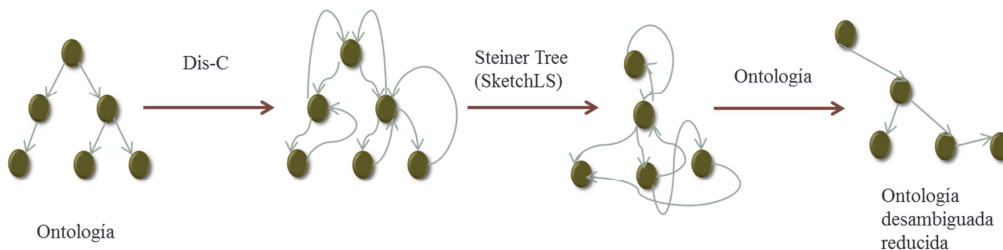


Figura 4.7 Algoritmo. SketchLs para Steiner Tree.

Un problema presentado fue que el algoritmo SketchLs está orientado a resolver un árbol Steiner Dirigido pero la presente metodología tuvo que enfrentar el problema del árbol Steiner Dirigido en Grupo, esto debido a que un concepto puede ser incluido en varios sentidos (*synsets*). Con el fin de resolver esta problemática, el conjunto de sentidos se agrupo en un solo nodo terminal por medio de una unión de vértices, es decir, se realiza un proceso de sustitución para que sea posible su procesamiento como un simple árbol D-Steiner y de este modo sea procesado por el algoritmo SkecthLS.

Se identifican, en primera instancia, los sentidos para cada terminal, posteriormente se sustituyen todos los sentidos por un nuevo nodo terminal que une al grupo de sentidos, esta nueva terminal hereda todas las relaciones del grupo de terminales. La siguiente fase es procesar el árbol Steiner con SkectcLs, al terminar el procesamiento de este algoritmo se identifica cual relación se encuentra en el árbol Steiner resultante, finalmente se re-sustituye la terminal artificial con uno del grupo de sentidos usando la relación que lo conecta. Como ejemplo de este proceso se muestra la figura 4.8 para este proceso de sustituciones.

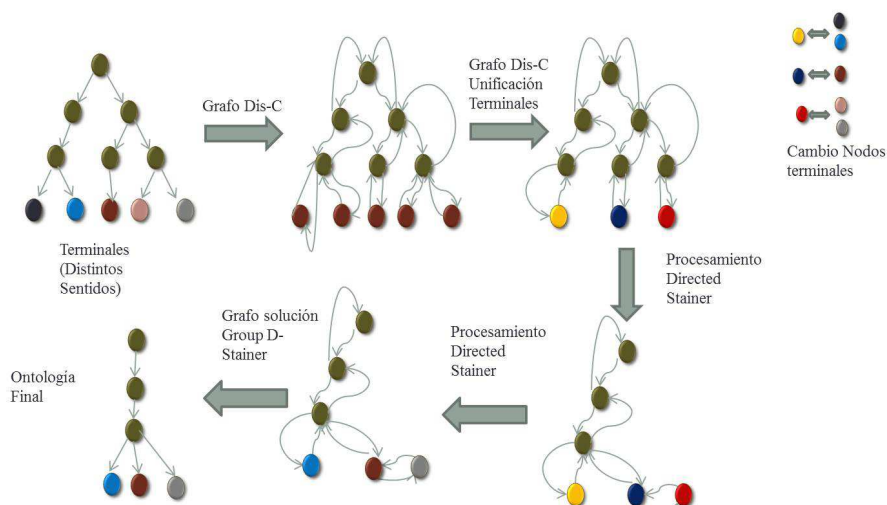


Figura 4.8 Procesamiento de D-Stainer a Group D-Stainer.

La descripción del algoritmo para agrupar la terminal y la preparación para el procesamiento de Group D-Steiner (AGST) se muestra en el algoritmo 3.

El resultado de esta etapa presenta un grafo mínimo desambiguado que representará a la ontología base final.

Algorithm 3 Agrupando Terminales (Group Steiner Tree a Directed Steiner Tree (AGST($O_b, ST, SenT$)))

Input: O_b : ontología creada por expansion sobre el dominio base
 ST : conjunto de conceptos terminales
 $SenT[size(ST)]$: arreglo de conjunto de conceptos que representan los sentidos de cada terminal respectivamente
Output: ontología steiner OST /* */

```

nT ← 1 /* nT numero de terminal siendo agrupado, se inicializa con el primer
terminal, */
foreach  $C_i \in ST$  do
  foreach  $c_j \in SenT[nT]$  do
    foreach  $c_k \in O_b$  do
      if  $c_k = c_j$  then
         $c_k \leftarrow -(nT)$  /* renombrar el concepto en  $O_b$  con el numero de concepto
        -nT, este numero se decrementa con la finalidad de distinguir
        esta terminal en la ontología base  $O_b$  */
      end
    end
  end
  nT ← nT + 1
end
ObST ← ST( $O_b$ ) /* calcular steiner tree a  $O_b$  y asignarlo a obST */
nT ← 1
foreach  $c_i \in ST$  do
  foreach  $c_j \in ObST$  do
    if  $c_j = -(nT)$  then
       $IdT \leftarrow Idp : C_j Idp C_z : C_j, C_z \in ObST$  /* obtener el identificador de
      relacion en obST */
       $c_s \leftarrow c_z : c_s IdT C_z : c_s, c_z \in Ob$  /* obtener el concepto destino de  $O_b$  basado
      en IdT */
       $c_j \leftarrow c_s$ 
    end
  end
  nT ← nT + 1
end
return ObST

```

Algoritmo 3 Group D-Steiner (AGST)

4.2.5. Fusión de la ontología base y ontologías externas

Esta etapa tiene la tarea de construir una ontología multi-dominio fusionando la expansión realizada sobre la ontología base y las expansiones sobre las ontologías externas, esta fase ejecuta la “expansión sobre la base de conocimientos multi-dominio” (figura 4.9). La

ontología base es fusionada con varios extractos de ontologías externas para componer de este modo una ontología final multi-dominio. El procedimiento completo se muestra en la figura.

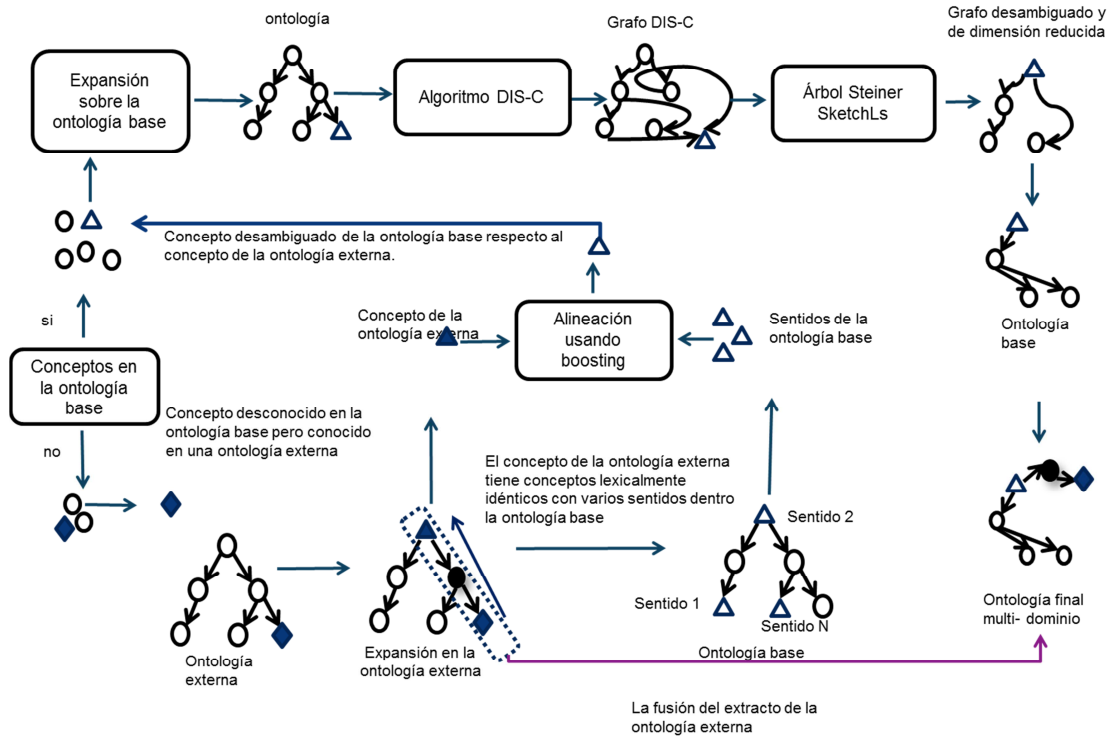


Figura 4.9 Ontología final multi-dominio.

Esta etapa comienza recibiendo cada concepto desconocido de la ontología base e intenta encontrarlos en ontologías externas (especializadas). Si el concepto se encuentra en la ontología externa con un único sentido este se expande hasta encontrar un concepto idéntico lexicalmente dentro de la ontología base. En contraste, puede existir el caso del sentido múltiple, donde un concepto tiene varios sentidos en la ontología externa y varios sentidos en la ontología base; este caso no será procesado en la alineación y será descartado.

Una vez que el concepto se encuentra en ambas ontologías, este puede ser ambiguo e incluido en varios sentidos dentro la ontología base. Con el fin de seleccionar sólo un sentido, se aplica el algoritmo de alineación basado en *boosting* [87] para encontrar el par de conceptos con la más alta similitud semántica entre ambas ontologías. Una vez que se selecciona un concepto de la ontología base este es incluido para formar parte de la

expansión sobre la ontología base y ayudar a definir el sentido de los demás conceptos durante la expansión. Es importante mencionar que el concepto de la ontología base encontrado y agregado por alineación ya se encuentra desambiguado, para la etapa de “desambiguación y reducción de la ontología base“, este concepto se compone de un solo sentido y suele omitir el procesamiento de agrupación de sentidos por lo que se incluirá directamente.

El proceso fusión tiene que ser ejecutado por cada concepto desconocido en la ontología base que fue encontrado en una ontología externa. La construcción de la ontología multi-dominio mediante la fusión de las ontologías base y externas se describe en el algoritmo 4.

Algorithm 4 Fusión de ontologías (MOBE(S_c, O, Ω, ρ))

```

Input:  $S_c$ : Conjunto de conceptos para crear la ontología multi-dominio
 $O(C, R)$ : La ontología base
 $\Omega = \{O_1, O_2, \dots, O_k\}$ : Conjunto de ontologías externas para  $k$  dominios
 $\rho$ : Las Relaciones a ser consideradas
Output: La ontología multi dominio  $O_m$ 
/* dividir conceptos internos  $C_i$  y conceptos externos  $C_x$  */
 $C_i \leftarrow \emptyset$ 
 $C_x \leftarrow \emptyset$ 
foreach  $c \in S_c$  do
  if  $c \in C$  then
     $C_i \leftarrow C_i \cup \{c\}$ 
  else
     $C_x \leftarrow C_x \cup \{c\}$ 
  end
end
/* Aplicar expansion para buscar cada concepto externo en cada ontología
externa */
 $E \leftarrow \emptyset$ 
foreach  $c \in C_x$  do
  foreach  $O_x \in \Omega$  do
     $E \leftarrow E \cup \{AOEE(c, O_x, \rho)\}$ 
  end
end
/* Aplicar expansion para cubrir todos los conceptos internos en la ontología
base */
 $O_e(C_e, R_e) \leftarrow AOEB(C_i, O, \rho)$ 
/* Aplicar desambiguación para obtener la ontología de cobertura mínima */
 $O_d(C_d, R_d) \leftarrow AGST(O_e)$ 
/* Unir cada partición de las ontologías externas con la ontología de
cobertura mínima para la construcción la ontología multi-dominio */
foreach  $E_i(C_j, R_j) \in E$  do
   $C_d \leftarrow C_d \cup C_j$ 
   $R_d \leftarrow R_d \cup R_j$ 
end
return  $O_m(C_d, R_d)$ 

```

Algoritmo 4. Fusión de las ontologías

4.3. Etapa de Síntesis

La etapa de síntesis tiene como entrada el conjunto de ontologías producidas por la fase de conceptualización, se genera un repositorio de ontologías y objetos espaciales, se proveerán los mecanismos de inclusión y extracción de ontologías y objetos geoespaciales como se describe en la figura 4.10. El repositorio se compone de conjunto de ontologías indexadas para su fácil gestión y recuperación, esta fase provee el módulo de extracción el

qual mediante una petición devuelve un subconjunto de ontologías o el total de ontologías del repositorio a la fase de análisis. Además se provee otro módulo de extracción de objetos espaciales que sirve a la fase de análisis, en este módulo son obtenidos los objetos geoespaciales mediante una petición que incluye los identificadores/índices de ontologías relacionados a los objetos.

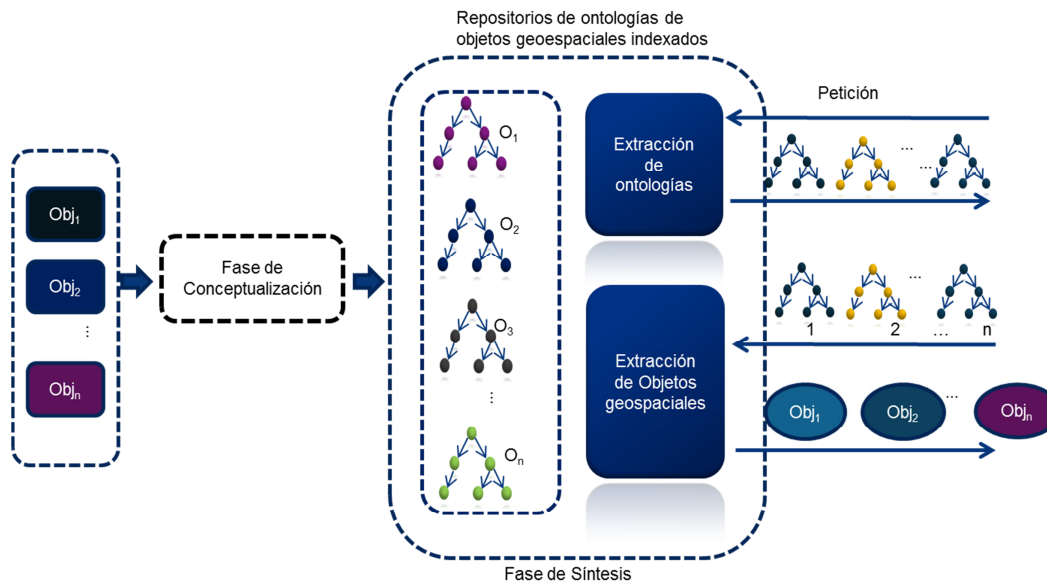


Figura 4.10 Fase de Síntesis.

El proceso de petición de ontologías a la fase de síntesis se presenta en el Algoritmo 5.

Algorithm 5 Sintesis_Peticion_Ontologias ($SPO(L_i, L_s)$)**Input:** l_i : Limite inferior del segmento de ontologias a recuperar l_s : Limite superior del segmento de ontologias a recuperar**Output:** Arreglo de ontologias recuperadas A_o

```

 $T_o \leftarrow size(RS)$  /* Obtener el numero total de ontologias en el repositorio de
sintesis RS */
 $AT_o[T_o] \leftarrow \{o|o \in RS\}$  /* Arreglo de de dimension  $T_o$  para todas las ontologias en
RS */
 $i \leftarrow 0$  /* Contador */
 $j \leftarrow 0$  /* Contador */
 $A_o[l_s - l_i] \leftarrow \phi$  /* Arreglo de ontologias de dimension  $l_s - l_i$  */
for  $i = l_i$   $i < l_s$   $i++$  do
     $A_o[j] \leftarrow AT_o[i]$  /* Agregar ontologia (conjunto de conceptos, relaciones y
propiedades) */
     $j \leftarrow j + 1$ 
end
return  $A_o$ 

```

Algoritmo 5 Síntesis petición de ontologías

El proceso de petición de objetos espaciales a la fase de síntesis se presenta en el Algoritmo 6.

Algorithm 6 Sintesis_Peticion_Objeto_Espaciales (SPOE(L_o))**Input:** l_o : Arreglo de índices de ontologías almacenadas en el repositorio de síntesis RS**Output:** Arreglo de objetos espaciales recuperados A_s

```

 $T_o \leftarrow size(RS)$  /* Obtener el numero total de ontologías en el repositorio de
    síntesis RS */
 $AT_s[T_o] \leftarrow \{S_o | S_o \in RS\}$  /* Arreglo de dimension  $T_o$  para todos los objetos
    espaciales en RS */
 $i \leftarrow 0$  /* Contador */
 $j \leftarrow 0$  /* Contador */
 $k \leftarrow 0$  /* Contador */
 $T_s \leftarrow size(l_o)$  /* Obtener el numero de ontologías del arreglo  $l_o$  */
 $A_s[T_s] \leftarrow \phi$  /* Arreglo de dimension  $T_s$  para objetos espaciales */
for  $i = 0$   $i < AT_s.size$   $i++$  do
    for  $j = 0$   $j < l_o.size$   $j++$  do
        if  $AT_s[i].id = l_o[j].id$  then
            /* Objeto espacial representado por una ontología) */
             $A_s[k] \leftarrow AT_s[i]$  /* Agregar objeto espacial (meta-informacion y mecanismos
                de recuperacion del objeto) */
             $k \leftarrow k + 1$ 
        end
    end
end
return  $A_s$ 

```

Algoritmo 6 Síntesis petición de objetos espaciales

4.4. Etapa de Análisis

En esta etapa se realiza el procesamiento de búsqueda y recuperación de objetos espaciales, (véase figura 4.11), se inicia con la obtención de una consulta la cual es representada por una ontología mediante un procesamiento de conceptualización, esta ontología de consulta resultante es contrastada mediante un proceso de comparación entre ontologías por alineación $\Gamma(O_x, O_y, \delta)$ aplicando una medida de similitud (δ), la ontología de consulta se compara con las ontologías de la fase de síntesis con la finalidad de ponderar objetos geoespaciales ligados a estas.

Para llevar a cabo este proceso de comparación se utilizan tres medida de similitud entre ontologías (δ) las cuales balancean la relación entre exactitud y rendimiento en la obtención de resultados.

Las medidas de similitud consideradas en la metodología son las siguientes:

Similitud léxica (Ψ). Esta similitud mide la similitud de los conceptos entre ontologías léxicamente. Sea dos ontologías O_x , O_y . la similitud léxica entre ontologías es la suma de las mayores valores de similitud por cada concepto de la ontología O_x dentro de la ontología O_y entre la suma del número conceptos de ambas ontologías .

$$\Psi(O_x, O_y) = \frac{\sum_{i=0}^n SimML(C_i, O_y)}{Card_C(O_x) \cup Card_C(O_y)}$$

Donde el valor de $SimML(C_x, O_y) = \text{MAX}(\text{Distancia_levenshtein}(C_x, SO))$, $SO = \{C_y | C_y \in O_y\}$, $\text{Distancia_levenshtein} \in [0, 1]$.

Similitud estructural (λ). Este tipo de similitud compara el tipo de relaciones que componen la estructura entre las ontologías. La medida es la relación del número de relaciones que comparten entre la suma del número de relaciones de ambas ontologías.

$$\lambda(O_x, O_y) = \frac{R(O_x) \cap R(O_y)}{Card_R(O_x) \cup Card_R(O_y)}$$

Similitud ontológica (Ω). En esta similitud es una híbrida de las anteriores similitudes, se consideran tres propiedades de una ontología las cuales son la similitud léxica (conceptos), similitud estructural (relaciones) y similitud de propiedades.

En primer lugar introducimos la similitud entre superclases (θ) como un proceso iterativo para seleccionar el par de conceptos con mayor similitud entre sus superclases.

$$\text{Similitud Superclases}(C_x, C_y) = \text{similitud_parcial}(C_x, C_y).$$

$$\text{Similitud parcial}(C_a, C_b) = \frac{\Psi(C_a, C_b) + \lambda(C_a, C_b)}{2}$$

$$\theta(O_x, O_y) = \text{MAX}(\text{Similitud_Superclases}(S_O, D_O)), S_O = \{C_x | C_x \in O_x\}, D_O = \{C_y | C_y \in O_y\},$$

Finalmente la similitud ontológica suma las similitudes léxica, estructural y de propiedades dando un factores de relevancia ajustables (α, β, δ).

$$\Omega((O_x, O_y)) = \alpha\{\Psi(O_x, O_y)\} + \beta\{\lambda(O_x, O_y)\} + \delta\theta(O_x, O_y)\}$$

Con base en las diferentes medidas de similitud entre ontologías, en la fase de recuperación los objetos son ordenados por relevancia y además se proveen los mecanismos de acceso a ellos.

Con fines de comparación de las medidas de similitud basadas en ontologías descritas anteriormente, se propuso una medida adicional como línea base la cual es totalmente léxica y solo comparan términos exactos.

Comparación Léxica (CL): Esta medida de similitud léxica calcula el promedio de términos compartidos léxicamente entre dos ontologías, con base en esta medida se genera una lista ordenada decreciente donde recuperan los objetos espaciales donde en sus descripciones se localicen términos de la consulta, es decir: $CL(O_x, O_y) = \frac{T(O_x) \cap T(O_y)}{\text{Card}\{T(O_x)\}}$.

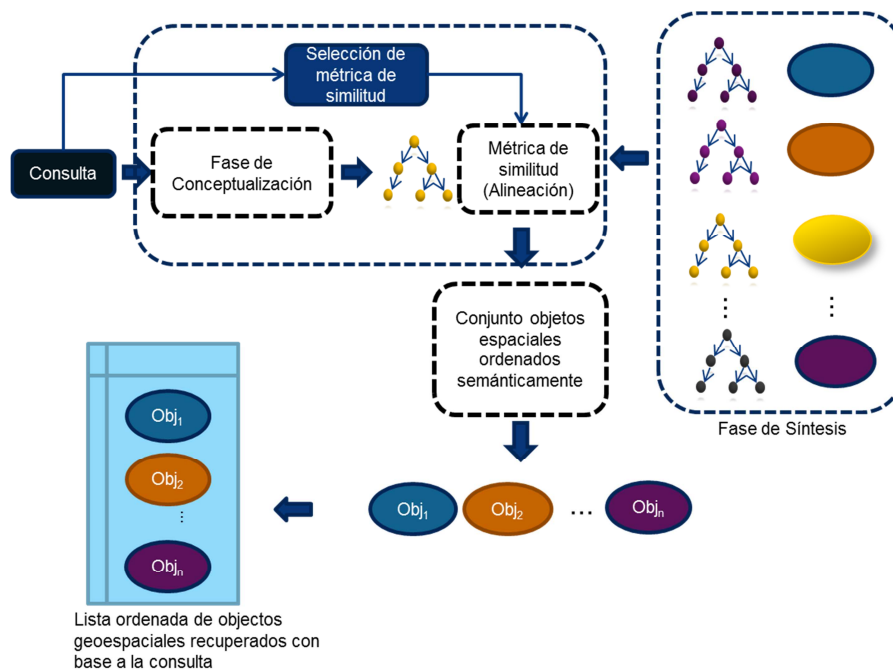


Figura 4.11 Fase de Análisis.

El procesamiento de la fase análisis para la recuperación de objetos espaciales es descrito en el Algoritmo 7.

Algorithm 7 Analisis ($A(S_c, \delta, S_n, Z_s)$)**Input:** S_c : Conjunto de conceptos de la consulta σ : Tipo de similitud (Ψ, λ, Ω) para la comparacion entre ontologias (Γ) S_n : Numero de segmento Z_s : Numero de objetos espaciales a recuperar por segmento**Output:** Arreglo de objetos espaciales A_o $O_c \leftarrow CO(S_c)$ /* creacion de ontologia (CO) con base al conjunto de conceptos S_c */ $T_o \leftarrow size(RS)$ /* Obtener el numero total de ontologias en el repositorio de sintesis RS */ $A_o[T_o] \leftarrow \phi$ /* Arreglo de dimension T_o para ontologias */ $AT_s[Z_s] \leftarrow \phi$ /* Arreglo de dimension Z_s para objetos espaciales */ $L_s[T_o][1] \leftarrow \phi$ /* Arreglo bidimensional de Γ y indice de ontologia en RS */ $O_i[Z_s] \leftarrow \phi$ /* Arreglo de indices de ontologias */ $i \leftarrow 0$ /* Contador */**for** $i = 0$ $i < A_o.size$ $i++$ **do** $L_s[i][0] \leftarrow \Gamma(O_c, A_o[i], \sigma)$ /* Comparacion entre ontologias Γ con respecto a la medida de similitud σ */ $L_s[i][1] \leftarrow i$ /* Guardar numero de indice de la ontologia en RS */**end** $OD(L_s)$ /* Ordenamiento decendente de L_s con base a Γ */ $l_i \leftarrow S_n * Z_s$ /* limite inferior */**for** $i = 0$ $i < Z_s$ $i++$ **do** $O_i[i] = L_s[l_i + i][1]$ /* obtener indices de ontologias */**end** $A_o \leftarrow SPOE(O_i)$ /* Obtener objetos espaciales */**return** A_o **Algoritmo 7 Análisis.**

5. Resultados Experimentales

5.1. Introducción

En este capítulo se presentan los resultados obtenidos en esta tesis. Los resultados están orientados a resolver un caso de estudio que se describe a continuación.

La metodología en general funciona sobre el análisis y comparación entre dos conjuntos de datos, es decir, un conjunto de términos de una consulta y otro conjunto de términos que describen una fuente, esto se muestra en 5.1 donde a partir de una consulta se obtiene la conceptos que la componen y por otro lado se obtiene los conceptos con la información descriptiva de un conjunto de objetos geospaciales, la metodología en general pondera el nivel de similitud entre la consulta con los objetos y devuelve un lista de relevancia de objetos espaciales, además provee un mecanismo de extracción de los objetos.

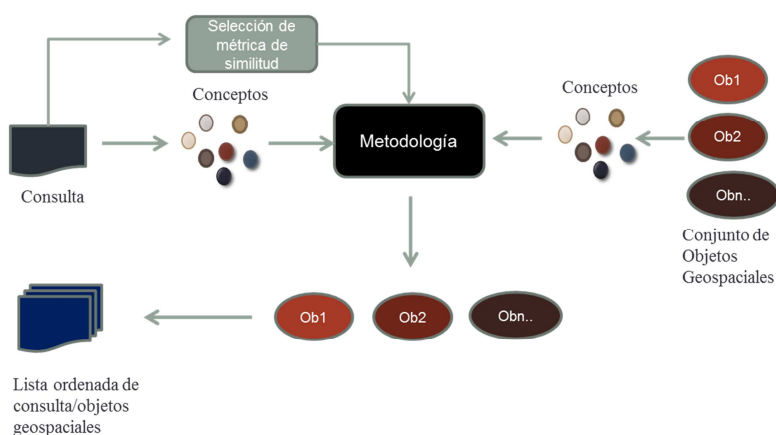


Figura 5.1 . Caso de estudio General.

Esta idea puede aplicarse para cualquier objeto espacial que tenga alguna información descriptiva. En la figura 5.2 se muestra como la metodología puede ser aplicada a objetos geospaciales que cuenten con algún metadato y se tenga acceso a esos recursos en internet, así de este modo se pueden obtener documentos de texto, archivos de audio, bases de datos, fotografías, etc. Todos estos objetos son localizables mediante un proceso de *web crawler*,

este proceso consulta un diccionario o configuración para saber dentro de la descripción del metadato cual información es la que describe al objeto y extraer los conceptos que sirven de entrada a la metodología junto con su localización en la *web*. Por otro lado un usuario que realiza una consulta ingresa los términos como la entrada a la metodología, como resultado se devuelve una lista ponderada semánticamente con relación a la consulta con todos los objetos geoespaciales.

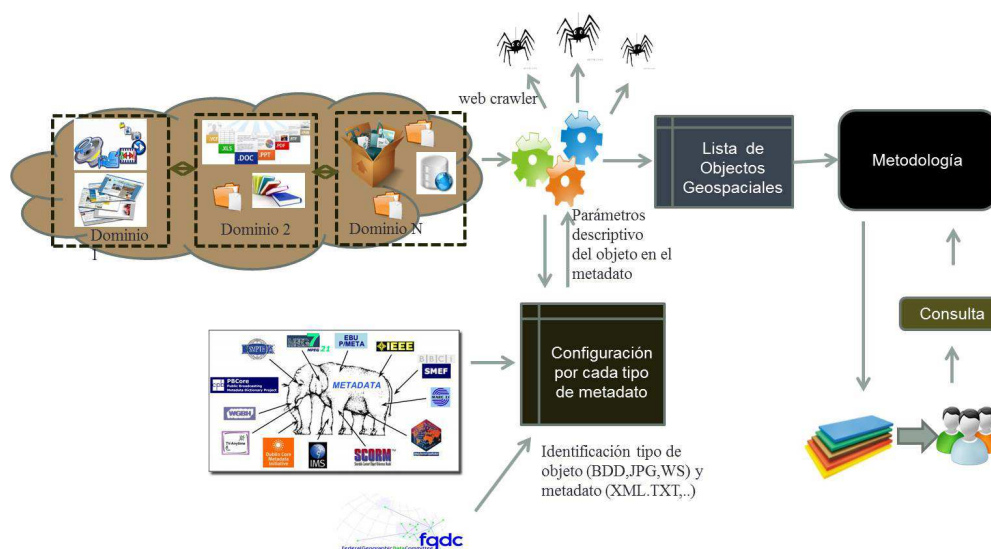


Figura 5.2 Caso de estudio general para metadatos.

Debido a tiempo de implementación se redujo el caso de estudio como se muestran en la figura 5.3 para un solo tipo de objeto espacial, estos son noticias dentro de canales GeoRss [88], estas noticias tienen un tipo específico de estructura de metadato, igual que en el caso de caso general de metadatos el proceso *web crawler* localiza y extrae las canales de noticias dentro de la *web*, extrae las noticias y alimenta a la metodología con objetos geoespaciales. En el Anexo 1 se muestran fragmentos de los canales de noticias utilizados en la implementación de la metodología. Por otro lado el usuario realiza una consulta sobre la metodología y obtiene las noticias relacionadas a su consulta ponderadas semánticamente y con opciones de visualización de estas.

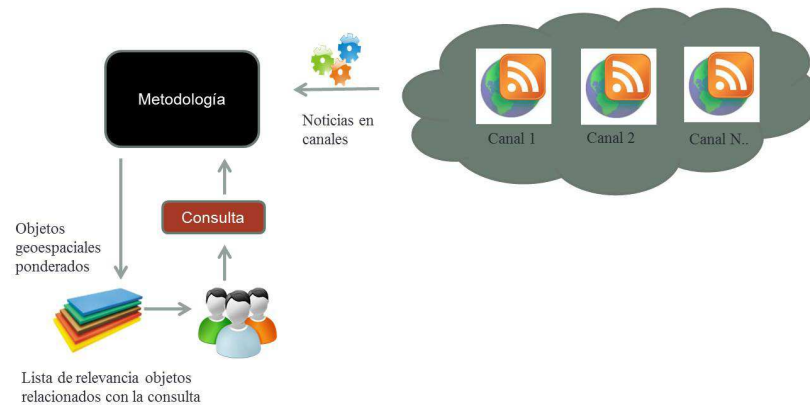


Figura 5.3 Caso de estudio particular delimitado al dominio de noticias.

5.2. Bases de conocimiento multi-dominio

La base de conocimiento multi-dominio está compuesta por una ontología base y varias ontologías externas. Estas son las siguientes:

5.2.1 Ontología base

- WordNet versión 2.0 [66] .WordNet es una gran base de datos léxica de Inglés. Sustantivos, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos (synsets), cada uno expresando un concepto distinto. Los synsets están vinculados entre sí por medio de las relaciones conceptuales semánticas y léxicas.

5.2.2 Ontologías externas

- Ontología ENVO [89]: Ontología de Medio Ambiente (ENVO) proporciona un vocabulario controlado y estructurado que está diseñado con los términos de biomas, las características ambientales y material ambiental.
- Ontología AGROVOC [90]. Se trata de un vocabulario controlado que cubre todas las áreas de interés de la Organización para la Agricultura y la Alimentación (FAO) de las Naciones Unidas, incluidos los alimentos, la nutrición, la agricultura, la pesca, la silvicultura, el medio ambiente, etc.
- Ontología Fauna-Taxón [91]. Esta ontología describe conceptos de fauna utilizando una taxonomía para el suroeste de los Estados Unidos

5.3. Resultados etapa de conceptualización

En esta sección se presentan los resultados obtenidos para la etapa de conceptualización proveyendo un texto o conjunto de términos para su transformación en una ontología. Los resultados se dividen en cinco subsecciones: "pre-procesamiento léxico", "expansión sobre la base de conocimiento multi-dominio", "medida de similitud", "desambiguación y la reducción de la dimensión de la ontología base" y "fusión de la ontología base y ontologías externas". Cada sección se muestra algunos de los procesos que intervienen y los resultados obtenidos.

5.3.1. Procesamiento léxico

El sistema recibió el texto "a mouse for an IBM computer". Los sistemas eliminan inicialmente los stop words : "a, for, an" (en el Anexo 2 se encuentra la lista completa de stop words utilizados). En el siguiente paso se obtienen las raíces de los términos: "mouse", que se reduce a "mouse", "computer" se reduce a "computer", finalmente "IBM" se mantuvo como "IBM". Después del procesamiento léxico el concepto "mouse" y "computer" pudieron ser mapeados en la base de conocimiento KB por medio de la ontología base, pero el concepto de IBM es desconocido en cualquier ontología (base y externas) y por lo tanto es descartado.

5.3.2. Expansión de la ontología

Para los conceptos "mouse" y "computer", los resultados de expansion se presentan en la tabla 5.1. Los conceptos se expanden en la ontología base hasta encontrar una raíz común que asegura un grafo completo y todos los conceptos conectados. La ontología por expansión creada tiene 112 conceptos con la raíz "device".

Concepto	Numero de iteración (nivel superior y inferior)		Total de conceptos	Raiz encontrada
	1	2		
Mouse	14	15		
Computer	27	53	112	device

Tabla 5.1 Resumen expansión ejemplo 1, computer y mouse

Después de la segunda iteración se encontró la raíz común. La figura 5.4 muestra gráficamente la expansión del concepto de "computer" y "mouse", el nodo negro representa el concepto que conecta la expansión de ambos conceptos. Se encuentra tres expansiones

inconexas que representan tres sentidos diferentes del concepto de "mouse" (M1,M2,M3). Sólo un sentido de "mouse" (M2) fue capaz de conectarse a la expansión de "computer" (C1) en la segunda iteración.

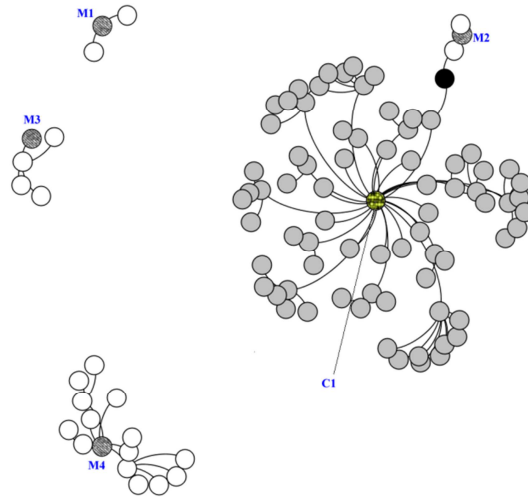


Figura 5.4 Expansión en la ontología.

Con respecto a la metodología, una vez que se crea la ontología base, el siguiente paso es procesar esta con el algoritmo DIS-C con el fin de establecer medidas de similitud semántica entre todos los conceptos (figura 5.5).

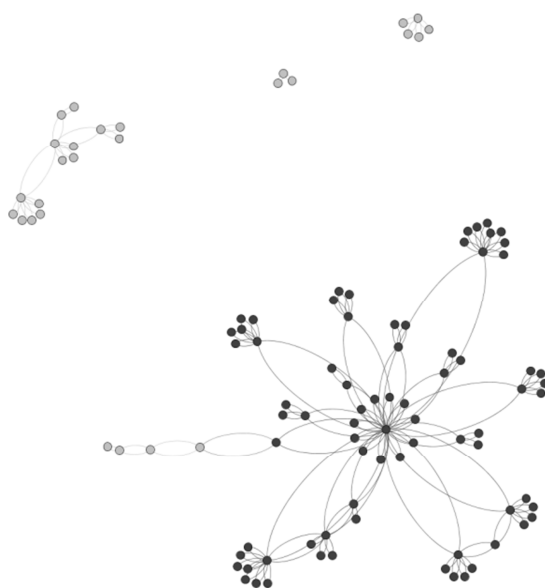


Figura 5.5 Grafo DIS-C para "mouse - computer".

Una vez que se ha creado el grafo DIS-C, es necesario transformarlo para agrupar los sentidos de un concepto a un solo nodo terminal con el fin de procesar este grafo y obtener el árbol Steiner. La figura 5.6 presenta la agrupación de los sentidos para el concepto "computer" (SSC) y "mouse" (SSM), a fin de distinguir el nodo de sentidos agrupados, los nodos creados por agrupación se nombran con el número de identificación negativa de -1, -2 respectivamente.

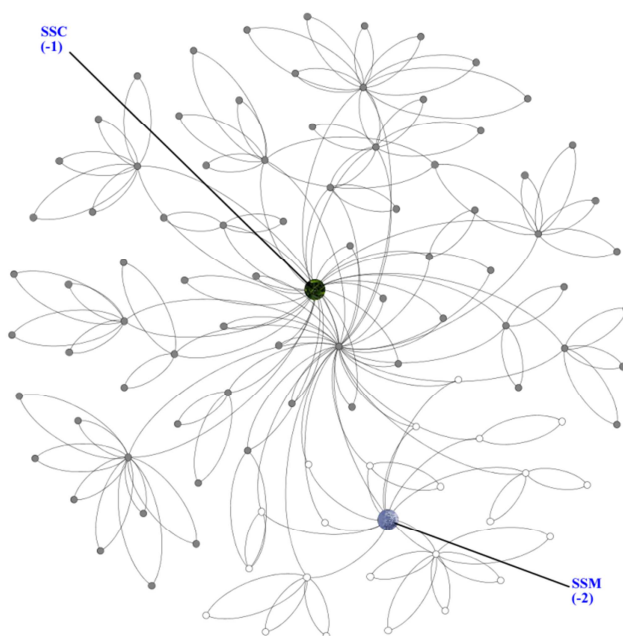


Figura 5.6 Agrupación de terminales del grafo DIS-C .

5.3.3. Reducción de dimensión y desambiguación.

En esta etapa el grafo DIS-C se procesa con el fin de obtener el árbol Steiner el cual produce una ontología desambiguada y reducida. La Figura 5.7 muestra los resultados de esta etapa en tres pasos. El primer grafo muestra el árbol de Steiner resultante del grafo DIS-C. El segundo grafo proporciona el sentido correcto para el terminal/concepto teniendo en cuenta la relación (ρ) de la ontología base. Finalmente todas las relaciones originales se recuperan a partir grafo DIS-C con el fin de construir la ontología final ,en este caso para la ontología base de las relaciones (R_a, \dots, R_d) cambiaron a ' \sim ' (hipónimo).

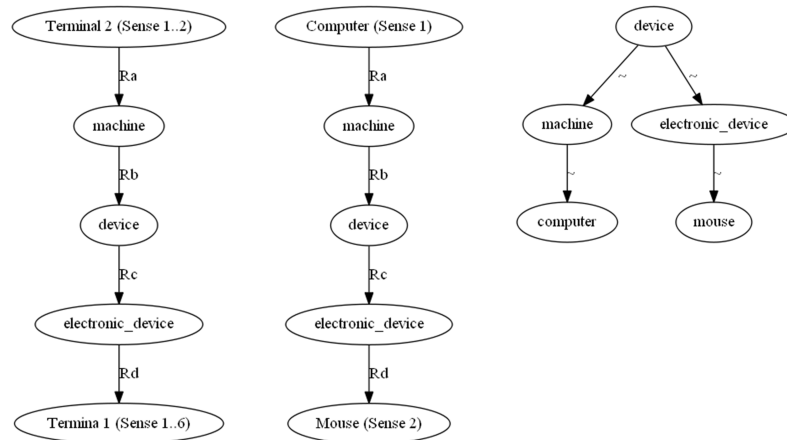


Figura 5.7 Ontología de “mouse - computer”.

Las siguientes dos tablas (tabla 5.2 y 5.3) presentan todos los sentidos para los conceptos “mouse” y “computer” donde se muestra los sentidos seleccionados en la ontología final los cuales eran la mejor selección según el contexto entre ambos conceptos.

any of numerous small rodents typically resembling diminutive rats having pointed snouts and small ears on elongated bodies with slender usually hairless tails	Sense 1
a hand-operated electronic device that controls the coordinates of a cursor on your computer screen as you move it around on a pad; on the bottom of the device is a ball that rolls on the surface of the pad; 'a mouse takes much more room than a trackball'	Sense 2
person who is quiet or timid	Sense 3
a swollen bruise caused by a blow to the eye	Sense 4
manipulate the mouse of a computer	Sense 5
to go stealthily or furtively; '...stead of sneaking around spying on the neighbor's house'	Sense 6

Tabla 5.2 Sentidos de mouse

El concepto “computer” tiene solo dos sentidos:

a machine for performing calculations automatically	Sense 1
an expert at calculation (or at operating calculating machines)	Sense 2

Tabla 5.3 Sentidos de computadora

5.3.4. Fusión de la ontología base y ontologías externas

Para mostrar claramente esta etapa, se procesa la construcción de un ontología para el texto "Tree on Lixisol", el concepto “Tree” se localiza en la ontología base y el concepto “Lixisol” únicamente en una ontología externa.

Antes de continuar, es necesario introducir algunas definiciones a fin de comprender algunos conceptos.

Lixisol: es un tipo de suelo con acumulación subterránea de baja actividad de arcillas y alta saturación de base. Este suelo se desarrolla bajo condiciones intensivas de clima tropicales.

Soil: Se define como la capa superior de la corteza terrestre. Está formado por partículas minerales, materia orgánica, agua, aire y organismos vivos.

El concepto Lixisol solamente se describe en la ontología externa ENVO la cual es una ontología con una estructura jerárquica (figura 5.8).

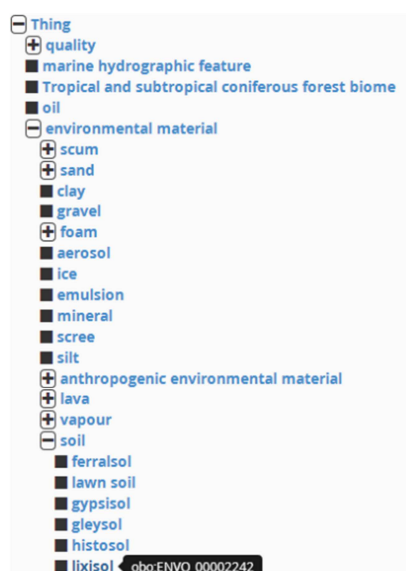


Figura 5.8 El concepto “lixisol” en la ontología ENVO.

El concepto Lixisol se encuentra en la ontología ENVO mediante el procesamiento de alineación mediante *boosting*. La salida del sistema muestra el alineamiento entre Envo y WordNet en la tabla 5.4. Durante la alineación del concepto “soil”, dicho concepto se comprueba si se trata de un sentido simple “single sense”, es decir que sólo un concepto con un único sentido en la ontología externa se refiere a un concepto con múltiples sentidos en la ontología base. Si se cumple este caso se continua con la alineación.

Concept : trees ; Steam : tree
Concept : lixisol Steam : lixisol
unknown concept in Base ontology: lixisol
Concept lixisol found in Envo ontology
Lexically matching between Wordnet and Envo ontology as :[Envo :[soil], WordNet : [dirt, soil]]
found by alignment: [[soil], [dirt, soil], ENVO ontology single Sense, Sense 4]

Tabla 5.4 Procesamiento léxico y alineación de tree y lixisol

El synset “Soil” que representa el sentido correcto con respecto a “tree” es seleccionado en WordNet por alineación (Tabla 5.5).

ID WordNet	Description
Sense 1	the geographical area under the jurisdiction of a sovereign state; 'American troops were stationed on Japanese soil'
Sense 2	material in the top layer of the surface of the earth in which plants can grow (especially with reference to its quality or use); 'the land had never been plowed'; 'good agricultural soil'
Sense 3	the state of being covered with unclean things
Sense 4	the part of the earth's surface consisting of humus and disintegrated rock
Sense 5	make soiled, filthy, or dirty; 'don't soil your clothes when you play outside!'

Tabla 5.5 Sentidos de soil

El concepto obtenido “Soil” por alineación es incluido en la expansión para la ontología base. Durante la expansión algunos conceptos se descartan debido a que durante las próximas iteraciones se repiten y pueden causar un bucle. En total, la ontología base creada incluye 1485 conceptos con raíz "entity" (tabla 5.6).

Concepto	Número de iteración						Total de Conceptos	Raiz encontrada
	1	2	3	4	5	6		
Tree	195	382	341	294	152	63		
Soil	33	19	2	2	3	3	1485	(Entity)

Tabla 5.6 Resumen expansión ejemplo 2, tree y lixisol

Durante el procesamiento de alineación por *boosting* el concepto “Soil” se encuentra ya desambiguado en la ontología base WordNet. De este modo, el concepto que requiere ser desambiguado es únicamente “Tree”. Después de obtener el árbol Steiner de la ontología base se obtienen los resultados como se muestra en la figura 5.9.

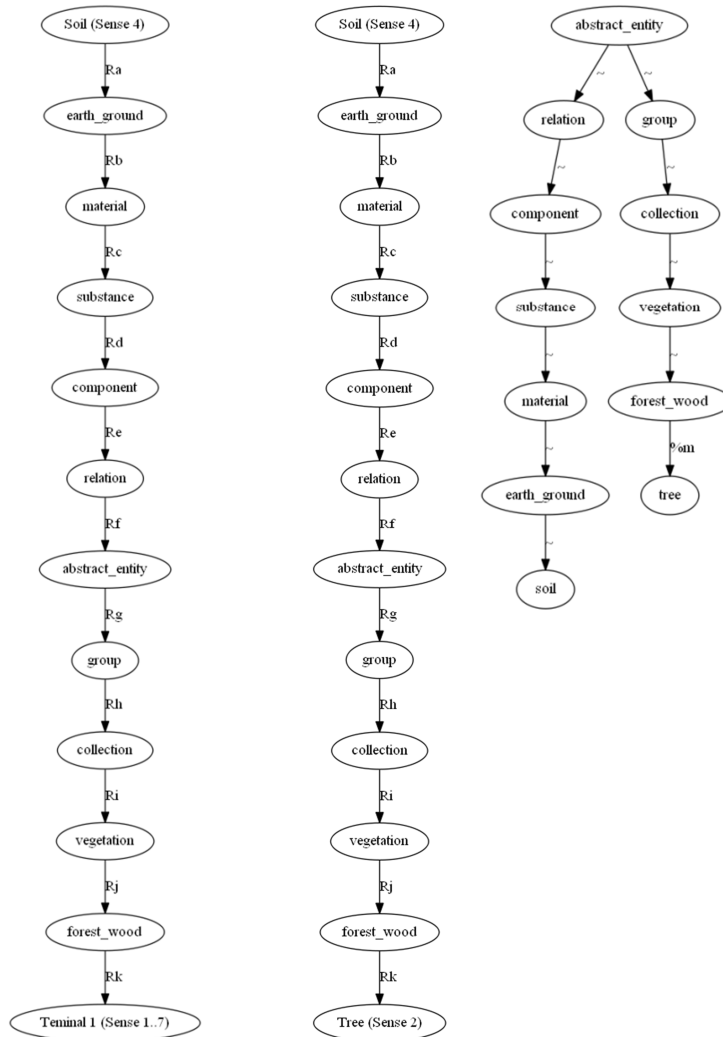


Figura 5.9 Grafo desambiguado-reducido y ontología para “tree-lixisol”.

Considerando la ontología base reducida por Steiner es posible ver el sentido seleccionado para el concepto "Tree". Esta selección se muestra en la tabla 5.7, este sentido es la mejor opción que corresponde al contexto de Lixisol (soil) .

ID Wordnet	Description
Sense 1	English actor and theatrical producer noted for his lavish productions of Shakespeare (1853-1917)
Sense 2	a tall perennial woody plant having a main trunk and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms
Sense 3	a figure that branches from a single root; genealogical tree

Sense 4	stretch (a shoe) on a shoetree
Sense 5	chase an animal up a tree; the hunters treed the bear with dogs and killed it . her dog likes to tree squirrels
Sense 6	plant with trees; this lot should be treed so that the house will be shaded in summer
Sense 7	force a person or an animal into a position from which he cannot escape

Tabla 5.7 Sentidos para tree

Finalmente la ontología multi-dominio es construida fusionando la ontología base y el extracto de la ontología externa (figura 5.10).

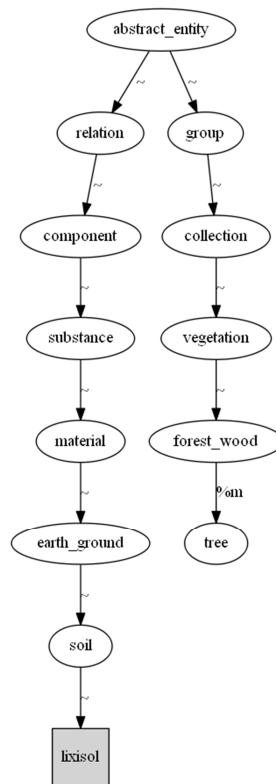


Figura 5.10 Ontología multi-dominio para "tree-lixisol".

5.4. Resultados etapa de síntesis

En la etapa de síntesis las ontologías son obtenidas de la fase de conceptualización y son concentradas en un repositorio que es implementado en una base de datos MySQL [92], aquí se almacenan y se crean los procesos de gestión de extracción del conjunto de

ontologías para la fase de análisis, esta etapa ofrece dos módulos: el primero es de extracción de conjuntos conjunto de ontologías por medio de una petición y en el segundo se recuperan las noticias indicando las ontologías que están ligadas a ellas, este proceso se muestran en la figura 5.11.

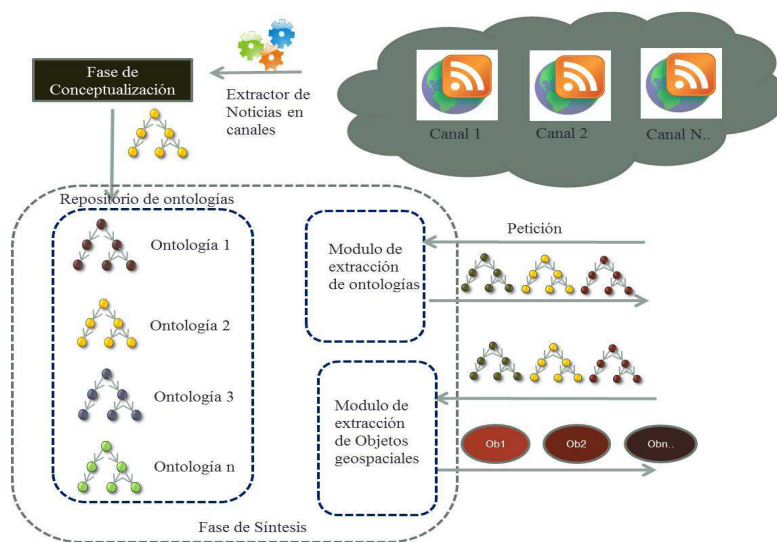


Figura 5.11 Fase de Síntesis.

A continuación en la tabla 5.8 se muestra una noticia como es guardada en la base de datos, los datos que incluye son el título, subtítulo, *link* de referencia a la noticia, descripción, categoría, fecha de publicación y una ubicación geoespacial (latitud ,longitud) en formato de punto de MySQL, este último servirá para realizar recuperación mediante un buffer de búsqueda.

Id Item	Id Channel	Item Title	Item Subtitle	link	Item_Description	category	GUID	Pub Date	Position
1	2	'Broken'		http://rss.cn	As he watched			2013	
		veterans		n.com/~r/rs	wounded troops being			-10-	
		come back		s/cnn_worl	brought inside the			03	
		strong		d/~3/5VLN	hospital on stretchers,			17:3	
				5wWJo24/i	Michael Conklin			9:38	
				ndex.html,"	couldn't believe what				
					he was seeing.",				

Tabla 5.8 Esquema de la base de datos donde se guardan noticias/canales

En la tabla 5.9, un fragmento de ontología ligada a noticia, de esta manera en cómo se almacenan las relaciones que componen su ontología.

idPartition	IdItem	IdRelation
1	1	265142
2	1	262069
3	1	22717
4	1	156028
5	1	265143
6	1	265140
7	1	262068
8	1	77726
9	1	19061

Tabla 5.9 Noticia/Id. Relación

5.5. Resultados etapa de análisis

En esta etapa los objetos geoespaciales son recuperados, es decir, con forme al caso de estudio las noticias en canales GeoRSS las noticias son recuperadas con base en una consulta, el proceso se inicia capturando la consulta del usuario la cual mediante una fase de conceptualización es transformada una ontología, por otro lado se obtienen de la fase de síntesis el conjunto de ontologías que representan cada una de las noticias, la ontología de consulta y las ontologías de noticias sirven de entrada al módulo de comparación por alineación que se encarga de medir el nivel de similitud que existen entre ellas, se genera una lista ponderada de ontologías con base a su similitud con la consulta, esta lista de identificadores de ontologías sirve al módulo de recuperación de objetos geoespaciales de la fase de síntesis para solicitar esas noticias, finalmente se recupera las noticias y son retornadas al usuario con opción de visualización y acceso a ellas, véase figura 5.12.

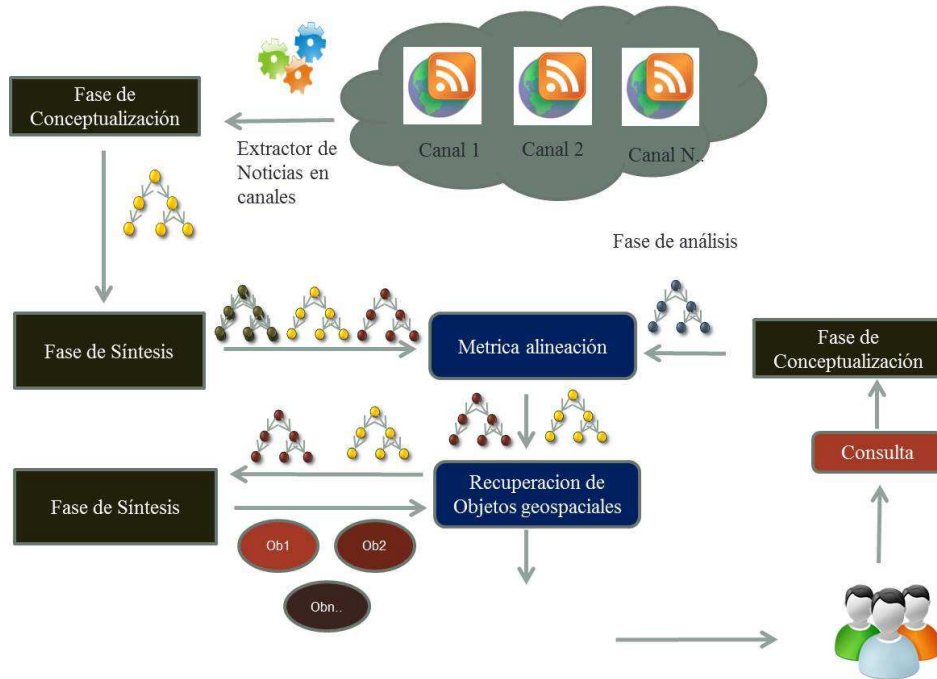


Figura 5.12 Fase de Análisis.

Cuando se obtiene una consulta del usuario, esta se almacena en la base de datos de MySQL llamada *ontologies_db_sources* en la cual se guardan temporalmente las consultas realizadas junto con el tiempo en que fueron realizadas con la finalidad de contar con un tiempo de expiración, las consultas expiradas son eliminadas, los resultados de las consultas también se almacenan temporalmente en donde se indica las noticias en una lista decreciente noticias relacionadas a ellas. De este modo en la figura 5.13 muestra la tabla *queryentry* donde se guarda la instancia de la consulta y *queryresults* el identificador de instancia de la consulta y su lista decreciente de noticias.

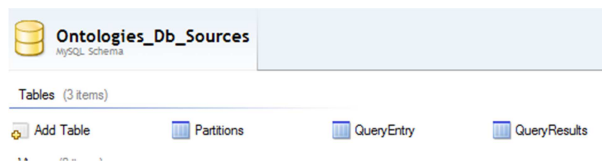


Figura 5.13 Esquema de consultas y noticia-ontología.

A continuación se muestra el sistema de consulta para noticias en Canales GeoRss mediante alineación de ontologías, para el acceso al sistema se muestra la pantalla de inicio en la figura 5.14 donde se inicia como usuario invitado.

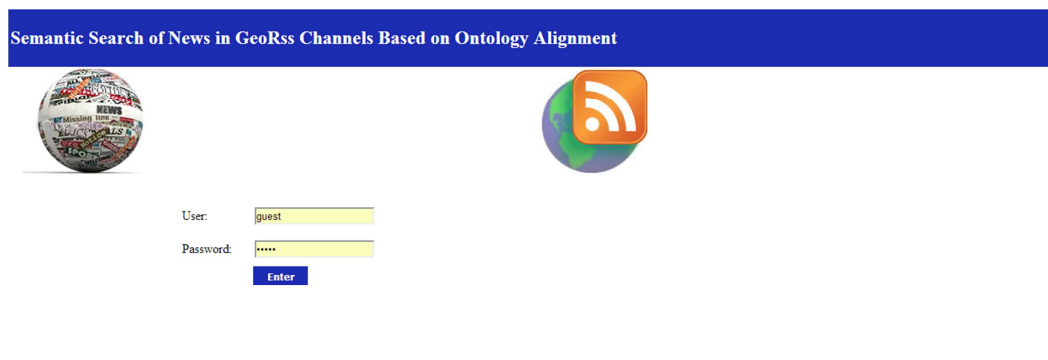


Figura 5.14 Entrada al sistema.

A continuación se muestran algunas consultas de ejemplo para la fase de análisis donde se muestran los resultados obtenidos.

Para la consulta “Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks” (figura 5.15), el procesamiento de la construcción de ontología para esta consulta es presentando en el anexo 3.

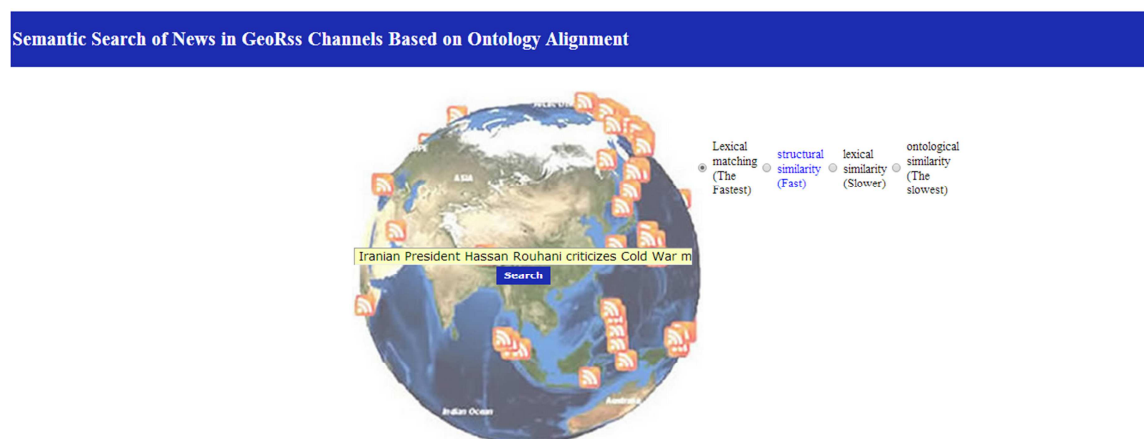


Figura 5.15 . Consulta al sistema.

Los resultados obtenidos para la consulta con base en comparación léxica, y las similitudes basadas en alineación de ontologías: similitud léxica, estructural y ontológica son mostrados en las siguientes figuras 5.16, 5.17, 5.18, 5.19 respectivamente, es resaltado en color rojo la noticia que incluye exactamente todos los conceptos de la consultado con un valor de similitud 1.0, es decir la consulta y la descripción del objeto espacial es el mismo texto.

El procesamiento para la construcción de consulta, la comparación de la ontología de consulta con las ontologías del repositorio de la etapa de síntesis es mostrados en el Anexo 4.

Item_Sim	Item Title	Item Description	Item Date	Channel Title	Item	Channel
1/1	Iran Willing to Begin 'Time-Bound,' Results-Oriented Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News	Item	Channel
2/0.25	Iran on Nuke Talks: 'It's a Question of Months, Not Years'	President Hasan Rouhani says 'everything is possible' following successful talks.	2013-09-26 10:27:00.0	U.S. News - News	Item	Channel
3/0.25	Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Obama is open to meeting with Iranian president to discuss nuclear program.	2013-09-23 06:22:00.0	U.S. News - News	Item	Channel
4/0.25	A new sense of optimism in Iran	From bakery to barber shop, cafe to carpet store, Iranians stroll their capital with a renewed step, uplifted by how their newly elected president seeks something remarkable after decades of cold war-like relations between their country and the West.	2013-09-26 08:32:25.0	CNN.com - World	Item	Channel
5/0.25	Report: Shoe thrown at Rouhani	Iranian President Hassan Rouhani's arrival in Tehran Saturday was marked by an incident during which a shoe was thrown by a protester in his direction, Iran's semi-official Fararu news agency said.	2013-09-30 09:12:03.0	CNN.com - World	Item	Channel
6/0.166667	Al Gore Headlines Opening of New Arm of Brookings Institution	The former vice president re-emerges as global warming talk ramps up.	2013-09-27 05:00:00.0	U.S. News - News	Item	Channel
7/0.0833333	Emmy 2013 Predictions: Drama	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-19 08:28:00.0	U.S. News - News	Item	Channel
8/0.0833333	U.N. Could Agree on Syria Chemicals Resolution by Friday	Security Council movement, 'energetic' Iranian delegate among Thursday highlights in New York.	2013-09-27 10:19:00.0	U.S. News - News	Item	Channel
9/0.0833333	Navy Yard Shooting Revives Violent Game Debate But Doesn't Threaten Grand Theft Auto Fans	Video game critics face obstacles blaming massacres on violent media.	2013-09-18 16:00:00.0	U.S. News - News	Item	Channel
10/0.0833333	Gunmen fire on sleeping students	Under the cover of darkness, gunmen approached a college dormitory in a rural Nigerian town and opened fire on students who were sleeping.	2013-09-30 09:10:55.0	CNN.com - World	Item	Channel

Figura 5.16 Primeros 10 resultados para la consulta por comparación léxica.

Item_Sim	Item_Title	Item_Description	Item_Date	Channel_Title	Item	Channel
1 1	Iran Willing to Begin 'Time-Bound,' Results-Oriented Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News	Item	Channel
2 0.459459	Obama's Former Defense Chiefs Criticize Him on Syria	President's handling of Syrian crisis has emboldened his critics.	2013-09-18 07:26:00.0	U.S. News - News	Item	Channel
3 0.325	Iranian Foreign Minister to Meet With Kerry at U.N. General Assembly	John Kerry among top delegates to meet with the Iranian foreign minister.	2013-09-23 12:59:00.0	U.S. News - News	Item	Channel
4 0.290323	Iran Cyber Chief Found Gunned Down, Reports Say	Reports indicate this could be another assassination against an influential Iranian official.	2013-10-03 11:20:00.0	U.S. News - News	Item	Channel
5 0.277778	Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Obama is open to meeting with Iranian president to discuss nuclear program.	2013-09-23 06:22:00.0	U.S. News - News	Item	Channel
6 0.258065	DCCC Says Syria Is Not An Election Issue	Steve Israel doesn't expect 2014 to be referendum on Syria.	2013-09-10 11:02:00.0	U.S. News - News	Item	Channel
7 0.252874	U.S. Draws Down Embassies in Lebanon and Turkey	A Syria strike may yield widespread Hezbollah retribution, a retired official warns.	2013-09-06 11:41:00.0	U.S. News - News	Item	Channel
8 0.247191	Vicki Kennedy: People Will Never Let Obamacare Be Taken Away	Ted Kennedy's widow talks health care and Edward M. Kennedy Institute	2013-09-23 16:30:00.0	U.S. News - News	Item	Channel
9 0.235294	Ron Paul Isn't Thrilled By 'Rather Annoying' Shutdown Debate	'The government doesn't really shut down,' says retired Texas congressman.	2013-10-01 16:04:00.0	U.S. News - News	Item	Channel
10 0.227273	U.N. Could Agree on Syria Chemicals Resolution by Friday	Security Council movement, 'energetic' Iranian delegate among Thursday highlights in New York.	2013-09-27 10:19:00.0	U.S. News - News	Item	Channel

[Next\(10\)](#)

Figura 5.17 Primeros 10 resultados para la consulta por similitud estructural.

Item_Sim	Item_Title	Item_Description	Item_Date	Channel_Title	Item	Channel
1 1	Iran Willing to Begin 'Time-Bound,' Results-Oriented Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News	Item	Channel
2 0.508876	Obama's Former Defense Chiefs Criticize Him on Syria	President's handling of Syrian crisis has emboldened his critics.	2013-09-18 07:26:00.0	U.S. News - News	Item	Channel
3 0.380952	Iran Cyber Chief Found Gunned Down, Reports Say	Reports indicate this could be another assassination against an influential Iranian official.	2013-10-03 11:20:00.0	U.S. News - News	Item	Channel
4 0.357899	Iranian Foreign Minister to Meet With Kerry at U.N. General Assembly	John Kerry among top delegates to meet with the Iranian foreign minister.	2013-09-23 12:59:00.0	U.S. News - News	Item	Channel
5 0.341463	Athletes Face Challenges in Speaking Out Against Russia's Anti-Gay Laws	The IOC and Russian officials have discouraged athletes from making political statements at the Olympics.	2013-10-04 10:22:00.0	U.S. News - News	Item	Channel
6 0.333333	Syrian Coalition Rejects Russian Proposal Without Prosecuting Assad	Syrian Coalition will not consider any proposal that does not bring the Assad regime to justice	2013-09-10 14:03:00.0	U.S. News - News	Item	Channel
7 0.316279	U.N. Could Agree on Syria Chemicals Resolution by Friday	Security Council movement, 'energetic' Iranian delegate among Thursday highlights in New York.	2013-09-27 10:19:00.0	U.S. News - News	Item	Channel
8 0.311688	Obama Bids Public for Trust in Syria Plan, Be it Diplomatic or Military	The president pledged diligence in decision-making on Syria.	2013-09-10 21:12:00.0	U.S. News - News	Item	Channel
9 0.310881	U.S. Draws Down Embassies in Lebanon and Turkey	A Syria strike may yield widespread Hezbollah retribution, a retired official warns.	2013-09-06 11:41:00.0	U.S. News - News	Item	Channel
10 0.310345	Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Obama is open to meeting with Iranian president to discuss nuclear program.	2013-09-23 06:22:00.0	U.S. News - News	Item	Channel

[Next\(10\)](#)

Figura 5.18 Primeros 10 resultados para la consulta por similitud léxica.

Item/Sim	Item Title	Item Description	Item Date	Channel Title	Item	Channel
1/1	Iran Willing to Begin 'Time-Bound,' 'Results-Oriented' Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News	Item	Channel
2/0.768876	Economic Issues Take Back Seat to Syria at G-20 Summit	Obama meets with Chinese and French on final day in Russia.	2013-09-06 08:07:00.0	U.S. News - News	Item	Channel
3/0.759223	DCCC Says Syria Is Not An Election Issue	Steve Israel doesn't expect 2014 to be referendum on Syria.	2013-09-10 11:02:00.0	U.S. News - News	Item	Channel
4/0.7343	Iran Cyber Chief Found Gunned Down, Reports Say	Reports indicate this could be another assassination against an influential Iranian official.	2013-10-03 11:20:00.0	U.S. News - News	Item	Channel
5/0.732136	Obama's Former Defense Chiefs Criticize Him on Syria	President's handling of Syrian crisis has emboldened his critics.	2013-09-18 07:26:00.0	U.S. News - News	Item	Channel
6/0.705422	Putin Wins a Round From Obama	Russian president outmaneuvers Obama with Syrian diplomacy.	2013-09-12 07:04:00.0	U.S. News - News	Item	Channel
7/0.690888	Athletes Face Challenges in Speaking Out Against Russia's Anti-Gay Laws	The IOC and Russian officials have discouraged athletes from making political statements at the Olympics.	2013-10-04 10:22:00.0	U.S. News - News	Item	Channel
8/0.690864	Committee Passes Resolution to Strike Syria	Committee Approves Resolution to Intervene in Syria.	2013-09-04 14:56:00.0	U.S. News - News	Item	Channel
9/0.67244	Liberal Group Tries to Make 'Flat Boehner' Happen	The group Americans United for Change got crafty.	2013-10-03 13:25:00.0	U.S. News - News	Item	Channel
10/0.671224	Ron Paul Isn't Thrilled By 'Rather Annoying' Shutdown Debate	The government doesn't really shut down,' says retired Texas congressman.	2013-10-01 16:04:00.0	U.S. News - News	Item	Channel

Figura 5.19 Primeros 10 resultados para la consulta por similitud ontológica.

A continuación en las figuras 5.20, 5.21 y 5.22 muestran las mediciones de precisión, *recall* y medida-F para la medida de comparación léxica y las tres medidas de similitud entre ontologías estructural, léxica y ontológica en cortes (*cut-offs*) de 5 noticias recuperadas, la lista completa de noticias recuperadas para estos cálculos se muestra en el anexo 5. Con base en estos resultados se muestra un ligero incremento de precisión por cada medida de similitud comprometiendo tiempo de procesamiento.

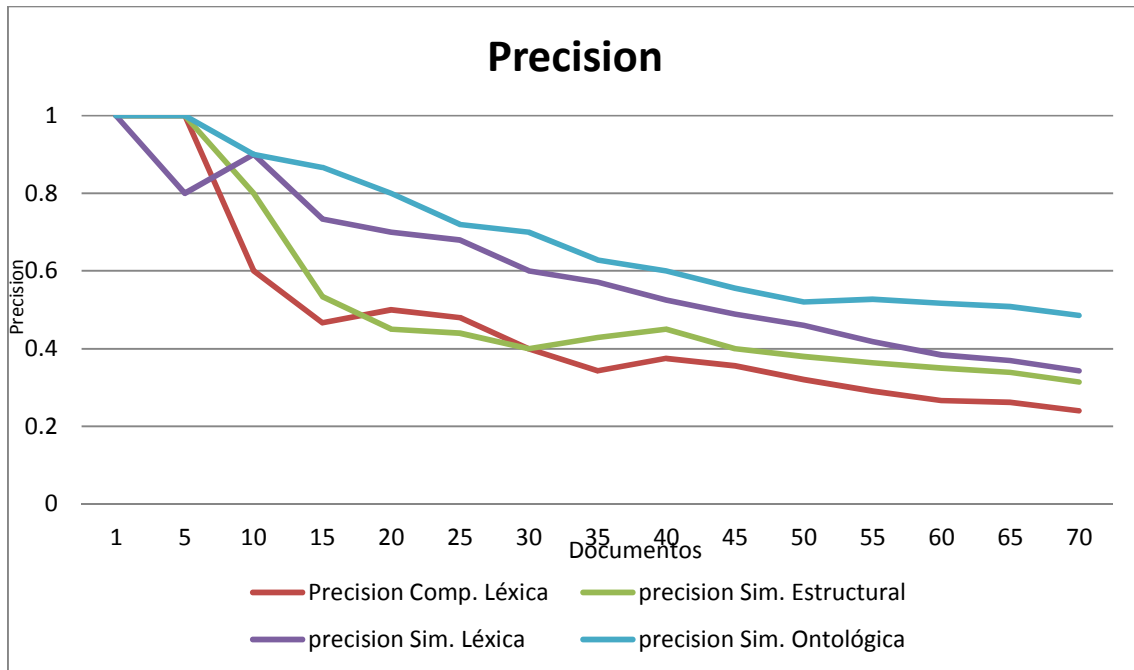


Figura 5.20 Precisión para las 3 medidas de similitud

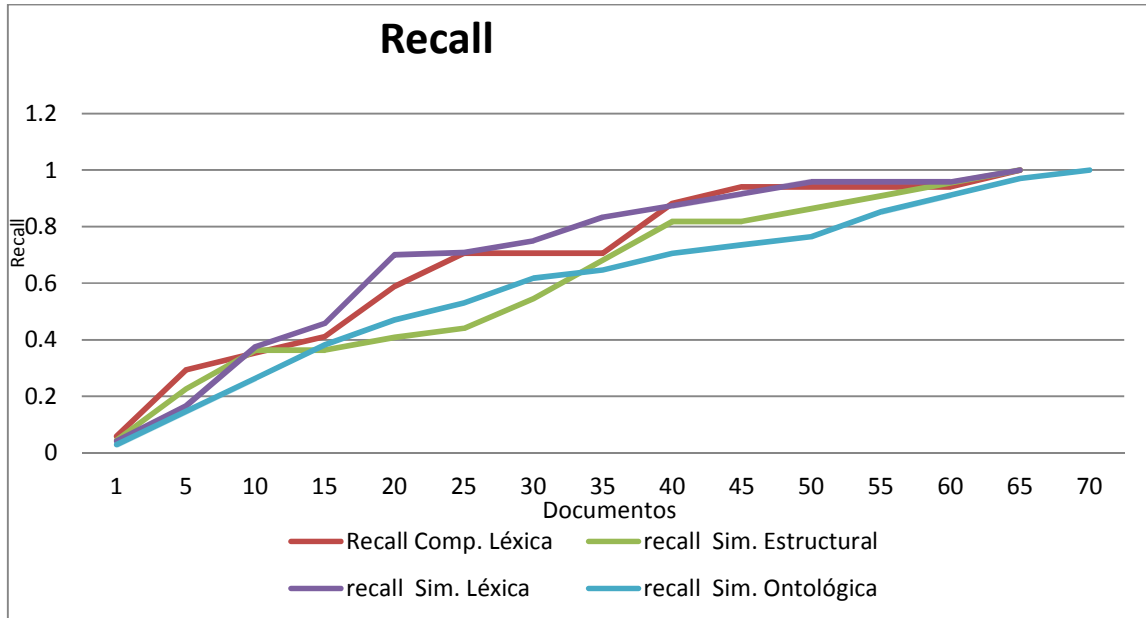


Figura 5.21 Recall para las 3 medidas de similitud

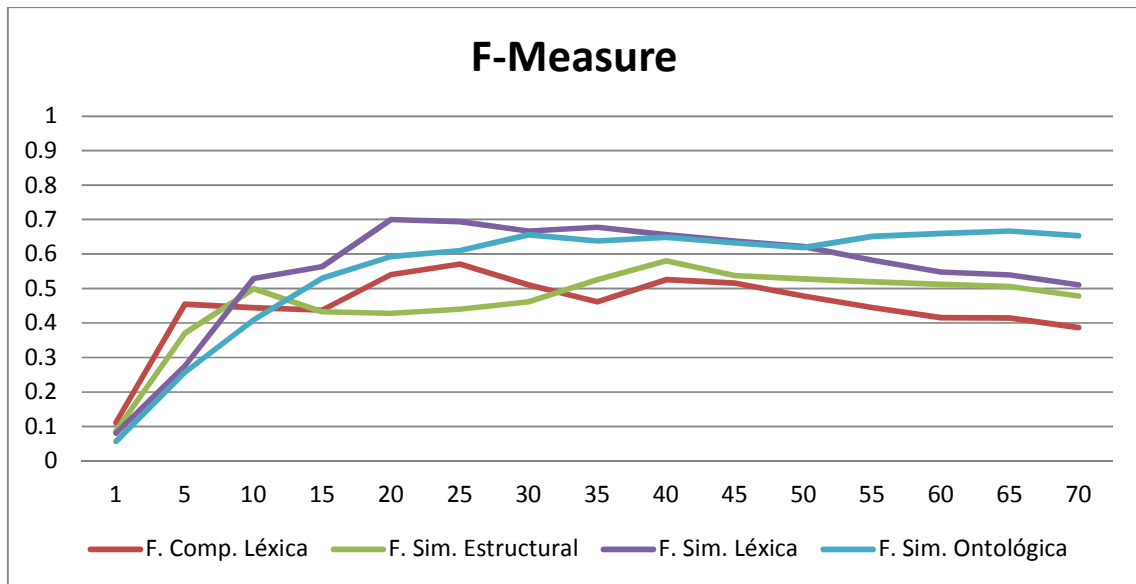


Figura 5.22. F-measure para las 3 medidas de similitud

Adicionalmente se presentan las figuras 5.23, 5.24, 5.25, donde se expone la relación precisión/recall solamente para las medidas de similitud entre ontologías, es decir; para las medidas estructural, léxica y ontológica respectivamente.

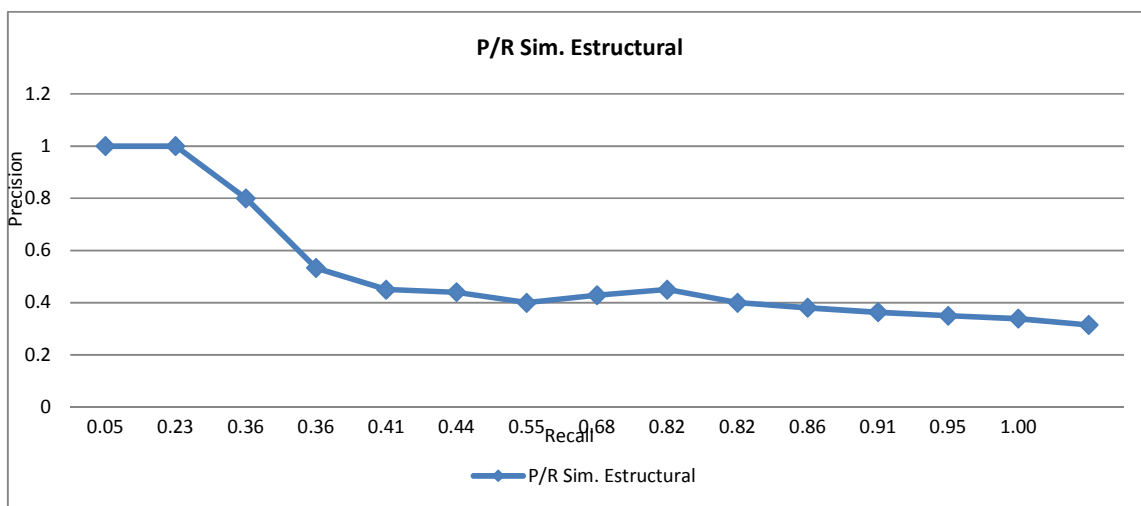


Figura 5.23 Relación Precisión/Recall para la de similitud estructural.

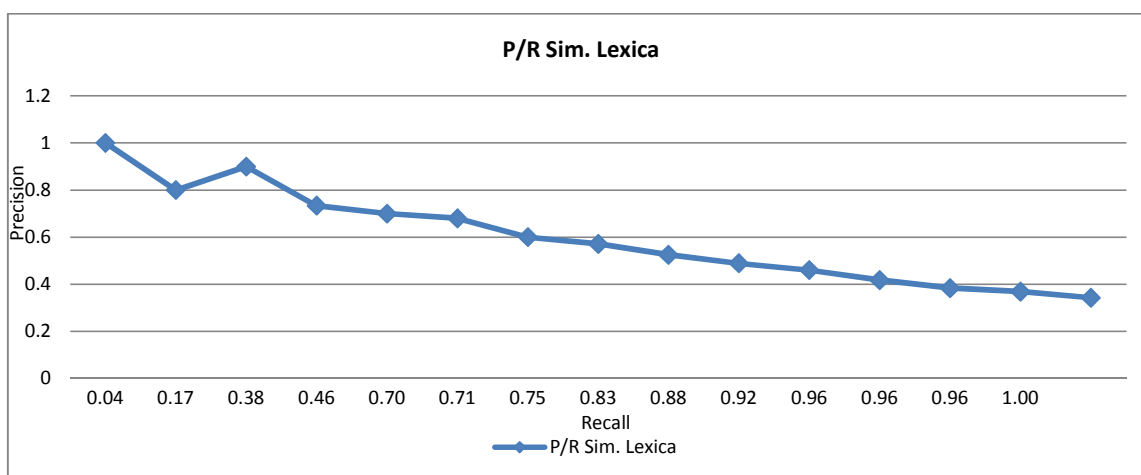


Figura 5.24 Relación Precisión/Recall para la de similitud léxica.

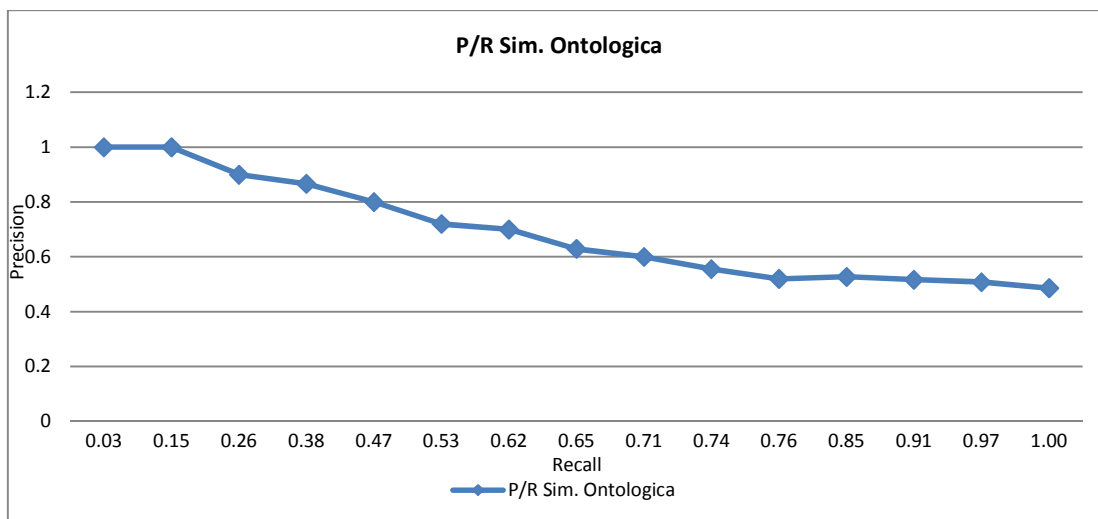


Figura 5.25 Relación Precision/Recall para la de similitud ontológica.

Presentando el sistema de la recuperación de noticias, se provee la posibilidad de localizar en un *web-mapping* las noticias debido a que se encuentran geo-referenciadas dentro del archivo GeoRSS (figura 5.26).

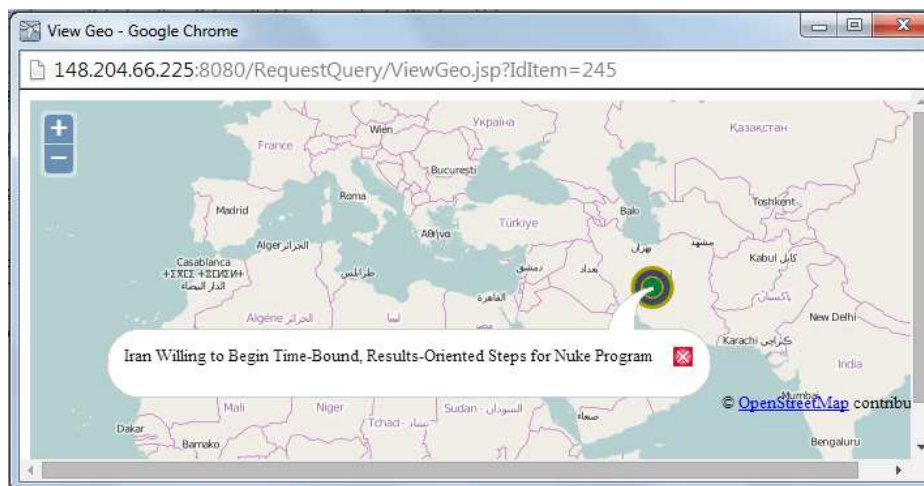


Figura 5.26 Localización geográfica de la noticia

Y además se obtiene noticias relacionadas espacialmente cercanas por medio de un buffer de búsqueda. (Figura 5.27).

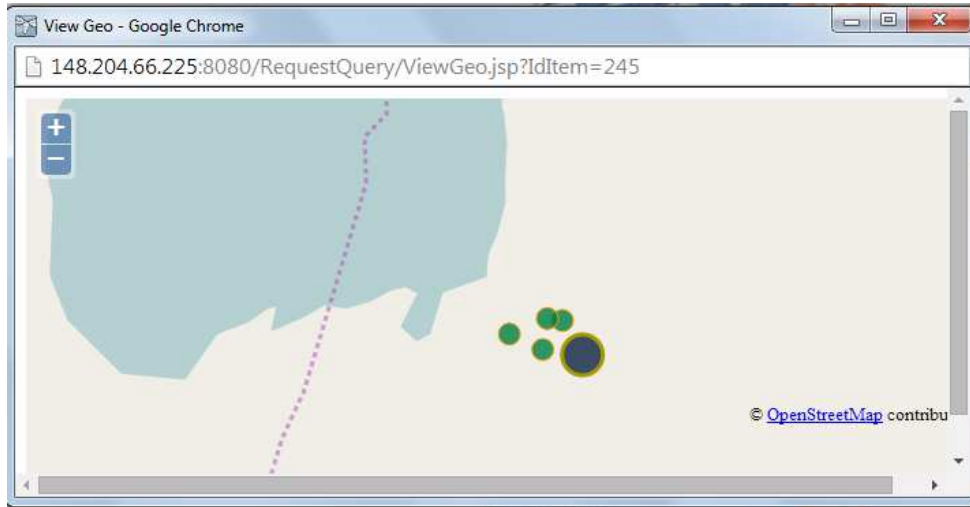


Figura 5.27 Noticias cercanas espacialmente a la noticia

Finalmente en la recuperación se presenta el canal donde fue emitida la noticia, además se muestra su información y el link de acceso (figura 5.28).

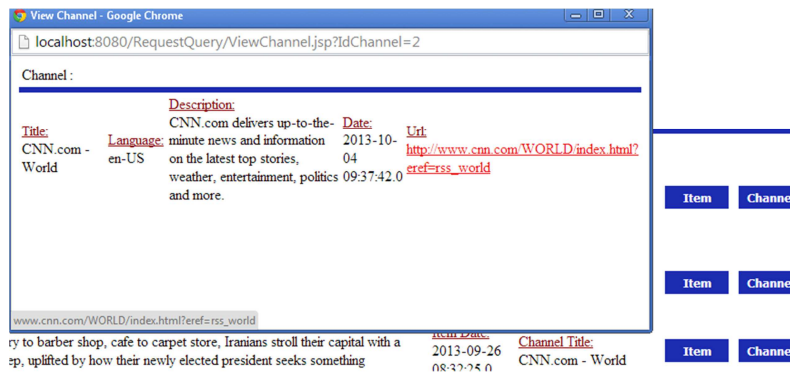


Figura 5.28 Ver canal emitido

5.6. Diagrama del Sistema

En esta sección se describe como fue implementada la metodología (véase figura 5.29), en la fase de conceptualización el *web crawler* encuentra una noticia, si tiene formato GeoRSS la agrega sino usa del servicio *web RssToGeorss* de Geonames.org, este servicio provee a cada noticia una ubicación en latitud, longitud con forma a su descripción, posteriormente este archivo GeoRSS se pasa a la fase de conceptualización para construir su ontología que será enviada a la fase de síntesis.

La fase de síntesis guarda la información del objeto geoespacial, es decir la noticia (Bds Noticias/canales) y su ontología relacionada (BDS ontología/noticia).

Finalmente la fase de análisis el sistema de visualización es construido bajo paginas JSP las cuales consultan un servicio web, este servicio recibe consultas y retorna noticias ponderadas semánticamente, este servicio usa la fase de análisis para la recuperación de noticias, además se lleva el control de las consultas con tiempo de expiración y guarda temporalmente el resultado de las noticias recuperadas, las consultas son gestionadas por un proceso monitor que elimina las consultas expiradas.

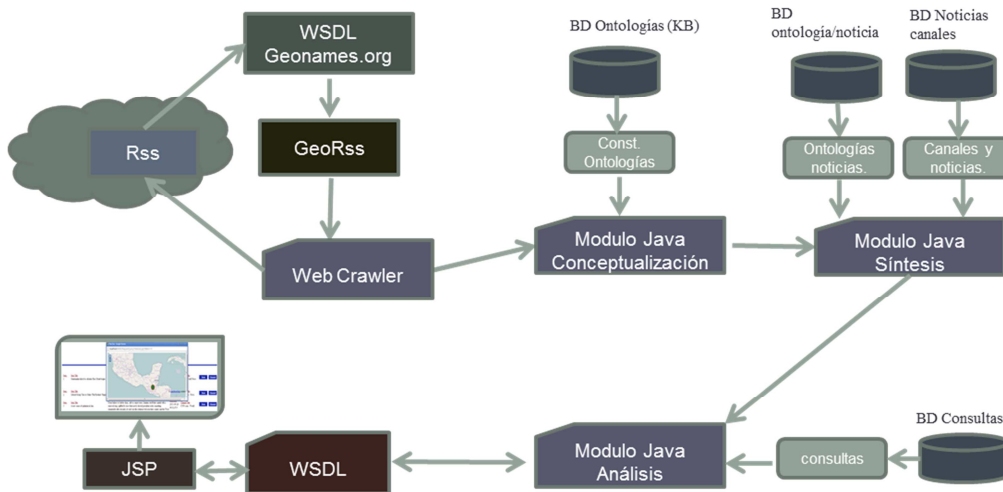


Figura 5.29 Implementación del sistema.

5.7. Consideraciones generales

En los experimentos de la implementación de la metodología:

- Para la búsqueda de objetos espaciales se generó una consulta que cuenta con 14 términos, 529 objetos espaciales (Noticias en GeorSS) en el repositorio de síntesis.
- La construcción de la ontología de consulta se ejecutó en un tiempo total de 26979 ms, de este tiempo 2164 ms fueron para pre-procesamiento léxico, la expansión sobre la

base de conocimiento se ejecutó en 24020 ms, esta última ejecución tomó gran parte del tiempo de ejecución debido básicamente al costo por acceso a la base de datos.

- Durante los experimentos de construcción de ontologías la mayoría de las expansiones fueron alrededor de 1000 a 3000 conceptos siendo la mayor expansión a 10000 de un total de 110 000 que cuenta WordNet. Para una computadora con 4 GB de memoria RAM se establecieron límites de expansión a 5000 conceptos lo cual consumía alrededor de (2 GB) de memoria.
- El texto de una noticia más grande son 67 términos convertido a ontología es el siguiente: “ CNN Films' "Girl Rising" tells the stories of girls across the globe and the power of education to change the world. Are you inspired to help the cause of girls' education around the world? You can make an impact in many ways, but by just being aware of the issue, you can spread the word. By acting in your own community, you can make a global difference”.
- En cuanto a ejecución, la limitante principal de la metodología propuesta es la memoria debido que el algoritmo DIS-C necesita de una tabla de al menos $N \times N$ donde N es el número de términos de entrada para calcular las distancias conceptuales, para la mayor ontología que se puede construir; es decir WordNet (el peor caso) de 100 000 términos se necesita al menos tabla en memoria de números flotante para el identificador de concepto de 4 bytes $\times 10^{10} \approx 40$ GB, por lo que se recurre a uso de memoria de disco y paginación.
- La complejidad computacional para la construcción de ontologías es: $O(\text{DIS-C}) = n \log(n) + O(\text{SkectLs}) = k^2 \log_2 |V| (D_{\max} + k \log |V|) \approx n \log(n)$.
- La complejidad para la recuperación es del orden $O(\text{Boosting}) = [\text{card}(O_y)] [\text{card}(O_y)] \approx N^2$, este es el caso más costoso que es el de similitud ontológica, es importante mencionar que en la mayoría de los casos el tamaño de las ontologías es reducida a su mínima dimensión, por lo tanto, la recuperación es un proceso relativamente rápido.
- Se realizaron optimizaciones tales como ejecución de algoritmo Disktra Fibonacci heap, se eliminó el procesamiento de expansión con conceptos desconocidos en la base de conocimiento, eliminación de ciclos de expansión, etc.

6. Conclusiones

En esta tesis se ha presentado una metodología para la recuperación de objetos espaciales con un enfoque basado en conocimiento. Este conocimiento está relacionado a las descripciones de cada objeto, y es representado por una ontología, la cual se genera de forma automática. La recuperación se realizó mediante métricas de alineación de ontologías: léxica, estructural e híbrida (ontológica). Dicha metodología se compone de las etapas de conceptualización, síntesis y análisis.

Como parte del caso de estudio, éste se limitó a la recuperación de objetos geospaciales (noticias) descritas por metadatos RSS y GeorSS dentro de la *web* a pesar de que la metodología es capaz de recuperar cualquier objeto geoespacial descrito por medio de metadatos.

Para la etapa de conceptualización, la base de conocimiento se compone por la ontología WordNet como la ontología base, debido a que puede definir una gran variedad de dominios usando el lenguaje cotidiano. En el caso de ontologías externas se seleccionaron ontologías especializadas en el dominio del medio ambiente, agricultura y fauna. Durante la construcción de ontologías se ejecutaron procesos de desambiguación mediante el algoritmo de Steiner, con este procesamiento se logró además la reducción de dimensión de la ontología.

En la etapa de síntesis se concentraron los objetos espaciales y sus respectivas ontologías que los representan, además se proporcionaron mecanismos de extracción, acceso y control para una óptima recuperación.

Finalmente en la etapa de análisis los objetos espaciales son recuperados por medio de comparaciones por alineación entre sus descripciones (ontologías) y la ontología que representan a la consulta, los resultados de la recuperación fueron evaluados por medio de las medidas precisión, *recall* y F. Además la recuperación basada en alineación fue comparada con una recuperación totalmente léxica como línea de base.

Considerando la complejidad y el tiempo de ejecución de la recuperación, ésta por lo general suele ser mayor a los métodos clásicos léxicos, ya que considera un enfoque basado

en conocimiento descrito por ontologías para representar la consulta o las fuentes de datos. Entre las ventajas de la metodología se encuentra la definición de un extracto de conocimiento para un conjunto de términos asignando el sentido correcto a cada uno y estableciendo un rango mínimo de conocimiento.

En los resultados obtenidos se mostró una correcta desambiguación y representación mínima en la construcción automática de ontologías multi-dominio. En la recuperación se obtuvieron buenos resultados considerando precisión, *recall* y medida F con la posibilidad de selección de medida de comparación por alineación; además de un método completamente léxico para el ajuste de precisión contra tiempo de ejecución.

6.1. Aportaciones

Haciendo un resumen de los logros obtenidos con el desarrollo de este trabajo, podemos puntualizar los siguientes:

- Se propuso una metodología para la creación automática de ontologías con un enfoque multi-dominio para texto no estructurado. Además, la ontología final generada, desambiguada y con un tamaño o dimensión mínima (conceptos y número de relaciones).
- Se creó un ambiente de interoperabilidad entre ontologías para completar y enriquecer el conocimiento en la etapa de conceptualización de objetos geoespaciales y la consulta, con ello se puede definir y desambiguar el contenido de un conjunto de términos.
- Se establecieron diferentes criterios de comparación entre ontologías considerando como base la alineación, es decir, las métricas de similitud: léxica, estructural e híbrida (ontológica).
- La construcción de las ontologías multi-dominio mostró la reducción del tamaño de la ontología final con los sentidos seleccionados correctamente con base en los términos de entrada. Además la interoperabilidad de ontologías integró extractos de otras ontologías de dominios especializados y sirvieron a su vez al proceso de desambiguación del resto de los conceptos.
- Los resultados mostraron un balance entre *precision/recall* con el tiempo de ejecución, para realizar la recuperación de objetos espaciales, seleccionando diferentes medidas de similitud en la recuperación.

6.2. Limitaciones

La metodología propuesta presenta las siguientes limitaciones, las cuales se generan por el enfoque propuesto en el diseño.

- En la implementación de la metodología solo se consideró el texto escrito en el idioma inglés, debido a que la base de conocimiento es descrita en inglés, es decir; la ontología base (WordNet) y las ontologías especializadas externas.
- El rendimiento en la construcción de ontologías es limitado a la granularidad y número de ontologías que componen la base de conocimiento, entre más conceptos se incluyan en la base de conocimiento, además de mayor número de ontologías se puede describir un mayor número de conceptos y relaciones entre ellos.
- Un gran número de conceptos de entrada para la construcción de ontologías o conceptos lejanos dentro de una ontología incrementa el tiempo de conceptualización; esto implica un incremento de iteraciones en la expansión sobre la base de conocimiento, se elevan el número de desambiguaciones y se aumenta la carga de procesamiento de los algoritmos: distancia conceptual DIS-C y Steiner tree SketchLs. Por otra parte, en la recuperación las ontologías que representan la consulta y los objetos geospaciales podrían ser de mayor tamaño, lo cual afecta el método de alineación.
- Conceptos desconocidos en la base de conocimiento son descartados, debido a que no están definidos previamente en la base de conocimiento y por lo tanto no existe una descripción y relaciones con otros conceptos.
- Enfoque basado en conocimiento en la metodología propuesta implica un mayor procesamiento y complejidad, comparado con otros enfoques de recuperación léxicos.

6.3. Trabajo a futuro

Como consecuencia de los alcances y limitaciones de la metodología, a continuación se presentan algunos trabajos futuros a ser desarrollados:

- Se propone establecer un radio variable de conocimiento a la ontología multi-dominio generada en la etapa de conceptualización, es decir un rango flexible de conceptos y relaciones para lograr un balance entre el conocimiento propuesto y la dimensión de la ontología.

- Incrementar el número de ontologías y aumentar la granularidad de las existentes para describir un número mayor de conceptos y reducir el número de conceptos desconocidos descartados.
- Considerando la implementación cambiar la ontología base de conocimiento. Se propone Multi-WordNet para la extensión multi-lenguaje y eliminar la limitación de procesamiento de la metodología en lenguaje inglés y de este modo también incluir ontologías externas especializadas en múltiples idiomas.
- Considerar conceptos desconocidos en la fase conceptualización, estos conceptos pueden ser útiles en la recuperación por ser muy específicos en las consultas.
- Posibles optimizaciones a la metodología dentro de los procesamientos de distancia conceptual DIS-C y Steiner Tree o si es el caso mejorarlos por otros relacionados. Dentro de la implementación es posible optimizar los tiempos de consulta de bases de datos, debido a que fueron muy notorios en la expansión sobre la base de conocimiento. Por otro lado, la base de conocimiento y el repositorio de la etapa de síntesis pueden ser optimizados a bases de datos *triple store*.
- Realizar optimizaciones de complejidad espacial para reducir la memoria necesaria, en especial para el algoritmo DIS-C, se proponen mejoras como son tablas de ruteo para aliviar el problema de complejidad espacial.

Referencias

1. Pérez González, D.; Dressler, M. & others, Tecnologías de la información para la gestión del conocimiento, OmniaScience, 2007.
2. Olaya, Vi., Sistemas de información geográfica, Cuadernos internacionales de tecnología para el desarrollo humano, Ingeniería Sin Fronteras, 2009, 15.
3. Gómez, Hé. F., Ontología para Sistemas de Información Geográfica, Universidad Nacional de Educación a Distancia. Universidad Técnica Particular de Loja.2004.
4. Cañas, A. J.; Ford, K. M.; Coffey, J.; Reichherzer, T.; Suri, N.; Carff, R.; Shamma, D.; Hill, G.; Hollinger, M. & Mitrovich, T., Herramientas para construir y compartir modelos de conocimiento, 99 Workshop Internacional sobre Educação Virtual, Fortaleza, Brasil, 1999. .
5. Fonseca, F.; Egenhofer, M. & Borges, K. A., Ontologias e interoperabilidade semântica entre SIGs, II Workshop Brasileiro em Geoinformática-GeoInfo2000, Proceedings. São Paulo, 2000.
6. Fonseca, F. T.; Egenhofer, M. J.; Agouris, P. & Câmara, G., Using ontologies for integrated geographic information systems, Transactions in GIS, Wiley Online Library, 2002, 6, 231-257.
7. Shvaiko, P. & Euzenat, Jé., Ontology matching: state of the art and future challenges, Knowledge and Data Engineering, IEEE Transactions on, IEEE, 2013, 25, 158-17.
8. Peis, E.; Herrera-Viedma, E.; Hassan-Montero, Y. & Herrera, J. C. Ontologias, metadatos y agentes: recuperación" semántica" de la información, Jornadas de Tratamiento y Recuperación de la Información, 2003.
9. Choo, C. W. & Diaz, D. R., La organización inteligente: el empleo de la información para dar significado, crear conocimiento y tomar decisiones, Oxford University Press México DF, 1999 .
10. Buccella, A.; Cechich, A. & Fillotrani, P. R., Integración de sistemas de información geográfica, IX Workshop de Investigadores en Ciencias de la Computación, 2007.
11. Vilches Blázquez, L. M.; Ramos Gargantilla, J. Á.; Corcho, O. & Capdevila i Subirana, J., Hacia una armonización semántica de la información geográfica, Treballs de la Societat Catalana de Geografia, 2009, 727-736.
12. Noy, N. F., Semantic integration: a survey of ontology-based approaches ACM Sigmod Record, ACM, 2004, 33, 65-70.
13. Tous, R., Data integration with XML and semantic web technologies, Universitat Pompeu Fabra Barcelona, 2006.
14. Gruninger, M. & Kopena, J. B., Semantic integration through invariants, AI magazine, 2005, 26, 11.
15. Batini, C.; Cappiello, C.; Francalanci, C. & Maurino, A., Methodologies for data quality assessment and improvement, ACM Computing Surveys (CSUR), ACM, 2009, 41, 16 .

16. Botello, A., Explotación de bases de datos heterogéneas mediante su integración parcial, Centro de Investigación en Computación, Instituto Politécnico Nacional, 2004.
17. Naumann, F.; Bilke, A.; Bleiholder, J. & Weis, M., Data Fusion in Three Steps: Resolving Schema, Tuple, and Value Inconsistencies., *IEEE Data Eng. Bull.*, 2006, 29, 21-31.
18. Hakimpour, F. & Geppert, A., Resolving semantic heterogeneity in schema integration, *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, 2001, 297-308.
19. Katayama, T.; Wilkinson, M. D.; Micklem, G.; Kawashima, S.; Yamaguchi, A.; Nakao, M.; Yamamoto, Y.; Okamoto, S.; Oouchida, K.; Chun, H.-W. & others, The 3rd DBCLS BioHackathon: improving life science data integration with Semantic Web technologies, *Journal of biomedical semantics*, Springer, 2013, 4, 1-17.
20. Zhan, Q.; Zhang, X. & Lic, D., Ontology-based semantic description model for discovery and retrieval of geospatial information, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Citeseer, 2008, 32, 141-146 .
21. Rinaldi, A. M., An ontology-driven approach for semantic information retrieval on the web, *ACM Transactions on Internet Technology (TOIT)*, ACM, 2009, 9, 10.
22. Wagh, K. & Kolhe, S., Information retrieval based on semantic similarity using information content, *IJCSI International Journal of Computer Science Issues*, Citeseer, 2011, 8 .
23. Barathi, M. & Valli, S., Context disambiguation based semantic web search for effective information retrieval, *Journal of Computer Science*, 2011, 7, 548.
24. Akmal, S.; Shih, L.-H. & Batres, R., Ontology-based similarity for product information retrieval, *Computers in Industry*, Elsevier, 2014, 65, 91-107.
25. Lipani, A.; Piroi, F.; Andersson, L. & Hanbury, A., An information retrieval ontology for information retrieval nanopublications, *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, Springer, 2014, 44-49.
26. Malik, S. K.; Prakash, N. & Rizvi, S., Semantic annotation framework for intelligent information retrieval using KIM architecture, *International Journal of Web & Semantic Technology (IJWest)*, Citeseer, 2010, 1, 12-26.
27. Kiryakov, A.; Popov, B.; Manov, D.; Ognyanoff, D.; Marinov, R. & Terziev, I. Automatic semantic annotation with KIM, *3rd International Semantic Web Conference (ISWC2004)*, 2004.
28. Davies, J.; Studer, R. & Warren, P., *Semantic Web technologies: trends and research in ontology-based systems*, John Wiley & Sons, 2006.
29. Motiee, S.; Nematzadeh, A. & Shamsfard, M., A Hybrid Ontology Based Approach for Ranking Documents, *Proceedings of World Academy of Science, Engineering and Technology*, 2006, 11.

30. Sridevi, U. & Nagaveni, N., Ontology based correlation analysis in information retrieval, *International Journal of Recent Trends in Engineering*, 2009, 2, 134-137.
31. Dhingra, V. & Bhatia, K. K., SemCrawl: Framework for crawling ontology annotated web documents for intelligent information retrieval, *Intelligent Distributed Computing*, Springer, 2015, 213-223 .
32. Rodriguez-Garcia, M. Á.; Valencia-Garcia, R.; Garcia-Sánchez, F. & Samper-Zapater, J. J., Ontology-based annotation and retrieval of services in the cloud, *Knowledge-Based Systems*, Elsevier, 2014, 56, 15-25.
33. Janowicz, K., Similarity-based retrieval for geospatial semantic web services specified using the web service modeling language (wsml-core), *The GeoSpatial Web*, Springer, 2007, 235-245.
34. Li, Y.; Nayak, R. & Tao, X., A knowledge retrieval model using ontology mining and user profiling, *Integrated Computer-Aided Engineering*, IOS Press, 2008, 15, 313-329.
35. Tanasescu, V.; Gugliotta, A.; Domingue, J.; Villarias, L. G.; Davies, R.; Rowlatt, M.; Richardson, M. & Stinčić, S., Geospatial data integration with semantic Web services: the eMerges approach, *The Geospatial Web*, Springer, 2007, 247-256.
36. Al-Sudairy, M. T. & Vasista, T., Semantic data integration approaches for e-governance, *International Journal of Web & Semantic Technology*, Academy & Industry Research Collaboration Center(AIRCC), 2011, 2.
37. Haav, H.M.; Ojamaa, A.; Grigorenko, P. & Kotkas, V., Ontology-Based Integration of Software Artefacts for DSL Development, *On the Move to Meaningful Internet Systems: OTM 2015 Workshops*, 2015, 309-318.
38. Vaccari, L.; Shvaiko, P. & Marchese, M., A geo-service semantic integration in spatial data infrastructures, *International Journal of Spatial Data Infrastructures Research*, 2009, 4, 24-51.
39. Liping, D.; Peng, Y. & others, Ontology-supported automatic service chaining for geospatial knowledge discovery, *American Society for Photogrammetry & Remote Sensing (ASPRS) 2007 Annual Conference*, Tampa Florida, 2007.
40. Palmonari, M., AERIA: Extending SKOS for the Practical, yet Well-founded, Representation and Integration of Web Schemas in the Large, *Journal of Emerging Technologies in Web Intelligence*, 2011, 3, 227-238.
41. Miles, A.; Matthews, B.; Wilson, M. & Brickley, D., SKOS core: simple knowledge organization for the web, *International Conference on Dublin Core and Metadata Applications*, 2005, pp-3..
42. Palmonari, M. & Batini, C., Abstract ER IA: a web language for conceptual metadata integration and abstraction in the large, *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, 2009, 18.

43. Artale, A.; Calvanese, D.; Kontchakov, R.; Ryzhikov, V. & Zakharyashev, M., Reasoning over extended ER models, *Conceptual Modeling-ER 2007*, Springer, 2007, 277-292.
44. Paul, M. & Ghosh, S., A service-oriented approach for integrating heterogeneous spatial data sources realization of a virtual geo-data repository, *International Journal of Cooperative Information Systems*, World Scientific, 2008, 17, 111-153.
45. Macário, C. G. N. & Medeiros, C. B., Specification of a framework for semantic annotation of geospatial data on the web, *SIGSPATIAL Special*, ACM, 2009, 1, 27-32.
46. Boley, H., Integrating positional and slotted knowledge on the semantic web *Journal of Emerging Technologies in Web Intelligence*, 2010, 2, 343-353.
47. Ge, J.; Li, Z. & Li, T., A Novel Chinese Domain Ontology Construction Method for Petroleum Exploration Information, *Journal of Computers*, 2012, 7, 1445-1452.
48. Kang, J.; Min, L.; Luan, Q.; Li, X. & Liu, J., Novel modified fuzzy c-means algorithm with applications, *Digital Signal Processing*, Elsevier, 2009, 19, 309-319 .
49. Dong, Z., Bigger Context and Better Understanding--Expectation on Future MT Technology, *Proceedings of International Conference on Machine Translation & Computer Language Information Processing*, 1999, 26-28.
50. de Azevedo, R. R.; Freitas, F.; Rocha, R.; Menezes, J. A. A. & Pereira, L. F. A., An Approach for Automatic Expressive Ontology Construction from Natural Language, *Computational Science and Its Applications--ICCSA 2014*, Springer, 2014, 746-759.
51. Sager, N., *Natural language information processing*, Addison-Wesley Publishing Company, Advanced Book Program, 1981. .
52. Touzi, A. G.; Massoud, H. B. & Ayadi, A., Automatic ontology generation for Data mining using FCA and clustering, *arXiv preprint arXiv:1311.1764*, 2013
53. Ganter, B. & Wille, R., *Formal concept analysis: mathematical foundations* Springer Science & Business Media, 2012.
54. Klir, G. & Yuan, B. *Fuzzy sets and fuzzy logic* Prentice Hall New Jersey, 1995, 4.
55. Dixit, P.; Sethi, S.; Sharma, A. & Dixit, A. Design of an automatic ontology construction mechanism using semantic analysis of the documents *Computational Intelligence and Communication Networks (CICN)*, 2012 Fourth International Conference on, 2012, 611-616.
56. Nakaya, A.; Katayama, T.; Itoh, M.; Hiranuka, K.; Kawashima, S.; Moriya, Y.; Okuda, S.; Tanaka, M.; Tokimatsu, T.; Yamanishi, Y. & others KEGG OC: a large-scale automatic construction of taxonomy-based ortholog clusters *Nucleic acids research*, Oxford Univ Press, 2013, 41, D353-D357
57. Küçük, D. & Arslan, Y. Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles *Renewable Energy*, Elsevier, 2014, 62, 484-489.

58. Xiong, J.; Liu, Y.; Wang, J. & Lan, Y. Research of Marine Organism Ontology Semi-Automatic Construction Open Cybernetics & Systemics Journal, 2014, 8, 984-989
59. Biddle, B. J. Role theory: Expectations, identities, and behaviors Academic Press, 2013.
60. Fernández-López, M.; Gómez-Pérez, A. & Juristo, N. Methontology: from ontological art towards ontological engineering, American Association for Artificial Intelligence, 1997 .
61. Uschold, M. & King, M., Towards a methodology for building ontologies, Citeseer, 1995 .
62. Suarez-Figueroa, M. C.; Gomez-Perez, A. & Fernandez-Lopez, M., The NeOn methodology for ontology engineering, Ontology engineering in a networked world, Springer, 2012, 9-34.
63. Gruber, T. R., A translation approach to portable ontology specifications Knowledge acquisition, Elsevier, 1993, 5, 199-220.
64. Guarino, N., Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy, IOS press, 1998, 46 .
65. Krovetz, R., Viewing morphology as an inference process Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, 1993, 191-202.
66. Miller, G. A., WordNet: a lexical database for English, Communications of the ACM, ACM, 1995, 38, 39-41.
67. Wordnet. <http://wordnet.princeton.edu> . consulted / 09/10/2015
68. Collins, A. M. & Quillian, M. R., Retrieval time from semantic memory, Journal of verbal learning and verbal behavior, Elsevier, 1969, 8, 240-247
69. Charikar, M.; Chekuri, C.; Cheung, T.-y.; Dai, Z.; Goel, A.; Guha, S. & Li, M. Approximation algorithms for directed Steiner problems, Journal of Algorithms, Elsevier, 1999, 33, 73-91 .
70. Hwang, F. K.; Richards, D. S. & Winter, P., The Steiner tree problem Elsevier, 1992.
71. Melzak, Z. A., On the problem of Steiner, Canad. Math. Bull, 1961, 4, 143-148.
72. Hanan, M., On Steiner's problem with rectilinear distance, SIAM Journal on Applied Mathematics, SIAM, 1966, 14, 255-265.
73. Garg, N.; Konjevod, G. & Ravi, R., A polylogarithmic approximation algorithm for the group Steiner tree problem, Journal of Algorithms, Elsevier, 2000, 37, 66-84.
74. Kasneci, G.; Ramanath, M.; Sozio, M.; Suchanek, F. M. & Weikum, G., Star: Steiner-tree approximation in relationship graphs, Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on, 2009, 868-879.
75. Kou, L.; Markowsky, G. & Berman, L., A fast algorithm for Steiner trees, Acta informatica, Springer, 1981, 15, 141-145.

76. Mehlhorn, K., A faster approximation algorithm for the Steiner problem in graphs, *Information Processing Letters*, Elsevier, 1988, 27, 125-128
77. Bhalotia, G.; Hulgeri, A.; Nakhe, C.; Chakrabarti, S. & Sudarshan, S., Keyword searching and browsing in databases using BANKS, *Data Engineering, 2002. Proceedings. 18th International Conference on*, 2002, 431-440.
78. Kacholia, V.; Pandit, S.; Chakrabarti, S.; Sudarshan, S.; Desai, R. & Karambelkar, H., Bidirectional expansion for keyword search on graph databases, *Proceedings of the 31st international conference on Very large data bases*, 2005, 505-516.
79. Ding, B.; Xu Yu, J.; Wang, S.; Qin, L.; Zhang, X. & Lin, X., Finding top-k min-cost connected trees in databases, *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, 2007, 836-845.
80. Ihler, E., Bounds on the quality of approximate solutions to the group Steiner problem, *Graph-theoretic concepts in computer science*, 1991, 109-118.
81. Li, W.-S.; Candan, K. S.; Vu, Q. & Agrawal, D., Query relaxation by structure and semantics for retrieval of logical web documents, *Knowledge and Data Engineering, IEEE Transactions on*, IEEE, 2002, 14, 768-791.
82. Dreyfus, S. E. & Wagner, R. A., The Steiner problem in graphs *Networks*, 1972, 1, 195-207.
83. Muhammad, R. B., A parallel local search algorithm for euclidean steiner tree problem, *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 2006. SNPD 2006. Seventh ACIS International Conference on, 2006, 157-164.
84. Faroe, O.; Pisinger, D. & Zachariasen, M., Local search for final placement in vlsi design, *Proceedings of the 2001 IEEE/ACM international conference on Computer-aided design*, 2001, 565-572.
85. Gubichev, A. & Neumann, T., Fast approximation of steiner trees in large graphs, *Proceedings of the 21st ACM international conference on Information and knowledge management*, 2012, 1497-1501.
86. Rodríguez Franco, H., Cálculo de la visibilidad de conceptos en ontologías Instituto Politécnico Nacional. Centro de Investigación en Computación, 2011.
87. Zagal, R., Alineación de ontologías usando el Método Boosting, Tesis de grado de Maestría en Ciencias de la Computación. Instituto Politécnico Nacional, Mexico, 2008.
88. Reed, C.; Singh, R.; Lake, R.; Lieberman, J. & Maron, M., An introduction to GeoRSS: A standards based approach for geo-enabling RSS feeds, *White Paper OGC*, 2006.
89. <http://environmentontology.org/home>. consulted on June 2015.
90. <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>. consulted on June 2015.

91. Fauna Taxon Ontology - Southwest US. Keith Kintigh, Katherine Spielmann. (tDAR ID: 376382) ; doi:10.6067/XCV8D21WWP, 2012.
92. DuBois, P., MySQL Cookbook: Solutions for Database Developers and Administrators, USA: O'Reilly Media, 2014.

Anexos

Anexo 1. Fragmento de Noticias

Dentro de este anexo se presentan fragmentos de canales GeoRSS para los sitios de noticias CNN, Reuters Y US News donde se muestra el formato y estructura de cada uno de ellos. Estos documentos son descritos en XML donde se lista un conjunto de noticias geo-referenciadas con posición latitud-longitud. Además se muestra meta información del canal y noticia como fecha de publicación, copyright, lenguaje, etc.

Fragmento del Canal CNN

```
<?xml version="1.0" encoding="UTF-8"?>
<rss xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:media="http://search.yahoo.com/mrss" xmlns:gml="http://www.opengis.net/gml"
xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/" xmlns:georss="http://www.georss.org/georss"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" version="2.0">
<channel>
<title>CNN.com - World</title>
<link>http://www.cnn.com/WORLD/index.html?eref=rss_world</link>
<description>
CNN.com delivers up-to-the-minute news and information on the latest top stories, weather, entertainment, politics and more.
</description>
<language>en-US</language>
<copyright>Copyright 2013 Cable News Network LP, LLLP.</copyright>
<pubDate>Fri, 04 Oct 2013 14:37:42 GMT</pubDate>
<dc:date>2013-10-04T14:37:42Z</dc:date>
<dc:language>en-US</dc:language>
<dc:rights>Copyright 2013 Cable News Network LP, LLLP.</dc:rights>
<image>
<title>CNN.com - World</title>
<url>
http://i.cdn.turner.com/cnn/e/img/1.0/logo/cnn.logo.rss.gif
</url>
<link>http://www.cnn.com/WORLD/index.html?eref=rss_world</link>
<description>
CNN.com delivers up-to-the-minute news and information on the latest top stories, weather, entertainment, politics and more.
</description>
</image>
<item>
<title>Malala's voice stronger, not silenced</title>
<link>
http://rss.cnn.com/~r/rss/cnn_world/~3/zoHCj0KLjRk/index.html
</link>
<description>
The attack was meant to silence the outspoken teenager who dared to defy the Taliban's ban against girls in school. Instead, it only
made Malala's voice more powerful. After a school year that started with a shooting, Malala now eyes a summer of speaking at the
```



```

U.N., telling her story in a new book and amplifying the issue of girls education.<div class="feedflare"> <a
href="http://rss.cnn.com/~ff/rss/cnn_world?a=zoHCj0KLjRk:dhTz76NGoi8:yII2AUoC8zA"></img></a>
href="http://rss.cnn.com/~ff/rss/cnn_world?a=zoHCj0KLjRk:dhTz76NGoi8:7Q72WNTAKBA"></img></a>
href="http://rss.cnn.com/~ff/rss/cnn_world?a=zoHCj0KLjRk:dhTz76NGoi8:V_sGLiPBpWU"></img></a>
href="http://rss.cnn.com/~ff/rss/cnn_world?a=zoHCj0KLjRk:dhTz76NGoi8:qj6IDK7rITs"></img></a>
href="http://rss.cnn.com/~ff/rss/cnn_world?a=zoHCj0KLjRk:dhTz76NGoi8:gIN9vFwOqvQ"></img></a>
</div>
</description>
<pubDate>Wed, 08 May 2013 21:03:44 GMT</pubDate>
<guid isPermaLink="false">
http://www.cnn.com/2013/04/30/world/malala-girls-education/index.html
</guid>
<dc:date>2013-05-08T21:03:44Z</dc:date>
<geo:lat>4.253611</geo:lat>
<geo:long>33.608333</geo:long>
</item>

```

Fragmento del Canal Reuters World News

```

<?xml version="1.0" encoding="UTF-8"?>
<rss xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:media="http://search.yahoo.com/mrss" xmlns:gml="http://www.opengis.net/gml"
xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/" xmlns:georss="http://www.georss.org/georss"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" version="2.0">
<channel>
<title>Reuters: World News</title>
<link>http://www.reuters.com</link>
<description>Reuters.com is your source for breaking news, business, financial and investing news, including personal finance and
stocks. Reuters is the leading global provider of news, financial information and technology solutions to the world's media, financial
institutions, businesses and individuals.</description>
<language>en-us</language>
<copyright>All rights reserved. Users may download and print extracts of content from this website for their own personal and non-
commercial use only. Republication or redistribution of Reuters content, including by framing or similar means, is expressly
prohibited without the prior written consent of Reuters. Reuters and the Reuters sphere logo are registered trademarks or trademarks
of the Reuters group of companies around the world. © Reuters 2013</copyright>
<pubDate>Tue, 19 Mar 2013 16:57:14 GMT</pubDate>
<dc:date>2013-03-19T16:57:14Z</dc:date>
<dc:language>en-us</dc:language>
<dc:rights>All rights reserved. Users may download and print extracts of content from this website for their own personal and non-
commercial use only. Republication or redistribution of Reuters content, including by framing or similar means, is expressly
prohibited without the prior written consent of Reuters. Reuters and the Reuters sphere logo are registered trademarks or trademarks
of the Reuters group of companies around the world. © Reuters 2013</dc:rights>
<image>
<title>Reuters News</title>
<url>http://www.reuters.com/resources_v2/images/reuters125.png</url>
<link>http://www.reuters.com</link>

```

```

</image>
<item>
  <title>Alleged chemical attack kills 25 in northern Syria</title>
  <link>http://feeds.reuters.com/~r/Reuters/worldNews/~3/K9rGko-PWbA/us-syria-crisis-chemical-
idUSBRE92I0A220130319</link>
  <description>BEIRUT (Reuters) - Syria's government and rebels accused each other of launching a deadly chemical attack near
the northern city of Aleppo on Tuesday in what would, if confirmed, be the first use of such weapons in the two-year-old
conflict.&lt;div class="feedflare"&gt;
&lt;a href="http://feeds.reuters.com/~ff/Reuters/worldNews?a=K9rGko-PWbA:LDox5ak6Yvw:yII2AUoC8zA"&gt;&lt;img
src="http://feeds.feedburner.com/~ff/Reuters/worldNews?d=yII2AUoC8zA" border="0"&gt;&lt;/img&gt;&lt;a&gt; &lt;a
href="http://feeds.reuters.com/~ff/Reuters/worldNews?a=K9rGko-PWbA:LDox5ak6Yvw:F7zBnMyn0Lo"&gt;&lt;img
src="http://feeds.feedburner.com/~ff/Reuters/worldNews?i=K9rGko-PWbA:LDox5ak6Yvw:F7zBnMyn0Lo"
border="0"&gt;&lt;/img&gt;&lt;a&gt; &lt;a href="http://feeds.reuters.com/~ff/Reuters/worldNews?a=K9rGko-
PWbA:LDox5ak6Yvw:V_sGLiPBpWU"&gt;&lt;img src="http://feeds.feedburner.com/~ff/Reuters/worldNews?i=K9rGko-
PWbA:LDox5ak6Yvw:V_sGLiPBpWU" border="0"&gt;&lt;/img&gt;&lt;a&gt;
&lt;/div&gt;&lt;img src="http://feeds.feedburner.com/~r/Reuters/worldNews/~4/K9rGko-PWbA" height="1"
width="1"&gt;&lt;/description>
  <category>worldNews</category>
  <pubDate>Tue, 19 Mar 2013 16:12:23 GMT</pubDate>
  <guid isPermaLink="false">http://www.reuters.com/article/2013/03/19/us-syria-crisis-chemical-
idUSBRE92I0A220130319?feedType=RSS&amp;feedName=worldNews</guid>
  <dc:date>2013-03-19T16:12:23Z</dc:date>
  <geo:lat>33.88894250077</dc:lat>
  <geo:long>35.4944229125977</dc:long>
</item>

```

Fragmento del Canal US News

```

<?xml version="1.0" encoding="UTF-8"?>
<rss xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:media="http://search.yahoo.com/mrss" xmlns:gml="http://www.opengis.net/gml"
xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/" xmlns:georss="http://www.georss.org/georss"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" version="2.0">
<channel>
<title>U.S. News - News</title>
<link>http://www.usnews.com/sections/news/index.html</link>
<description/>
<language>en-en</language>
<copyright>2013 USNews.com</copyright>
<pubDate>Fri, 04 Oct 2013 17:33:00 GMT</pubDate>
<dc:date>2013-10-04T17:33:00Z</dc:date>
<dc:language>en-en</dc:language>
<dc:rights>2013 USNews.com</dc:rights>
<item>
<title>A Brewing Storm</title>
<link>
http://www.usnews.com/blogs/at-the-edge/2013/10/4/a-brewing-storm?s_cid=rss:at-the-edge:a-brewing-storm
</link>
<description>
If a storm is coming, how will we know, since watchers are on furlough?
</description>
<pubDate>Fri, 04 Oct 2013 17:33:00 GMT</pubDate>

```

```

<guid isPermaLink="false">
http://www.usnews.com/blogs/at-the-edge/2013/10/4/a-brewing-storm
</guid>
<dc:date>2013-10-04T17:33:00Z</dc:date>
</item>
<item>
<title>
Vo Nguyen Giap, Military Genius Who Drove U.S. Out of Vietnam, Dies at 102
</title>
<link>
http://www.usnews.com/news/newsgram/articles/2013/10/4/vo-nguyen-giap-military-genius-who-drove-us-out-of-vietnam-dies-at-102?s_cid=rss:vo-nguyen-giap-military-genius-who-drove-us-out-of-vietnam-dies-at-102
</link>
<description>
The legendary general is credited with writing the blueprint of modern-day terrorism.
</description>
<pubDate>Fri, 04 Oct 2013 17:08:00 GMT</pubDate>
<guid isPermaLink="false">
http://www.usnews.com/news/newsgram/articles/2013/10/4/vo-nguyen-giap-military-genius-who-drove-us-out-of-vietnam-dies-at-102
</guid>
<dc:date>2013-10-04T17:08:00Z</dc:date>
<geo:lat>16.1666667</geo:lat>
<geo:long>107.8333333</geo:long>
</item>
<item>
<title>Boehner: 'This Isn't Some Damn Game'</title>
<link>
http://www.usnews.com/news/articles/2013/10/4/boehner-this-isnt-some-damn-game?s_cid=rss:boehner-this-isnt-some-damn-game
</link>
<description>
Republicans and Democrats not giving an inch on the shutdown or the debt ceiling.
</description>
<pubDate>Fri, 04 Oct 2013 16:56:00 GMT</pubDate>
<guid isPermaLink="false">
http://www.usnews.com/news/articles/2013/10/4/boehner-this-isnt-some-damn-game
</guid>
<dc:date>2013-10-04T16:56:00Z</dc:date>
<geo:lat>64.0</geo:lat>
<geo:long>26.0</geo:long>
</item>

```

Anexo 2. Listado de palabras usadas como Stop Words en inglés

En la siguiente tabla se presenta el conjunto de términos que fueron considerados como stop words (palabras ignoradas) en el idioma inglés para el procesamiento léxico en la etapa de conceptualización, este conjunto de stop words es definido en las búsquedas web para el motor Google.

a	about	above	after	again	against	all
am	an	and	any	are	aren't	as
at	be	because	been	before	being	below
between	both	but	by	can't	cannot	could
couldn't	did	didn't	do	does	doesn't	doing
don't	down	during	each	few	for	from
further	had	hadn't	has	hasn't	have	haven't
having	he	he'd	he'll	he's	her	here
here's	hers	herself	him	himself	his	how
how's	i	i'd	i'll	i'm	i've	if
in	into	is	isn't	it	it's	its
itself	let's	me	more	most	mustn't	my
myself	no	nor	not	of	off	on
once	only	or	other	ought	our	ours
ourselves	out	over	own	same	shan't	she
she'd	she'll	she's	should	shouldn't	so	some
such	than	that	that's	the	their	theirs
them	themselves	then	there	there's	these	they
they'd	they'll	they're	they've	this	those	through
to	too	under	until	up	very	was
wasn't	we	we'd	we'll	we're	we've	were
weren't	what	what's	when	when's	where	where's
which	while	who	who's	whom	why	why's
with	won't	would	wouldn't	you	you'd	you'll
you're	you've	your	yours	yourself	yourselves	-
%	&	/	!	=	?	*
+	()	'	[]	{

}	-	.	,	;	<	>
---	---	---	---	---	---	---

Anexo 3. Construcción de la ontología de consulta en el sistema de recuperación

En este anexo se muestra la traza completa para la construcción de la ontología de la **consulta** “Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks”.

La traza describe los siguiente procesos:

- Se muestra el pre-procesamiento léxico donde se obtiene el lexema (raíz) de cada término, posteriormente se buscan los lexemas en la base de conocimiento donde algunos conceptos no se encuentran en ninguna ontología y son descartados.
- Una vez los términos de entrada fueron pre-procesados lexicalmente se procede a expandir cada término iterativamente (nivel superior e inferior) dentro de la ontología base (WordNet), el proceso se detiene hasta encontrar un concepto raíz común para todas las sub-expansiones, como resultado se obtiene una ontología que es a su vez un grafo conexo. Durante le expansión se suele expandir conceptos que en una iteración anterior fueron encontrados, por lo cual formarían un ciclo, estos conceptos se descartan de la iteración actual.
- Posteriormente se procesa la ontología creada por expansión como un grafo con el algoritmo DIS-C donde solo toma 4 iteraciones para establecer las distancias conceptuales entre todos los conceptos.
- Finalmente se procesa el grafo DIS-C con Steiner SketchLs para crear su árbol Steiner y de este modo la ontología final.

Info: iranian steam : iran
Info: president steam : president
Info: hassan steam : hassan
Info: rouhani steam : rouhani
Info: criticizes steam : critic
Info: cold steam : cold
Info: war steam : war

Info: mentality steam : mental
Info: expresses steam : express
Info: openness steam : open
Info: nuke steam : nuke
Info: talks steam : talk
Info: unknown concept or not a noun:hassan
Info: unknown concept or not a noun:mental
Info: unknown concept or not a noun:rouhani
Info: Iteration Expansion :0
Info: Concept nuke Move Base concepts:3
Info: Concept critic Move Base concepts:19
Info: Concept iran Move Base concepts:23
Info: Concept war Move Base concepts:22
Info: Concept cold Move Base concepts:14
Info: Concept express Move Base concepts:7
Info: Concept talk Move Base concepts:18
Info: Concept open Move Base concepts:8
Info: Concept president Move Base concepts:15
Info: Iteration Expansion :1
Info: Thread[nuke,5,main] concepts : 3
Info: Thread[critic,5,main] concepts : 19
Info: Thread[iran,5,main] concepts : 23
Info: Thread[war,5,main] concepts : 22
Info: Thread[cold,5,main] concepts : 14
Info: Thread[express,5,main] concepts : 7
Info: Thread[talk,5,main] concepts : 18
Info: Thread[open,5,main] concepts : 8
Info: Thread[president,5,main] concepts : 15

Info: running threads level: 1
Info: Concept nuke Move Up :3
Info: running threads level: 1
Info: Concept critic Move Up :4
Info: Concept critic Move Down :4
Info: running threads level: 1
Info: Concept iran Move Up :3
Info: Concept iran Move Down :3
Info: running threads level: 1
Info: Concept war Move Up :4
Info: *****CICLE DOWN*****
Info: Concept war Move Down :10
Info: running threads level: 1
Info: Concept cold Move Up :6
Info: running threads level: 1
Info: Concept express Move Up :5
Info: running threads level: 1
Info: Concept talk Move Up :6
Info: Concept talk Move Down :1
Info: running threads level: 1
Info: Concept open Move Up :4
Info: running threads level: 1
Info: Concept president Move Up :7
Info: Iteration Expansion :2
Info: running threads level: 2
Info: Concept nuke Move Up :4
Info: running threads level: 2
Info: Concept critic Move Up :4

Info: Concept critic Move Down :2
Info: running threads level: 2
Info: Concept iran Move Up :4
Info: Concept iran Move Down :1
Info: running threads level: 2
Info: Concept war Move Up :4
Info: Concept war Move Down :2
Info: running threads level: 2
Info: Concept cold Move Up :5
Info: running threads level: 2
Info: *****CICLE UP*****
Info: Concept express Move Up :6
Info: running threads level: 2
Info: Concept talk Move Up :6
Info: running threads level: 2
Info: Concept open Move Up :4
Info: running threads level: 2
Info: *****CICLE UP*****
Info: Concept president Move Up :6
Info: Iteration Expansion :3
Info: running threads level: 3
Info: *****CICLE UP*****
Info: Concept nuke Move Up :5
Info: running threads level: 3
Info: *****CICLE UP*****
Info: Concept critic Move Up :4
Info: Concept critic Move Down :2
Info: running threads level: 3

Info: Concept iran Move Up :3
Info: running threads level: 3
Info: *****CICLE UP*****
Info: Concept war Move Up :3
Info: running threads level: 3
Info: Concept cold Move Up :5
Info: running threads level: 3
Info: Concept express Move Up :6
Info: running threads level: 3
Info: *****CICLE UP*****
Info: Concept talk Move Up :4
Info: running threads level: 3
Info: Concept open Move Up :4
Info: running threads level: 3
Info: *****CICLE UP*****
Info: Concept president Move Up :9
Info: Iteration Expansion :4
Info: running threads level: 4
Info: *****CICLE UP*****
Info: Concept nuke Move Up :4
Info: running threads level: 4
Info: Concept critic Move Up :5
Info: running threads level: 4
Info: Concept iran Move Up :4
Info: running threads level: 4
Info: *****CICLE UP*****
Info: Concept war Move Up :2
Info: running threads level: 4

Info: *****CICLE UP*****
Info: Concept cold Move Up :4
Info: running threads level: 4
Info: Concept express Move Up :6
Info: running threads level: 4
Info: *****CICLE UP*****
Info: Concept talk Move Up :3
Info: running threads level: 4
Info: *****CICLE UP*****
Info: Concept open Move Up :3
Info: running threads level: 4
Info: *****CICLE UP*****
Info: Concept president Move Up :8
Info: Iteration Expansion :5
Info: running threads level: 5
Info: Concept nuke Move Up :4
Info: running threads level: 5
Info: *****CICLE UP*****
Info: Concept critic Move Up :4
Info: running threads level: 5
Info: *****CICLE UP*****
Info: Concept iran Move Up :3
Info: running threads level: 5
Info: *****CICLE UP*****
Info: Concept war Move Up :1
Info: running threads level: 5
Info: Concept cold Move Up :2
Info: running threads level: 5

Info: *****CICLE UP*****
Info: Concept express Move Up :3
Info: running threads level: 5
Info: *****CICLE UP*****
Info: Concept talk Move Up :1
Info: running threads level: 5
Info: *****CICLE UP*****
Info: Concept open Move Up :2
Info: running threads level: 5
Info: *****CICLE UP*****
Info: Concept president Move Up :8
Info: Iteration Expansion :6
Info: running threads level: 6
Info: *****CICLE UP*****
Info: Concept nuke Move Up :4
Info: running threads level: 6
Info: *****CICLE UP*****
Info: Concept critic Move Up :3
Info: running threads level: 6
Info: Concept iran Move Up :4
Info: running threads level: 5
Info: running threads level: 6
Info: Concept cold Move Up :2
Info: running threads level: 6
Info: *****CICLE UP*****
Info: Concept express Move Up :2
Info: running threads level: 6
Info: running threads level: 5

Info: running threads level: 6
Info: *****CICLE UP*****
Info: Concept president Move Up :5
Info: Iteration Expansion :7
Info: Root Found: 7958392.0
Info: Total of Relation Item :299
Info: *****DIS-C*****
Info: It:0
Info: It:1
Info: It:2
Info: It:3
Info: It:4
Info: total dis-c 650 milliseconds
Info: Graph DISC Created
Info: Seed Sampling
Info: M: 6 Nodes:264
Info: [[2], [6, 10], [15, 19, 23, 27], [31, 36, 40, 44, 48, 52, 57, 61], [65, 69, 73, 78, 82, 86, 90, 94, 98, 103, 107, 111, 115, 119, 124, 128], [132, 136, 140, 145, 149, 153, 157, 161, 166, 170, 174, 178, 182, 186, 191, 195, 199, 203, 207, 212, 216, 220, 224, 228, 233, 237, 241, 245, 249, 254, 258, 262]]
Info: Dijkstra :
Info: running threads
Info: Sketch
Info: Skectch of terminals :
[[2, 10, 19, 52, 73, 182], [2, 10, 27, 36, 128, 249], [10, 23, 48, 103, 233], [2, 10, 15, 40, 107, 241], [2, 10, 15, 52, 90, 157], [2, 10, 19, 44, 111, 199], [2, 10, 15, 57, 82, 216], [2, 10, 15, 52, 78, 149], [2, 10, 15, 44, 65, 195]]
Info: Stainer
Info: *****Steiner Created*****, returning Ontology Steiner

Anexo 4. Comparación por alineación de la ontología de consulta en el sistema de recuperación

En este anexo se presentan tres fragmentos del procesamiento para la comparación basada en alineación, se compara la ontología de la consulta con las primeras ontologías de los objetos espaciales localizados en el repositorio de la etapa de síntesis.

En el procesamiento de similitud estructural se comparan números de identificador de las relaciones entre ambas ontologías. Para la similitud léxica se comparan los conceptos que comparten cadenas de texto. Finalmente en la similitud ontológica se compara toda la topología de sus ontologías (léxica, estructural y propiedades), de esta última similitud solo se muestran los valores de similitud.

Procesamiento de similitud estructural

Info: Comparison ontology O:1 of 529
Info: Stop: [WordNet:1241, WordNet:21166, WordNet:149982, WordNet:82639, WordNet:21500, WordNet:196430, WordNet:151708, WordNet:82677, WordNet:1245, WordNet:146016, WordNet:18848, WordNet:83122, WordNet:994, WordNet:65448, WordNet:1329, WordNet:277, WordNet:59392, WordNet:70513, WordNet:21469, WordNet:20953, WordNet:13, WordNet:15, WordNet:133311, WordNet:74497, WordNet:152158, WordNet:147334, WordNet:178979, WordNet:1330, WordNet:75815, WordNet:20961, WordNet:196640, WordNet:205, WordNet:147335, WordNet:145851, WordNet:1313, WordNet:204, WordNet:59384, WordNet:137208, WordNet:134074, WordNet:1319, WordNet:131942, WordNet:262, WordNet:137343, WordNet:170730, WordNet:178998, WordNet:152146, WordNet:178713]
Info: Item:3 Relations Shared : [WordNet:1329, WordNet:152158, WordNet:145851] Sim: 0.06
Info: Comparison ontology O:2 of 529
Info: Stop: [WordNet:1241, WordNet:21166, WordNet:149982, WordNet:82639, WordNet:21500, WordNet:196430, WordNet:151708, WordNet:82677, WordNet:1245, WordNet:146016, WordNet:18848, WordNet:83122, WordNet:994, WordNet:65448, WordNet:1329, WordNet:277, WordNet:59392, WordNet:70513, WordNet:21469, WordNet:20953, WordNet:13, WordNet:15, WordNet:133311, WordNet:74497, WordNet:152158, WordNet:147334, WordNet:178979, WordNet:1330, WordNet:75815, WordNet:20961, WordNet:196640, WordNet:205, WordNet:147335, WordNet:145851, WordNet:1313, WordNet:204, WordNet:59384, WordNet:137208, WordNet:134074, WordNet:1319, WordNet:131942, WordNet:262, WordNet:137343, WordNet:170730, WordNet:178998, WordNet:152146, WordNet:178713]
Info: Item:4 Relations Shared : [WordNet:196430, WordNet:204, WordNet:1319, WordNet:131942, WordNet:262, WordNet:178713] Sim: 0.13043478
Info: Comparison ontology O:3 of 529
Info: Stop: [WordNet:1241, WordNet:21166, WordNet:149982, WordNet:82639, WordNet:21500, WordNet:196430, WordNet:151708, WordNet:82677, WordNet:1245, WordNet:146016, WordNet:18848, WordNet:83122, WordNet:994, WordNet:65448, WordNet:1329, WordNet:277, WordNet:59392, WordNet:70513, WordNet:21469, WordNet:20953, WordNet:13, WordNet:15, WordNet:133311, WordNet:74497, WordNet:152158, WordNet:147334, WordNet:178979, WordNet:1330, WordNet:75815, WordNet:20961, WordNet:196640, WordNet:205, WordNet:147335, WordNet:145851, WordNet:1313, WordNet:204, WordNet:59384, WordNet:137208, WordNet:134074, WordNet:1319, WordNet:131942, WordNet:262, WordNet:137343, WordNet:170730, WordNet:178998, WordNet:152146, WordNet:178713]

Info: Item:5 Relations Shared : [WordNet:13, WordNet:262] Sim: 0.036697246

Procesamiento de similitud léxica

Info: Comparison ontology O:1 of 529

Info: Item:3 Relations Shared : [OPEC, international_organisation, grouping, somebody, world_organization, global_organization, people, organization, Organization_of_Petroleum-Exporting_Countries, organisation, group, individual, person, soul, mortal, someone, world_organisation, international_organization] Sim: 0.14229248

Info: Comparison ontology O:2 of 529

Info: Item:4 Relations Shared : [leader, presiding_officer, somebody, address, chair, human_activity, chairman, people, speech, act, chairperson, causal_agency, individual, person, speech_act, soul, cause, human_action, deed, causal_agent, mortal, chairwoman, someone, president] Sim: 0.21052632

Info: Comparison ontology O:3 of 529

Info: Item:5 Relations Shared : [leader, competition, somebody, abstract_entity, psychological_feature, individual, person, soul, mortal, abstraction, someone] Sim: 0.08661418

Info: Comparison ontology O:4 of 529

Info: Item:6 Relations Shared : [leader, professional, warfare, struggle, international_organisation, somebody, world_organization, professional_person, coalition, global_organization, alinement, group_action, human_activity, battle, organization, alliance, organisation, act, individual, person, alignment, professional_organization, conflict, soul, human_action, deed, mortal, war, someone, world_organisation, professional_organisation, international_organization] Sim: 0.30331755

Info: Comparison ontology O:5 of 529

Info: Item:7 Relations Shared : [somebody, event, group_action, human_activity, social_event, act, individual, person, soul, human_action, deed, mortal, someone] Sim: 0.1092437

.....

Procesamiento de similitud ontológica

Info: Comparison ontology O:1 of 529	>Item :2 with SimVal:0.34167194
Info: Comparison ontology O:2 of 529	>Item :3 with SimVal:0.40718114
Info: Comparison ontology O:3 of 529	>Item :4 with SimVal:0.46718565
Info: Comparison ontology O:4 of 529	>Item :5 with SimVal:0.34866846
Info: Comparison ontology O:5 of 529	>Item :6 with SimVal:0.52414095
Info: Comparison ontology O:6 of 529	>Item :7 with SimVal:0.37394428
Info: Comparison ontology O:7 of 529	>Item :8 with SimVal:0.47445294

Info: Comparison ontology O:8 of 529	>Item :9 with SimVal:0.40449563
Info: Comparison ontology O:9 of 529	>Item :10 with SimVal:0.40546483
.....	

Anexo 5. Objetos geoespaciales recuperados

En este anexo se listan los objetos geoespaciales (noticias) recuperados para la consulta “Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks”.

Los resultados mostrados son para las primeras 70 noticias ordenadas decrecientemente por orden de similitud estructural, léxica, y ontológica con la ontología de la consulta. Con base en estos resultados fueron creados las gráficas de *precision*, *recall* y medida *F*.

Comparación léxica

Item_Sim	Item_Title	Item_Description	Item_Date	Channel_Title	Item	Channel
1/1	Iran Willing to Begin 'Time-Bound,' 'Results-Oriented' Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News	Item	Channel
2/0.25	Iran on Nuke Talks: 'It's a Question of Months, Not Years'	President Hasan Rouhani says 'everything is possible' following successful talks.	2013-09-26 10:27:00.0	U.S. News - News	Item	Channel
3/0.25	Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Obama is open to meeting with Iranian president to discuss nuclear program.	2013-09-23 06:22:00.0	U.S. News - News	Item	Channel
4/0.25	A new sense of optimism in Iran	From bakery to barber shop, cafe to carpet store, Iranians stroll their capital with a renewed step, uplifted by how their newly elected president seeks something remarkable after decades of cold war-like relations between their country and the West.	2013-09-26 08:32:25.0	CNN.com - World	Item	Channel
5/0.25	Report: Shoe thrown at Rouhani	Iranian President Hassan Rouhani's arrival in Tehran Saturday was marked by an incident during which a shoe was thrown by a protester in his direction, Iran's semi-official Fararu news agency said.	2013-09-30 09:12:03.0	CNN.com - World	Item	Channel
6/0.16667	Al Gore Headlines Opening of New Arm of Brookings Institution	The former vice president re-emerges as global warming talk ramps up.	2013-09-27 05:00:00.0	U.S. News - News	Item	Channel
7/0.0833333	Emmy 2013 Predictions: Drama	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-19 08:28:00.0	U.S. News - News	Item	Channel
8/0.0833333	U.N. Could Agree on Syria Chemicals Resolution by Friday	Security Council movement, 'energetic' Iranian delegate among Thursday highlights in New York.	2013-09-27 10:19:00.0	U.S. News - News	Item	Channel
9/0.0833333	Navy Yard Shooting Revives Violent Game Debate But Doesn't Threaten Grand Theft Auto Fans	Video game critics face obstacles blaming massacres on violent media.	2013-09-18 16:00:00.0	U.S. News - News	Item	Channel
10/0.0833333	Gunmen fire on sleeping students	Under the cover of darkness, gunmen approached a college dormitory in a rural Nigerian town and opened fire on students who were sleeping.	2013-09-30 09:10:55.0	CNN.com - World	Item	Channel

[Next\(10\)](#)

Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
11/0.0833333	Benghazi One Year Later: Risk to Diplomats Remains High	Security experts criticize reforms since Sept. 11, 2012 attacks as diplomats continue risky work.	2013-09-11 06:00:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
12/0.0833333	Obama Begg Public for Trust in Syria Plan, Be it Diplomatic or Military	The president pledged diligence in decision-making on Syria.	2013-09-10 21:12:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
13/0.0833333	Obama, Blaming Shutdown, Cancels Asia Trip	President had planned to visit Indonesia and Brunei.	2013-10-04 06:36:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
14/0.0833333	Obama: GOP Trying to 'Extort' Him on Budget	President tells business leaders Republicans are behaving irresponsibly over threat to defund Obamacare.	2013-09-19 06:42:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
15/0.0833333	Obama's EPA Makes New Carbon Limits Official	EPA announcement furthers image of Obama's 'war on coal.'	2013-09-20 10:05:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
16/0.0833333	Emmy 2013 Predictions: Miniseries or Movie	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 08:46:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
17/0.0833333	Netanyahu: I'd talk to Rouhani	Israeli Prime Minister Benjamin Netanyahu says he would take a phone call from Iran's president, if he called, but warns that he would have tough words for the man he described this week as "a wolf in sheep's clothing."	2013-10-04 08:14:33	CNN.com - World	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
18/0.0833333	Obama, Advisers Go All Out to Make Syria Points	The president takes to TV, as does Syria's Assad.	2013-09-09 07:16:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
19/0.0833333	Emmy 2013 Predictions: Reality Competition and Variety	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 10:37:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
20/0.0833333	Hagel to Congress: Keep Military Option in Syria 'Real and Credible' Despite Diplomatic Breakthrough	Hagel tells Congress diplomatic breakthrough should not distract them from authorizing war footing.	2013-09-10 11:35:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
21/0.0833333	Ex-Liberian President Charles Taylor Loses Appeal	The U.N.-backed court's decision to imprison Taylor for 50 years on war crimes still stands.	2013-09-26 13:28:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
22/0.0833333	Obama Looks For Lifelines in Push for Syria Attack	The president continues his efforts to drum up support for a military strike against Syria.	2013-09-10 08:30:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
23/0.0833333	Netanyahu: 'Not averse to calling' Rouhani	Israeli Prime Minister Benjamin Netanyahu says he would take a phone call from Iran's president, if he called, but warns that he would have tough words for the man he described this week as "a wolf in sheep's clothing."	2013-10-04 08:02:16	CNN.com - World	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
24/0.0833333	Obama Will Win the Government Shutdown Battle	Presidents usually win in shutdowns with Congress.	2013-09-26 06:36:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
25/0.0833333	Ex-Lung Association Chief Leases His Lungs to E-Cigarette Industry	'Science should triumph over opinion,' says past president of the American Lung Association.	2013-09-27 15:10:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
26/0.0833333	Shutdown Forces Obama to Cancel Part of Asia Trip	President says he will reschedule stops in Malaysia and the Philippines.	2013-10-02 08:10:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
27/0.0833333	Democrats, Old and New, Have the Edge on the GOP	The Clintons and President Obama post better ratings than the Republicans.	2013-09-17 06:34:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
28/0.0833333	The President Will 'Shake His Groove Thing,' Says Michelle Obama	FLOTUS gets frank when talking about White House concerts.	2013-09-16 16:55:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
29/0.0833333	Kenyan Journalist Suspected of Bribing International Criminal Court Witnesses	Journalist accused of bribing witness in ICC trial of Kenyan Deputy President.	2013-10-02 11:23:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
30/0.0833333	Assad Agrees to U.N. Chemical Weapons Treaty	Chief diplomat criticizes all sides in Syria following Assad's agreement to U.N. treaty.	2013-09-13 09:38:00	U.S. News - News	Item	Channel

[Back\(10\)](#) [Next\(30\)](#)

Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
31/0.0833333	What You Need to Know About Terrorism in East Africa	A guide to the groups in and around Somalia contributing to the ongoing war on terrorism.	2013-09-30 14:37:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
32/0.0833333	Michelle Obama. White House Concerts for 'Rich' and 'Fancy' People	FLOTUS gets frank when talking about White House concerts.	2013-09-16 14:20:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
33/0.0833333	Wall Street Asks: What Shutdown?	Stocks rose slightly as trading opened on Tuesday.	2013-10-01 10:50:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
34/0.0833333	Obama to Hang Tough Over Shutdown	President cannot afford to show weakness, strategists say.	2013-10-01 06:40:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
35/0.0833333	The Issue of Race Follows New York Fashion Week Down the Carwalk	Model-agent launches campaign criticizing fashion industry's lack of diversity.	2013-09-13 16:50:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
36/0.0833333	Intensive Care Units Do Better Under Pressure, Study Finds	A new study shows critical care providers treat patients more effectively when put under pressure.	2013-10-01 13:49:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
37/0.0833333	Obama Heads Into the Fall Weakened	Syria, the economy and the budget dog the president as he deals with Congress.	2013-09-16 06:20:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
38/0.0833333	Putin Wins a Round From Obama	Russian president outmaneuvers Obama with Syrian diplomacy.	2013-09-12 07:04:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
39/0.0833333	Kerry Arrives in Geneva as Rebels Condemn Russian Proposal	Opposition condemns Russian proposal as bilateral talks begin in Geneva.	2013-09-12 08:00:00	U.S. News - News	Item	Channel
Item.Sim	Item.Title	Item.Description	Item.Date	Channel.Title	Item	Channel
40/0.0833333	Starbucks CEO: Don't Bring Guns to Our Stores	The Seattle-based chain has long-honored 'open carry' laws that allow residents to carry concealed guns.	2013-09-18 10:13:00	U.S. News - News	Item	Channel

[Back\(20\)](#) [Next\(40\)](#)

Recuperación e integración de datos espaciales mediante alineación de ontologías

<u>Item_Sim</u>	<u>Item_Title</u>	<u>Item_Description</u>	<u>Item_Date</u>	<u>Channel_Title</u>	<u>Item</u>	<u>Channel</u>
41/0.0833333	U.S. Seizes NYC Skyscraper With Ties to Iran	Iran had secret skyscraper on U.S. soil	2013-09-18 12:15:00.0	U.S. News - News	Item	Channel
42/0.0833333	Iran Cyber Chief Found Gunned Down, Reports Say	Reports indicate this could be another assassination against an influential Iranian official.	2013-10-03 11:20:00.0	U.S. News - News	Item	Channel
43/0.0833333	Declassified Documents Reveal NSA Privacy Violations	Declassified documents spur criticism of government surveillance programs.	2013-09-11 17:40:00.0	U.S. News - News	Item	Channel
44/0.0833333	Mental Health Legislation, Not Gun Control, Focus of Congress After Navy Yard Shooting	Republicans and Democrats seek mental health care legislation to curb gun violence.	2013-09-17 16:57:00.0	U.S. News - News	Item	Channel
45/0.0833333	Joaquin Castro Explains How to Tell Him Apart From Twin Brother	The Castro in Congress talked Ted Cruz and Wendy Davis.	2013-10-02 14:13:00.0	U.S. News - News	Item	Channel
46/0.0833333	Roland Martin is Washington's Preeminent 'Scandal' Correspondent	The former CNN talking head hosts a party for the show.	2013-10-02 13:47:00.0	U.S. News - News	Item	Channel
47/0.0833333	Al Gore Recalls Needing Newt Over Government Shutdown	The vice president had some fun with former House Speaker.	2013-09-27 17:01:00.0	U.S. News - News	Item	Channel
48/0.0833333	Obama's Former Defense Chiefs Criticize Him on Syria	President's handling of Syrian crisis has emboldened his critics.	2013-09-18 07:26:00.0	U.S. News - News	Item	Channel
49/0.0833333	Emmy 2013 Predictions: Comedy	Critics, fans and Emmy voters don't always agree who should go home with the big prize.	2013-09-18 06:00:00.0	U.S. News - News	Item	Channel
50/0.0833333	Norquist: Obama Will Be Weaker Because of Syria	President will use political capital he needs for other fights with Congress, Norquist says.	2013-09-05 06:52:00.0	U.S. News - News	Item	Channel
ack(30) Next(50)						
51/0.0833333	Bill Clinton Calls for NSA Transparency, Defends Obama on Syria	Former president said Obama did not need Congress for Syria bombings.	2013-09-23 04:00:00.0	U.S. News - News	Item	Channel
52/0.0833333	Afghan government hits back at NATO chief, says war aimless	KABUL (Reuters) - The Afghan government has hit back at remarks by the head of NATO who said Kabul must recognize the sacrifices made by other states, calling the alliance's war on terrorism in Afghanistan "aimless and unwise".	2013-03-19 09:42:37.0	Reuters: World News	Item	Channel
53/0.0833333	Obama Counting on Democratic Loyalty on Syria	President appeals to party faithful not to turn him into a lame duck.	2013-09-06 06:58:00.0	U.S. News - News	Item	Channel
54/0.0833333	A gift of U.S. diplomacy to Iran?	A 2,700 year-old silver chalice may be a new token of friendship between the United States and Iran, at least that's the way Iran's cultural heritage chief sees it.	2013-09-30 08:00:07.0	CNN.com - World	Item	Channel
55/0.0833333	Egyptian Court Bans Muslim Brotherhood	All assets to be seized from conservative Islamist party of former President Mohammed Morsi.	2013-09-23 12:58:00.0	U.S. News - News	Item	Channel
56/0.0833333	Guatemalan trial of ex-dictator Rios Montt begins	GUATEMALA CITY (Reuters) - The trial of former Guatemalan dictator Efraim Rios Montt on charges of genocide and crimes against humanity opened on Tuesday, the first time a country has prosecuted an ex-head of state in a national court on such charges.	2013-03-19 10:15:45.0	Reuters: World News	Item	Channel
57/0.0833333	Putin Puts Pressure on U.S. to Release Proof of Syrian Chemical Attacks	U.S. Will have to be as open with Russia as it is with Congress to prove point.	2013-09-04 15:46:00.0	U.S. News - News	Item	Channel
58/0.0833333	Syria Strike Protesters Hope to Harness Momentum Saturday in D.C.	Organizer says turnout will likely be lower than at similar protests against Iraq War.	2013-09-06 12:24:00.0	U.S. News - News	Item	Channel
59/0.0833333	Vicki Kennedy: People Will Never Let Obamacare Be Taken Away	Ted Kennedy's widow talks health care and Edward M. Kennedy Institute	2013-09-23 16:30:00.0	U.S. News - News	Item	Channel
60/0.0833333	Iranian Foreign Minister to Meet With Kerry at U.N. General Assembly	John Kerry among top delegates to meet with the Iranian foreign minister.	2013-09-23 12:59:00.0	U.S. News - News	Item	Channel
ack(40) Next(60)						

Recuperación e integración de datos espaciales mediante alineación de ontologías

Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
61/0.0833333	Obama Likely to Win War of Words Over Health Care	President presents united front, unlike his GOP opposition.	2013-09-24 06:36:00.0	U.S. News - News	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
62/0.0833333	Ted Cruz Marathon Continues on Senate Floor	Cruz keeps talking against the health care bill.	2013-09-25 07:52:00.0	U.S. News - News	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
63/0.0833333	Girls flee bombs, start over in U.S.	Blair Brettschneider didn't experience war or genocide as a young girl in suburban Michigan. She has never received a death threat, and she has never seen a family member murdered. But the 49 teenagers she helps every week have.	2013-08-23 08:08:36.0	CNN.com - World	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
64/0.0833333	Obama at U.N.: Mideast Peace, Preventing Nuclear Iran Are Top U.S. Goals	Obama asked for international support of diplomatic paths in Syria and Iran.	2013-09-24 10:48:00.0	U.S. News - News	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
65/0.0833333	Jewels turn up atop French glacier	It reads like the opening scene of an "Indiana Jones" movie.	2013-09-27 10:21:24.0	CNN.com - World	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
66/0.0833333	Zimbabwe approves constitution on path to elections	HARARE (Reuters) - Zimbabweans have approved a new constitution that curbs the powers of the president and puts the turbulent southern African country a step closer to holding elections, the election commission said on Tuesday.	2013-03-19 10:01:34.0	Reuters: World News	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
67/0.0833333	Saving orphans or child trafficking?	As international adoption becomes more difficult, a growing number of voices are pushing to reduce restrictions. That's a mistake, critics say.	2013-09-20 07:22:09.0	CNN.com - World	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
68/0.0833333	Capriles barbs unite "Chavistas" in Venezuela: strategist	CARACAS (Reuters) - Supporters of late president Hugo Chavez will punish opposition leader Henrique Capriles for his anti-government barbs at a vote next month likely to produce a bigger win for the socialists than last year, a campaign strategist for Capriles' rival forecast.	2013-03-19 08:26:18.0	Reuters: World News	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
69/0	Jennifer Hudson Joins the Obamacare Wars With Funny or Die Video	Hudson spoofs the hit show 'Scandal' in a pro-Affordable Care Act campaign.	2013-09-30 11:12:00.0	U.S. News - News	Item	Channel
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:	Item	Channel
70/0	This Time, Debt Ceiling Debate May Not Be a Golden Opportunity	In 2011, the debt ceiling fight helped push gold prices to record levels, but the market looks calmer now.	2013-09-24 14:40:00.0	U.S. News - News	Item	Channel

Similitud estructural

Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
1/1	Iran Willing to Begin 'Time-Bound,' 'Results-Oriented' Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
2/0.459459	Obama's Former Defense Chiefs Criticize Him on Syria	President's handling of Syrian crisis has emboldened his critics.	2013-09-18 07:26:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
3/0.325	Iranian Foreign Minister to Meet With Kerry at U.N. General Assembly	John Kerry among top delegates to meet with the Iranian foreign minister.	2013-09-23 12:59:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
4/0.290323	Iran Cyber Chief Found Gunned Down, Reports Say	Reports indicate this could be another assassination against an influential Iranian official.	2013-10-03 11:20:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
5/0.277778	Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Obama is open to meeting with Iranian president to discuss nuclear program.	2013-09-23 06:22:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
6/0.258065	DCCC Says Syria Is Not An Election Issue	Steve Israel doesn't expect 2014 to be referendum on Syria.	2013-09-10 11:02:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
7/0.252874	U.S. Draws Down Embassies in Lebanon and Turkey	A Syria strike may yield widespread Hezbollah retribution, a retired official warns.	2013-09-06 11:41:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
8/0.247191	Vicki Kennedy: People Will Never Let Obamacare Be Taken Away	Ted Kennedy's widow talks health care and Edward M. Kennedy Institute	2013-09-23 16:30:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
9/0.235294	Ron Paul Isn't Thrilled By 'Rather Annoying' Shutdown Debate	'The government doesn't really shut down,' says retired Texas congressman.	2013-10-01 16:04:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
10/0.227273	U.N. Could Agree on Syria Chemicals Resolution by Friday	Security Council movement, 'energetic' Iranian delegate among Thursday highlights in New York.	2013-09-27 10:19:00.0	U.S. News - News

[Next\(10\)](#)

Item/Sim	Item Title	Item Description	Item Date	Channel Title
11 0.222222	Athletes Face Challenges in Speaking Out Against Russia's Anti-Gay Laws	The IOC and Russian officials have discouraged athletes from making political statements at the Olympics.	2013-10-04 10:22:00.0	U.S. News - News
12 0.219512	Pope Francis Announces Canonization of Pope John Paul II and Pope John XXIII Set for April 2014	The Vatican looks forward to lots of visitors for April's canonization ceremony.	2013-09-30 11:34:00.0	U.S. News - News
13 0.213592	Plane crash kills 14 at Nigeria airport	A plane carrying 20 people crashed on takeoff from Nigeria's Murtala Muhammed Airport in Lagos on Thursday, killing 14, the nation's aviation minister said.	2013-10-04 08:01:09.0	CNN.com - World
14 0.213333	Emmy 2013 Predictions: Miniseries or Movie	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 08:46:00.0	U.S. News - News
15 0.213333	Emmy 2013 Predictions: Drama	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-19 08:28:00.0	U.S. News - News
16 0.213333	Emmy 2013 Predictions: Comedy	Critics, fans and Emmy voters don't always agree who should go home with the big prize.	2013-09-18 06:00:00.0	U.S. News - News
17 0.213333	Emmy 2013 Predictions: Reality Competition and Variety	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 10:37:00.0	U.S. News - News
18 0.211765	Wall Street Asks: What Shutdown?	Stocks rose slightly as trading opened on Tuesday.	2013-10-01 10:50:00.0	U.S. News - News
19 0.210526	Drumbeat Intensifies for Electronic Cigarette Regulations	E-cigarette advocates say hit the brakes, act on evidence.	2013-09-25 17:08:00.0	U.S. News - News
20 0.209302	Obama Lays Everything on the Line for Syria	Tuesday's address on Syria has consequences for Obama's entire agenda.	2013-09-06 16:13:00.0	U.S. News - News

[Back\(0\) Next\(20\)](#)

Item/Sim	Item Title	Item Description	Item Date	Channel Title
21 0.207792	Cumings Calls for AIG CEO Benmosche's Resignation	The Democratic congressman calls Benmosche's comments about bonuses and lynch mobs 'unbelievably appalling.'	2013-09-24 17:21:00.0	U.S. News - News
22 0.205128	Bill de Blasio, Joe Lhota Appear in Control in New York City Mayoral Primary Races	Candidates must get 40 percent to avoid a runoff.	2013-09-10 06:26:00.0	U.S. News - News
23 0.202899	Joaquin Castro Explains How to Tell Him Apart From Twin Brother	The Castro in Congress talked Ted Cruz and Wendy Davis.	2013-10-02 14:13:00.0	U.S. News - News
24 0.2	Committee Passes Resolution to Strike Syria	Committee Approves Resolution to Intervene in Syria.	2013-09-04 14:56:00.0	U.S. News - News
25 0.2	In Montana, Democrats Might Have Found Their Guy	Montana Senate race not off the Democrats' map yet.	2013-09-16 06:00:00.0	U.S. News - News
26 0.197183	India Sentences Convicted Rapists to Death	India Gives the Death Penalty	2013-09-13 11:30:00.0	U.S. News - News
27 0.197183	Lawmakers Creating a Blue Christmas	Washington's fiscal shenanigans may restrain holiday spending this year.	2013-10-03 15:35:00.0	U.S. News - News
28 0.195652	Your solstice traditions	If Stonehenge is anything to go by, summer celebrations have a millennia-old history. While the ancient druids may have commemorated the fertility of the season, many of us are just happy to get outdoors.	2013-06-21 05:35:52.0	CNN.com - World
29 0.194444	Antonin Scalia Says Judicial Activism Could Be Worse	Outspoken jurist says Germans live under a more activist court.	2013-09-16 17:27:00.0	U.S. News - News
30 0.191781	Poll: Biden, Kerry Beat Obama in Popularity	Public punishes Obama for Syria as well as IRS and Benghazi, Libya, controversies.	2013-09-16 07:14:00.0	U.S. News - News

[Back\(10\) Next\(30\)](#)

<u>Item Sim</u>	<u>Item Title</u>	<u>Item Description</u>	<u>Item Date</u>	<u>Channel Title</u>
31 0.191781	Syrian Coalition Rejects Russian Proposal Without Prosecuting Assad	Syrian Coalition will not consider any proposal that does not bring the Assad regime to justice.	2013-09-10 14:03:00.0	U.S. News - News
32 0.190476	Afghan government hits back at NATO chief, says war aimless	KABUL (Reuters) - The Afghan government has hit back at remarks by the head of NATO who said Kabul must recognize the sacrifices made by other states, calling the alliance's war on terrorism in Afghanistan "aimless and unwise".	2013-03-19 09:42:37.0	Reuters: World News
33 0.190476	Cuba to allow athletes abroad	Ever since the Cuban Revolution more than 50 years ago, athletes on the communist island have been able to garner gold medals and glory, but not big paychecks.	2013-09-30 08:06:50.0	CNN.com - World
34 0.189189	Gunmen Attack Russian Embassy in Tripoli	No injuries as attackers attempt to storm Russian compound.	2013-10-02 17:04:00.0	U.S. News - News
35 0.189189	Obama Blitzes Press to Make Case for Syria; Assad Threatens Back	Syria decision threatens to color Obama's legacy.	2013-09-09 07:00:00.0	U.S. News - News
36 0.186667	In Arkansas Senate Race, Syria Emerges As Defining Issue	Sen. Pryor's decision to turn away from Obama on Syria, could help in Senate race.	2013-09-09 13:02:00.0	U.S. News - News
37 0.184211	Freed Hostages Reveal Information on Chemical Attacks in Syria	Hostages share potentially new details about Syria's attacks.	2013-09-10 13:00:00.0	U.S. News - News
38 0.181818	Rockefeller Expands Investigation on Consumer Data Brokers	Senate chairman mails 12 websites requesting privacy information.	2013-09-24 23:00:00.0	U.S. News - News
39 0.179487	Bill Clinton Calls for NSA Transparency, Defends Obama on Syria	Former president said Obama did not need Congress for Syria bombings.	2013-09-23 04:00:00.0	U.S. News - News
40 0.179487	Political Ping Pong, 'The West Wing' Await Furloughed Federal Workers	A Washington synagogue has planned entertainment for the shutdown.	2013-09-30 16:50:00.0	U.S. News - News

[Back\(20\)](#) [Next\(40\)](#)

41 0.179104	Egyptian Court Bans Muslim Brotherhood	All assets to be seized from conservative Islamist party of former President Mohammed Morsi.	2013-09-23 12:58:00.0	U.S. News - News
42 0.179104	Assad Agrees to U.N. Chemical Weapons Treaty	Chief diplomat criticizes all sides in Syria following Assad's agreement to U.N. treaty.	2013-09-13 09:38:00.0	U.S. News - News
43 0.176471	Ted Cruz: U.S. Should Turn Up the Heat on China, Russia	Cruz says Obama is mishandling Syria and should push China, Russia.	2013-09-11 17:30:00.0	U.S. News - News
44 0.173913	Republicans, Democrats Prepare for Shutdown Blame Game	Republicans and Democrats aren't blinking ahead of government shutdown.	2013-09-30 10:34:00.0	U.S. News - News
45 0.17284	Veteran Advocates Caution Against Pinning Navy Yard Shooting on PTSD	Combat stress likely did not cause a man to kill 12 in Navy Yard shooting, experts say.	2013-09-24 10:42:00.0	U.S. News - News
46 0.172414	Liberal Group Tries to Make 'Flat Boehner' Happen	The group Americans United for Change got crafty.	2013-10-03 13:25:00.0	U.S. News - News
47 0.172043	Kenyan Journalist Suspected of Bribing International Criminal Court Witnesses	Journalist accused of bribing witness in ICC trial of Kenyan Deputy President.	2013-10-02 11:23:00.0	U.S. News - News
48 0.171420	Iran on Nuke Talks: 'It's a Question of Months, Not Years'	President Hasan Rouhani says 'everything is possible' following successful talks.	2013-09-26 10:27:00.0	U.S. News - News
49 0.170732	Jennifer Hudson Joins the Obamacare Wars With Funny or Die Video	Hudson spoofs the hit show 'Scandal' in a pro-Affordable Care Act campaign.	2013-09-30 11:12:00.0	U.S. News - News
50 0.170732	Consumers Earned, Spent, Saved More in August	A little extra income meant Americans both spent and saved more last month.	2013-09-27 09:40:00.0	U.S. News - News

[Back\(30\)](#) [Next\(50\)](#)

<u>Item/Sim</u>	<u>Item Title</u>	<u>Item Description</u>	<u>Item Date</u>	<u>Channel Title</u>
51 0.169014	Obama Muddles Through on Syria	Presidential address fails to resolve the inconsistencies of Obama's Syria policy.	2013-09-11 06:58:00.0	U.S. News - News
52 0.169014	Putin Wins a Round From Obama	Russian president outmaneuvers Obama with Syrian diplomacy.	2013-09-12 07:04:00.0	U.S. News - News
53 0.169014	Verizon Challenges Open Internet Rules in Court	Court hears arguments in Verizon lawsuit against net neutrality.	2013-09-10 14:15:00.0	U.S. News - News
54 0.169014	D.C. Mayor Vincent Gray Says At Least 13 Dead In Navy Yard Shooting	Police continue to search for one potential suspect who remains at-large.	2013-09-16 11:51:00.0	U.S. News - News
55 0.164706	Obama Heads Into the Fall Weakened	Syria, the economy and the budget dog the president as he deals with Congress.	2013-09-16 06:20:00.0	U.S. News - News
56 0.164706	Syria Strike Protesters Hope to Harness Momentum Saturday in D.C.	Organizer says turnout will likely be lower than at similar protests against Iraq War.	2013-09-06 12:24:00.0	U.S. News - News
57 0.164384	India's obsession with lighter skin	In India, many women are obsessed with lightening their skin. CNN writer Moni Basu recalls her own family's pressures to not get "too dark."	2013-09-25 12:02:27.0	CNN.com - World
58 0.164384	Crossfire's Cutter Slaps Sarah Palin Over Let's Move Diss	An Obama aide hinted Palin knocked the program for ratings.	2013-09-26 13:17:00.0	U.S. News - News
59 0.164384	Study: The American Family Is No More	A new study suggests immigrants now most closely resemble the 'typical' American family.	2013-09-11 15:12:00.0	U.S. News - News
60 0.162162	Navy Yard Shooting Reminiscent of Past Massacres	America has had a long history with massacres since the turn of the century.	2013-09-16 16:20:00.0	U.S. News - News

[back\(40\)](#) [Next\(60\)](#)

<u>Item/Sim</u>	<u>Item Title</u>	<u>Item Description</u>	<u>Item Date</u>	<u>Channel Title</u>
61 0.157895	Flash Floods Force Thousands of Evacuations in Colorado	Colorado National Guard is assisting in search and rescue efforts statewide, helping evacuate residents.	2013-09-13 15:32:00.0	U.S. News - News
62 0.157895	NBC, CNN Cancel Hillary Clinton Specials	The RNC calls on NBC to do the same for its planned Hillary Clinton miniseries.	2013-09-30 14:23:00.0	U.S. News - News
63 0.157895	Al Gore Headlines Opening of New Arm of Brookings Institution	The former vice president re-emerges as global warming talk ramps up.	2013-09-27 05:00:00.0	U.S. News - News
64 0.155844	Democrats Think They Are Winning the Shutdown War	Polls show voters more likely to blame Republicans for the mess in Washington.	2013-10-02 06:48:00.0	U.S. News - News
65 0.153846	Obama Bids Public for Trust in Syria Plan, Be it Diplomatic or Military	The president pledged diligence in decision-making on Syria.	2013-09-10 21:12:00.0	U.S. News - News
66 0.153846	Ex-Lung Association Chief Leases His Lungs to E-Cigarette Industry	'Science should triumph over opinion,' says past president of the American Lung Association.	2013-09-27 15:10:00.0	U.S. News - News
67 0.153846	'Nashville' Recap: Season 2 Premiere	'Nashville' struggles to regain its mojo with star Rayna sidelined.	2013-09-26 06:46:00.0	U.S. News - News
68 0.153846	Report: Health Care System in Crisis Over Cancer Treatment	An eroding oncology workforce, rising annual diagnoses and astronomical costs plague the industry.	2013-09-11 16:38:00.0	U.S. News - News
69 0.153846	'Top Chef' Tom Colicchio Serves Up Side of Food-Stamp Politics	Chef lobbies on Capitol Hill against GOP cuts in food stamp program.	2013-09-18 09:56:00.0	U.S. News - News
70 0.151515	Lemur Hibernation Patterns Shed Light Onto Why Humans Sleep	Researchers say their findings could one day allow scientists to induce hibernation states in humans.	2013-09-04 16:32:00.0	U.S. News - News

[back\(50\)](#) [Next\(70\)](#)

Similitud léxica

Item/Sim 1 1	Item Title: Iran Willing to Begin 'Time-Bound,' 'Results-Oriented' Steps for Nuke Program	Item Description: Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	Item Date: 2013-09-24 17:41:00.0	Channel Title: U.S. News - News
Item/Sim 2 0.508876	Item Title: Obama's Former Defense Chiefs Criticize Him on Syria	Item Description: President's handling of Syrian crisis has emboldened his critics.	Item Date: 2013-09-18 07:26:00.0	Channel Title: U.S. News - News
Item/Sim 3 0.380952	Item Title: Iran Cyber Chief Found Gunned Down, Reports Say	Item Description: Reports indicate this could be another assassination against an influential Iranian official.	Item Date: 2013-10-03 11:20:00.0	Channel Title: U.S. News - News
Item/Sim 4 0.357895	Item Title: Iranian Foreign Minister to Meet With Kerry at U.N. General Assembly	Item Description: John Kerry among top delegates to meet with the Iranian foreign minister.	Item Date: 2013-09-23 12:59:00.0	Channel Title: U.S. News - News
Item/Sim 5 0.341463	Item Title: Athletes Face Challenges in Speaking Out Against Russia's Anti-Gay Laws	Item Description: The IOC and Russian officials have discouraged athletes from making political statements at the Olympics.	Item Date: 2013-10-04 10:22:00.0	Channel Title: U.S. News - News
Item/Sim 6 0.333333	Item Title: Syrian Coalition Rejects Russian Proposal Without Prosecuting Assad	Item Description: Syrian Coalition will not consider any proposal that does not bring the Assad regime to justice.	Item Date: 2013-09-10 14:03:00.0	Channel Title: U.S. News - News
Item/Sim 7 0.316279	Item Title: U.N. Could Agree on Syria Chemicals Resolution by Friday	Item Description: Security Council movement, 'energetic' Iranian delegate among Thursday highlights in New York.	Item Date: 2013-09-27 10:19:00.0	Channel Title: U.S. News - News
Item/Sim 8 0.311688	Item Title: Obama Begs Public for Trust in Syria Plan, Be it Diplomatic or Military	Item Description: The president pledged diligence in decision-making on Syria.	Item Date: 2013-09-10 21:12:00.0	Channel Title: U.S. News - News
Item/Sim 9 0.310881	Item Title: U.S. Draws Down Embassies in Lebanon and Turkey	Item Description: A Syria strike may yield widespread Hezbollah retribution, a retired official warns.	Item Date: 2013-09-06 11:41:00.0	Channel Title: U.S. News - News
Item/Sim 10 0.310345	Item Title: Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Item Description: Obama is open to meeting with Iranian president to discuss nuclear program.	Item Date: 2013-09-23 06:22:00.0	Channel Title: U.S. News - News
Next(10)				
Item/Sim 11 0.308571	Item Title: In Arkansas Senate Race, Syria Emerges As Defining Issue	Item Description: Sen. Pryor's decision to turn away from Obama on Syria, could help in Senate race.	Item Date: 2013-09-09 13:02:00.0	Channel Title: U.S. News - News
Item/Sim 12 0.306011	Item Title: Freed Hostages Reveal Information on Chemical Attacks in Syria	Item Description: Hostages share potentially new details about Syria's attacks.	Item Date: 2013-09-10 13:00:00.0	Channel Title: U.S. News - News
Item/Sim 13 0.303318	Item Title: Afghan government hits back at NATO chief, says war aimless	Item Description: KABUL (Reuters) - The Afghan government has hit back at remarks by the head of NATO who said Kabul must recognize the sacrifices made by other states, calling the alliance's war on terrorism in Afghanistan "aimless and unwise".	Item Date: 2013-03-19 09:42:37.0	Channel Title: Reuters: World News
Item/Sim 14 0.300518	Item Title: Pope Francis Announces Canonization of Pope John Paul II and Pope John XXIII Set for April 2014	Item Description: The Vatican looks forward to lots of visitors for April's canonization ceremony.	Item Date: 2013-09-30 11:34:00.0	Channel Title: U.S. News - News
Item/Sim 15 0.300429	Item Title: Vicki Kennedy: People Will Never Let Obamacare Be Taken Away	Item Description: Ted Kennedy's widow talks health care and Edward M. Kennedy Institute	Item Date: 2013-09-23 16:30:00.0	Channel Title: U.S. News - News
Item/Sim 16 0.297872	Item Title: Committee Passes Resolution to Strike Syria	Item Description: Committee Approves Resolution to Intervene in Syria.	Item Date: 2013-09-04 14:56:00.0	Channel Title: U.S. News - News
Item/Sim 17 0.296296	Item Title: Ron Paul Isn't Thrilled By 'Rather Annoying' Shutdown Debate	Item Description: 'The government doesn't really shut down,' says retired Texas congressman.	Item Date: 2013-10-01 16:04:00.0	Channel Title: U.S. News - News
Item/Sim 18 0.291892	Item Title: Ex-Lung Association Chief Leases His Lungs to E-Cigarette Industry	Item Description: 'Science should triumph over opinion,' says past president of the American Lung Association.	Item Date: 2013-09-27 15:10:00.0	Channel Title: U.S. News - News
Item/Sim 19 0.291139	Item Title: Obama at U.N.: Mideast Peace, Preventing Nuclear Iran Are Top U.S. Goals	Item Description: Obama asked for international support of diplomatic paths in Syria and Iran.	Item Date: 2013-09-24 10:48:00.0	Channel Title: U.S. News - News
Item/Sim 20 0.290909	Item Title: Obama Muddles Through on Syria	Item Description: Presidential address fails to resolve the inconsistencies of Obama's Syria policy.	Item Date: 2013-09-11 06:58:00.0	Channel Title: U.S. News - News
Back(0) Next(20)				

Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
21 0.289157	Obama Blitzes Press to Make Case for Syria; Assad Threatens Back	Syria decision threatens to color Obama's legacy.	2013-09-09 07:00:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
22 0.282209	Assad Agrees to U.N. Chemical Weapons Treaty	Chief diplomat criticizes all sides in Syria following Assad's agreement to U.N. treaty.	2013-09-13 09:38:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
23 0.281938	Drumbeat Intensifies for Electronic Cigarette Regulations	E-cigarette advocates say hit the brakes, act on evidence.	2013-09-25 17:08:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
24 0.27451	Plane crash kills 14 at Nigeria airport	A plane carrying 20 people crashed on takeoff from Nigeria's Murtala Muhammed Airport in Lagos on Thursday, killing 14, the nation's aviation minister said.	2013-10-04 08:01:09.0	CNN.com - World
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
25 0.27381	Antonin Scalia Says Judicial Activism Could Be Worse	Outspoken jurist says Germans live under a more activist court.	2013-09-16 17:27:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
26 0.273381	DCCC Says Syria Is Not An Election Issue	Steve Israel doesn't expect 2014 to be referendum on Syria.	2013-09-10 11:02:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
27 0.272189	In Montana, Democrats Might Have Found Their Guy	Montana Senate race not off the Democrats' map yet.	2013-09-16 06:00:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
28 0.269231	Hackers Attack NASA's Website to Protest NSA	Brazilian group targets incorrect U.S. agency.	2013-09-12 11:15:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
29 0.268156	Poll: Biden, Kerry Beat Obama in Popularity	Public punishes Obama for Syria as well as IRS and Benghazi, Libya, controversies.	2013-09-16 07:14:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
30 0.265957	Flash Floods Force Thousands of Evacuations in Colorado	Colorado National Guard is assisting in search and rescue efforts statewide, helping evacuate residents.	2013-09-13 15:32:00.0	U.S. News - News
back(10) Next(30)				
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
31 0.26506	Gunmen Attack Russian Embassy in Tripoli	No injuries as attackers attempt to storm Russian compound.	2013-10-02 17:04:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
32 0.264901	John McCain Uses Pravda to Bash Vladimir Putin	McCain says Putin's support of Bashar Assad hurts Russia.	2013-09-19 08:10:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
33 0.264151	Ted Cruz: U.S. Should Turn Up the Heat on China, Russia	Cruz says Obama is mishandling Syria and should push China, Russia.	2013-09-11 17:30:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
34 0.258427	Study: The American Family Is No More	A new study suggests immigrants now most closely resemble the 'typical' American family.	2013-09-11 15:12:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
35 0.257669	Putin Wins a Round From Obama	Russian president outmaneuvers Obama with Syrian diplomacy.	2013-09-12 07:04:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
36 0.255556	Bill Clinton Calls for NSA Transparency, Defends Obama on Syria	Former president said Obama did not need Congress for Syria bombings.	2013-09-23 04:00:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
37 0.252874	Joaquin Castro Explains How to Tell Him Apart From Twin Brother	The Castro in Congress talked Ted Cruz and Wendy Davis.	2013-10-02 14:13:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
38 0.252632	Gun Activists Oust 2 Colorado Lawmakers in Election Recall	Voters force out gun control advocates in Colorado recall election.	2013-09-11 12:15:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
39 0.251497	Bill de Blasio, Joe Lhota Appear in Control in New York City Mayoral Primary Races	Candidates must get 40 percent to avoid a runoff.	2013-09-10 06:26:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
40 0.25	Aaron Alexis Used Shotgun, Handgun to Kill at D.C. Navy Yard, Police Say	The alleged suspect also had valid access to the Washington facility.	2013-09-17 15:43:00.0	U.S. News - News
back(20) Next(40)				

Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
41(0.248756)	Your solstice traditions	If Stonehenge is anything to go by, summer celebrations have a millennia-old history. While the ancient druids may have commemorated the fertility of the season, many of us are just happy to get outdoors.	2013-06-21 05:35:52.0	CNN.com - World
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
42(0.247191)	Lawmakers Creating a Blue Christmas	Washington's fiscal shenanigans may restrain holiday spending this year.	2013-10-03 15:35:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
43(0.245614)	India's obsession with lighter skin	In India, many women are obsessed with lightening their skin. CNN writer Moni Basu recalls her own family's pressures to not get "too dark."	2013-09-25 12:02:27.0	CNN.com - World
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
44(0.245614)	A new sense of optimism in Iran	From bakery to barber shop, cafe to carpet store, Iranians stroll their capital with a renewed step, uplifted by how their newly elected president seeks something remarkable after decades of cold war-like relations between their country and the West.	2013-09-26 08:32:25.0	CNN.com - World
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
45(0.245399)	Kerry and Hagel Get Earful From House About Syria	House hearing reveals strong convictions against Syrian intervention.	2013-09-04 16:29:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
46(0.244898)	Egyptian Court Bans Muslim Brotherhood	All assets to be seized from conservative Islamist party of former President Mohammed Morsi.	2013-09-23 12:58:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
47(0.242718)	Obama Lays Everything on the Line for Syria	Tuesday's address on Syria has consequences for Obama's entire agenda.	2013-09-06 16:13:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
48(0.240506)	India Sentences Convicted Rapists to Death	India Gives the Death Penalty	2013-09-13 11:30:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
49(0.240437)	John Kerry Touts U.N. Syria Chemical Weapons Report Ahead of U.N. Meeting	The secretary of state says the world needs to 'stand up' in Syria.	2013-09-19 16:02:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
50(0.240437)	Veteran Advocates Caution Against Pinning Navy Yard Shooting on PTSD	Combat stress likely did not cause a man to kill 12 in Navy Yard shooting, experts say.	2013-09-24 10:42:00.0	U.S. News - News

[ack\(30\) Next\(50\)](#)

Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
51(0.239521)	Iran on Nuke Talks: 'It's a Question of Months, Not Years'	President Hasan Rouhani says 'everything is possible' following successful talks.	2013-09-26 10:27:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
52(0.23913)	Consumers Earned, Spent, Saved More in August	A little extra income meant Americans both spent and saved more last month.	2013-09-27 09:40:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
53(0.238636)	Rockefeller Expands Investigation on Consumer Data Brokers	Senate chairman mails 12 websites requesting privacy information.	2013-09-24 23:00:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
54(0.238095)	Republicans, Democrats Prepare for Shutdown Blame Game	Republicans and Democrats aren't blinking ahead of government shutdown.	2013-09-30 10:34:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
55(0.237113)	Starbucks CEO: Don't Bring Guns to Our Stores	The Seattle-based chain has long-honored 'open carry' laws that allow residents to carry unconcealed guns.	2013-09-18 10:13:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
56(0.236686)	Obama Counting on Democratic Loyalty on Syria	President appeals to party faithful not to turn him into a lame duck.	2013-09-06 06:58:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
57(0.236559)	Political Ping Pong, 'The West Wing' Await Furloughed Federal Workers	A Washington synagogue has planned entertainment for the shutdown.	2013-09-30 16:50:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
58(0.236559)	Grisly Video, Putin Criticism Further Tarnishes Image of Syrian Rebels	Russian leader calls Kerry a liar; Video depicts execution of regime forces.	2013-09-05 10:06:00.0	U.S. News - News
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
59(0.234043)	Qatar: We're no slave state	There's nearly a decade to go before a ball is kicked at soccer's 2022 World Cup in Qatar. But the tiny Gulf state is embroiled in controversy over the treatment of the huge migrant labor force within its borders.	2013-10-04 08:34:26.0	CNN.com - World
Item/Sim	Item Title:	Item Description:	Item Date:	Channel Title:
60(0.234043)	Qatar confronts 'slave state' accusations	There's nearly a decade to go before a ball is kicked at soccer's 2022 World Cup in Qatar. But the tiny Gulf state is embroiled in controversy over the treatment of the huge migrant labor force within its borders.	2013-10-04 08:16:34.0	CNN.com - World

[ack\(40\) Next\(60\)](#)

Item/Sim	Item Title	Item Description	Item Date	Channel Title
61/0.233918	Al Gore Headlines Opening of New Arm of Brookings Institution	The former vice president re-emerges as global warming talk ramps up.	2013-09-27 05:00:00.0	U.S. News - News
62/0.233577	Liberal Group Tries to Make 'Flat Boehner' Happen	The group Americans United for Change got crafty.	2013-10-03 13:25:00.0	U.S. News - News
63/0.233333	FBI says it has caught shadowy Silk Road creator	The FBI caught the man accused of creating Silk Road -- the shadowy e-commerce site it describes as "the most sophisticated and extensive criminal marketplace on the Internet today" -- after he allegedly posted his Gmail address online, according to court documents.	2013-10-04 07:41:14.0	CNN.com - World
64/0.232804	Obama Heads Into the Fall Weakened	Syria, the economy and the budget dog the president as he deals with Congress.	2013-09-16 06:20:00.0	U.S. News - News
65/0.230769	More Americans Want Less Strict Gun Laws	Fewer Americans today than in December 2012 see gun control issues as the top cause of shootings.	2013-09-20 09:45:00.0	U.S. News - News
66/0.228571	Verizon Challenges Open Internet Rules in Court	Court hears arguments in Verizon lawsuit against net neutrality.	2013-09-10 14:15:00.0	U.S. News - News
67/0.225352	Obama Likely to Win War of Words Over Health Care	President presents united front, unlike his GOP opposition.	2013-09-24 06:36:00.0	U.S. News - News
68/0.224719	Cummings Calls for AIG CEO Benmosche's Resignation	The Democratic congressman calls Benmosche's comments about bonuses and lynch mobs 'unbelievably appalling.'	2013-09-24 17:21:00.0	U.S. News - News
69/0.222222	Emmy 2013 Predictions: Drama	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-19 08:28:00.0	U.S. News - News
70/0.222222	Emmy 2013 Predictions: Reality Competition and Variety	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 10:37:00.0	U.S. News - News

[ack\(50\) Next\(70\)](#)

Similitud ontológica

Item/Sim	Item Title	Item Description	Item Date	Channel Title
1/1	Iran Willing to Begin 'Time-Bound,' 'Results-Oriented' Steps for Nuke Program	Iranian President Hassan Rouhani criticizes Cold War mentality, but expresses openness to nuke talks.	2013-09-24 17:41:00.0	U.S. News - News
2/0.768876	Economic Issues Take Back Seat to Syria at G-20 Summit	Obama meets with Chinese and French on final day in Russia.	2013-09-06 08:07:00.0	U.S. News - News
3/0.759223	DCCC Says Syria Is Not An Election Issue	Steve Israel doesn't expect 2014 to be referendum on Syria.	2013-09-10 11:02:00.0	U.S. News - News
4/0.7343	Iran Cyber Chief Found Gunned Down, Reports Say	Reports indicate this could be another assassination against an influential Iranian official.	2013-10-03 11:20:00.0	U.S. News - News
5/0.732136	Obama's Former Defense Chiefs Criticize Him on Syria	President's handling of Syrian crisis has emboldened his critics.	2013-09-18 07:26:00.0	U.S. News - News
6/0.705422	Putin Wins a Round From Obama	Russian president outmaneuvers Obama with Syrian diplomacy.	2013-09-12 07:04:00.0	U.S. News - News
7/0.690888	Athletes Face Challenges in Speaking Out Against Russia's Anti-Gay Laws	The IOC and Russian officials have discouraged athletes from making political statements at the Olympics.	2013-10-04 10:22:00.0	U.S. News - News
8/0.690864	Committee Passes Resolution to Strike Syria	Committee Approves Resolution to Intervene in Syria.	2013-09-04 14:56:00.0	U.S. News - News
9/0.67244	Liberal Group Tries to Make 'Flat Boehner' Happen	The group Americans United for Change got crafty.	2013-10-03 13:25:00.0	U.S. News - News
10/0.671224	Ron Paul Isn't Thrilled By 'Rather Annoying' Shutdown Debate	'The government doesn't really shut down,' says retired Texas congressman.	2013-10-01 16:04:00.0	U.S. News - News

[sxt\(10\)](#)

<u>Item Sim</u>	<u>Item Title</u>	<u>Item Description</u>	<u>Item Date</u>	<u>Channel Title</u>
11 0.67109	D.C. Navy Yard Shooting Kills At Least 13	Motive unclear but an alleged shooter is among the dead.	2013-09-16 17:36:00.0	U.S. News - News
12 0.668477	Kerry Arrives in Geneva as Rebels Condemn Russian Proposal	Opposition condemns Russian proposal as bilateral talks begin in Geneva.	2013-09-12 08:00:00.0	U.S. News - News
13 0.665041	Iranian Foreign Minister to Meet With Kerry at U.N. General Assembly	John Kerry among top delegates to meet with the Iranian foreign minister.	2013-09-23 12:59:00.0	U.S. News - News
14 0.661961	Obama Hopes for Syria-Russia Diplomatic Save	Russia says it will pursue effort to secure Syrian chemical weapons.	2013-09-10 11:59:00.0	U.S. News - News
15 0.65808	Joaquin Castro Explains How to Tell Him Apart From Twin Brother	The Castro in Congress talked Ted Cruz and Wendy Davis.	2013-10-02 14:13:00.0	U.S. News - News
16 0.655988	Israel accuses man of spying for Iran	Israel has arrested an Iranian-born man, accusing him of spying for Tehran.	2013-09-30 09:19:18.0	CNN.com - World
17 0.655645	Obama Heads to U.N. Assembly in New York City to Address Syria, Mideast Peace	Obama is open to meeting with Iranian president to discuss nuclear program.	2013-09-23 06:22:00.0	U.S. News - News
18 0.65492	In Arkansas Senate Race, Syria Emerges As Defining Issue	Sen. Pryor's decision to turn away from Obama on Syria, could help in Senate race.	2013-09-09 13:02:00.0	U.S. News - News
19 0.653322	In Montana, Democrats Might Have Found Their Guy	Montana Senate race not off the Democrats' map yet.	2013-09-16 06:00:00.0	U.S. News - News
20 0.649231	Rand Paul Pitches Mandatory Minimum Sentence Reform	Paul's stance could raise his stock with black voters in 2016.	2013-09-18 08:58:00.0	U.S. News - News

jack(0) Next(20)

<u>Item Sim</u>	<u>Item Title</u>	<u>Item Description</u>	<u>Item Date</u>	<u>Channel Title</u>
21 0.649009	Siege at Nairobi, Kenya, Mall Enters Third Day	'There will be no negotiations at Westgate,' extremist group al-Shabaab affirms.	2013-09-23 08:30:00.0	U.S. News - News
22 0.648359	Syrian Coalition Rejects Russian Proposal Without Prosecuting Assad	Syrian Coalition will not consider any proposal that does not bring the Assad regime to justice.	2013-09-10 14:03:00.0	U.S. News - News
23 0.643993	This Is How Ralph Nader Does Stand-Up Comedy	The consumer advocate tried out some laughs.	2013-09-26 14:48:00.0	U.S. News - News
24 0.642849	Hackers Attack NASA's Website to Protest NSA	Brazilian group targets incorrect U.S. agency.	2013-09-12 11:15:00.0	U.S. News - News
25 0.64196	Antonin Scalia Says Judicial Activism Could Be Worse	Outspoken jurist says Germans live under a more activist court.	2013-09-16 17:27:00.0	U.S. News - News
26 0.640648	Obama Muddles Through on Syria	Presidential address fails to resolve the inconsistencies of Obama's Syria policy.	2013-09-11 06:58:00.0	U.S. News - News
27 0.639472	What You Need to Know About Terrorism in East Africa	A guide to the groups in and around Somalia contributing to the ongoing war on terrorism.	2013-09-30 14:37:00.0	U.S. News - News
28 0.63946	Republicans, Democrats Prepare for Shutdown Blame Game	Republicans and Democrats aren't blinking ahead of government shutdown.	2013-09-30 10:34:00.0	U.S. News - News
29 0.638466	John McCain Uses Pravda to Bash Vladimir Putin	McCain says Putin's support of Bashar Assad hurts Russia.	2013-09-19 08:10:00.0	U.S. News - News
30 0.634741	It's Bud Light For Sen. Heidi Heitkamp	The senator is a favorite of the liquor industry.	2013-09-25 16:50:00.0	U.S. News - News

jack(10) Next(30)

Item/Sim	Item Title	Item Description	Item Date	Channel Title
31 0.633233	Freed Hostages Reveal Information on Chemical Attacks in Syria	Hostages share potentially new details about Syria's attacks.	2013-09-10 13:00:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
32 0.629276	Emmy 2013 Predictions: Miniseries or Movie	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 08:46:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
33 0.629276	Emmy 2013 Predictions: Drama	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-19 08:28:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
34 0.629276	Emmy 2013 Predictions: Comedy	Critics, fans and Emmy voters don't always agree who should go home with the big prize.	2013-09-18 06:00:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
35 0.629276	Emmy 2013 Predictions: Reality Competition and Variety	Emmy voters, critics and fans don't always agree who should go home with the big prize.	2013-09-20 10:37:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
36 0.628422	Obama Likely to Win War of Words Over Health Care	President presents united front, unlike his GOP opposition.	2013-09-24 06:36:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
37 0.626655	Lawmakers Creating a Blue Christmas	Washington's fiscal shenanigans may restrain holiday spending this year.	2013-10-03 15:35:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
38 0.625171	U.S. Draws Down Embassies in Lebanon and Turkey	A Syria strike may yield widespread Hezbollah retribution, a retired official warns.	2013-09-06 11:41:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
39 0.621873	Study: The American Family Is No More	A new study suggests immigrants now most closely resemble the 'typical' American family.	2013-09-11 15:12:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
40 0.621862	Poll: Biden, Kerry Beat Obama in Popularity	Public punishes Obama for Syria as well as IRS and Benghazi, Libya, controversies.	2013-09-16 07:14:00.0	U.S. News - News

[ack\(20\) Next\(40\)](#)

Item/Sim	Item Title	Item Description	Item Date	Channel Title
41 0.621473	Most States Spend Less Per Student Now Than Before the Recession	Most states are spending less per student this year than when the recession hit.	2013-09-12 13:33:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
42 0.620933	Lawsuit Demands Obama Fully Enforce Obamacare	Obama is 'lawlessly delaying' employer mandate, lawsuit claims.	2013-10-01 13:20:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
43 0.620277	India Sentences Convicted Rapists to Death	India Gives the Death Penalty	2013-09-13 11:30:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
44 0.6202	White House Adviser Susan Rice Invokes Motherhood to Justify Syrian Strikes	Former U.N. ambassador dismisses multilateral diplomatic solution as unrealistic.	2013-09-09 14:09:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
45 0.619875	Conservative Republicans Look to Delay Obamacare in CR	Republicans coalescing around plan to delay Obamacare.	2013-09-27 15:38:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
46 0.619119	Cummings Calls for AIG CEO Benmosche's Resignation	The Democratic congressman calls Benmosche's comments about bonuses and lynch mobs 'unbelievably appalling.'	2013-09-24 17:21:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
47 0.618684	Verizon Challenges Open Internet Rules in Court	Court hears arguments in Verizon lawsuit against net neutrality.	2013-09-10 14:15:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
48 0.618071	Pope Francis Announces Canonization of Pope John Paul II and Pope John XXIII Set for April 2014	The Vatican looks forward to lots of visitors for April's canonization ceremony.	2013-09-30 11:34:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
49 0.617087	Gunmen Attack Russian Embassy in Tripoli	No injuries as attackers attempt to storm Russian compound.	2013-10-02 17:04:00.0	U.S. News - News
Item/Sim	Item Title	Item Description	Item Date	Channel Title
50 0.616225	Lawmakers Put Pay Aside During Shutdown	Many lawmakers defer pay or donate it to charity.	2013-10-03 07:16:00.0	U.S. News - News

[ack\(30\) Next\(50\)](#)

Recuperación e integración de datos espaciales mediante alineación de ontologías

<u>Item/Sim</u>	<u>Item Title</u>	<u>Item Description</u>	<u>Item Date</u>	<u>Channel Title</u>
51/0.613274	Egyptian Court Bans Muslim Brotherhood	All assets to be seized from conservative Islamist party of former President Mohammed Morsi.	2013-09-23 12:58:00.0	U.S. News - News
52/0.612974	Obama Blitzes Press to Make Case for Syria; Assad Threatens Back	Syria decision threatens to color Obama's legacy.	2013-09-09 07:00:00.0	U.S. News - News
53/0.611869	Oy Vey! Anthony Weiner Unloads on Jewish Voter	Well, there goes the Jewish vote.	2013-09-04 13:56:00.0	U.S. News - News
54/0.611605	Assad 'Very Concerned' Missile Strikes Would 'Tip the Balance'	International leaders weigh in on military action in Syria.	2013-09-09 10:49:00.0	U.S. News - News
55/0.610983	NSA Surveillance Practices 'Violate the Constitution,' Issa Says	Team Amash anticipates a decisive victory in the next NSA-restraining vote.	2013-09-10 13:13:00.0	U.S. News - News
56/0.61042	Obama Accuses Republicans of 'Blackmail' for Pitting Obamacare Against Government Shutdown	Obama predicts public opinion shift in favor of Obamacare.	2013-09-26 13:07:00.0	U.S. News - News
57/0.609951	Assad Agrees to U.N. Chemical Weapons Treaty	Chief diplomat criticizes all sides in Syria following Assad's agreement to U.N. treaty.	2013-09-13 09:38:00.0	U.S. News - News
58/0.609309	TV's Best Fictional Government Shutdowns	Maybe D.C. politicians can take a word of advice from Leslie Knope and Leo McGarry.	2013-10-01 13:25:00.0	U.S. News - News
59/0.607804	Obama at U.N.: Mideast Peace, Preventing Nuclear Iran Are Top U.S. Goals	Obama asked for international support of diplomatic paths in Syria and Iran.	2013-09-24 10:48:00.0	U.S. News - News
60/0.60761	Bill de Blasio, Joe Lhota Win New York Mayoral Primaries	Weiner, Spitzer fall flat in post-scandal comebacks.	2013-09-11 06:44:00.0	U.S. News - News

lack(40) Next(60)

61/0.607251	White House Meeting Over, Obama Back to Blaming Shutdown on Boehner	Obama mocked Republicans for not knowing what they want.	2013-10-03 11:54:00.0	U.S. News - News
62/0.606893	Facebook CEO Talks Immigration Reform, Future of Social Networks	Mark Zuckerberg meets with lawmakers on immigration, wants government transparency.	2013-09-19 12:45:00.0	U.S. News - News
63/0.606737	Chris Christie May Be Too Popular for His Own Good	Can-do governor appeals to everyday folk, but angers hard-line conservatives.	2013-09-16 07:24:00.0	U.S. News - News
64/0.605597	Study: Anti-Bullying Programs May Have Opposite Effect	Students in schools with anti-bullying programs are more likely to become victims than their peers.	2013-09-13 11:34:00.0	U.S. News - News
65/0.605581	Obama Bids Public for Trust in Syria Plan, Be it Diplomatic or Military	The president pledged diligence in decision-making on Syria.	2013-09-10 21:12:00.0	U.S. News - News
66/0.605499	Ailing BlackBerry Finds a Buyer to Go Private	Buyout would shift BlackBerry from consumers to business, government.	2013-09-23 15:51:00.0	U.S. News - News
67/0.604539	Bill de Blasio, Joe Lhota Appear in Control in New York City Mayoral Primary Races	Candidates must get 40 percent to avoid a runoff.	2013-09-10 06:26:00.0	U.S. News - News
68/0.604488	Shaq Makes Michelle Obama Feel Short	Michelle Obama and NBA great Shaquille O'Neal teamed up to promote children's physical education.	2013-09-06 16:50:00.0	U.S. News - News
69/0.604306	Why Kerry Washington Got Involved in the New Jersey Governor's Race	The 'Scandal' star attended a campaign rally Sunday.	2013-09-09 13:10:00.0	U.S. News - News
70/0.603999	Obama Counting on Democratic Loyalty on Syria	President appeals to party faithful not to turn him into a lame duck.	2013-09-06 06:58:00.0	U.S. News - News

lack(50) Next(70)

