



Polibits

ISSN: 1870-9044

polibits@nlp.cic.ipn.mx

Instituto Politécnico Nacional

México

Loza-Pacheco, Eduardo; Torres-Ruiz, Miguel; Guzmán-Lugo, Giovanni
Razonamiento espacial para determinar el dominio de un conjunto de etiquetas que
representan objetos geográficos
Polibits, vol. 45, 2012, pp. 37-52
Instituto Politécnico Nacional
Distrito Federal, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=402640459006>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Razonamiento espacial para determinar el dominio de un conjunto de etiquetas que representan objetos geográficos

Eduardo Loza-Pacheco, Miguel Torres-Ruiz y Giovanni Guzmán-Lugo

Resumen—Actualmente, existe una gran cantidad de información geográfica, proveniente de diversas fuentes como imágenes de satélite, fotografías aéreas, mapas, bases de datos, entre otras. Estas fuentes proporcionan una descripción exhaustiva de los objetos geográficos. Sin embargo, la tarea de identificar el dominio geográfico al que pertenecen involucra un procesamiento semántico, el cual está basado en la conceptualización de un dominio, lo que permite interpretarlo de una manera similar a como los seres humanos reconocen a las entidades geográficas y evitar así la vaguedad. Este trabajo propone un método para realizar un proceso de razonamiento espacial cualitativo en representaciones geográficas. El método se basa en el conocimiento *a priori* del dominio, el cual se encuentra explícitamente formalizado a través de una ontología. El conocimiento descrito en la ontología se valora de acuerdo con un conjunto de etiquetas que pertenecen a algún tipo de dominio geográfico para realizar el análisis semántico correspondiente y mapear esas etiquetas con los conceptos definidos en la ontología. Como resultado, se obtiene un conjunto de dominios geográficos ordenados por su relevancia, para proporcionar un concepto general relacionado directamente con las etiquetas de entrada, simulando la forma en que cognitivamente percibimos algún dominio geográfico en el mundo real.

Palabras Clave—Inteligencia artificial y computacional, sistemas inteligentes, adquisición de conocimiento, representación de conocimiento.

Spatial Reasoning for Determining the Domain of the Set of Tags that Represent Geographic Objects

Eduardo Loza-Pacheco, Miguel Torres-Ruiz
and Giovanni Guzmán-Lugo

Abstract—Nowadays, there is much geospatial information from different sources such as satellite images, aerial photographs, maps, databases, and so on. It provides a comprehensive description of geographic objects. However, the task to identify the geographic domain to which it belongs is not simple, because this task involves semantic processing based on a conceptualization of a domain. It allows us to understand information in a way similar to how humans recognize the geographic entities and avoid vagueness. We propose a method for qualitative spatial reasoning in geospatial representations.

Manuscrito recibido el 9 de marzo de 2012, manuscrito aceptado el 7 de mayo de 2012.

Los autores trabajan en el Centro de Investigación en Computación del Instituto Politécnico Nacional, Av. Juan de Dios Bátiz, s/n, UPALM-Zacatenco, 07738, México D.F., México (email: eduardo.loza@gmail.com, mtorres@cic.ipn.mx, jguzman@cic.ipn.mx).

The method is based on *a priori* knowledge, which is explicitly formalized through an ontology. The knowledge described in the ontology is assessed according to a set of tags that belong to any geographical domain for semantic analysis, to map those tags to concepts defined in the ontology. As a result, a set of geographic domains in order of relevance is obtained, for providing a general concept directly related to the input tags, simulating the way in which humans cognitively perceive a geographic domain in the real world.

Index Terms—Computational and Artificial Intelligence, Intelligent Systems, Knowledge Acquisition, Knowledge Representation.

I. INTRODUCCIÓN

HOY en día, los Sistemas de Información Geográfica (SIG) se han convertido en herramientas muy populares para la representación y razonamiento de datos geográficos [1]. Estas aplicaciones requieren de métodos de razonamiento acerca de las entidades geográficas y de las relaciones que se definen entre sí, incluyendo todas aquellas relaciones espaciales [2]. Los algoritmos de razonamiento son ampliamente utilizados en el campo de la Inteligencia Artificial, cuyas tareas más relevantes son la capacidad de verificar la consistencia de los conjuntos de datos, actualizar el conocimiento compartido, derivar nuevo conocimiento y encontrar una representación mínima [3, 4].

No obstante, antes de realizar alguna tarea de razonamiento, es necesario tomar en cuenta una representación formal que nos permita conceptualizar el conocimiento del dominio de interés [5]. En este caso, las ontologías son herramientas muy poderosas para conceptualizar cualquier contexto, describiendo sus conceptos y expresando sus relaciones. Además, las ontologías han sido referidas como un método para llevar a cabo este tipo de razonamiento [6–8]. Sin embargo, existen algunas metodologías que no manejan adecuadamente la vaguedad inherente de los datos espaciales.

Las entidades geográficas son frecuentemente dependientes del contexto en el cual residen, con un conocimiento local que afecta a las definiciones [9]. Asimismo, los objetos geográficos no demarcan claramente en algunas ocasiones a una entidad, ya que pueden formar parte de algún otro objeto [10]. Por lo tanto, la individualización de las entidades es más importante, con respecto a los dominios geográficos que representan o pueden pertenecer.

De acuerdo con [11], la *vaguedad* es inherente a los dominios geográficos, con relación a muchos elementos que

son dependientes del contexto, así como carentes de definiciones y límites precisos. La vaguedad no es un defecto de nuestro lenguaje de comunicación sino más bien una parte útil e integral. Como una consecuencia, los SIG no manejan adecuadamente múltiples interpretaciones, por lo cual la carencia de esta característica implica la creación de nuevas técnicas que permitan el manejo de múltiples significados, así como la inferencia basada en razonamiento.

En este trabajo se propone un método que permite llevar a cabo un proceso de razonamiento espacial cualitativo, sobre un conjunto de objetos geográficos que están representados como etiquetas de entrada y pertenecen a algún dominio geográfico. Para el proceso de inferencia se proponen tres algoritmos que realizan un razonamiento espacial, los cuales consideran el conocimiento *a priori*, el cual se define por medio de una ontología de aplicación y marcos conceptuales. El proceso de razonamiento se basa fundamentalmente en el procesamiento de las relaciones topológicas, las cuales se encargan de describir la interacción o comportamiento de un objeto geográfico con respecto a otros.

Este artículo está organizado de la siguiente manera: en la Sección 2 se presenta el estado del arte con respecto a los trabajos más relevantes en esta área. Asimismo, la Sección 3 describe la metodología propuesta para llevar a cabo el razonamiento espacial cualitativo. La Sección 4 muestra los resultados obtenidos. Por último, las conclusiones son presentadas en la Sección 5.

II. TRABAJOS RELACIONADOS

La geometría de puntos y líneas es una de las más ancestrales ramas del razonamiento espacial [12]. La abstracción de un punto sin dimensión es el elemento básico de todas las entidades espaciales que tienen que ser construidas con base en los puntos. Una de las teorías más antiguas es la Geometría Euclidiana, cuyo sistema axiomático se utiliza hoy en día. La idea de todo el razonamiento geométrico puede estar basada, en las aplicaciones orientadas a razonamiento matemático.

Los trabajos de razonamiento espacial cualitativo son precedidos por un conjunto de representaciones espaciales, en donde se busca que puedan ser leídos e interpretados por una máquina [8]. En [13], se menciona la importancia de que exista una correcta representación de la realidad para llevar a cabo un proceso de razonamiento espacial, ya que una de las razones para representar el conocimiento acerca de un dominio más que el mismo dominio en sí mismo, es a través del conocimiento, por lo cual el mundo sería mucho más accesible por medios formales. Esto se debe a que las máquinas están sujetas al uso de enfoques formales si se quiere realizar algún proceso de razonamiento. Sin embargo, la información capturada debe contener descripciones lo más cercanas a como el ser humano percibe su entorno [Egenhofer, 1995]. Además de que uno de los principales objetivos del razonamiento espacial cualitativo es encontrar maneras adecuadas de representar propiedades continuas del mundo, utilizando un sistema basado en símbolos discretos [14], [15].

Las representaciones espaciales cualitativas han estado evolucionado rápidamente, esto se puede observar en [16], en donde se propone un conjunto de relaciones binarias $C(x,y)$, la cual se puede leer como “x conecta y”, y se demuestra que esta relación cumple con las propiedades de simetría y reflexividad. Este tipo de relaciones han sido definidas para el trabajo con regiones espaciales, en donde se puede presentar la mayor ambigüedad entre las entidades geográficas. Para el caso del *razonamiento espacial temporal*, en [17], se propone un conjunto de relaciones que se encargan de capturar el comportamiento entre dos intervalos, siendo este conjunto explícito muy utilizado actualmente en los modelos espacio-temporales. En la Figura 1 se muestra el conjunto de las relaciones de Allen.

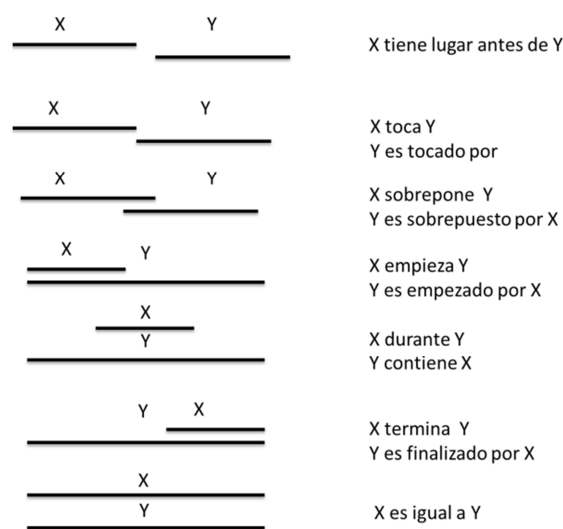


Fig. 1. Relaciones temporales propuestas por Allen.

El uso de estas relaciones se extiende más allá del razonamiento espacial temporal, como ha sido descrito en trabajos como [18], donde este conjunto de relaciones puede expandirse a un dominio \mathbb{R}^2 . Por otro lado, existen nuevos modelos de representación como el modelo 9-intersección de Egenhofer [19], el cual descompone una región en términos de su interior y frontera. Este modelo define un conjunto de relaciones espaciales topológicas entre dos regiones que se describen por una tupla de cuatro elementos, donde el interior se denota por el símbolo ($^\circ$) y la frontera se define por (∂). Las relaciones espaciales topológicas se denotan por una tupla de cuatro elementos $\langle a,b,c,d \rangle$. Las entradas corresponden al orden de los valores topológicos invariantes, asociados a las cuatro intersecciones, las cuales se denominan *frontera-frontera*, *interior-interior*, *frontera-interior* y la cuarta *interior-frontera*. Además, se restringen las relaciones espaciales topológicas definidas por los valores vacío (\emptyset) y no-vacío ($\neg\emptyset$) a las entradas de la tupla [20].

Asimismo, se ha propuesto el modelo RCC8, el cual es una representación del espacio que provee un conjunto de ocho relaciones topológicas entre dos regiones [21]. En la Figura 2 se muestra el conjunto de relaciones propuesto. Este modelo

describe las relaciones: dentro, afuera, toca, sobrepone, contiene, cubre, disjunto e igual para objetos que se definen como conjuntos regulares. Por tanto, el conjunto RCC8 cuenta con una semántica bien formada y basada en la Geometría Euclidiana [22].

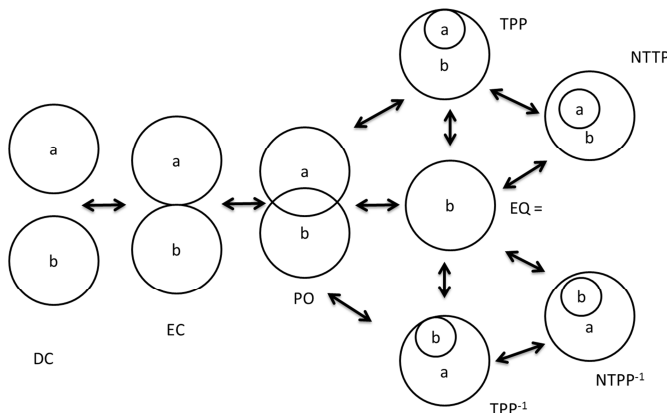


Fig. 2. Modelo de relaciones topológicas RCC8.

Por otra parte, existen diversas propuestas para motores de razonamiento espacial, entre ellas se encuentra el trabajo de El-Geresy [23], el cual propone SPARQS. Este modelo es un motor de razonamiento espacial cualitativo, cuyo enfoque se basa en la generación de relaciones espaciales que se definen mediante tablas de composición de manera automática. El formalismo presentado en este trabajo está dividido en dos partes que consisten en las restricciones generales que rigen a las relaciones espaciales entre los objetos en el espacio y las reglas generales para propagar las relaciones entre los objetos. Tanto las restricciones y las reglas están basadas en una representación uniforme de la topología de los objetos, el espacio que lo rodea y la representación de las relaciones entre los objetos. Un trabajo similar es el propuesto por Grutter [24], el cual propone un conjunto de reglas y restricciones a partir del modelo RCC8. Asimismo, Wang [25] propone nuevos modelos de relaciones espaciales que agrupan a las relaciones principales (dirección, topología, distancia y forma), con la finalidad de implementar una aplicación que permita recuperar esta información en una base de datos. Por otra parte, Eagleson en [26], enumera las desventajas de las fronteras en los mapas geográficos y en especial con relación a las fronteras políticas. La utilidad de definir las de manera adecuada para sus posibles usos en razonamiento espacial jerárquico utilizando agregaciones espaciales.

En el caso de las relaciones de dirección, en [Frank, 1996] se describe que los SIG y en Geografía en general, se tiene que lidiar de manera común con espacios de gran escala, y es ahí donde las relaciones espaciales de dirección tienen su mayor utilidad, puesto que éstas son utilizadas casi exclusivamente para estos casos. De igual forma, se han planteado trabajos que utilizan el razonamiento espacial enfocado en aplicaciones como la navegación, robótica y planificación de rutas [27]. Este trabajo consiste en recibir un conjunto de descripciones del arreglo espacial de los objetos, en la forma de un lenguaje

definido por el usuario, con el propósito de que sea entendible. Además, se utiliza un subconjunto de relaciones espaciales que son las relaciones de dirección (arriba, abajo, derecha, izquierda).

La propuesta de Clementini [28], tiene como objetivo definir una taxonomía para emplearse en aplicaciones de razonamiento espacial. Este método define tres tipos de marcos de referencia, los cuales son: el *intrínseco* que es establecido con un objeto de referencia que determina el origen del sistema de coordenadas, así como la orientación. El *extrínseco* que puede heredar su origen de un objeto de referencia; sin embargo, su orientación se determina por factores externos como podría ser un objeto convencional o una marca. Finalmente, el *diético* que envuelve los tres objetos, el objeto primario que es una relación particular con respecto al objeto de referencia y el punto de vista. La relación es impuesta en el objeto de referencia de acuerdo con su punto de vista.

Existen varios trabajos que han sugerido la combinación de diversas relaciones geográficas como el descrito en [8], el cual combina las relaciones topológicas y las de dirección, o bien propuestas que combinan las relaciones de Allen, utilizadas normalmente en \mathbb{R}^1 para proyectarlas en \mathbb{R}^2 y usar esta forma de representación para relaciones topológicas y de dirección [18]. Otros esfuerzos están orientados directamente a combinar representaciones [29], en donde se aplican técnicas de razonamiento espacial para solucionar problemas donde existe una cantidad pequeña de datos y cuyos formalismos utilizan el modelo RCC8 y el álgebra de Allen. Por otro parte, están los problemas con una gran cantidad de información por ejemplo, la información que se recaba para los servicios meteorológicos. Por lo anterior, es necesario el uso de agregaciones espaciales. En este sentido, Shultz [30] utiliza la combinación de dos modelos el de proximidad cualitativa y el RCC, con la finalidad de determinar la distancia cualitativa entre dos objetos. Utilizando el modelo RCC, el primer paso es determinar la relación entre los objetos geográficos. Después utilizando distancias cualitativas, se toman las distancias entre cada par de objetos, mediante la fórmula de distancia entre dos puntos para un espacio en dos dimensiones. Aquí se combinan los dos enfoques el razonamiento espacial cuantitativo y el cualitativo.

Actualmente han surgido aplicaciones orientadas a la recuperación de imágenes para comprender que tipo de información envuelve a una imagen [31], [32]. Asimismo, es necesario entender como se conceptualiza una imagen o bien que información contiene y si es posible definirla, en caso contrario se pueden presentar problemas relacionados con la ambigüedad [33]. En [34] se menciona que una imagen es una descripción ontológica de sí misma. Como es el caso de [35], cuyo trabajo está enfocado en la obtención de un formalismo que sea lo suficientemente expresivo para incrementar la utilidad de una descripción pictográfica. Se menciona que los mecanismos de recuperación deben trabajar con información textual, imágenes, sonidos, videos, etc. Además se establece que existen dos tipos de recuperación de imágenes. El primero es del tipo sintáctico, donde las imágenes son guardadas en

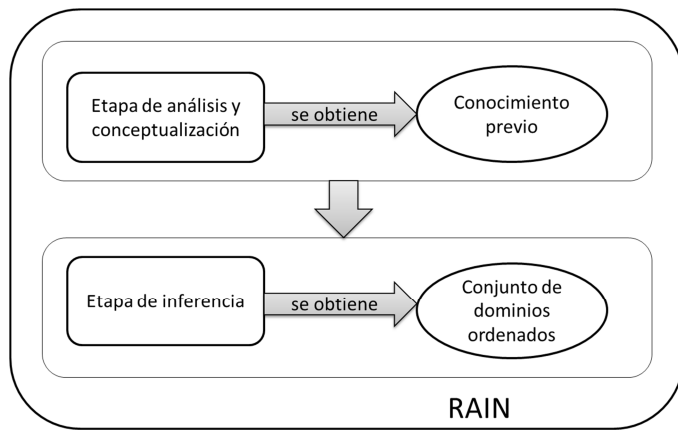


Fig. 3. Marco general de la metodología RAIN.

una base de datos, tomando como criterio algún valor físico de ésta. El segundo tipo es el enfoque basado en su significado, se propone que la imagen sea descrita en términos del modelo RCC8 para expresar las relaciones topológicas entre los objetos [35].

En [36] se presenta una metodología para la recuperación de imágenes para la búsqueda en grandes colecciones de datos. La aproximación propuesta emplea un algoritmo de segmentación no supervisado, el cual permite dividir la imagen en regiones. A partir de cada una de estas regiones, se hace una extracción de características de bajo nivel como son: color, posición, tamaño y forma, con el objeto de emplear un conjunto de descriptores que permitan describir dichas regiones, asociando a cada descriptor un vocabulario simple conocido como “objeto ontológico”. En [37], se propone utilizar una herramienta más semántica, la cual pretende realizar una búsqueda y anotación semántica dentro de una colección de imágenes, empleando ontologías como la *Art and Architecture Thesaurus* (AAT), *WordNet*, *Union List of Artist Names* (ULAN) e *IconClass*. Finalmente, en [38] se presenta una técnica basada en razonamiento espacial, la cual está compuesta por dos partes: la extracción de los objetos de la imagen remota y la construcción de una representación gráfica que sirve para realizar una etapa de razonamiento. Para la parte de extracción de objetos se propone una segmentación multiespacial de algoritmos de segmentación cóncavo, convexo, luminoso y oscuro e imágenes a diferente escala. Posteriormente, se utiliza el modelo RCC8 para construir un grafo basado en las relaciones entre las regiones y se construye un árbol de decisión binario.

III. METODOLOGÍA RAIN

RAIN es una técnica cuyo enfoque principal consiste en establecer un conjunto de técnicas para realizar un proceso de razonamiento espacial, en descripciones semánticas del contexto geográfico, tomando en consideración el conocimiento *a priori* del dominio geoespacial. Para ello, se formaliza este conocimiento por medio de una estructura conceptual (ontología), con la finalidad de que sea legible para una máquina.

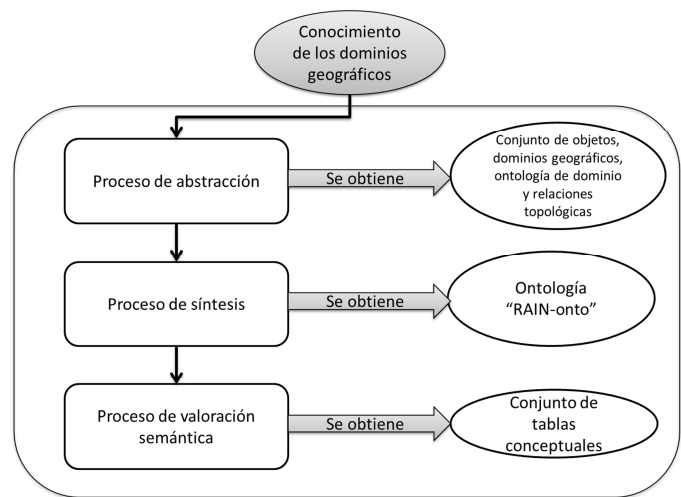


Fig. 4. Etapa de análisis y conceptualización.

Este método está orientado en la conceptualización de los conceptos y las relaciones que existen en un dominio dado, con la finalidad de responder a una pregunta sobre a qué dominio pertenecen un conjunto de descripciones y cuál es la relevancia de los conceptos en ese dominio. RAIN permite conocer el contexto de un conjunto de descripciones que a simple vista parecen inconexas. Esta técnica se compone de dos etapas: 1) *Análisis y conceptualización* y 2) *Inferencia*. En la primera, se obtiene un conocimiento previo, definido con base en las necesidades del razonamiento. En la etapa de Inferencia se obtiene un conjunto de dominios ordenados, de acuerdo con la cercanía o similitud de las descripciones recibidas como entrada. En la Figura 3 se muestra el marco general de la metodología RAIN.

A. Etapa de análisis y conceptualización

El proceso de conceptualización (ver Figura 4) está basado en el método para diseñar ontologías del dominio geográfico GEONTO-MET, propuesto en [39].

Con base en esta técnica, se han utilizado los dos conjuntos de relaciones axiomáticas base que han sido definidos: $A_1 = \{es, tiene, hace\}$ y $A_2 = \{preposiciones\}$ para traducir directamente las relaciones entre los conceptos como parte de la conceptualización; es decir, la esencia fundamental es reducir las relaciones axiomáticas en la ontología, con ello por ejemplo, relaciones topológicas como conecta, cruza, contiene, entre otras; son definidas como conceptos del tipo relación en la conceptualización, con lo cual se obtiene una mayor expresividad, granularidad y riqueza semántica en la representación [40]. Asimismo, estas relaciones axiomáticas son utilizadas para definir a los conceptos y clases en la ontología.

Proceso de abstracción: Esta tarea tiene como propósito realizar una revisión exhaustiva sobre los objetos geográficos que se encuentran involucrados en el conjunto de dominios geográficos, con la finalidad de llevar a cabo un proceso de abstracción que defina un conocimiento *a priori*. Para esta tarea la información fue recopilada de una ontología de dominio *Kaab* (significa Tierra en el lenguaje maya), definida

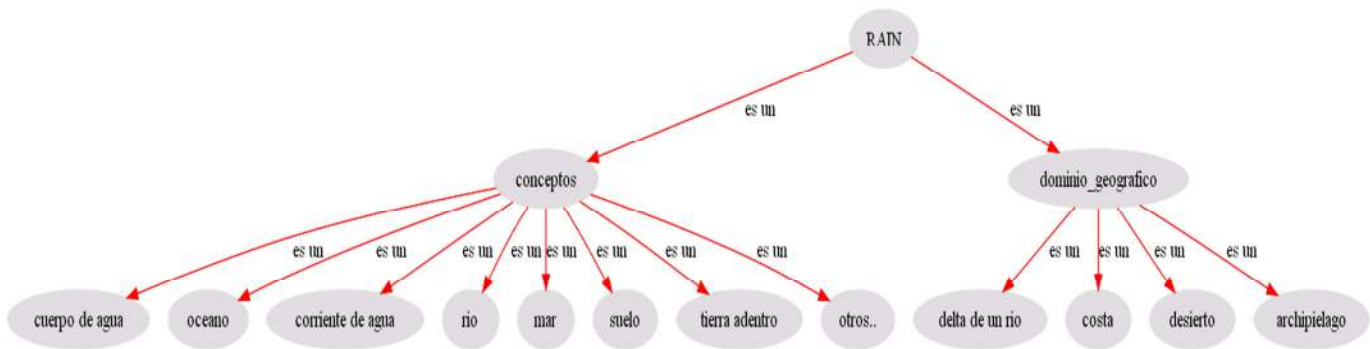


Fig. 5. Fragmento de la ontología RAIN.

en [40], cuya ventaja radica en que se pueden encontrar entidades abstractas y sus relaciones [39]. Se utilizaron definiciones del diccionario de datos del Instituto Nacional de Estadística, Geografía e Informática (INEGI) [41] y del Centro Nacional para Ontologías Biomédicas [42]. Como resultado de esta etapa, se obtuvo la información necesaria de todos aquellos objetos y dominios geográficos que nos interesa identificar en la etapa de razonamiento o inferencia. En la Figura 5 se muestra la información recabada y un fragmento de la ontología construida a partir de *Kaab*.

Proceso de síntesis: Para el proceso de síntesis ya se cuenta con los objetos y dominios que intervienen en el proceso de razonamiento. Sin embargo, esta información no se encuentra estructurada.

Por tanto, es necesario definir cada dominio, de acuerdo con la pertenencia y relación de los objetos geográficos que intervienen en él.

Entonces, se define la *relación topológica* que existe entre los objetos geográficos, la cual se representa por medio de una relación jerárquica descrita entre ellos, junto con sus propiedades y sinónimos o *alias*, tanto de los dominios como de los objetos geográficos.

En este proceso se lleva a cabo un *mapeo* de los objetos geográficos con respecto a los conceptos definidos en la ontología, con la finalidad de iniciar el poblado de la estructura conceptual.

La ontología *Kaab* cuenta con un conjunto de entidades abstractas del dominio geográfico que ayudan a delimitar el dominio y restringir el número de conceptos que están involucrados en cada uno.

En la Figura 6 se muestra el proceso de síntesis, en donde se puede observar que el conjunto de conceptos y relaciones topológicas permiten la definición del conjunto de dominios, los cuales interactúan con las ontologías para extraer las instancias que generalizarán el proceso; es decir, a partir de la descripción de muchos conceptos especializados, se obtiene un concepto general que describe al dominio en cuestión. Asimismo, en la Figura 7 se presenta el proceso de mapeo de la información recabada con la ontología *Kaab*.

Proceso de valoración semántica: El proceso de valoración semántica utiliza la ontología propuesta, junto con un conjunto de tablas para obtener la información con respecto a la

definición de conceptos, sus relaciones con otros conceptos como con los dominios a las que pertenecen, sus reglas y restricciones. Esto con la finalidad de refinar y asignar un valor semántico a cada dominio, con base en la construcción de las tablas de conceptos que contienen las propiedades, lugar en la jerarquía, nombres y sinónimos; así como una tabla de sinónimos de los conceptos, una tabla de dominios, una tabla de sinónimos de los dominios, una tabla de frecuencia de conceptos en los dominios, una tabla de composición de relaciones topológicas ordenadas, de acuerdo con su relevancia y por último, una tabla de refinación semántica para mejorar el proceso de inferencia, basada en la retroalimentación. A continuación se definen las tablas mencionadas para la valoración semántica.

Tabla de conceptos: En esta tabla aparecen todos los posibles conceptos que forman parte de cada uno de los dominios. Esta tabla permite expresar de manera explícita las características que contiene cada uno de los conceptos. La tabla contiene la información acerca de los sinónimos o de los *alias* con que se conoce a los conceptos y sus propiedades. Adicionalmente, permite conocer la ubicación de cada concepto en la jerarquía, identificando los conceptos padre e hijo de cada concepto. Cada concepto puede tener n número de hijos; sin embargo, solo puede tener un padre.

Asimismo, se expresan las relaciones que se pueden operar sobre ese concepto. Finalmente, se obtiene una ontología con todo el conjunto de propiedades, conceptos que contienen los dominios y la jerarquía de cada concepto (ver Figura 8). De

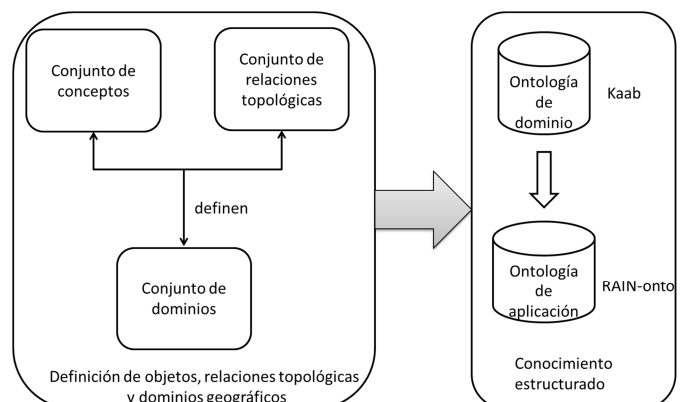


Fig. 6. Representación del proceso de síntesis.

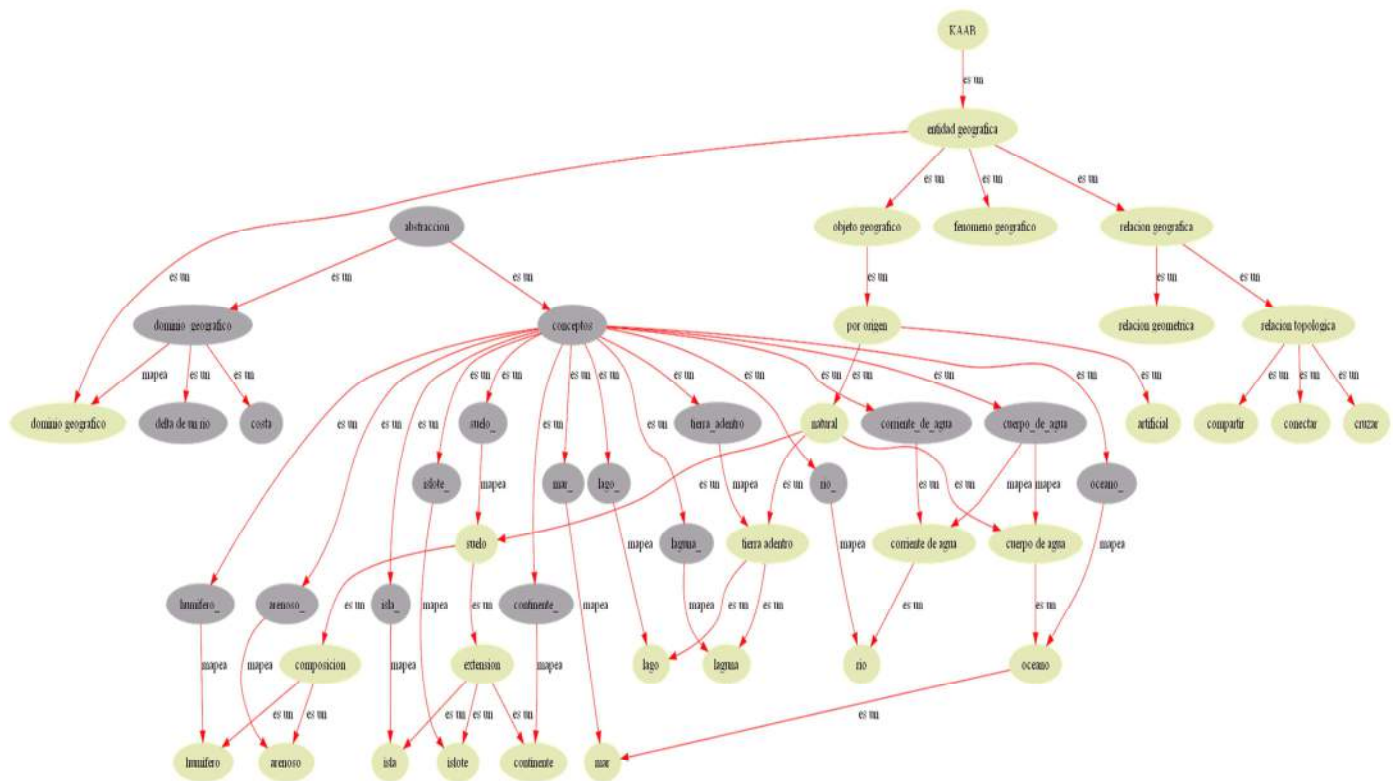


Fig. 7. Proceso de mapeo.

igual forma, se describen las definiciones para la generación de la tabla de conceptos (ver tabla I).

Para una tabla de conceptos, sea un concepto $c_{i,j}$ que pertenece al conjunto C , por lo cual se propone la función $father(c_{i,j})$ que devuelve el concepto $c_{p,q}$, definiendo como el padre de $c_{i,j}$. Por lo tanto, $father(c_{i,j}) = c_{p,q}$ es definida.

Asimismo, sea un concepto $c_{i,j}$ relacionado con la función $son(c_{i,j})$ que devuelve el (los) concepto (s) $c_{p,q}$ que es (son) hijo (s) de $c_{i,j}$, por tanto, la función queda definida como sigue: $son(c_{i,j}) = c_{p,q}$.

De igual manera, sea $P_{i,j}$ el conjunto de las propiedades que pertenecen directamente a un concepto $c_{i,j}$, de un dominio en particular $d_i | d_i \in D$, definido como se expresa a continuación: $P_{i,j} = \{p(i,j,1), p(i,j,2), \dots, p(i,j,n)\}$.

Con respecto a las relaciones, sea $R_{i,j}$ un conjunto de relaciones topológicas que aplican a un concepto $c_{i,j}$ de un dominio en particular $d_i | d_i \in D$, definido de la siguiente forma: $R_{i,j} = \{r(i,j,1), r(i,j,2), \dots, r(i,j,n)\}$.

TABLA I
TABLA DE CONCEPTOS

Nombre Concepto	Sinónimo	Propiedad	Padre	Hijo	Relación Topológica
$c_{i,1}$	$S_{i,1}$	$P_{i,1}$	$father(c_{i,1})$	$son(c_{i,1})$	$R_{i,1}$
$c_{i,2}$	$S_{i,2}$	$P_{i,2}$	$father(c_{i,2})$	$son(c_{i,2})$	$R_{i,2}$
...
$c_{i,n}$	$S_{i,n}$	$P_{i,n}$	$father(c_{i,n})$	$son(c_{i,n})$	$R_{i,n}$
...
$c_{m,n}$	$S_{m,n}$	$P_{m,n}$	$father(c_{m,n})$	$son(c_{m,n})$	$R_{m,n}$

De acuerdo con lo anterior, la función $exist_concept$, recibe una etiqueta et_i , la cual regresa como resultado el número de concepto de interés o relevancia. En caso contrario, devuelve un valor de falso si no existe el concepto. Esta función se define de la siguiente forma:

$$exist_concept(et_i) = \begin{cases} c_j & | 0 \leq j \leq n \\ falso & \end{cases}$$

De igual forma, para la generación de la tabla de conceptos, es necesario definir la función $exist_synonymous$, la cual recibe una etiqueta, regresando como resultado el número de concepto de interés o relevancia. En caso contrario, devuelve un valor de falso cuando no existe algún sinónimo.

$$exist_synonymous(et_i) = \begin{cases} c_j & | 0 \leq j \leq n \\ falso & \end{cases}$$

Por lo tanto, de acuerdo con las definiciones anteriores la tabla de conceptos se define como sigue:

Tabla de frecuencia de conceptos en dominios: De acuerdo con la información recopilada de cada dominio que fue definido previamente, existe un conjunto de conceptos pertenecientes a un dominio en particular, con la finalidad de obtener la frecuencia de los conceptos en los dominios geográficos.

Por tanto, sea D el conjunto de dominios, donde d es un dominio geográfico en particular que ha sido definido en una conceptualización. Entonces ese conjunto se define de la siguiente forma: $D = \{d_1, d_2, \dots, d_n\}$.



Polibits (45) 2012

Sea la función *concept*, la cual recibe dos parámetros que describen el número de dominio y el número de concepto, regresando como resultado el número de concepto de interés o falso en caso de que no exista el concepto. Por tanto, la función es: $concept(d, c) = c_{j,k} | 0 \leq j \leq m, 0 \leq k \leq n$.

Tabla de sinónimos de dominios: Una tabla de sinónimos, permite conocer los posibles *alias* que puede tener un dominio en particular, relacionado directamente con su búsqueda (ver Tabla III). Para ello, sea S_i el conjunto de sinónimos para un determinado dominio d_i , el cual está definido por: $S_i = \{s_{(i,1)}, s_{(i,2)}, \dots, s_{(i,n)}\}$.

Tabla de composición de relaciones topológicas: Como parte del procesamiento semántico, es necesario obtener el conjunto de relaciones existentes entre los conceptos que intervienen directamente en cada dominio y definir su relevancia para

Dominio	Conceptos
D_1	$C_1 = \{c_{(1,1)}, c_{(1,2)}, \dots, c_{(1,n_1)}\}$
D_2	$C_2 = \{c_{(2,1)}, c_{(2,2)}, \dots, c_{(2,n_2)}\}$
D_3	$C_3 = \{c_{(3,1)}, c_{(3,2)}, \dots, c_{(3,n_3)}\}$
\dots	\dots
D_m	$C_m = \{c_{(m,1)}, c_{(m,2)}, \dots, c_{(m,n_m)}\}$

Dominio	Conceptos
D_1	$S_1 = \{s_{(1,1)}, s_{(1,2)}, \dots, s_{(1,n_1)}\}$
D_2	$S_2 = \{s_{(2,1)}, s_{(2,2)}, \dots, s_{(2,n_2)}\}$
D_3	$S_3 = \{s_{(3,1)}, s_{(3,2)}, \dots, s_{(3,n_3)}\}$
\dots	\dots
D_m	$S_m = \{s_{(m,1)}, s_{(m,2)}, \dots, s_{(m,n_m)}\}$

TABLA IV
TABLA DE COMPOSICIÓN DE RELACIONES TOPOLÓGICAS

Dominio	Etiquetas de Relevancia		
	Necesario	Común	Raro
D_{t1}	N_{D1}	CO_{D1}	RA_{D1}
D_{t2}	N_{D2}	CO_{D2}	RA_{D2}
D_{t3}	N_{D3}	CO_{D3}	RA_{D3}
...
D_{ti}	N_{Di}	CO_{Di}	RA_{Di}

cada dominio. Por tal motivo, se propone la generación de una tabla de composición de relaciones topológicas, la cual se encuentra ordenada con base en su relevancia o importancia dentro de algún dominio en particular (ver la tabla IV).

Para este caso, sea r_t una relación de la tabla de composición integrada por los conceptos c_i & $c_j \in C$ & $r_t \in R$, denotada por $r_t = c_i r_i c_j$.

Entonces, sea N el conjunto de relaciones necesarias presentes en un dominio geográfico i , en donde la presencia de estas relaciones en el dominio indican que estos conceptos siempre están ligados de esta forma e interactúan en un dominio en particular, de la forma: $N_{Di} = \{r_{tN1}, r_{tN2}, \dots, r_{tNn}\}$.

Para este caso, entonces sea C un conjunto de relaciones comunes presentes en un dominio geográfico i . La presencia de esta relaciones en el dominio nos indica que estos conceptos son comunes encontrarlos en este dominio en particular, de acuerdo con: $C_{Di} = \{r_{tC1}, r_{tC2}, \dots, r_{tCn}\}$.

Para el caso de las relaciones, sea RA un conjunto de relaciones “raras” presentes en un dominio geográfico. La presencia de estas relaciones en el dominio indican que es poco probable encontrarlas en ese dominio; sin embargo, es posible encontrarlas en algunas ocasiones y en algunos dominios. Por lo tanto, el conjunto se define como sigue: $RA_{Di} = \{R_{tRA1}, R_{tRA2}, \dots, R_{tRAn}\}$.

El dominio de la tabla de composición D_{ti} está compuesto por la unión de los conjuntos N_{Di} , CO_{Di} y RA_{Di} que se denota por: $D_{ti} = \{N_{Di} \cup CO_{Di} \cup RA_{Di}\}$.

Después de obtener la tabla de composición, se procede a ordenar los conceptos junto con sus relaciones topológicas, de acuerdo con su relevancia en el dominio. Esto indica que las relaciones del conjunto N son una parte fundamental para definir el dominio. En tanto, las relaciones comunes no lo son, por lo que éstas tendrán una relevancia menor en la definición del dominio, por lo que no es requisito contar con estas relaciones. De igual forma, las relaciones del conjunto RA , no son indispensables en la definición del dominio geográfico, además de que su relevancia está por debajo del conjunto C y además tiene menor relevancia que el conjunto N .

Para este caso, se define la función *relevant*, la cual recibe tres parámetros, una triplete de conceptos y un dominio. Estos elementos describen el número de dominios y el número de conceptos en el caso de la triplete, regresando como resultado la relación de relevancia de la triplete de nuestro interés o falso en caso de no existir una relación de relevancia para el dominio determinado, lo cual se define de la siguiente forma: $relevant(d, c_a, c_b) = r_{tj} | 0 \leq j \leq m$.

TABLA V
TABLA DE REFINACIÓN SEMÁNTICA

Etiquetas de Entrada	Dominios de Salida
$E_1 = \{e_{11}, e_{12}, e_{13}, \dots, e_{1n}\}$	$Salida_1 = (D)$
$E_2 = \{e_{21}, e_{22}, e_{23}, \dots, e_{2n}\}$	$Salida_2 = (D)$
...	...
$E_m = \{e_{31}, e_{32}, e_{33}, \dots, e_{3n}\}$	$Salida_m = (D)$

Tabla de refinación semántica: Se define para almacenar primeramente las etiquetas que se reciben como entrada, con la finalidad de que se validen en algunos de los algoritmos propuestos en la *etapa de inferencia*, con el propósito de asignar el dominio al que pertenece, siempre y cuando exista una validación por parte del usuario (ver la tabla V).

De acuerdo con lo anterior, sea *return_ref* una función que recibe el conjunto etiquetas de entrada $E_i = \{e_{11}, e_{12}, \dots, e_{1n}\}$ que describen semánticamente ya sea uno o varios conjuntos de dominios en particular, regresando como resultado un conjunto de dominios de salida ordenados, de acuerdo con la cercanía definida por su *similitud semántica* con respecto a la descripción de las etiquetas de $Salida_i(D)$. Por tanto, la función se define de la siguiente manera: $return_res(E_1) = Salida_i(D) | 0 \leq i \leq m$.

B. Etapa de inferencia

La etapa de *inferencia* está compuesta por cuatro tareas que interactúan de manera conjunta para analizar, describir y deducir a qué dominio pertenece un conjunto de objetos geográficos representados por etiquetas. La primera tarea consiste en establecer un método de mapeo, el cual recibe como entrada un conjunto de etiquetas. Con este conjunto se busca si existe un concepto definido en la ontología y que esté directamente relacionado con esas etiquetas. Posteriormente, con este mapeo se obtiene un conjunto de conceptos que están relacionados con la base de conocimiento definida en la ontología.

La segunda tarea consiste en la definición de un conjunto de *técnicas de razonamiento espacial cualitativo*, descrita por tres algoritmos definidos para tal fin, a saber: 1) algoritmo de frecuencias, 2) algoritmo de relevancia y 3) algoritmo de genealogía semántica. Con estos algoritmos se lleva a cabo el proceso de inferencia, tomando como entrada un conjunto de conceptos descritos en la base de conocimiento (ontología).

La tercera tarea consiste en definir un *selector de salida*, el cual se encarga de proporcionar la salida de uno de los algoritmos de razonamiento, de acuerdo con el grado de efectividad que se ha definido para cada uno. Este grado de efectividad varía con la interacción del usuario. Finalmente, la tarea de *refinación semántica* consiste en brindar una respuesta al usuario, con base en un proceso de validación, en donde la retroalimentación juega un papel preponderante para examinar la riqueza semántica de la representación. La inferencia obtenida se almacena en la base de conocimiento para formar parte de la conceptualización y determinar el grado de efectividad del algoritmo de razonamiento empleado. En la Figura 9 se muestra el marco de trabajo y los elementos que forman parte de la etapa de inferencia.

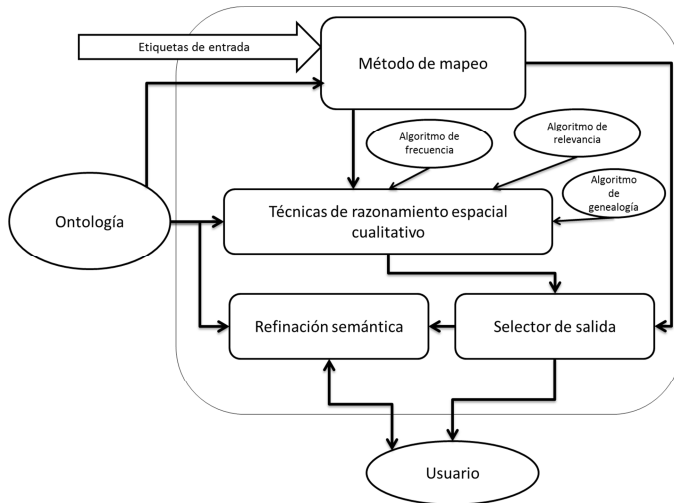


Fig. 9. Marco de trabajo de la etapa de inferencia.

Método de mapeo: El método de mapeo es un algoritmo diseñado para recibir las etiquetas de entrada y localizar si existen conceptos relacionados con esas etiquetas para almacenarlas en la ontología como *instancias* o *individuos* de un concepto clase (proceso de síntesis). Sin embargo, antes de esto es necesario verificar si es que ya existe un resultado previo para este conjunto de etiquetas, utilizando la tabla de refinación semántica. Como salida de este método se obtiene un conjunto de entradas existentes dentro del conocimiento previo o bien el conjunto de dominios D ordenado. En caso de que las etiquetas ya hayan sido agregadas anteriormente y validadas por un usuario, se almacenan como parte de la conceptualización (ontología O). A continuación se presenta el algoritmo referente al método de mapeo propuesto (ver la tabla VI).

TABLE VI
ALGORITMO PROPUESTO PARA EL MÉTODO DE MAPEO

Entrada:

Se tiene como entrada un conjunto de etiquetas que representan objetos geográficos: $EtE = \{ete_1, ete_2, \dots, ete_n\}$, una tabla T_c de conceptos y una tabla de refinación semántica T_{re} .

Salida:

Un vector de mapeo VM conformado por una bandera llamada SC que toma dos valores posibles: falso o verdadero, un conjunto de conceptos de entrada $CE = \{ce_1, ce_2, \dots, ce_n\}$ cuando SC sea falso, o bien un conjunto de dominios ordenados $D = \{d_1, d_2, \dots, d_n\}$ cuando SC sea verdadero.

Variable $SC \leftarrow falso$

Si $return_res(EtE)$

Entonces $SC \leftarrow verdadero$

Devuelve (SC, D)

Sino

Entonces

Por cada $c_i \in C$ donde $i = 1:n$

Si $exist_concept(c_i)$

Entonces $c_i \rightarrow ce_i$

Sino $exist_synonymous(c_i)$

Entonces $c_i \rightarrow ce_i$

Fin de Ciclo

Devuelve (SC, CE)

Fin del Algoritmo

C. Técnicas de razonamiento espacial cualitativo

En este trabajo se proponen y describen tres algoritmos enfocados a realizar razonamiento espacial. El primero es un algoritmo de *frecuencia conceptual* que cuenta las ocurrencias de cada concepto en un dominio geográfico. El segundo es un algoritmo de *relevancia* que busca la relación que existe entre los conceptos que recibe como entrada y su importancia en el dominio geográfico y el tercero es un algoritmo de *genealogía semántica* que obtiene a los conceptos padre, calculando la jerarquía de conceptos que involucran a un objeto geográfico representado por un concepto en el dominio, con lo cual se asume que este método podría utilizarse para proporcionar una definición del contexto.

Algoritmo de frecuencia conceptual: La finalidad de este algoritmo radica en contar el número de veces o calcular la frecuencia con la que aparece un conjunto de conceptos en un determinado dominio, a lo que se le ha denominado como *frecuencia conceptual*.

Asimismo, la salida se ordena de mayor a menor de acuerdo con su repetición. Este algoritmo recibe un conjunto de conceptos que previamente se han verificado, por tanto, existen en la base del conocimiento (esta verificación se realiza por medio del algoritmo de mapeo).

Posteriormente, es necesario utilizar la tabla de frecuencias de conceptos en los dominios, con la finalidad de buscar cuantas veces aparece cada concepto en cada dominio geográfico que tiene la base de conocimiento. En la Tabla 7 se presenta el algoritmo de frecuencia conceptual.

TABLE VII
ALGORITMO DE FRECUENCIA CONCEPTUAL

Entrada:

Un conjunto de conceptos que existen en la ontología $CE = \{ce_1, ce_2, \dots, ce_m\}$, una tabla Tf de frecuencias de conceptos en dominios y un vector de mapeo VM .

Salida:

Un conjunto de dominios ordenados $D = \{d_1, d_2, \dots, d_n\}$

Variables: $frecuenciadelDominio = \{fd_1, fd_2, \dots, fd_n\}$

Inicializar el conjunto de variables frecuencia de dominio en 0

Si SC es falso

Entonces

Por cada $d_i \in D$, donde $i = 1:n$

Por cada $ce_j \in CE$, donde $j = 1:m$

Si $concept(d_i, ce_j)$

Entonces incrementa fd_i

Fin de ciclo

Fin de ciclo

Ordena resultados $frecuenciadelDominio = \{fd_{d1}, fd_{d2}, \dots, fd_{dN}\}$

Sino entonces salir

Fin del Algoritmo

Algoritmo de relevancia: El algoritmo de relevancia recibe como parámetro un conjunto de entradas existentes del algoritmo de mapeo y un conjunto de tablas de composición de relevancia en el dominio. Por tanto, como salida se obtiene un conjunto de dominios ordenados, de acuerdo con su relevancia.

TABLA VIII
ALGORITMO DE RELEVANCIA

Entrada:
Un conjunto de conceptos que existen en la ontología $CE = \{ce_1, ce_2, \dots, ce_m\}$, una Tc de composición y un vector de mapeo VM
Salida:
Un conjunto de dominios ordenados $D = \{d_1, d_2, \dots, d_n\}$, junto con un conjunto de etiquetas de relevancia $ERD = \{ERd_1, ERd_2, \dots, ERd_i\}$ para cada dominio $ERd_i = \{N_{Dti}, CO_{Dti}, RA_{Dti}\}$, donde $i = 1:n$
Variables: Un conjunto de etiquetas de relevancia $ERD = \{ERd_1, ERd_2, \dots, ERd_i\}$ para cada dominio.
Si SC es falso
Entonces
Por cada $d_i \in D$, donde $i = 1:n$
Por cada $ce_j \in CE$, donde $j = 1:m$
Por cada $ce_k \in CE$, donde $k = 1:m$
Si $relevant(d_i, ce_j, ce_k)$
Entonces incrementa la etiqueta correspondiente a
$ERd_i = \{N_{Dti}, CO_{Dti}, RA_{Dti}\}$
Fin de ciclo
Fin de ciclo
Fin de ciclo
Ordena resultados de ERD
Sino entonces salir
Fin del Algoritmo

La utilidad de este algoritmo es que permite clasificar el grado de *importancia* de los conceptos en un dominio, y así proporcionar mayor riqueza semántica a los conceptos en cada dominio en particular. El algoritmo de relevancia realiza una búsqueda de todos los conceptos que recibe como entrada en la tabla de composición de relevancia de cada dominio, con el fin de saber si esos conceptos están relacionados. Finalmente, se obtiene como salida un conjunto de dominios. En la tabla VII se describe el algoritmo de relevancia para aplicar razonamiento espacial a un conjunto de objetos geográficos.

Algoritmo de genealogía semántica: Las operaciones que realiza el algoritmo de genealogía semántica se dividen en tres tareas. La primera es buscar los conceptos padres e hijos de cada uno de los conceptos de entrada. La segunda consiste en utilizar el algoritmo de relevancia para obtener el dominio resultante (solo se tomará el primer resultado obtenido por el algoritmo de relevancia). La última tarea consiste en realizar una suma de las salidas para mostrar el dominio que haya aparecido más con las entradas dadas al algoritmo de relevancia. La salida del algoritmo de genealogía, puede ser un dominio o bien la salida puede ser un conjunto vacío, lo cual indica que no se encontró ningún dominio.

El proceso del algoritmo de genealogía es el siguiente: primero, si el vector de mapeo VM es falso, entonces recibe los conceptos de entrada. Luego, busca la clase padre de cada uno de los n -conceptos de entrada y sustituye el i -ésimo concepto por el concepto padre, para invocar el algoritmo de relevancia con ese nuevo conjunto de entradas. Del resultado obtenido por el algoritmo de relevancia se guarda solo el primer elemento obtenido en un vector de salida, el cual contiene el dominio resultante de las veces que se repite dicho dominio. Este proceso es iterativo y se repite de acuerdo con el número de conceptos padres que se hayan encontrado.

TABLA IX
ALGORITMO DE GENEALOGÍA SEMÁNTICA

Entrada:
Un conjunto de conceptos que existen en la ontología $CE = \{ce_1, ce_2, \dots, ce_m\}$ y un vector de mapeo VM
Salida:
Un conjunto de dominios ordenados $D = \{d_1, d_2, \dots, d_i\}$
Si SC es falso
Variables: Una copia temporal CET de los conceptos de entrada CE , Un conjunto de dominios $Dominios = \{d_1, d_2, \dots, d_n\}$.
Entonces
Por cada $i \in \mathbb{N}$, donde $i = 1:n$
$CE \rightarrow CET$
Si $father(cet_i) \neq 0$
Entonces $father(cet_i) \rightarrow cet_i$
Algoritmo de relevancia(CET)
$d_1 \rightarrow Dominios$
Fin de ciclo
Por cada $son(ce_i)$
$CE \rightarrow CET$
Si $son(cet_i) \neq 0$
Entonces $son(cet_i) \rightarrow cet_i$
Algoritmo de relevancia(CET)
$d_1 \rightarrow Dominios$
Fin de ciclo
Eliminar_Repetidos($Dominios$)
Sino entonces salir
Fin del Algoritmo

TABLA X
ALGORITMO ELIMINAR CONCEPTOS REPETIDOS

Entrada:
Un conjunto de dominios $Dominios = \{d_1, d_2, \dots, d_n\}$.
Salida:
Un conjunto de dominios ordenados $D = \{d_1, d_2, \dots, d_i\}$
Variables: Un conjunto de vectores de salida dominios conformado por el nombre del dominio sd y las ocurrencias en el dominios osd
$SD = \{<sd_1, osd_1>, <sd_1, osd_1>, \dots, <sd_n, osd_n>\}$
Por cada $d_i \in Dominios$, donde $i = 1:n$
Por cada $d_j \in Dominios$, donde $j = 1:m$
Si $d_j \neq nulo$
Si $d_i = d_j$ & $i \neq j$
Entonces $d_j \rightarrow nulo$
Por cada sd_i
Si $d_i = sd_i$
Entonces incrementa osd_i una unidad
Sino Entonces $di \rightarrow sdi$,
incrementa osd_i una unidad
Fin de ciclo
Fin de ciclo
Ordena(SD)
Fin del Algoritmo

El proceso para los conceptos hijos de cada concepto de entrada es similar, con la diferencia de que cada concepto de entrada puede tener un número m de conceptos hijos; por lo que cada concepto hijo será una entrada individual al algoritmo de frecuencia. Este algoritmo se describe en la Tabla 9, así como el algoritmo que representa una función para eliminar conceptos repetidos en la tabla X.

D. Selector de salida

El algoritmo selector de salida recibe los resultados del conjunto de algoritmos y define cual es el resultado de

razonamiento más apropiado, de acuerdo con el grado de éxito de cada algoritmo de razonamiento implementado.

La efectividad o éxito del algoritmo depende directamente de las interacciones que se realicen con el usuario, ya que es necesario validar este conocimiento de una manera cognitiva para que pueda servir como un mecanismo de retroalimentación a la base de conocimiento y se tenga mayor *granularidad semántica*.

La función principal de este mecanismo es la comparación entre un grupo de tres resultados, si los tres conjuntos de resultados son iguales, se desplegará cualquiera de los conjuntos, si encuentra al menos dos resultados iguales de los tres que recibió, se mostrarán los resultados que fueron iguales. En caso de que ninguno de los resultados fue igual se visualiza el que tenga mayor grado de efectividad. Este procedimiento es descrito en el algoritmo mostrado en la Tabla 11.

TABLA XI
ALGORITMO SELECTOR DE SALIDA

Entrada:
Un vector de mapeo VM , Un conjunto de dominios ordenados por cada algoritmo de razonamiento $Car = \{Ar_1, Ar_2, \dots, Ar_n\}$
Salida:
Un vector de salida $VS = \{A, CARS\}$ con los siguientes datos. Un conjunto de dominios ordenados $A = \{D_1, D_2, \dots, D_n\}$, y el conjunto de algoritmos de razonamiento espacial cualitativo de salida utilizados en la forma $CARS = \{A_1, A_2, A_3, salidaConocida\}$
Variables: Tres conjuntos A_1, A_2, A_3 que denotan a los algoritmos con mayor grado de éxito de la forma $A = \{D_1, D_2, \dots, D_n\}$
Si SC es verdadero
Entonces entrega el conjunto $\{(D), (salidaConocida)\}$ del vector de mapeo VM como salida.
Sino
Entonces $exito(Car) \rightarrow \{A_1, A_2, A_3\}$
Si $compara(A_1, A_2)$ es verdadero
& $compara(A_1, A_3)$ es verdadero
Entonces despliega $\{(A_1), (A_1, A_2, A_3)\}$
Si $compara(A_1, A_2)$ es falso &
$compara(A_1, A_3)$ es verdadero
Entonces despliega $\{(A_1), (A_1, A_3)\}$
Si $compara(A_1, A_2)$ es verdadero &
$compara(A_1, A_3)$ es falso
Entonces despliega $\{(A_1), (A_1, A_2)\}$
Si $compara(A_1, A_2)$ es falso &
$compara(A_1, A_3)$ es falso
Entonces
Si $compara(A_2, A_3)$ es verdadero
Entonces despliega $\{(A_2), (A_2)\}$
Sino
Entonces despliega $\{(A_1), (A_1)\}$
Fin del Algoritmo
Inicio Función Compara
Entrada:
Un par de conjuntos ordenados A_1 y A_2
Salida:
<i>verdadero</i> en caso de que sean iguales o <i>falso</i> en caso de que sean diferentes
Por cada $DA1_i \in DA1$, donde $i = 1:n$
Si $DA1_i \neq DA2_i$
Entonces devuelve falso
Fin de ciclo
devuelve verdadero
Fin de función compara

IV. RESULTADOS OBTENIDOS

En la presente sección se presentan los resultados obtenidos de acuerdo con cada una de las etapas definidas en este trabajo.

A. Arquitectura del sistema

La arquitectura del sistema implementado se muestra en la Figura 10 y está compuesta de tres partes. La primera es el repositorio donde se procesan los marcos conceptuales. La segunda realiza la carga del modelo persistente de base de datos construido a partir de la ontología.

Finalmente, se encuentra la entrada del sistema que se relaciona directamente con la etapa de inferencia. Esta etapa se encarga de consultar la información del modelo persistente para devolver como resultado un concepto general al dominio en cuestión o al contexto geográfico.

B. Marcos conceptuales

Para representar el conocimiento y que éste pueda ser leído por la máquina, sin que todo el conocimiento tenga que ser ingresado de una sola vez, se selecciona una estructura básica.

Para este caso, se ha elegido el uso de los marcos conceptuales [43], [44]. Para su implementación, se eligió utilizar el metalenguaje XML [45], ya que de acuerdo con su estructura se puede realizar una apropiada clasificación taxonómica y organización del conocimiento *a priori*. Los marcos conceptuales están clasificados en dos tipos: *particulares* y *generales*.

Los marcos *particulares* (ver Figura 11) contienen la información de todos los conceptos que integran directamente y representan un dominio geográfico en particular, con datos referentes a su nombre, sinónimos, concepto padre e hijo relacionado. Los marcos *generales* (ver Figura 12) contienen la definición del dominio dado por las relaciones topológicas existentes entre los conceptos que componen al dominio, así como sus sinónimos. Este tipo de marcos definen un contexto para los conceptos y sinónimos utilizados.

A. Diseño del modelo persistente basado en la ontología

El modelo persistente ha sido construido a partir de la ontología. Con el propósito de almacenar las instancias de la base de conocimiento se ha convertido este modelo a una base

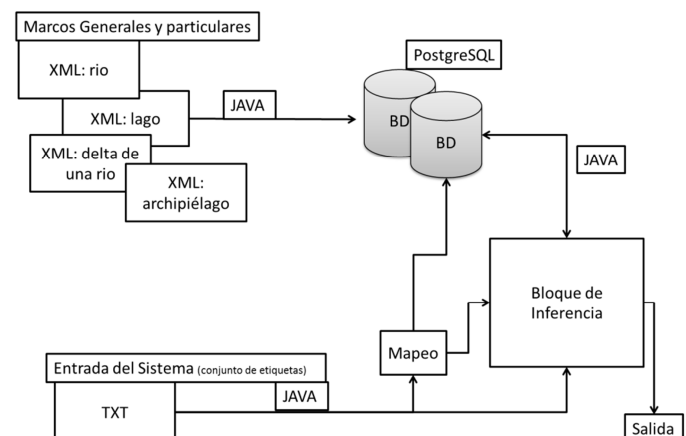


Fig. 10. Arquitectura del sistema.

```

1      <?xml version="1.0"?>
2      [-] <particular>
3      [-] <concepto>
4          <nombre>rio</nombre>
5          <sinonimos>river,flujo,flujo de agua</sinonimos>
6          <propiedades>anamosado,meandrico,rectilineo</propiedades>
7          <padre>caudal</padre>
8          <hijos>perene,estacionario,transitorio</hijos>
9      [-] <relacion_topologica>
10          <cruza>area verde,tierra</cruza>
11          <conecta>lago,mar</conecta>
12          <comparte>isla</comparte>
13      </relacion_topologica>
14  </concepto>
15  [-] <concepto>
16      <nombre>lago</nombre>
17  .

```

Fig. 11. Marco conceptual particular.

```

1      <?xml version="1.0"?>
2      [-] <general>
3      [-] <concepto>
4          <nombre>delta de un rio</nombre>
5          <sinonimos>river,flujo,flujo de agua</sinonimos>
6      [-] <conceptos_que_lo_componen>rio,lago, area verde,tierra,mar,
7          cuerpo de agua,isla,arena</conceptos_que_lo_componen>
8      [-] <definiciones>
9          <necesario>rio,conecta,mar</necesario>
10         <comun>isla,comparte,mar</comun>
11         <raro>arena,comparte,mar</raro>
12     </definiciones>
13 </concepto>
14 [-] <concepto>
15     <nombre>costa</nombre>
16 .

```

Fig. 12. Marco conceptual general.

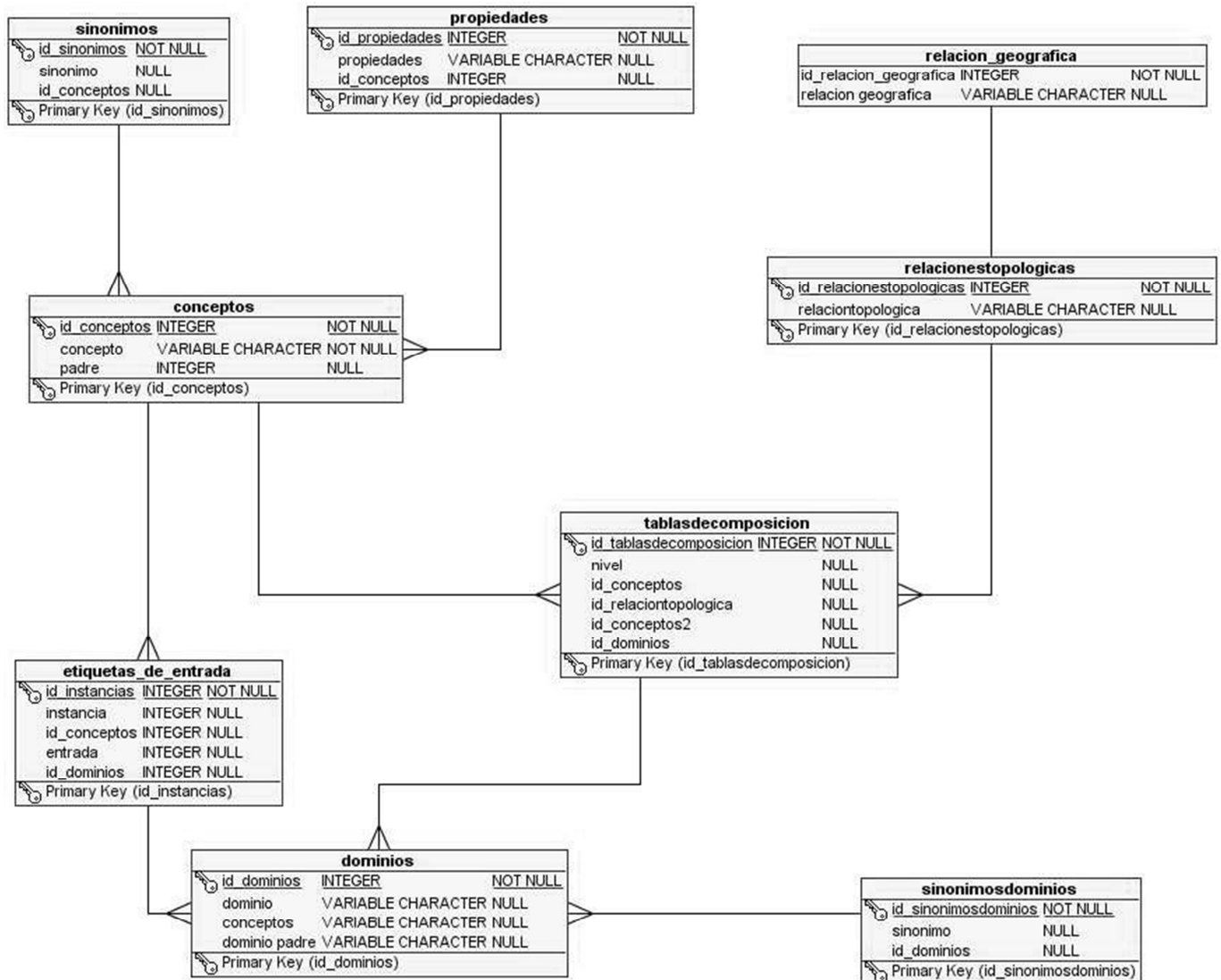


Fig. 13. Modelo persistente generado a partir de la ontología RAIN.



Polibits (45) 2012

la definición de los objetos geográficos y las relaciones topológicas que intervienen en los dominios geográficos y está basada en la ontología de dominio *Kaab*, descrita en [40].

B. Aplicación enfocada al razonamiento espacial

Por último, se tiene la tabla de etiquetas de entrada que contiene información acerca de las etiquetas que recibe la etapa de inferencia, para que así estas etiquetas sean relacionadas a través del algoritmo de mapeo con un concepto existente en la tabla de conceptos.

La información contenida en el modelo persistente está basada en la ontología RAIN (ver Figura 14), la cual contiene

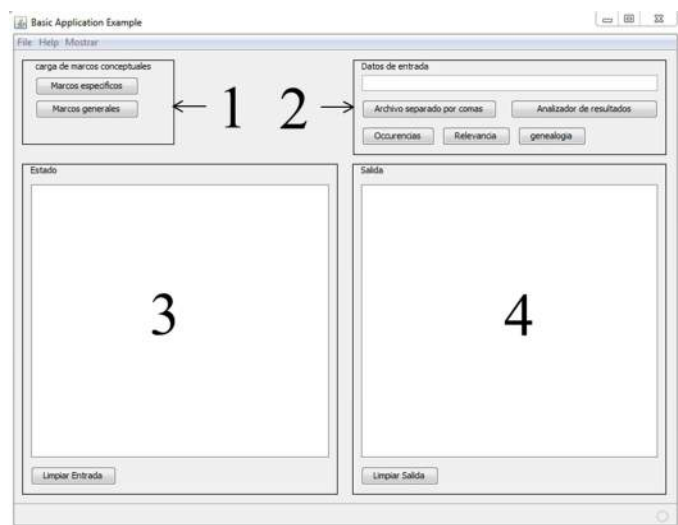


Fig. 15. Aplicación enfocada al razonamiento espacial.

C. Pruebas realizadas a los algoritmos de razonamiento

A continuación se muestran los resultados obtenidos con el algoritmo de razonamiento basado en frecuencia conceptual, utilizando los dominios geográficos *delta de un río* y *costa*. Primeramente, se inicia con la carga de los marcos conceptuales particulares y generales a la base de datos. En la tabla XII, se pueden observar las frecuencias en los dominios para este algoritmo.

TABLA XII
TABLA FRECUENCIAS DE CONCEPTOS EN DOMINIOS

Dominio	Algoritmo selector de salida
Delta de un río	{ <i>río, lago, área verde, tierra, mar, cuerpo de agua, isla, arena</i> }
Costa	{ <i>mar, arena, río</i> }

Por lo tanto, las etiquetas de entrada son: <mar, arena, río, isla, área verde>. La salida que se obtuvo fue “delta de un río”, debido a que fue el dominio que más ocurrencias tuvo y en segundo lugar “costa”, ya que solo se encontraron tres conceptos en ese dominio.

En la Figura 16, se muestra la salida de la aplicación, donde se puede observar que el algoritmo de frecuencia conceptual contabiliza el número de ocurrencias de cada dominio por concepto, directamente de la representación conceptual.

Sin embargo, se presenta un problema que está asociado directamente entre estos dos dominios. Esto es que para el caso en que la entrada de información se introducen etiquetas que se encuentran en ambos dominios, se obtendría la misma salida (por ejemplo, *mar* y *arena*). Para resolver este conflicto, se hace una distinción entre las etiquetas, tomando en consideración su grado de importancia dentro de un dominio. Para ello, se utiliza el algoritmo de relevancia, en donde es necesario contar con la definición de los conceptos con base en sus relaciones topológicas que intrínsecamente poseen, lo cual es descrito de manera explícita en la tabla XIII.

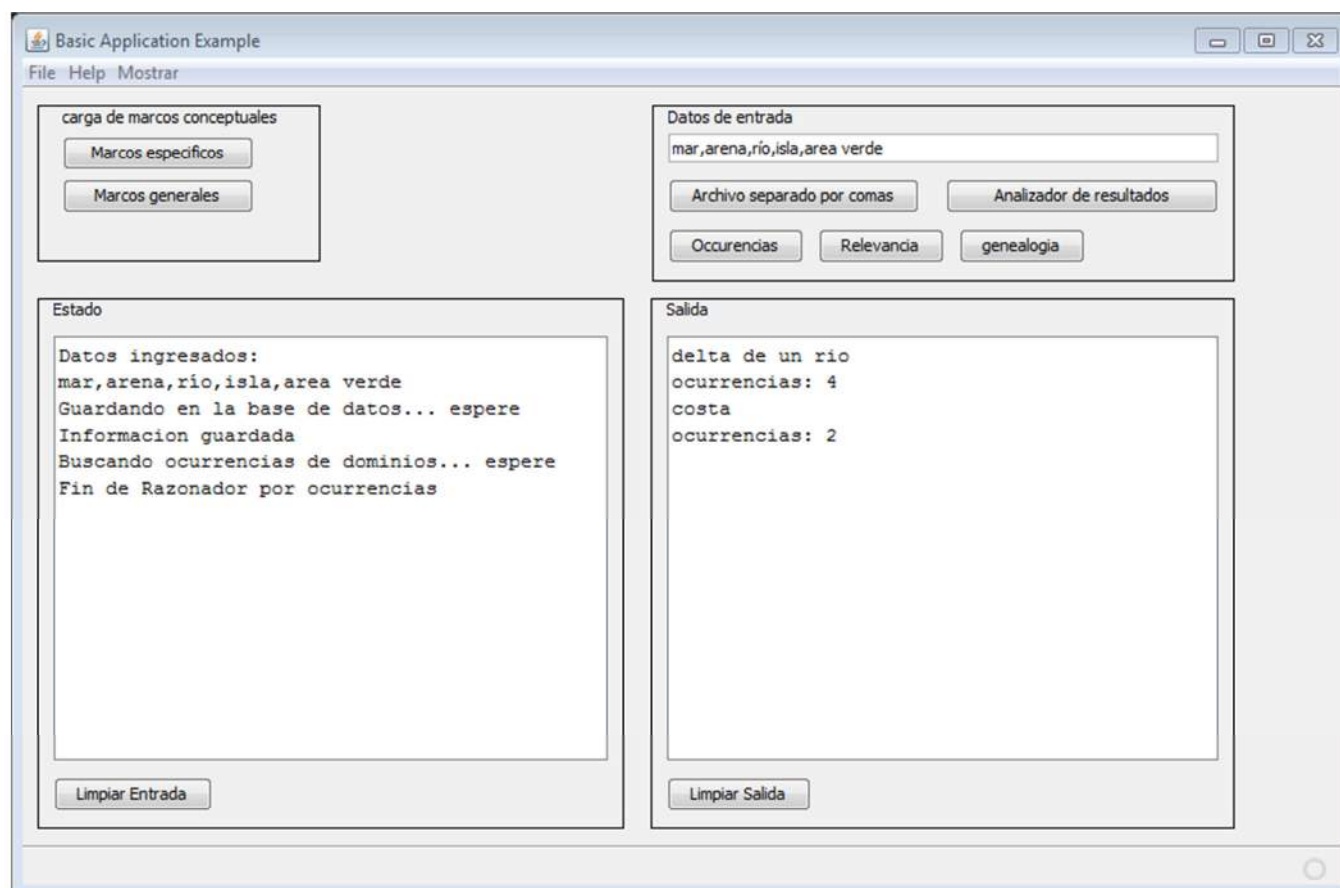


Fig. 16. Salida del algoritmo de frecuencia conceptual.

TABLA XIII
TABLA DE COMPOSICIÓN DE RELACIONES TOPOLÓGICAS

Dominio	Concepto1	Relación Topológica	Concepto2	Nivel de Relevancia
Delta de un río	Río	Conecta	Mar	Necesario
Delta de un río	Isla	Comparte	Mar	Común
Delta de un río	Arena	Comparte	Mar	Raro
Costa	Arena	Comparte	Mar	Necesario
Costa	Río	Cruza	Area verde	Común
Costa	Río	Conecta	Mar	Común

Con base en la tabla anterior, ahora se puede observar que cuando se introduce la entrada “mar” y “arena” y además se utiliza el algoritmo de relevancia, se obtiene en primer lugar “costa” y en segundo “delta de un río”. Esto se debe a que en la definición de cada uno de los dominios, ambos conceptos existen; pero solo en “costa” se presenta una relación necesaria que está definida por “arena comparte mar”, mientras que la relevancia en el dominio “delta de un río” es menor, ya que es raro encontrar este tipo de relación en ese dominio o contexto. En la figura 17, se muestra el resultado de aplicar el algoritmo de relevancia.

Por otra parte, cuando se escriben conceptos que no están explícitamente en las definiciones de los dominios; por ejemplo, sea el caso de la entrada <perenne, mar>, donde “perenne” pertenece a una subclase de “río”. Entonces, cuando esto sucede es necesario utilizar el algoritmo de *genealogía semántica* para buscar al padre de este concepto y ejecutar posteriormente el *algoritmo de relevancia*, con lo cual se debe almacenar el resultado. Consecuentemente, se deben buscar las clases hijos y verificar si existen definiciones. Como último paso, se repetirá este procedimiento para el concepto “mar”. En la figura 18 se puede observar la salida de este algoritmo.

```

Salida
costa
siempre: 1
comun: 0
raro: 0
####
delta de un río
siempre: 0
comun: 0
raro: 1
####

```

Fig. 17. Resultado de aplicar el algoritmo de relevancia.

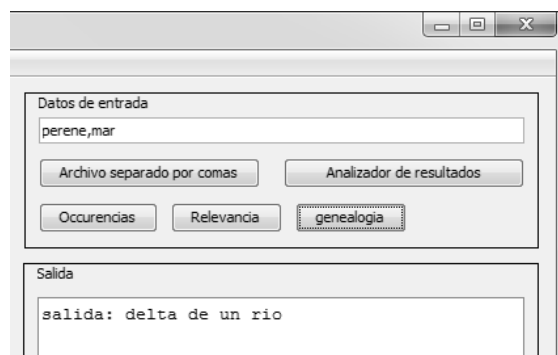


Fig. 18. Salida del algoritmo de genealogía semántica.

V. CONCLUSIONES

En este trabajo, se presenta un método compuesto básicamente por tres algoritmos que realizan un proceso de inferencia, con los cuales se lleva a cabo un proceso de razonamiento espacial cualitativo en descripciones de objetos geográficos.

Para ello, se genera una base de conocimiento, por medio de una ontología de aplicación que ha sido construida en OWL, bajo la metodología de GEONTO-MET. Esta base de conocimiento representa el conocimiento *a priori* de diferentes dominios geográficos.

Asimismo, se propone el uso de marcos conceptuales para representar de forma explícita el conocimiento de algún dominio, con la finalidad de estructurar la semántica del contexto geográfico.

Por otra parte, se puede argumentar que todo algoritmo de razonamiento espacial cualitativo, está compuesto por dos bloques: un bloque asociado con la *búsqueda* y otro relacionado con el *ordenamiento* basado en relevancia.

Asimismo, podemos afirmar que el conocimiento previo es una estructura formal, la cual contiene el vocabulario, las reglas para el lenguaje y un conjunto de proposiciones que son verdaderas (ya sean hechos o restricciones) que permiten solucionar problemas de ambigüedad y vaguedad, fundamentalmente para datos geoespaciales que por su naturaleza presentan estas deficiencias. Además, se propone una definición de razonamiento espacial, la cual consiste en el proceso de transformar una representación descriptiva en otra más general con base en el uso y procesamiento semántico de las relaciones topológicas. Cabe señalar que el proceso que se realiza en este trabajo es intentar *generalizar* hacia una clase o concepto superior, en un sentido inverso a la granularidad semántica (particularidad de objetos geográficos) que puede tener una representación conceptual, con lo cual se busca que la máquina procese las entidades geográficas de una manera similar a como los seres humanos procesamos cognitivamente e interpretamos el mundo real.

AGRADECIMIENTOS

Este trabajo ha sido auspiciado y apoyado por el Instituto Politécnico Nacional (IPN), el Consejo Nacional de Ciencia y Tecnología (CONACYT) y la Secretaría de Investigación y Posgrado del IPN, a través de los proyectos: 20113712, 20113757, 20120563 y 20120482.

REFERENCIAS

- [1] G. A. Elmes et al., “GIS and Society: Interrelation, Integration, and Transformation,” *A Research Agenda for Geographic Information Science*, vol. 3, CRC, pp. 287, 2005.
- [2] J. Hobbs, J. Blythe, H. Chalupsky, and T.A. Russ, *A Survey of Geospatial Resources, Representation and Reasoning*. Public Distribution of the University of Southern California, 2006.
- [3] M. Donnelly, T. Bittner, and C. Rosse, “A formal theory for spatial representation and reasoning in biomedical ontologies,” *Artificial Intelligence in Medicine*, vol. 36, Elsevier, pp. 1-27, 2006.
- [4] D. Hernandez, and A. Mukerjee, “Representation of spatial knowledge,” *Information Systems Magazine*, 1995.

- [5] J. Renz, *Qualitative spatial reasoning with topological information*. New York: Springer-Verlag, ch. 1, 2002.
- [6] D. M. Mark, "Geographic Information Science: Defining the field," *Foundations of Geographic Information Science*, New York: Taylor and Francis, pp. 3-18, 2003.
- [7] M. Egenhofer, and D. M. Mark, "Naive Geography," *Spatial Information Theory A Theoretical Basis for GIS*, New York: Springer-Verlag, 1995, pp. 1-15.
- [8] J. Sharma, *Integrated spatial reasoning in geographic information systems: combining topology and direction*, University of Maine, 1996.
- [9] B. Smith, "Mereotopology: a theory of parts and boundaries," *Data & Knowledge Engineering*, vol. 20, no. 3, Elsevier, pp. 287-303, 1996.
- [10] L. Jiming and L. K. Daneshmand, *Spatial reasoning and planning: geometry, mechanism, and motion*, New York: Springer-Verlag, ch. 1, 2004.
- [11] B. Bennett, "Physical objects, identity and vagueness," *Principles of Knowledge Representation and Reasoning International Conference*, Morgan Kaufmann, pp. 395-408, 2002.
- [12] B. Bennett, *Logical representations for automated reasoning about spatial relationships*, University of Leeds, 1997.
- [13] C. Freksa, "Qualitative spatial reasoning," *Cognitive and Linguistic Aspects of Geographic Space*, vol. 63, Springer-Verlag, pp. 361-372, 1991.
- [14] A.G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts, "Qualitative spatial representation and reasoning with the region connection calculus," *GeoInformatica*, vol. 3, no. 3, Springer-Verlag, 1997, pp. 275-316.
- [15] A.G. Cohn and S. M. Hazarika, "Qualitative spatial representation and reasoning: An overview," *Fundamenta Informaticae*, vol. 46, no. 1-2, IO Press, pp. 1-29, 2001.
- [16] L. Clarke, "A calculus of individuals based on connection," *Notre Dame Journal of Formal Logic*, vol. 22, no. 3, pp. 204-218, 1981.
- [17] J. F. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, no. 11, ACM, pp. 832-843, 1983.
- [18] D. Papadias and M. J. Egenhofer, "Algorithms for hierarchical spatial reasoning," *GeoInformatica*, vol. 1, no. 3, Springer, pp. 1-23, 1997.
- [19] M. J. Egenhofer and K. Al-Taha, "Reasoning about gradual changes of topological relationships," *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, New York: Springer-Verlag, pp. 196-219, 1992.
- [20] M. Egenhofer and R. Franzosa, "Point-set topological spatial relations," *International Journal of Geographical Information Science*, vol. 5, no. 2, Taylor & Francis, pp. 161-174, 1991.
- [21] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection," in *Third International Conference on Knowledge Representation and Reasoning*, vol. 92, Morgan Kaufmann, pp. 165-176, 1992.
- [22] J. R. Hobbs and S. Narayanan, "Spatial representation and reasoning," *Encyclopedia of Cognitive Science*, John Wiley, Online Library, 2001.
- [23] El-Geresy and A. I. Abdelmoty, "SPARQS: a qualitative spatial reasoning engine," *Knowledge-Based Systems*, vol. 17, no. 2, Elsevier, pp. 89-102, 2004.
- [24] R. Grütter, B. Bauer-Messmer, and M. Hägeli, "Extending an ontology-based search with a formalism for spatial reasoning," *Proceedings of the 2008 ACM Symposium on Applied Computing*, ACM, pp. 2266-2270, 2008.
- [25] S. Wang, D. Liu, X. Wang, and J. Liu, "Spatial reasoning based spatial data mining for precision agriculture," *Advanced Web and Network Technologies, and Applications*, New York: Springer-Verlag, pp. 506-510, 2006.
- [26] F. J. Escobar, S. Eagleson, and I. P. Williamson, "Automating the Administration Boundary Design Process using Hierarchical Spatial Reasoning Theory and Geographic Information Systems," *International Journal of Geographical Information Science*, vol. 17, no. 2, Taylor & Francis, pp. 99-118, 2003.
- [27] S. Wiebrock, L. Wittenburg, U. Schmid, and F. Wysotzki, "Inference and visualization of spatial relations," *Spatial Cognition II*, Springer-Verlag, pp.212-214, 2000.
- [28] E. Clementini, "Directional relations and frames of reference," *GeoInformatica*, Springer-Verlag, pp. 1-21, 2011.
- [29] Bailey-Kellogg and F. Zhao, "Qualitative spatial reasoning extracting and reasoning with spatial aggregates," *Artificial Intelligence Magazine*, vol. 24, no. 4, pp. 47-60, 2003.
- [30] P. L. Schultz, H. W. Guesgen, and R. Amor, "Computer-human interaction issues when integrating qualitative spatial reasoning into geographic information systems," in *Proceedings of the 7th ACM SIGCHI, New Zealand Chapter's International Conference on Computer-Human Interaction: Design Centered HCI*, ACM, pp. 43-51, 2006.
- [31] J. Brennan, and A. Sowmya, "Satellite image interpretation using spatial reasoning," in *Australasian Remote Sensing Photogrammet. Conf.*, Sydney Australia, vol. 1, 1998.
- [32] T. Barkowsky, S. Bertel, D. Engel, and C. Freksa, "Design of an architecture for reasoning with mental images," in *International Workshop on Spatial and Visual Components in Mental Reasoning about Large-Scale Spaces*, pp. 01-02, 2003.
- [33] B. Bennett, "What is a forest? On the vagueness of certain geographic concepts," *Topoi*, Springer-Verlag, vol. 20, no. 2, pp. 189-201, 2003.
- [34] G. Câmara, M. J. Egenhofer, F. Fonseca, and A. M. Vieira, "What's in an Image?" in *Conference on Spatial Information Theory*, New York: Springer-Verlag, pp.474-488, 2001.
- [35] M. Aiello, C. Areces, and M. D. Rijke, "Spatial reasoning for image retrieval," *International Workshop on Description Logics*, vol. 22, New York: Springer-Verlag, 1999.
- [36] Mezaris, I. Kompatsiaris, and M. G. Strintzis, "An ontology approach to object-based image retrieval," *International Conference on Image Processing*, vol. 2, IEEE, pp. 2-51, 2003.
- [37] L. Hollink, G. Schreiber, J. Wielemaker, and B. Wielinga, "Semantic annotation of image collections," *Knowledge Capture*, pp. 41-48, 2003.
- [38] J. Inglada and J. Michel, "Qualitative spatial reasoning for high-resolution remote sensing image analysis," *Geoscience and Remote Sensing, IEEE Transactions*, vol. 47, no. 2, 2009, pp. 599-612.
- [39] M. Torres and S. Levachkine, "Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos," *Computación y Sistemas*, vol. 12, no.3, pp.356-371, 2009.
- [40] M. Torres, R. Quintero, M. Moreno-Ibarra, R. Menchaca-Méndez, and G. Guzmán, "GEONTO-MET: An approach to conceptualizing the geographic domain," *International Journal of Geographical Information Science*, vol. 25, no. 10, Taylor & Francis, pp. 1633-1657, 2011.
- [41] Instituto Nacional de Estadística y Geografía, *Diccionario de Datos del INEGI*, INEGI, 2011.
- [42] National Center for Biomedical Ontology, *Ontology of BIOMES*, 2011.
- [43] M. Minsky, "K-lines: A Theory of Memory," *Cognitive Science*, vol. 4, no. 2, Elsevier, pp. 117-133, 1980.
- [44] M. Minsky, *A framework for representing knowledge*, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Memo No. 306, June 1974.
- [45] A. Salminen and F. Tompa, *Communicating with XML*, New York: Springer-Verlag, ch. 1, 2011.