

Integración de técnicas browsing con ontologías en un modelo de recuperación de información geográfica

Manuel E. Puebla Martínez
José M. Perea-Ortega
Alfredo Simón-Cuevas

El enorme volumen de datos geoespaciales que existen así como el incremento de las necesidades de usarlos en disímiles aplicaciones, han hecho imprescindible el surgimiento de mecanismos de Recuperación de Información Geográfica que posibiliten el acceso oportuno y preciso a dicha información. En este sentido, las técnicas más empleadas están basadas en la interrogación o querying y en la exploración o browsing. Sin embargo, las propuestas actuales de Recuperación de la Información Geográfica no proporcionan buenos resultados en presencia de ambigüedad en los términos empleados en las consultas, y su semántica no es tratada, y cuando las necesidades de información del usuario final suelen estar, poco o mal definidas. En el trabajo se propone una primera aproximación de un Modelo de Recuperación de Información Geográfica en el cual se integran el uso de la técnica de browsing y de una ontología geográfica (geontología), como recurso de conocimiento que brinda soporte al proceso de recuperación. La geo-ontología describe y formaliza la información semántica asociada a los datos geoespaciales y constituye el núcleo fundamental del modelo propuesto. Adicionalmente, se propone incluir en el modelo un mecanismo semiautomático para el enriquecimiento o mantenimiento de la geo-ontología a partir de diferentes fuentes, así como para la generación de consultas visuales, el cual estará soportado también por la geo-ontología.

Palabras clave: recuperación de información geográfica; browsing; semántica espacial; geontología.

RESUMEN

ABSTRACT

The enormous volume of geospatial data existent and the increment of the necessities to using them in dissimilar applications, have made indispensable the emergence of mechanisms of Geographic Information Retrieval that facilitates the opportune and specify access to this information. In this sense, more used techniques are based on querying and browsing. However, current proposals of Geographic Information Retrieval don't provide good results when the ambiguity in the terms used in the query is present, and their semantics is not treated, and when the information necessities of the users are not well defined. In this paper, a first approach of a Recovery of Geographic Information Model is proposed, in which the use of the browsing technique and a geographical ontology (geo-ontology), as a knowledge resource for supporting the retrieval process, are integrated. The geo-ontology describes and formalizes the semantic information associated to the geospatial data and constitutes the core of the proposed model. Additionally, in the model is included a semiautomatic for enrichment or maintenance the geo-ontology from different sources, as well as a mechanism for visual query generation, which also be supported by the geo-ontology.

Keywords: hill climbing; evolutionary strategy; process mining; metaheuristic; discovery models

Introducción

La información geográfica se refiere a la información relativa a entes o eventos en la que se incluye la referencia a su localización o posición sobre, o en las inmediaciones de la superficie de la tierra (Garea, 2007). El enorme volumen de datos geospaciales existentes y la necesidad creciente de acceso oportuno y preciso a dicha información, ha hecho imprescindible el surgimiento de técnicas de Recuperación de Información Geográfica (RIG). En este mismo escenario están los Sistemas de Información Geográfica (SIG), como una herramienta integradora que permiten el almacenamiento eficiente, la recuperación y visualización de información geográfica, así como compartir el enorme volumen de datos geospaciales. En la literatura se reportan diferentes modelos de recuperación de información (RI) (Andrés, Padilla, & Iribarne, 2013); (Larson, 1996); (Luna, 2001), algunos de los cuales han sido utilizados en el ámbito de la RIG. En este sentido, la RIG se ha abordado a partir de dos estrategias fundamentales: la interrogación o *querying* y la exploración o *browsing* (Herrero & Hassan, 2006). En años recientes, diferentes grupos de investigación han enfocado sus esfuerzos en tratar de mejorar los resultados de la RIG. La precisión promedio de los sistemas de RIG evaluados dentro del CLEF fue de 26.31 %, 27.25 %, 25.55 % y 23.70% en los años 2005, 2006, 2007 y 2008 (última convocatoria de evaluación). Se pueden apreciar que los valores de precisión son relativamente bajos y los avances en su incremento son poco significativos; de ahí la necesidad de seguir abordando este problema (Tello, 2010).

La técnica de *querying*, se sustenta en una consulta (pregunta) que debe ser realizada por el usuario, en cuya formulación el usuario debe dominar el lenguaje de consulta y estar consciente de lo que realmente necesita para poder generar una pregunta de la cual se pueda obtener una respuesta. El lenguaje natural utilizado es el comúnmente empleado para la formulación de consultas, ya que es más intuitivo para las personas. Sin embargo, el utilizar este tipo de lenguaje incorpora a los modelos de RI, problemas derivados de la ambigüedad inherente al lenguaje natural. Un ejemplo de esta problemática lo encontramos si lanzamos la consulta «Ríos en Granma» en el buscador Google y observamos como

los resultados pertenecen a diferentes áreas y contextos, sin estar relacionados con los ríos de dicha provincia cubana.

Una solución alternativa a esto lo constituye la incorporación del procesamiento semántico a los mecanismos de RI, incluyendo la definición y formalización de conceptos y relaciones entre ellos, para de esta manera reducir la ambigüedad al considerar que una palabra puede ser ambigua mientras que un concepto no lo es. De esta manera el proceso de RI estaría guiado por el conjunto de conceptos y relaciones que formalmente se definan, para lo cual las ontologías constituyen una prometedora alternativa (Mata, 2009).

En la técnica de *browsing* el usuario no tiene la necesidad de formular una pregunta, el mismo explora en el banco de información seleccionando sobre lo que necesita, sin la obligación de concientizar previamente sus necesidades de información. Sucede que en múltiples ocasiones cuando se busca información sobre un tema no específico o no se es capaz de expresarlo con precisión, bajo estas circunstancias la estrategia del *browsing* se muestra más prometedora que la técnica *querying*, afectada por el desconocimiento del usuario (Herrero & Hassan, 2006). Los nuevos modelos teóricos de RI aceptan al *browsing* como la estrategia básica e imprescindible para la búsqueda de información, especialmente en sistemas que van dirigidos a usuarios finales, cuyas necesidades de información suelen estar en un principio poco o mal definidas (Herrero, 2000).

Un tema fundamental es cómo diseñar capacidades de visualización que, como proceso, faciliten el pensamiento creativo para descubrir previamente información nueva en grandes bases de datos; lo cual es identificado esto como geobrowsing (Peuquet & Kraak, 2002). El geobrowsing también ha sido empleado para ayudar a los usuarios en la formulación de sus consultas, elección de vocabulario e identificación de relaciones entre los datos (Tufté, 1990). Sin embargo, las herramientas de visualización geoespacial actuales continúan basándose en marcadores con hipervínculos a los documentos, por lo que no se favorece la formulación de consultas y el pensamiento creativo de los usuarios (Buchel, 2013). Otras limitaciones que presentan estas técnicas son: las que se derivan de la no utilización de información semántica en el

tratamiento de los datos geográficos, la exploración está dirigida a través de una única capa del mapa, la gestión de un gran número de marcadores que causan los llamados «mapas parásitos» o que la selección y recuperación de los documentos se limita exclusivamente a los relacionados con los mapas (Buchel, 2013). Al no tratar y almacenar información semántica en las propuestas actuales de geobrowsing, limita la generación de un nuevo conocimiento que puede incrementar la satisfacción de usuarios finales en la RIG.

En este trabajo se propone una primera aproximación de un modelo para la RIG en el cual se integran el uso de la técnica de browsing con el empleo de una ontología geográfica, como recurso de soporte a la RI. En la ontología geográfica se encontrará formalizada información semántica de datos provenientes de diferentes fuentes, a saber: cartografías, bases de datos espaciales, conocimiento de los usuarios, relaciones espaciales e información geográfica que se encuentra disponible en Internet; a diferencia de otras propuestas que solo consideran información proveniente de fuentes textuales. Esto permitirá obtener descripciones semánticas de los datos geográficos más precisas en el sistema y así tener una mayor capacidad de respuesta, y al mismo tiempo resolver limitaciones presentes en las propuestas actuales de geobrowsing. Se propone incluir en este modelo mecanismos para el enriquecimiento o mantenimiento de la ontología geográfica desde las diferentes fuentes mencionadas, así como para la generación de consultas visuales.

Materiales y métodos

La definición del modelo de RIG que se propone se sustenta en estudios realizados sobre varias propuestas reportadas en la literatura, haciendo especial énfasis a las evaluadas en el marco de GeoCLEF. GeoCLEF constituye un espacio para la evaluación de técnicas de RIG, organizado por el CLEF desde el año 2005 hasta el 2008. En el estudio se centró la atención en la aplicación la técnica *browsing*, así como en la semántica espacial para la RIG.

Browsing en la RIG

Actualmente, se reporta varias propuestas basadas en browsing para la RIG (Larson,

1996)(Buchel, 2013)(Indratmo, 2010) (Buchel, 2014)(Abascal-Mena, López-Ornelas, & Zepeda, 2013), sin embargo, no consideran el almacenamiento y el procesamiento de la semántica de los datos espaciales; los más avanzados se limitan a inferir información semántica y tematizarla. También es común en estas propuestas el uso de grandes colecciones de textos existentes en Internet como fuente de búsqueda para la RIG. En (Buchel, 2013) se propone un modelo conceptual para el geobrowsing en colecciones de textos, en el que se visualizan tres espacios: el geográfico, el semántico y el cognitivo. Sin embargo, no se visualiza las relaciones semánticas existentes entre los documentos de la colección, tampoco define un soporte para almacenar y generar nueva información semántica a partir de la existente en el banco de información. Según se reporta en (Lin, 1997) y (Marcos, 2002) la estrategia de browsing posibilita representar una gran cantidad de información en un espacio visual limitado; tener la capacidad de revelar relaciones semánticas entre documentos y términos; facilitar la exploración visual así como la inferencia perceptiva desde a la interfaz; plantear un marco de búsqueda de información similar al empleado en el mundo real; y permitir el hallazgo de información de interés de manera fortuita. Por otro lado, se reportan limitaciones de esta técnica a partir de ineficiencias para resolver que una consulta bien definida (Marchionini, 1995).

Semántica Espacial en la RIG

En el ámbito de la RI, las ontologías resultan las herramientas más usadas para almacenar y generar conocimiento sobre un tema en particular. Estas fueron desarrolladas en el marco de la inteligencia artificial para facilitar el compartimiento del conocimiento y su reutilización (Euzenat, 2002). En (Garea, 2007) se plantea que la razón de la gran popularidad de las ontologías es que ofrecen un entendimiento compartido y común de algunos dominios que pueden ser comunicados entre las personas y los sistemas aplicados. Uno de los tipos de ontologías existentes son las ontologías geográficas o geontologías, consideradas como una extensión de una ontología convencional (Hess et al., 2007). Sobre la base de esta definición, se reporta en (Vera & Garea, 2009), entre otras, tres diferencias entre las geo-ontologías y las ontologías convencionales, a saber:

1. Las relaciones espaciales tienen una semántica predefinida y estandarizada, mientras que las relaciones convencionales son definidas de forma abierta según los conceptos que relaciona.
2. Cada concepto geográfico tiene, al menos, una geometría asociada que los representa la que juega un papel fundamental en la definición de posibles relaciones espaciales que pueda tener el concepto.
3. Una instancia geográfica tiene un número de pares de coordenadas (x, y) , expresadas en un sistema de coordenadas dado, que

representa su posición espacial sobre la superficie.

Son varias las investigaciones que apuestan por el uso de ontologías geográficas en el proceso de RIG (Andrés, Padilla, & Iribarne, 2013)(Mata, 2009)(Buscaldi, Rosso, & García, 2006)(Ping & Yong, 2009)(Koo, Lim, & Lee, 2003)(Heæimoviaæ & Ciceli, 2013). La mayoría coinciden en el uso de las ontologías para almacenar la semántica de los datos, para luego utilizarla en la generación de nuevo conocimiento. En la mayor parte de las propuestas la ontología se obtiene de forma semi-automática(Koo, Lim, & Lee, 2003)(Buscaldi, Rosso&García, 2006)(Ping & Yong, 2009)(Andrés, Padilla, & Iribarne, 2013)(Heæimoviaæ & Ciceli, 2013), y en otros se utiliza una ontología ya existente (Mata, 2009). La información disponible en Internet predomina como fuente de información para alimentar la automatización de la construcción de la ontología, tal es el caso de Wikipedia, aunque también es usado WordNet y diccionarios de datos como GNS y el GNIS.

En la Tabla 1 se resume una caracterización de nueve modelos de RIG, según: fuente de información sobre la que se realiza la recuperación, técnica para capturar necesidades de búsqueda de usuarios, estrategia de indexación de la fuente de información, y la estrategia de análisis de datos. Este último aspecto esta asociado al

Tabla 1. Caracterización de modelos de RIG del estado del arte

| Propuestas | Fuente de Información | Captura de Necesidades de Búsqueda | Indexación de la Fuente de Información | Estrategia de Analisis de Datos |
|-------------------------------|---|---|---|--|
| (Larson, 1996) | Bibliotecas Digitales. | Browsing | Basado en los nombres que aparecen en los documentos | Interfaces gráficas, diccionarios geográficos, |
| (Jones, Alani, Tudhope, 2001) | Tesauros, Fuentes de la historia cultural. | Querying | A través de la Ontología | Ponderación: basado en cercanía geográfica y cercanía semántica con respecto al tema de interés. Expansión de Consultas |
| (Walker, 2007) | Información del Usuario, relaciones causales entre los conjuntos de datos de SIG. | Querying | A través de un Mapa | Inferencia bayesiana, conocimiento experto, Browsing |
| (Mata, 2009) | Ontologías, diccionarios de datos y archivos vectoriales. | Querying | No existe. Recupera directamente de las tres fuentes de información | Ponderación de los resultados a partir del análisis de la semántica espacial y la intención del usuario basado en su perfil. |

| Propuestas | Fuente de Información | Captura de Necesidades de Búsqueda | Indexación de la Fuente de Información | Estrategia de Analisis de Datos |
|-----------------------------|---|------------------------------------|--|---|
| (Indratmo, 2010) | Historialde la interacción social:trazas de la interacción entre los usuarios en un espacio de información. | Browsing | A través de un Mapa | Browsing, Minería de Uso: historia interacción social |
| (Perea , 2010) | Internet. | Querying | Motores de búsqueda | Enfoque sintáctico, Expansión de Consultas |
| (Tello, 2010) | Internet. | Querying | Motores de búsqueda | Enfoque sintáctico, Re-ordenamiento vía retroalimentación de relevancia |
| (Abascal-Mena et al., 2013) | Documentos no estructurados en línea: narrativas de viaje y periódicos en línea | Browsing | A través de un Mapa | Browsing |
| (Buchel, 2013) | Colecciones de Textos | Querying, Browsing | A través de un Mapa | Browsing, Diccionarios Geográficos, Expansión de Consultas |

proceso de descubrimiento de conocimiento (a priori desconocido) sobre repositorios de datos complejos, mediante la extracción de información ‘oculta’ y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre datos.

Los estudios realizados también arrojaron que es imprescindible hacer uso de recursos externos de información geográfica, ej.: tesauro, así como es conveniente el empleo de WordNet o Wikipedia. Es recomendable contar con un desambiguador de topónimos para reducir las ambigüedades geográficas y no se aprecia un acuerdo en lo que respecta a la conveniencia de aplicar técnicas de expansión de consultas en este contexto. El uso de interfaces de usuarios multimodal que proporcionan entrada de textos y mapas de retroalimentación interactiva en el contexto de la información recuperada se considera una forma efectiva de interactuar con la información geográfica, aunque aún no se han realizado investigaciones en esta área para la RIG, según se reporta en (Jones, 2002). Según lo anteriormente expuesto, se ha considerado conveniente definir un modelo de RIF en el que tenga en cuenta diferentes fuentes de información, como entorno de búsqueda, por ejemplo: Internet, bases de datos espaciales, archivos vectoriales, conocimiento a suministrar por usuarios, entre otras, así como la aplicación de técnicas browsing, soportada en el una ontología como herramienta para la

indexación, y mecanismos de análisis de datos con enfoque semántico, combinado con la generación de consultas visuales.

Resultados y discusión

Modelo de RIG Propuesto

El enfoque RIG propuesto en el trabajo integra en un único modelo el uso de la semántica espacial de los datos y la técnica del *browsing*. Mediante esta integración se pretende resolver las limitaciones que presentan los modelos basados en un enfoque sintáctico y los basados en *querying*. Además, la incorporación de la estrategia del *browsing*, no solo se limitará a RI, sino que también podrá ser utilizado como patrón de observación o técnica de visualización. También se pretende resolver la limitación que se le atribuye al *browsing* relacionada con la ineficiencia para resolver una consulta bien definida.

La propuesta del modelo parte de una ontología básica del dominio geográfico que se enriquece de forma semi-automática con el objetivo de modelar el conocimiento de una zona geográfica en particular. El proceso de enriquecimiento se hará a partir del procesamiento de varias fuentes de información que serán posteriormente analizadas. Diferentes investigadores han abordado la construcción de una ontología

del dominio geográfico que sirva de guía para la construcción de ontologías de menor nivel (Ping & Yong, 2009), (Tomai & Spanaki, 2005), (Torres, 2007), (Lopez-Pellicer, Silva, & Chaves, 2010), (Smart et al., 2007). Por tal motivo, en esta propuesta no se pretende construir un nuevo modelo ontológico para el dominio geoespacial, sino más bien seleccionar uno de que fueron desarrollados y tomarlo como punto de partida. Luego de analizar varias propuestas se la ontología del dominio geográfico *Kaab-Ontology* especificada en (Torres, 2007). Con esta ontología se materializa la metodología *GEONTO-MET*. Como caso de estudio se diseñaron en (Torres, 2007) dos ontologías de aplicación basadas en la metodología *GEONTO-MET*, una en el contexto turístico y otra en el contexto topográfico; las cuales pudieran ser un punto de partida más preciso para la futura evaluación del modelo propuesto en esta investigación. La ontología geográfica resultante será gestionada a través de un SIG que también incluirá una base de datos cartográfica del área geográfica objetivo, a partir del cual el usuario podrá explorar el mapa de la zona o la ontología según sus necesidades.

Esta propuesta está basada en la hipótesis de que a partir de la interacción entre el mapa y la geo-ontología se obtendrán mejores resultados en la generación de consultas visuales por parte del usuario final, así como en la aplicación de la estrategia del *browsing*.

El objetivo de construir consultas visuales es lograr que el usuario pueda realizar consultas geográficas con un soporte semántico, sin necesidad de ser experto en semántica espacial o ingeniería ontológica. De este modo, el usuario simplemente seleccionará en la geo-ontología o en el mapa del área geográfica objetivo, los elementos que deben participar en la consulta, incluyendo las relaciones espaciales y axiomas presentes en la geo-ontología. La Fig. 1 muestra un esquema del modelo propuesto, donde se distinguen como procesos principales: enriquecimiento de la geo-ontología a partir de diferentes fuentes de información; y la RIG basada en *browsing*, donde en esta última se incluye la generación de consultas visuales y la clasificación de la información recuperada.

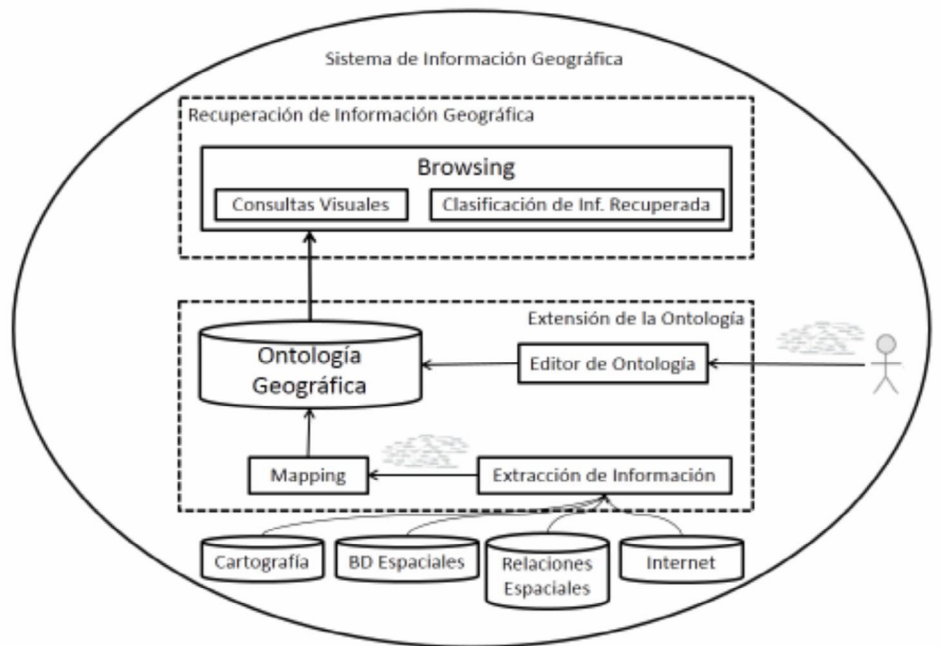


Figura 1: Esquema de componentes del modelo propuesto

El modelo es genérico, es decir, aplicable a cualquier zona geográfica pero estará limitado por la información que se pueda extraer de la zona seleccionada a partir del análisis de cinco fuentes de información: Bases de datos espaciales que modelan la zona; Cartografía de la zona; Conocimiento de usuarios finales en la zona; Relaciones espaciales entre los datos geográficos ya identificados en la zona; Información geográfica disponible en Internet sobre la zona seleccionada. Según a cada tipo de fuente se desarrollarán mecanismos de extracción de información.

Cartografía de la zona: A partir de esta fuente serán identificadas instancias de conceptos y propiedades de datos, cuyo proceso dependerá de la calidad y el nivel de detalle de la cartografía disponible. Esta cartografía debe estar en formato vectorial, específicamente en formato *Shapefile*, para luego ser cargada en un gestor de bases de datos relacionales. A través de la integración de la información extraída en

esta fuente, combinada con otras, se identificarán relaciones topográficas, así como otros elementos de información útiles, según lo reportado en (Martínez & Moreno, 2006).

Bases de datos espaciales (BDE): Este tipo de base de datos permite describir los objetos espaciales que la forman a través de tres características básicas: atributos, localización y topología (Sarabia, 2008). El modelo no está concebido para usar una BDE específica, para lo cual se define como estructura organizativa genérica la que se muestra en la Fig. 2, donde TObjeto1, Tobjeto2 y TObjetoN representan el universo de posibles objetos espaciales que pudieran estar presente en cualquier BDE (los campos son definidos y expuestos en la Fig. 2 no son los únicos posibles).

En (Baglioni et al., 2007) se reportan los

primeros pasos de una metodología para la construcción automática de una ontología espacial a partir de una BDE. Esta metodología está basada en una serie de reglas para la generación de la ontología a partir de bases de datos que se encuentren al menos en tercera forma normal, e incluyen elementos para el almacenamiento de posición espacial y la geometría del objeto representado pero no toman en cuenta las relaciones espaciales y topológicas existentes. Con el objetivo de potenciar el enriquecimiento de las geontologías, se reporta en (Larín & Garea, 2011) un método que incluye las relaciones topológicas, sin analizar el resto de las relaciones espaciales. En el modelo se integran estas estrategias, y se incorporan las relaciones espaciales implícitas en los datos.

Información geográfica en Internet: En la actualidad son varias las fuentes de

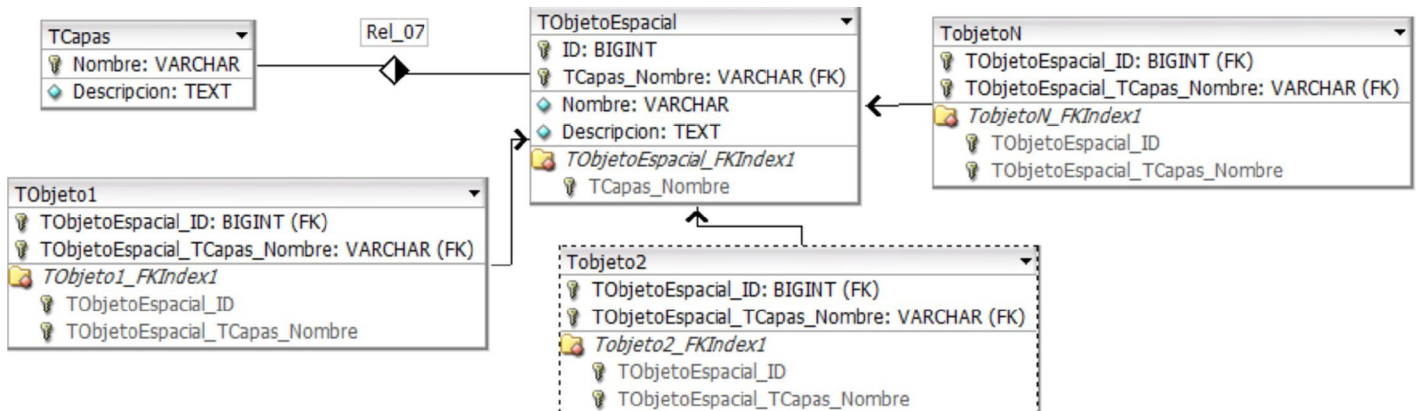


Figura 2: Estructura organizativa genérica.

de información geográfica disponibles en Internet, una de las más populares es GeoNames. Toda la información geográfica que se recolecte debe estar en el perímetro de la zona geográfica modelada en cada caso. Son varias las investigaciones que abordan la automatización

del proceso de construcción de ontologías geográficas a partir de fuentes disponibles en Internet. Entre ellas destaca (Buscaldi et al., 2006). Otras como (Abascal-Mena et al., 2013), (Buscaldi & Rosso, 2009) y (Overell, 2009) no se concentran en la construcción de la ontología de forma automática pero sí en la extracción de información geográfica desde fuentes de Internet. En (Buscaldi et al., 2006) se describe cómo integrar, en una geontología implementada con cláusulas del lenguaje Prolog, información extraída de WordNet, de los diccionarios geográficos GNS y GNIS, además de Wikipedia. En (Abascal-Mena et al., 2013) se describe un servicio web de geo-enriquecimiento que ayuda a los desarrolladores en la determinación de la localidad en contenidos no estructurados y atómicos, haciendo uso de Internet basada en localización. Es decir, ofrece la posibilidad de convertir documentos de texto en bases de datos geoespaciales. Este proceso se realiza en dos etapas: 1) extracción de entidad y 2) desambiguación, que también se conoce como conexión a tierra o geotiquetado. La extracción de entidades geoespaciales usa técnicas de procesamiento de lenguaje natural para identificar nombres de lugares en el texto, mientras que desambiguación establece el vínculo de un nombre con su correcta ubicación. En (Buscaldi & Rosso, 2009) se presenta un método que utiliza GeoWordNet para la RIG. Durante la fase de indexación, son desambiguados todos los lugares y se les asigna coordenadas geográficas en el Mundo.

Relaciones Espaciales: Teniendo todas las localizaciones geográficas y el perímetro de un grupo de objetos geográficos, es posible automatizar la identificación de las relaciones espaciales existentes entre dichos objetos apoyándose en la suite JTS (Java Topology Suite, por sus siglas en inglés). El objetivo que se pretende en esta fuente de información es identificar de manera automatizada la existencia relaciones espaciales implícitas en las instancias presentes en la geontología, e incorporarlas a la ontología geográfica. El tratamiento de esta fuente posibilita integrar

información recuperada a partir otras fuentes, y es útil para la generación de nuevo conocimiento.

La indexación de la información es uno de los componentes más importantes en cualquier modelo de RI. Según (Luna, 2001), el primer paso que debe dar un sistema de RI antes de estar totalmente preparado para que los usuarios lo utilicen es el de procesar la base de datos de documentos, dejándola en un formato cuya manipulación por parte del sistema sea fácil y rápida. En esta propuesta, la secuencia de términos de indexación que representará a cada dato será la descripción semántica que estará almacenada en la geo-ontología. Esto permite afirmar que el enriquecimiento de la geontología constituirá el mecanismo de indexación de la información manipulada por el modelo propuesto. Esto está soportado en un proceso de *mapping*, que se ejecuta luego que es extraída la información de las diferentes fuentes descritas. El *mapping* no es más que un mecanismo que identifica correspondencia entre la información que es extraída y los conceptos representados en la ontología del dominio geográfico inicial. Una primera variante pudiera ser incorporarlos a la geo-ontología sin ser clasificados para que luego se clasifiquen de forma manual por el usuario final a través de las facilidades de edición que podrá brindar el editor de ontologías. Otra variante sería automatizar la clasificación de conceptos a partir del análisis de los nombres y características recuperados y los conceptos presentes en la geo-ontología, aplicando algunas medidas de similitud entre entidades ontológicas descritas en (Vera & Garea, 2009).

Otra tipología de información que ha sido considerada en el modelo propuesto para el enriquecimiento de la geo-ontología es el conocimiento que pueda tener una persona sobre la zona geográfica a modelar en esta propuesta. Para capturar este conocimiento se emplea un *editor de ontologías*, a través del cual se brindan interfaces amigables que permitan a usuarios no expertos en ingeniería ontológica poder registrar el conocimiento que poseen de una manera fácil e intuitiva. En este sentido, se intenta diseñar y construir un editor propio de ontologías geográficas, que se apoyará en un visor SIG para simular la visualización de la información geográfica presente en el mapa, con la edición de la geo-ontología. Según la literatura consultada,

y en particular los estudios en (Tomai & Spanaki, 2005) (Albrecht, Derman, & Ramasubramanian, 2008) no se ha reportado un editor de ontologías geográficas con las características mencionadas.

Mecanismo de RIG basado en la estrategia del Browsing

Una vez que se termina el proceso de enriquecimiento de la geo-ontología, ésta constituirá la plataforma sobre la que descansa el sistema de RIG basado en *browsing*. En el reconocimiento de la información que realiza el usuario a través *del browsing* hay que garantizar que no se manifieste el problema de los «mapas parásitos», segundo problema planteado por (Buchel, 2013). Para ello se le asignará a cada dato geográfico un rango de valores del «zoom» en el cual él será visible en el mapa. Los datos geográficos solo serán visibles en el mapa cuando sean visibles las coordenadas y se esté en el nivel de «zoom» adecuado (Larson, 1996). Además, la visualización de información geográfica en el mapa será regulada por el componente de la geo-ontología que esté seleccionado en cada momento. Por ejemplo, si tenemos una geo-ontología correspondiente a la Universidad de las Ciencias Informáticas (UCI) y seleccionamos el concepto «Edificio Docente», el modelo solo debe visualizar los edificios docentes cuyas coordenadas estén en la zona que visualiza en ese momento el visor SIG y en correspondencia con el nivel de zoom definido.

Generación de Consultas Visuales

Una línea de investigación habitual para facilitar al usuario la expresión de consultas es el uso de lenguajes visuales. Se procura facilitar el uso de conceptos predefinidos en la ontología, ya que estas representaciones reducen el nivel de abstracción (Moya & Macías, 2006). El modelo propuesto además de explorar el espacio geográfico y semántico, debe permitir el diseño de consultas visuales a partir de la selección, por parte de los usuarios finales, de elementos de la geo-ontología y elementos del visor SIG. Esta forma de generar consultas al modelo posee tres puntos significativos:

1. Los usuarios finales no necesitan ser grandes conocedores de los lenguajes de consultas a las ontologías, solo deben

dominar la información disponible en la zona geográfica que se modela en correspondencia con lo que plantea Félix Mata en (Mata, 2009), sobre la «necesidad de mecanismos que permitan a personas no expertas en ingeniería ontológica, obtener resultados relevantes para las búsquedas que realizan».

2. El clásico problema de la ambigüedad y la vaguedad de los modelos de RI basados en *querying* se minimiza. En los modelos clásicos de RI basados en *querying*, la palabra «cerca» pudiera referirse a una distancia o a un tiempo. Consideremos los siguientes ejemplos (primer ejemplo se refiere a distancia y el segundo a tiempo):
 - a) 'El Hospital Militar está cerca de la vivienda de María',
 - b) 'La fecha del cumpleaños de Pedro está cerca'

3. La incapacidad del *browsing* para poder resolver consultas bien definidas, carencia históricamente señalada a esta estrategia (Marchionini, 1995), quedará resuelta. De este modo la integración del *browsing* con la estrategia de consultas visuales permitirá resolver la problemática antes planteada.

Clasificación de los resultados

El objetivo de la clasificación en la RIG es el de presentar al usuario una lista ordenada de los documentos que cumplan los dos criterios en la consulta, el no geográfico y el geográfico (Andogah, 2010). En RI se requiere de un acercamiento a un resultado cuando no existen instancias que correspondan con el criterio de consulta. En este enfoque, las respuestas aproximadas tienen un grado de relevancia (Mata, 2009). Si se considera la consulta 'Hospitales en el Oriente Cubano' en la geo-ontología, cuya conceptualización parcial se representa en la Fig. 3. Un resultado relevante sería 'Hospital Celia Sánchez Manduley en Manzanillo', mientras que otro con menor relevancia sería 'Policlínico René Vallejo en Manzanillo'. En el primero, 'Policlínico René Vallejo en Manzanillo' aparece porque el concepto 'Policlínico' (concepto al que pertenece la instancia 'René Vallejo') está semánticamente relacionado con el concepto 'Hospital' por medio de la relación de hiperonimia y además está relacionado espacialmente mediante la

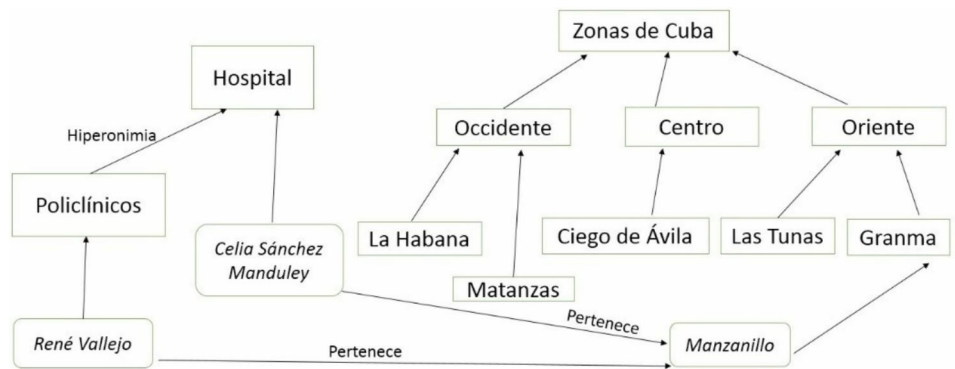


Figura 3: Conceptualización parcial de la Geontología

relación topológica 'pertenece' con 'Oriente Cubano', el municipio de 'Manzanillo' 'pertenece' al 'Oriente Cubano'. En este caso, la geo-ontología constituye el soporte para evaluar la relevancia de los resultados obtenidos luego de la ejecución de las consultas visuales. Hasta el momento solo se ha valorado la utilización de este método de evaluación de los resultados; otros métodos como los analizados e (Andogah, 2010) serán analizados en trabajos futuros.

Conclusiones

Se ha presentado una aproximación inicial de un modelo para la RIG que integra el uso de la técnica de browsing con el uso de una ontología geográfica, como recurso para el conocimiento. La integración del browsing a la geo-ontología posibilita fortalecer aspectos positivos y mitigar los negativos de ambas técnicas, para mejorar los resultados de precisión en la RIG. En la geo-ontología son descritos semánticamente datos geoespaciales a partir de información provenientes de diferentes fuentes, tales como: cartografías, bases de datos espaciales, conocimiento de los usuarios, relaciones espaciales e información geográfica disponible en Internet; para obtener descripciones semánticas más completas y precisas. El empleo de una geo-ontología en este sentido, brinda la posibilidad de usar mecanismos de inferencia sobre este tipo de ontologías, para generar conocimiento útil en el mejoramiento de los resultados de la RIG, tal es el caso de nuevas relaciones semánticas entre datos geoespaciales no explícitas inicialmente. También visualiza el espacio semántico y geográfico de la zona geográfica que modela, facilitando así el pensamiento creativo y la generación o descubrimiento de nuevo conocimiento por parte del usuario final. Minimiza la

ambigüedad en el procesamiento de las consultas de RIG. En el modelo se han incluido mecanismos para enriquecer o mantener la ontología geográfica, que posibilita incrementar las capacidades de soporte al proceso de RIG. También se ha incluido un mecanismo de generación de consultas visuales, soportado en la geo-ontología, como otra alternativa para facilitar la RIG. La vinculación de la geo-ontología con la generación de consultas visuales, posibilita que el usuario final pueda ser orientado, a través de los conceptos y relaciones semánticas representadas en la geo-ontología, en la formulación de su consulta, sin tener la necesidad de ser un especialista en la manipulación de información semántica y espacial.

Bibliografía

Abascal, R., López, E., & Zepeda, J. S. (2013). Geographic Information Retrieval and Visualization of Online Unstructured Documents. *International Journal of Computer Information Systems and Industrial Management Applications*, 5, 89-97.

Albrecht, J, Derman, B. & Ramasubramanian, L. (2008). Geo-ontology Tools: The Missing Link. *Transactions in GIS*, 12(4), 409-424.

Andogah, G. (2010). Geographically Constrained Information Retrieval. Groningen: Center for Language and Cognition Groningen. Tesis de doctorado.

Andrés, J., Padilla, N., & Iribarne, L. (2013). Information retrieval using an Ontological Web-

- Trading model (pp. 243 – 249).
Presentado en Computer Science and Information Systems (FedCSIS), Finlandia: IEEE.
- Baglioni, M., Masserotti, V. & Renso, Ch. (2007). Building geospatial ontologies from geographical databases. In Proceedings of the Second International Conference on Geospatial Semantics. México.
- Buchel, O. (2013). Redefining geobrowsing. Proceedings of the American Society for Information Science and Technology, 50(1), 1–10.
- Buchel, O. (2014). How Can Geographic Information Retrieval Benefit from Geovisualization Principles?, En Proceedings of iConference 2014, pp. 1066–1071.
- Buscaldi, D., & Rosso, P. (2009). Using GeoWordNet for Geographical Information Retrieval. En Proceedings of the 9th Cross-language Evaluation Forum Conference on Evaluating Systems for Multilingual and Multimodal Information Access, pp. 863–866. Berlin, Heidelberg: Springer-Verlag.
- Buscaldi, D., Rosso, P., & García, P. P. (2006). Inferring geographic ontologies from multiple resources for geographic information retrieval. En Proceedings of SIGIR Workshop on Geographic Information Retrieval, pp. 52–55, Seattle: USA.
- Euzenat, J. (2002). Research challenges and perspectives of the Semantic Web. *Intelligent Systems*, 17(5), 86–88.
- Garea, E. (2007). Estado actual de la interpretación semántica de datos espaciales. Blue Series. Pattern Recognition. Disponible http://www.cenatav.co.cu/doc/RTecnicos/RT%20SerieAzul_001web.pdf
- Heäimoviä, Ž., & Ciceli, T. (2013). Spatial Intelligence and Toponyms. Trabajo presentado en 26th Int. Cartographic Conf. (ICC 2013). From Pole to Pole, Dresden, Germany.
- Herrero, V. (2000). Modelos de Representación Visual de la Información Bibliográfica (p. 82). Granada, España.,: Dpto de Biblioteconomía y Documentación de la Universidad de Granada.
- Herrero, V., & Hassan, Y. (2006). Metodologías para el desarrollo de interfaces visuales de recuperación de información: análisis y comparación. *Information Research*, 11(3), 258.
- Hess, G., Iochpe, C., Ferrara, A., & Castano, S. (2007). Towards Effective Geographic Ontology Matching. En F. Fonseca, M. A. Rodríguez, & S. Levashkin (Eds.), *GeoSpatial Semantics*, Vol. 4853, pp. 51–65. Springer.
- Indratmo, I. (2010). Supporting exploratory browsing with visualization of social interaction history. Department of Computer Science University of Saskatchewan Saskatoon
- Jones, C., Alani, H., & Tudhope, D. (2001). Geographical Information Retrieval with Ontologies of Place. In D. Montello (Ed.), *Spatial Information Theory*, Vol. 2205, pp. 322–335. Springer Berlin Heidelberg.
- Jones, C., & R., R. P. (2002). Spatial information retrieval and geographical ontologies an overview of the SPIRIT project. New York, Computer Science & Informatics.
- Koo, S. O., Lim, S. Y., & Lee, S. J. (2003). Building an Ontology on Hub Words for Information Retrieval. En Proceedings of International Conference on Web Intelligence (WI'03), pp. 466–469. Halifax, Canada.
- Larín, R., & Garea, E. (2011). Relaciones topológicas para la generación automática de ontologías de aplicación en sistemas de información geográfica. *Ciencia. Tierra Y Esp.*, 2(12), 47–57.
- Larson, R. R. (1996). *Geographic Information Retrieval and Spatial Browsing* (pp. 81–124). Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
- Lin, X. (1997). Map Displays for Information Retrieval. *J. Am. Soc. Inf. Sci.*, 48(1), 40–54.
- Lopez, F. J., Silva, M. J., & Chaves, M. (2010). Linkable Geographic Ontologies. En Proceedings of the 6th Workshop on Geographic Information Retrieval, pp. 1-8. New York, NY, USA: ACM.
- Luna, J. M. F. (2001). Modelos de Recuperación de Información Basados en Redes de Creencias. Tesis Doctoral. Departamento Ciencias de la Computación e Inteligencia Artificial, Granada.
- Marchionini, G. (1995). *Information Seeking In Electronic Environments*. United States of America: Cambridge University Press. Disponible <http://assets.cambridge.org/97805214/43722/sample/9780521443722ws.pdf>
- Marcos, M. del C. (2002). Interacción en interfaces de recuperación de información: conceptos, metáforas y visualización. Ediciones Trea, S.L.
- Martínez, M. A., & Moreno, M. A. (2006). Descriptor Topológico para Mapas Topográficos. Instituto Politécnico Nacional. Tesis de Maestría, Centro de investigación en computación, Laboratorio de geoprociamiento, México.
- Mata, M. F. (2009). Recuperación y Ponderación de Información Geográfica desde Repositorios no Estructurados Conducidos por Ontologías. Instituto Politécnico Nacional. Tesis Doctoral. Mexico.
- Moya, D., & Macías, J. A. (2006). Autoría de Consultas para la Web Semántica Orientada al Usuario Final. Madrid: Ingeniería Informática. Disponible <http://ir.ii.uam.es/s5t/publications/>

interaccion06.pdf

México.

- Overell, S. E. (2009). Geographic Information Retrieval: Classification, Disambiguation and Modelling. University of London. Tesis Doctoral. London.
- Perea, J. M. (2010). Recuperación de Información Geográfica basada en múltiples formulaciones y motores de búsqueda. Universidad de Jaén. Tesis Doctoral. Departamento de Informática. Jaén. España.
- Peuquet, D. J., & Kraak, M.-J. (2002). Geobrowsing: creative thinking and knowledge discovery using geographic visualization. *Information Visualization*, 1(1), 80-91.
- Ping, D., & Yong, L. (2009). Building Place Name Ontology to Assist in Geographic Information Retrieval, En *Proceedings of FCSTA '09*, Vol. 1, pp. 306-309, Chongqing, China: IEEE Press.
- Sarabia, G. (2008). Búsqueda y ponderación de información contenida en bases de datos espaciales, utilizando jerarquías. Instituto Politécnico Nacional de México. Tesis de Maestría. México.
- Smart, P., Jones, C., Abdelmoty, A., & Hall, M. (2007, de diciembre de). Toponym ontology Design.
- Tello, E. V. (2010). Ordenamiento Basado en Ejemplos para la Recuperación de Información Geográfica. INAOE, Tonantzintla. Tesis Doctoral. Puebla, México.
- Tomai, E., & Spanaki, M. (2005). From ontology design to ontology implementation: A web tool for building geographic ontologies. En *Proceedings of AGILE*, pp. 281-290.
- Torres, M. J. (2007). Representación Ontológica Basada en Descriptores Semánticos Aplicada a Objetos Geográficos. Instituto Politécnico Nacional. Tesis Doctoral. Centro de Investigación en Computación,
- Tufte, E. R. (1990). *Envisioning Information*. Kansas, United States of America: GRAPHICS PRESS.
- Vera, F., & Garea, E. (2009). Alineamiento de Ontologías en el Dominio Geoespacial. Reporte Técnico Reconocimiento de Patrones. Serie Azul.
- Walker, A. R. (2007). Automated spatial information retrieval and visualisation of spatial data. University of Technology the Queensland. Tesis Doctoral. Queensland. Disponible <http://eprints.qut.edu.au/17258/>

Recibido: 25 de diciembre de 2014.

Aprobado en su forma definitiva:
19 de febrero de 2014

Manuel E. Puebla Martínez

Universidad de las Ciencias Informáticas
(UCI), La Habana, Cuba
Correo-e.: aizquierdo@ceis.cujae.edu.cu

José M. Perea-Ortega

Joint Research Centre (JRC), Institute for the
Protection and Security of the Citizen (IPSC),
Ispra, Italy
Correo electrónico:
jose-manuel.perea-ortega@jrc.ec.europa.eu

Alfredo Simón-Cuevas

Instituto Superior Politécnico José Antonio
Echeverría (CUJAE), La Habana, Cuba.
Correo-e.: asimon@ceis.cujae.edu.cu
