

# Acercamiento a un modelo de recuperación de información geográfica basado en técnicas de exploración y descripción semántica de datos geoespaciales

Manuel E. Puebla-Martínez<sup>1</sup>, José M. Perea-Ortega<sup>2</sup>, Alfredo Simón-Cuevas<sup>3</sup>

<sup>1</sup> Universidad de las Ciencias Informáticas. Carretera a San Antonio de los Baos, Km. 2 1/2. Torrens, La Lisa. La Habana, Cuba

`mpuebla@uci.cu`

<sup>2</sup> European Commission, Joint Research Centre (JRC) Institute for the Protection and Security of the Citizen (IPSC), Ispra, Italy.

`jose-manuel.perea-ortega@jrc.ec.europa.eu`

<sup>3</sup> Instituto Superior Politécnico José Antonio Echeverría. Calle 114, No. 11901. e/ Ciclovía y Rotonda, Marianao, La Habana, Cuba. CP 19390

`asimon@ceis.cujae.edu.cu`

**Resumen** Este artículo realiza una actualización del estado del arte sobre el modelo de Recuperación de Información (RI) basado en browsing, en particular su utilización en la Recuperación de la Información Geográfica (RIG) donde se le conoce como geobrowsing o browsing espacial. De la misma forma, se indagará en el aprovechamiento de la semántica de los datos geoespaciales en los modelos de RI. Finalmente, se expone una primera aproximación a una propuesta de modelo para la RIG basado en la estrategia del browsing y en la descripción semántica de los datos geoespacial. El modelo propuesto posee como base fundamental de conocimiento a una ontología del dominio geográfico o geontología, la misma será enriquecida por mecanismos semiautomáticos y transformada en una geontología de aplicación con el objetivo de modelar el conocimiento de una zona geográfica en particular.

**Palabras Claves:** Recuperación de Información Geográfica, Browsing o exploración, Semántica Espacial, y Ontología

**Abstract** This paper presents a survey on the current state of art of Information Retrieval models based on browsing and, specifically, on those used for Geographical Information Retrieval. The use of semantics from geospatial data applied to Information Retrieval models is also studied. This area is known by the research community as geo-browsing or spatial browsing. Moreover, this paper proposes a preliminary approach based on spatial browsing and semantic descriptions from spatial data to address the Geographical Information Retrieval task. This approach makes use of a specific ontology as geographic knowledge base which will be enriched semi-automatically. Then, it will be transformed into a geo-ontology application in order to model the knowledge of a particular

geographic area.

**Keywords:** Geographical Information Retrieval, Browsing, Spatial Semantic, Ontology

## 1. Introducción

El concepto de información espacial (o de información geográfica, o georreferenciada o geodatos, términos prácticamente sinónimos) es en la actualidad suficientemente conocido. Nos referimos con él a la información de todo tipo relativa a entes o eventos en la que se incluye la referencia a la localización de la misma sobre, o en las inmediaciones de la superficie de la Tierra. La referencia a la posición que ocupan dichos entes o donde suceden tales eventos puede adoptar distintas formas, como pueden ser las coordenadas geográficas (latitud y longitud), cartesianas en algún sistema de referencia cartográfico o, simplemente, una dirección postal que permita ubicar dicha posición en el espacio de forma inequívoca [1].

En el pasado, intercambiar la IG era tan simple como enviar mapas de papel o datos, a través del correo [2]. En la actualidad debido al enorme volumen de datos geoespaciales existentes, se ha hecho imprescindible el surgimiento de técnicas de RIG como mecanismos que permiten el acceso oportuno y preciso a dicha información. Por otra parte, como respuesta a la necesidad compartir el enorme volumen de datos geoespaciales por el mayor número de usuarios, se ha incrementado el uso de los Sistemas de Información Geográfica (SIG), como una herramienta integradora que permiten el almacenamiento eficiente, la recuperación y la visualización de información geográfica.

En la actualidad existe abundante literatura escrita sobre los diferentes modelos de RI, algunos de las cuales han sido utilizados en la RIG. No obstante, la investigación en técnicas de RI ha abordado esta problemática dividiéndola para su estudio en dos grandes estrategias: la interrogación o querying y la exploración o browsing [3].

Los modelos de RI basados en querying, se sustentan en una pregunta que debe ser formulada por el usuario final del sistema de RI. Para formular la interrogante el usuario debe dominar el lenguaje de consulta y estar consciente de lo que realmente necesita para poder generar una pregunta asequible al sistema.

Es interesante analizar el caso particular en el que el lenguaje de consulta fuese el Lenguaje Natural (LN). En este caso estaría solucionado el inconveniente relacionado con el dominio del lenguaje de consulta. Sin embargo, a pesar de que el LN brinda grandes posibilidades para describir la consulta, puede ser impreciso y ambiguo su procesamiento desde el punto de vista computacional. La complejidad inherente a este procesamiento y la precisión de los resultados dependen en gran medida de la habilidad y experiencia del usuario que construye la consulta. En los modelos de RI basados en querying también está presente la ambigüedad y vaguedad del lenguaje natural. Analicemos el siguiente ejemplo de consultas a un sistema de RI basado en querying. Tomemos como ejemplo al buscador Google y hagamos la siguiente consulta: Ríos en Granma.

Al analizar los resultados de Google se muestra que los mismos pertenecen a diferentes áreas y contextos, sin estar asociados directamente con los ríos que pertenecen a la provincia Granma. Por ejemplo: Dos Ríos - Provincia Granma (Pueblos de Cuba), Granma (Definición en EcuRed ). En (Fig.1) se muestran los resultados usando el motor de búsqueda Google.

The image shows a screenshot of a Google search for "Ríos en Granma". The results are as follows:

- Granma - EcuRed**: A result from EcuRed with the URL [www.ecured.cu/index.php/Granma](http://www.ecured.cu/index.php/Granma). The text describes the elevation of Granma and mentions rivers. A red circle highlights the title, and a red arrow points to the text "¿Coincedencia exacta?".
- Ríos de Cuba - EcuRed**: A result from EcuRed with the URL [www.ecured.cu/index.php/Rios\\_de\\_Cuba](http://www.ecured.cu/index.php/Rios_de_Cuba). The text discusses river lengths. A red circle highlights the title.
- Imágenes de ríos en granma**: A section showing four small images of rivers. A red arrow points from the "¿Coincedencia exacta?" text to this section.
- ¿Ríos?**: A red question mark icon.
- Dos Ríos Provincia Granma - Pueblos de cuba**: A result from [www.turismocursos.com/pueblos/cuba/pueblo.php?id=2836](http://www.turismocursos.com/pueblos/cuba/pueblo.php?id=2836). The text provides information about the town of Dos Ríos. A red circle highlights the title, and a red arrow points to the text "Pueblos de Cuba".

**Figura 1.** Resultados de la consulta Ríos en Granma usando el buscador Google.

En (Fig.1) se resalta el problema de la ambigüedad, donde Río es analizado como una palabra y no como un concepto geográfico descrito por propiedades y relaciones. Google utiliza la técnica de coincidencia de palabra o keyword-matching. Es decir se basan en encontrar palabras idénticas a las presentes en la consulta y como resultado se pueden obtener resultados irrelevantes en ocasiones, sobre todo cuando es usado de forma libre el LN para la construcción de las consultas.

Una alternativa a la problemática planteada en el ejemplo (Fig.1) la encontramos al hacer procesamiento semántico (concepto y relaciones) lo cual permite una recuperación más precisa, ya que evita la ambigüedad (una palabra puede ser ambigua mientras que un concepto no lo es). Por lo tanto, se requiere una estructura de datos que maneje conceptos. Para que a través de ellos se guíe hacia una mejor recuperación. Este tipo de estructuras son las ontologías [4].

En años recientes, diferentes grupos de investigación han enfocado sus esfuerzos en tratar de resolver el problema de la RIG. La precisión promedio de los sistemas de recuperación de información geográfica evaluados dentro del CLEF fue de 26.31 %, 27.25 %, 25.55 % y 23.70 % en los años 2005, 2006, 2007 y 2008 respectivamente. Los resultados muestran que el avance ha sido prácticamente nulo. De aquí la necesidad de nuevas alternativas para abordar el problema. [5]

En los modelos de RI basados en browsing, en oposición a los basados en querying, el usuario no tiene la necesidad de formular una pregunta, es decir, no tiene que dominar el lenguaje de consulta porque no lo hay. El mismo explora o navega en el banco de información seleccionando qué es lo que necesita, sin la obligación de concientizar previamente sus necesidades de información. En múltiples ocasiones nos sucede que estamos buscando información sobre un tema que no es específico o no somos capaces de expresarlo con precisión, bajo estas circunstancias la estrategia del browsing parece superar a la estrategia del querying debido al desconocimiento del usuario [3] .

Los nuevos modelos teóricos de la recuperación de la información aceptan al browsing como una estrategia básica e imprescindible para la búsqueda de información, especialmente en sistemas dirigidos a usuarios finales, cuyas necesidades de información suelen estar en un principio poco o mal definidas. Sin un intermediario que pueda ayudarle a clarificar y formular el problema, el usuario se verá obligado a llevar a cabo una larga y compleja modificación de su estrategia de búsqueda original a través de la interacción con el sistema y en este contexto el empleo del browsing es sin duda la mejor solución [6].

Para algunos tipos de información, el browsing parece ser la única alternativa de búsqueda. Por ejemplo, cuando deseamos hallar una imagen o un gráfico determinado dentro de un banco de imágenes, nos enfrentamos a una gran problemática: Cómo caracterizar la información almacenada en cada una de las imágenes del banco? Cualquiera de los modelos basados en querying exigirá indexar las imágenes a partir de la información que visualizan.

En el contexto de la informática moderna, el mapa ya no es sólo un producto final. Los mapas ahora se están utilizando de una manera diferente como una herramienta de autodirigido para derivar la información deseada de los datos geográficos. Esto, junto con la evolución de las Ciencias de los SIG y gráficos por ordenador, ha llevado a un nuevo campo: la visualización geográfica. Un tema central es cómo diseñar capacidades de visualización que, como proceso, faciliten el pensamiento creativo para descubrir previamente nueva información de grandes bases de datos. Los autores proponen el término geobrowsing para designar este proceso. [7]

Varios investigadores, entre ellos Tufte [8] y Peuquet [7], previeron que el geobrowsing puede ayudar a los usuarios en la formulación de consultas, elección de vocabulario, identificación de relaciones entre los datos y facilitación del pensamiento creativo para descubrir nueva información en grandes bases de datos. Sin embargo, en [9] se plantea que a pesar de los esfuerzos, las visualizaciones geoespaciales existentes continúan soportando solo el modelo más simple de lo que implica geobrowsing: marcadores con hipervínculos a los documentos. Este modelo simple, no favorece la formulación de consultas y el pensamiento creativo de los usuarios. Otras limitaciones que se reportan son las siguientes [9]:

1. No prestan mucha atención a la semántica presente en los documentos a recuperar. Documentos vinculados a los mapas están asociados a espacios semánticos que se pueden explorar y buscar de forma no dirigida. El modelo actual de geo-browsing no aborda esta cuestión y trata a los mapas como

”llanuras”, como caracterizaría Tufte en [8], donde todos los marcadores se pegan a una sola capa de mapa.

2. No explica cómo hacer frente a un gran número de marcadores que causan mapa de parásitos.
3. Se limitan a la recuperación y la selección de documentos relacionados con los mapas.

Los mapas han cambiado dramáticamente en la última década. Interfaces de programación de aplicaciones de mapas ahora cuentan con amplias bibliotecas de técnicas de interacción que permiten las visualizaciones para apoyar una variedad más amplia de tareas y actividades cognitivas. Estas actividades no se limitan a la recuperación y la selección de documentos relacionados con los mapas. Como dijera Chang y Rice en [10]: El browsing no se limita a la búsqueda de libros en los anaqueles de la biblioteca o registros bibliográficos en una base de datos, sino que también es aplicable a otros dominios humanos de la vida cotidiana. Es concebido como una estrategia de búsqueda, un patrón de observación, una técnica de visualización, e incluso una actividad recreativa.

El Browsing no sólo tiene el objetivo de ayudar al usuario a navegar y recuperar información, sino también el de ofrecer una visión global del conjunto documental. A través de esta vista global el usuario podrá predecir qué puede encontrar, e incluso adquirir nuevo conocimiento mediante la interiorización de la visualización. [11]

## 2. Modelos de Recuperación de Información

### 2.1. Modelos de RI basados en browsing

En [9] se propone un modelo conceptual para el geobrowsing en colecciones de textos. Dicho modelo visualiza tres espacios: el geográfico, el semántico y el cognitivo.

Por ejemplo, para la recuperación de lugares geográficos solo recuperan nombres de estados, provincias o ciudades con ayuda de un diccionario geográfico. Luego el diccionario geográfico es utilizado para expandir las consultas, generando nuevas consultas con nuevos términos.

Los nombres de lugares tienen dos representaciones en la visualización que hacen del espacio geográfico: una lista de los lugares con hipervínculos al mapa y los marcadores en el mapa. Mientras que el mapa muestra las distancias espaciales, la lista muestra las distancias en un orden alfabético. No se visualizan relaciones espaciales existentes entre los lugares recuperados.

En la visualización del espacio semántico se limitan a representar en diferentes tipos de gráficos la dimensión del banco de información con el que trabajan, por ejemplo, un gráfico de columnas que muestra por años la cantidad de publicaciones sobre el cáncer en China. Mientras que en la visualización del espacio cognitivo construyen representaciones de las acciones realizadas por los usuarios al navegar por los espacios geográfico y semánticos.

En resumen, en el modelo propuesto en [9] 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.

## 2.2. Utilización de la Semántica Espacial en la RIG

**2.2.1. Sang Ok et al (2003)** En este artículo Sang OK et al [12] proponen un método semiautomático para la construcción de una ontología usando agrupaciones de palabras (hub words).

El enfoque tiene como hipótesis la definición de ciertas palabras (hub words) las cuales están relacionadas con un conjunto de otras palabras. Estas hub words son determinadas por la frecuencia de aparición de los términos en un documento. Un detalle interesante es que se propone un proceso de construcción para una ontología usando hub words y un método automático para la extensión de la ontología (agregando relaciones). Además de que resaltan que la ontología propuesta se puede utilizar como un archivo de indexado en recuperación de la información, justificando esta afirmación en el hecho de que una ontología puede ofrecer mayor información semántica que los archivos de índices.

El artículo describe la construcción de la ontología y su extensión, así como el proceso de recuperación de información.

Este trabajo no explica cómo se establecieron las reglas de extracción para las ontologías (las cuales serían muy valiosas). Además de que las técnicas usadas son estrictamente de análisis de texto. [4]

**2.2.2. Miguel Félix Mata 2009** En esta tesis doctoral Miguel Felix Mata [4] propone una metodología integral para RIG y un método para ponderar los resultados de la recuperación. El éxito de la recuperación se basa en la integración de tres criterios de búsqueda para tres fuentes de datos geográficas heterogéneas: Ontologías geográficas, diccionarios de datos y archivos vectoriales.

Los resultados obtenidos demuestran que al combinar las tres fuentes de información y los criterios de búsqueda, se puede obtener información relevante que podría ser omitida al recuperar independientemente por cada fuente.

Un detalle importante es que esta investigación no aborda la construcción de las tres fuentes de datos geográficos, asume que existen y solo con el objetivo de validar la investigación construyen un caso de estudio.

**2.2.3. Jones, Christopher Bernard et al (2002)** Jones en [13] hace una breve reseña de las facilidades existentes para la RIG en la web antes de describir un conjunto de herramientas y técnicas que se están desarrollando en el proyecto SPIRIT: Recuperación de Información Espacial en Internet. Del grupo de técnicas que se están desarrollando en el proyecto SPIRIT hay dos que serán abordadas en el modelo propuesto, ellas son:

1. Ontologías para modelar la terminología geográfica: La ontología geográfica dentro del proyecto SPIRIT, modela el vocabulario y la estructura espacial de los lugares con el objetivo de ser utilizadas en la RIG.
2. Interfaces de usuarios multimodal que proporcionan entrada de textos y mapas de retroalimentación interactiva en el contexto de los documentos recuperados.

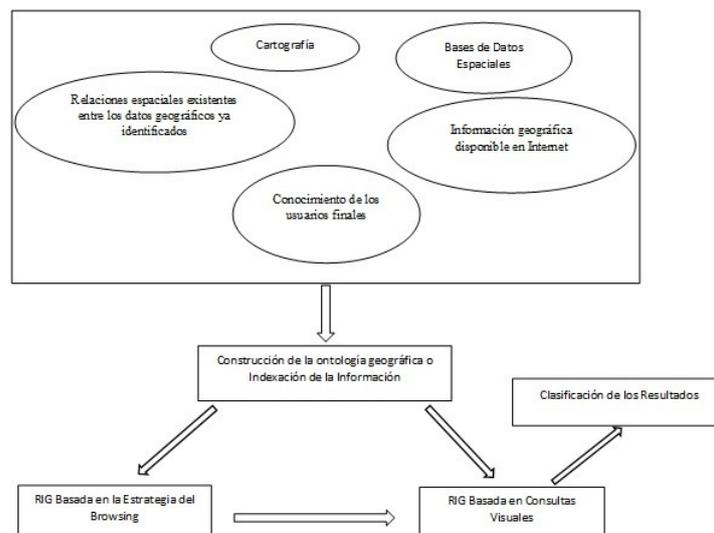
Se plantea que las interfaces multimodales son conocidas por ser una forma efectiva de interactuar con la información geográfica, aunque aún no se han realizado investigaciones en esta área para la RIG. Además, afirman que las ontologías ayudarán a eliminar la ambigüedad de la consulta y que el nivel de detalle del mapa se adaptará a los intereses del usuario y reflejará las estructuras jerárquicas del espacio geográfico codificado en la ontología [13]. Ambas ideas son defendidas en esta propuesta.

### 3. Modelo propuesto

Son varias las investigaciones que apuestan por el uso de la semántica espacial en el proceso de recuperación de información geográfica [4], [12], [13] ; mientras otros proponen el modelo del browsing con el mismo objetivo [9]. No obstante, en la bibliografía consultada hasta la fecha no se han identificado investigaciones que integren verdaderamente ambos enfoques en la RIG.

La integración de ambos enfoque en su totalidad, conducirá a un modelo de RIG que resolverá las insuficiencias antes sealadas a los modelos basados en un enfoque sintáctico y a los basados en una interrogante o querying. Además, con la incorporación de la estrategia del browsing, el modelo propuesto no solo se limitará a la recuperación de información sino que también podrá ser utilizado como un patrón de observación, una técnica de visualización, e incluso una actividad recreativa según Chang y Rice en [10]. También eliminará la ineficiencia para resolver una consulta bien definida, limitación que se le atribuye al browsing desde sus orígenes [14].

El modelo propuesto posee como base fundamental de conocimiento a una ontología del dominio geográfico que será enriquecida por mecanismos semiautomáticos con el objetivo de modelar el conocimiento de una zona geográfica en particular. La ontología geográfica interactuará con un SIG, en él se modelará la zona geográfica y se ubicarán geográficamente las instancias de conceptos de la geontología. El usuario final podrá explorar el mapa o la ontología según sus necesidades. La interacción entre el mapa y la geontología brindará interesantes resultados en la generación de consultas visuales por parte del usuario final, en la formulación de dichas consultas se podrán vincular los componentes de la geontología con características particulares de la zona geográfica. En la (fig.2) ilustramos un esquema que representa los componentes del modelo que se propone.



**Figura 2.** Esquema de componentes del modelo propuesto.

### 3.1. Conformación del Banco de Conocimiento o Indexación de la Información Geográfica

El objetivo general de la propuesta es recuperar la mayor cantidad de IG existente en una zona geográfica definida con antelación. 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:

1. La cartografía o bases cartográficas de la zona.
2. Bases de datos espaciales que modelan la zona.
3. Conocimiento de los usuarios finales sobre la zona.
4. Relaciones espaciales entre los datos geográficos ya identificados en la zona.
5. Información geográfica disponible en Internet sobre la zona seleccionada.

El modelo propuesto se fundamenta en el diseño y mantenimiento semiautomatizado de una ontología geográfica o geontología. Consiste en iniciar con una ontología del dominio geográfico e ir especializándola en una geontología de aplicación que conceptualice la zona geográfica seleccionada. La especialización de la ontología del dominio geográfico se hará a partir del análisis de las cinco fuentes de información mencionadas anteriormente. Hasta la fecha, con la excepción de la fuente de información número tres, se pretende automatizar el procesamiento de las restantes fuentes de información.

A continuación se profundiza en las ideas relacionadas con el procesamiento de las cinco fuentes de información:

1. Se pretende automatizar el procesamiento de la cartografía en formato vectorial, con el objetivo de extraer la mayor cantidad de información geográfica

relacionada con la zona que modelamos. La información extraída será incorporada a la geontología.

2. El objetivo final del modelo propuesto es ser aplicado en una plataforma para el desarrollo de SIG en la Web, para luego ser desplegado en las personalizaciones que se hagan de dicha plataforma SIG. A partir de la experiencia adquirida en las personalizaciones de las plataformas SIG, se ha identificado que las entidades o empresas que solicitan una SIG generalmente ya poseen una base de datos espacial. En caso de no ser así, esta fuente de información no estaría disponible para el modelo propuesto.
3. En el procesamiento de la tercera fuente de información se pretende diseñar y construir un editor de ontologías geográficas que facilite en gran medida el enriquecimiento semántico y sintáctico de la geontología a partir del conocimiento de los usuarios finales. Dicho editor se apoyará en visor SIG que permitirá simultáneamente la visualización de la geontología que se edita con las ubicaciones espaciales de las instancias de conceptos presente en dicha geontología, que al mismo tiempo representan información geográfica disponible en la zona con la que trabajaría el modelo propuesto. Un elemento importante que se pretende con el editor de geontologías es potenciar el conocimiento de la información geográfica disponible en la zona, por encima del conocimiento que deba tener los usuarios finales sobre la teoría de ontologías y el diseño de las mismas. Es decir, lograr que la edición de la ontología pueda ser realizada por usuarios finales que tienen un enorme conocimiento de la información geográfica existente en la zona pero no dominan en profundidad la teoría relacionada con las ontologías y sus lenguajes de diseño. El mantenimiento y actualización de las ontologías es uno de los grandes problemas que enfrenta hoy la comunidad científica relacionada con el tema. El editor de ontologías geográficas facilitará la solución de esta problemática. El modelo iniciará con una ontología del dominio geográfico que podrá ser detallada a través del editor de ontologías geográficas. El visor SIG también permitirá visualizar la topografía de la zona seleccionada, facilitando la identificación de las relaciones espaciales existentes entre los datos geográficos ya identificados en la zona: cuarta fuente de información, cuyo análisis será automatizado.
4. Un punto a considerar en los esquemas conceptuales para el dominio geográfico, son las relaciones espaciales, las cuales son abstracciones que ayudan a entender cómo en el mundo real los objetos se relacionan con otros [15]. Teniendo 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 Interfaz de Programación de Aplicaciones (API, por sus siglas en inglés) JTS (Java Topology Suite, por sus siglas en inglés).
5. En la actualidad son varias las fuentes de información geográfica disponibles en Internet, una de las más populares es GeoNames. El objetivo de esta fuente de información es enriquecer la geontología a partir de las diversas fuentes de información geográfica que contengan información valiosa para la zona que se modelará.

La indexación de la información es uno de los componentes más importantes en cualquier modelo de RI. Según [16], el primer paso que debe dar un S.R.I. 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. Por tanto, a partir de un documento se generará una representación del mismo, formada por una secuencia de términos de indexación, los cuales mantendrán lo más fielmente posible el contenido original del documento. A este proceso general se le denomina análisis automático de textos. En el modelo propuesto la información a recuperar no son textos sino información geográfica, es decir, información con coordenadas geográficas. La secuencia de términos de indexación que representará a cada dato será la descripción semántica que estará almacenada en la geontología. Resumiendo: la información estará indexada a través de descripciones semánticas del contenido de la misma.

### **3.2. Recuperación Basada en Browsing**

Una vez concluido el proceso de enriquecimiento de la geontología, la misma pasa a constituir la base sobre la que descansa el modelo de RI basado en browsing. Los usuarios finales podrán explorar el potencial semántico y sintáctico que posee la conceptualización de la zona geográfica, es decir, la geontología. El modelo propuesto debe garantizar la posibilidad de mostrar de manera simultánea la geontología y la cartografía que manipula el visor SIG, es decir, el espacio geográfico y el semántico. Las instancias de conceptos de la geontología deben permitir la posibilidad de ser geo referenciadas en el visor SIG a petición de los usuarios finales cuando se explore la geontología.

En la exploración de la información hay que garantizar que no se manifieste el segundo problema planteado por [9], 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 las coordenadas sean visibles y se esté en el nivel de zoom adecuado [17]. Además, la visualización de información geográfica en el mapa estará regulada por el componente de la geontología que esté seleccionado en cada momento. Por ejemplo, si tenemos una geontología de 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. Con esta estrategia se minimiza uno de los inconvenientes sealados al browsing en [14].

### **3.3. Generación de Consultas Visuales**

Una línea de investigación habitual para facilitar al usuario la expresión de consultas es el uso de lenguajes visuales. Se persigue facilitar el uso de los conceptos predefinidos en la ontología, ya que estas representaciones reducen el nivel de abstracción [18].

Además de explorar el espacio geográfico y el semántico, el modelo propuesto debe permitir el diseo de consultas visuales a partir de la selección, por parte de los usuarios finales, de elementos de la geontología y elementos del visor SIG.

Por ejemplo, se pretende recuperar todas la carreteras que se intersecten dentro de la provincia La Habana. Para resolver esta problemática se marcaría en la ontología el concepto carretera y la relación intersecten, además delimitaríamos en el visor SIG la zona que comprende la provincia La Habana. Con estos elementos seleccionados, el modelo propuesto debe generar una consulta que permita retornar la información esperada. Otro ejemplo de consultas sería el siguiente: Supongamos que pretendemos conocer los hospitales que se encuentra a una distancia cercana al lugar donde vive una persona con el objetivo de seleccionar uno de ellos para realizarse un examen. Marcaríamos en la geontología el concepto hospital y la relación cerca, en el visor SIG marcaríamos las coordenadas geográficas donde vive la persona. Luego el modelo deberá ser capaz de generar una consulta que retorne la información esperada. 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 acuerdo con lo que plantea Felix Mata en [4], 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. Note como en el segundo de los ejemplos anteriormente planteados, la relación espacial cerca debió ser definida en la geontología, ya sea por el usuario final del modelo a través del editor propuesto o de manera automatizada. En los modelos clásicos de RI basados en querying, la palabra cerca pudiera referirse a una distancia o a un tiempo. Para comprender la idea expresada en el párrafo anterior consideremos los siguientes ejemplos (1) El Hospital Militar está cerca de la vivienda de María y (2) La fecha del cumpleaños de Pedro está cerca. El primer ejemplo se refiere a distancia y el segundo a tiempo.
3. La incapacidad del browsing para resolver consultas bien definidas, carencia históricamente sealada a esta estrategia [14], quedará resuelta. Noten como la integración del Browsing con la estrategia de consultas visuales permite resolver la problemática antes planteada.

### 3.4. 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 [19].

En RI se requiere de una aproximación a un resultado cuando no existen instancias que correspondan con el criterio de consulta. Bajo este enfoque, las respuestas aproximadas tienen un grado de relevancia [4].

Consideremos la consulta Hospitales en el Oriente Cubano en la geontología de ejemplo que se ilustra en la (fig. 3). Un resultado relevante sería Hospital Celia Sánchez Manduley en Manzanillo, ver (fig. 4). Mientras que otro con menor relevancia sería Policlínico René Vallejo en Manzanillo, ver (fig. 5). En el ejemplo anterior Policlínico René Vallejo en Manzanillo aparece porque el concepto Policlínico (concepto al que pertenece la instancia René Vallejo) está relacionado semánticamente con el concepto Hospital mediante la relación semántica de Hiperonimia y además está relacionado espacialmente mediante la relación topológica pertenece con Oriente Cubano, el municipio Manzanillo pertenece al Oriente Cubano.

Hasta el momento solo se ha valorado la utilización de este método de evaluación de los resultados. En el futuro se analizarán el resto de los métodos propuestos en [19].

### Clasificación de los Resultados

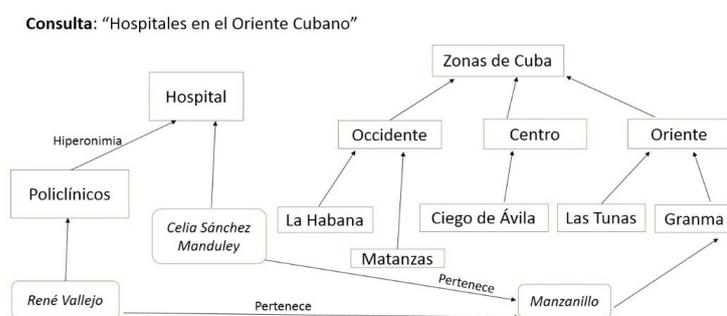


Figura 3. Geontología de ejemplo.

### Clasificación de los Resultados

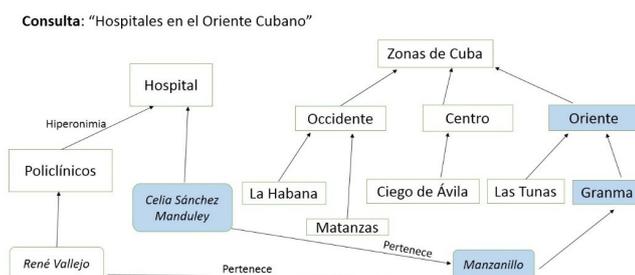
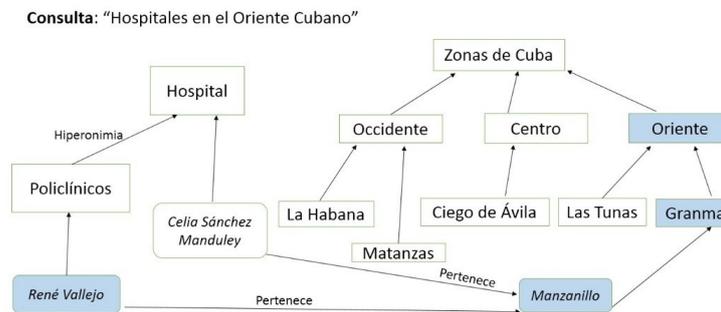


Figura 4. Resultado "relevante" de la consulta "Hospitales en el Oriente Cubano".

## Clasificación de los Resultados



**Figura 5.** Resultado “de menor relevancia” de la consulta “Hospitales en el Oriente Cubano”.

## 4. Conclusiones

A partir del análisis del estado del arte se concluye que los actuales modelos de RIG basados en la estrategia del browsing o exploración, no aprovechan las verdaderas potencialidades implícitas en la semántica de los datos espaciales. No permiten almacenar las relaciones semánticas existentes entre los datos y menos aún generar nuevas relaciones semánticas a partir de las existentes. Los investigadores que más profundizan en la semántica de los datos, como es el caso de Olga Bchel en [9], se limitan a interpretar el espacio geográfico y construir gráficas de información que constituyen una interpretación semántica del espacio geográfico. Otros como Miguel Felix Mata [4] sí aprovechan la semántica de los datos para resolver la consulta pero no abordan el proceso de conformación de la ontología, asumen que existe. Además, no utilizan la estrategia del browsing sino la estrategia del querying. El modelo de RIG propuesto posee las siguientes fortalezas:

1. Integra la exploración geográfica y la descripción semántica de los datos espaciales, favoreciéndose de los puntos positivos y mitigando los negativos de ambas técnicas.
2. Se alimenta de varias fuentes de información geográfica.
3. Brinda un soporte para almacenar y generar nueva información semántica asociada a los datos espaciales.
4. Utiliza una ontología geográfica como archivo de indexación de la información en el modelo de RI.
5. Automatiza gran parte del mantenimiento a la base de conocimiento del modelo de RIG.
6. Visualiza el espacio semántico y geográfico de la zona geográfica que modela. Facilitando el pensamiento creativo y la generación o descubrimiento de nuevo conocimiento por parte del usuario final. Por lo que se puede afirmar que no se limita a la RIG.

7. Propone un sistema de consultas visuales donde el usuario final solo debe seleccionar los elementos semánticos y sintácticos que formarán parte de su consulta, sin necesidad de ser un especialista en la manipulación de información semántica y espacial.
8. Extiende el enfoque sintáctico de los modelos de RI hacia el semántico y espacial.
9. Minimiza el clásico problema de la ambigüedad en las consultas de RIG.
10. Brinda resultados aproximados a la consulta realizada, apoyándose en la sem12. Disminuye la sobrecarga de información, otra de las deficiencias sealadas al browsing en [14].
11. Elimina la ineficiencia para resolver una consulta bien definida, limitación que se le atribuye al browsing desde sus orígenes según Marchionini [14].
12. Disminuye la sobrecarga de información, otra de las deficiencias sealadas al browsing en [14].

## Referencias

1. Garea-Llano, E.: Estado actual de la Interpretación Semántica de datos espaciales, Blue Series. Pattern Recognition. La Habana, Cuba. (2007)
2. Garea-Llano E., Oliva-Santos, R.: Integración Ontológica de Datos Metadatos y Conocimiento en Sistemas de Información Geográfica como Herramienta para la Interpretación Semántica de la Información Espacial. En VI Congreso Internacional de Geomática 2009. (2009)
3. Herrero-Solana, V., Hassan, Y.: Metodologías para el desarrollo de interfaces visuales de recuperación de información: análisis y comparación. Information Research, vol. 11, no. 3, 34. (2006)
4. M. F. Mata Rivera, Recuperación y Ponderación de Información Geográfica desde Repositorios no Estructurados Conducidas por Ontologías, México, 2009.
5. Villatoro-Tello, E.: Ordenamiento Basado en Ejemplos para la Recuperación de Información Geográfica, INAOE, Tonantzintla, Puebla. (2010)
6. Herrero-Solana, V.: Modelos de Representación Visual de la Información Bibliográfica, Granada, Espaa.(2000)
7. D. J. Peuquet y M.-J. Kraak, Geobrowsing: creative thinking and knowledge discovery using geographic visualization, Information Visualization, vol. 1, no. 80, 13. (2002)
8. E. R. Tufte, Envisioning Information, GRAPHICS PRESS. (1990)
9. O. Buchel, Redefining Geobrowsing, Ontario. (2013)
10. C. S. J. y R. R. E. , Browsing: A Multidimensional Framework, Annual review of information science and technology (ARIST), vol. 28, pp. 231-76. (1993)
11. Y. Hassan Montero, Visualización y Recuperación de Información, de II Encontro de Cincias e Tecnologias da Documentao e Informao., Escola Superior de Estudos Industriais e de Gesto (Vila do Conde). (2006)
12. S. Ok Koo, . S. Yeon Lim y S. Jo Lee, Building an Ontology on Hub Words for Information Retrieval, International Conference on Web Intelligence (WI'03). (2003)
13. C. B. Jones y R. R. Purves, Spatial information retrieval and geographical ontologies an overview of the SPIRIT project de Computer Science & Informatics, New York. (2002)

14. G. Marchionini, *Information Seeking In Electronic Environments*, United States of America: Cambridge University Press. (1995)
15. M. J. Torres Ruiz, *Representación Ontológica Basada en Descriptores Semánticos Aplicada a Objetos Geográficos*, Ciudad de México. (2007)
16. J. M. Fernández Luna, *Modelos de Recuperación de Información Basados en Redes de Creencias*, Granada. (2001)
17. R. R. Larson, *Browsing, Geographic Information Retrieval and Spatial*, Escuela de Postgrado de Biblioteca y Documentación de la Universidad de Illinois en Urbana-Champaign, California. (1996)
18. D. Moya y J. A. Macías.: *Autoría de Consultas para la Web Semántica Orientada al Usuario Final*, de VII Congreso de Interacción Persona Ordenador, Madrid. (2006)
19. G. Andogah, *Geographically Constrained Information Retrieval*. (2010)