

Hacia métodos de análisis de datos espaciales raster en el nivel semántico

Towards Raster Spatial Analysis Methods at the Semantic Level

Eydel Jaime, Rainer Larín y Eduardo Garea

Centro de Aplicaciones de Tecnologías de Avanzada,

La Habana, Cuba

{ejaim, rlarin, egarea}@cenatav.co.cu

Artículo recibido el 14 de febrero de 2011; aceptado el 30 de junio de 2011

Resumen. Los Sistemas de Información Geográfica (SIG), debido a sus aplicaciones en el análisis y recuperación de la información geo-gráfica, se han convertido en una necesaria herramienta para la ayuda a la toma de decisiones.

Con el surgimiento de la Geosemántica, un nuevo enfoque en la recuperación y análisis de información espacial se impone, recuperar y analizar la información por su significado semántico. El presente trabajo pretende hacer una síntesis de las diferentes técnicas y métodos a través de los cuales los SIG han venido recuperando la información espacial, llegando hasta las nuevas pro-puestas que incluyen la semántica, enfocándose en el modelo de datos raster, específicamente en la técnica de superposición de mapas y la problemática de la naturaleza heterogénea de los datos geográficos. Por lo que se plantea la necesidad de encontrar métodos que tomen en cuenta esta limitante y permitan trabajar con los datos mezclados de manera simultánea.

Palabras clave. Recuperación y análisis, ontologías y datos raster.

Abstract. Geographic Information Systems (GIS), due to their applications in analysis and geographic information retrieval, have become a necessary tool to aid decision making.

After the emergence of Geosemantic a new approach in the retrieval and analysis of spatial information is imposed, retrieve and analyze information by its semantic meaning. This paper aims to summarize the different techniques and methods by which GIS has been recovering spatial information reaching new proposals that include semantics, focusing on the raster data model specifically the map overlay technique and the problem of the heterogeneous nature of geographic data. As there is a need to find methods that takes into account this limitation and allows working with mixed data simultaneously.

Keywords. Retrieval and analysis, ontologies and raster data.

1 Introducción

Desde el surgimiento de los Sistemas de Información Geográfica (SIG) hasta la actualidad, estos sistemas han ido cobrando paulatinamente cada vez más importancia y se han convertido en una necesaria herramienta de análisis y recuperación de la información geográfica en el proceso de toma de decisiones a todos los niveles.

Con el acelerado crecimiento del volumen de datos almacenados en las bases de datos espaciales, cada vez se hace más difícil y complejas las operaciones de recuperación y análisis de la información de forma rápida y precisa.

Los métodos y algoritmos convencionales se centran sólo en las coordenadas o proyecciones geográficas, además de que se necesita de un buen dominio de la información geográfica por parte de los usuarios que operan estos sistemas. Por lo tanto no cuentan con un nivel de abstracción que les permita asemejarse al razonamiento que un humano realizaría a la hora de extraer una determinada información de entre un gran volumen de datos. Las tendencias actuales para el tratamiento de la información geoespacial están enfocadas en el desarrollo de técnicas basadas en la representación y recuperación de estos datos desde un punto de vista semántico.

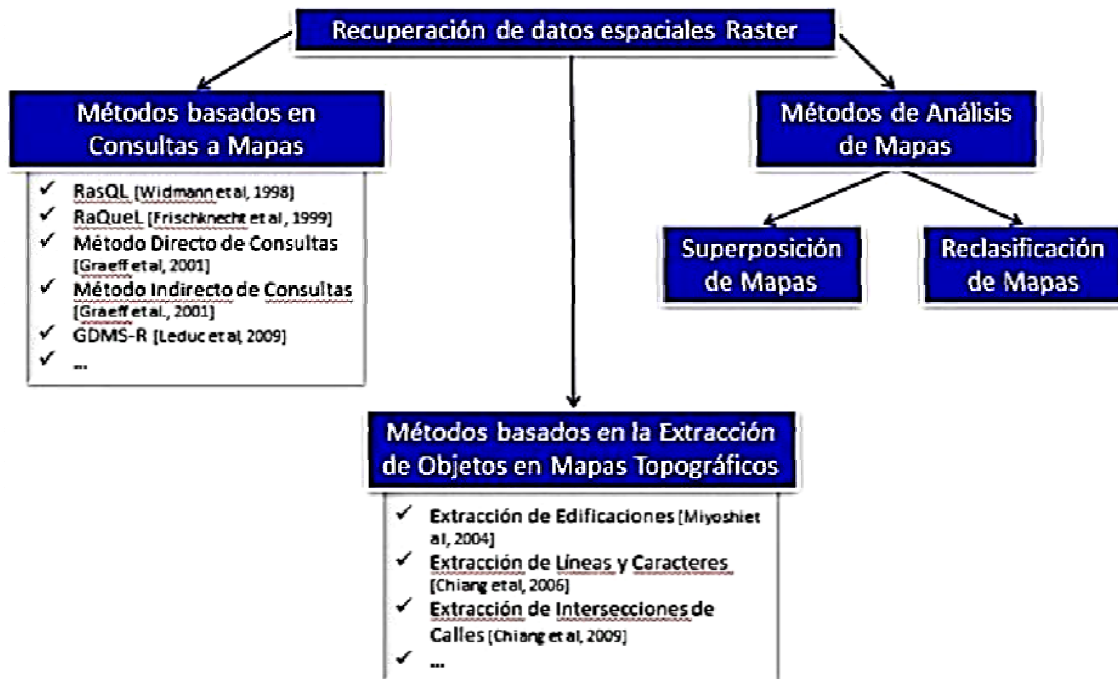


Fig. 1. Taxonomía sobre métodos de recuperación de datos espaciales raster

Sin embargo aunque las herramientas por excelencia para el tratamiento de la información geoespacial son los Sistemas de Información Geográfica (SIG) actualmente estos sistemas no disponen de las herramientas necesarias para el procesamiento de la naturaleza semántica subyacente en este tipo de información.

Con el lanzamiento de la Web Semántica, y posteriormente el surgimiento de la Geosemántica, un nuevo enfoque en la Recuperación de la Información espacial se impone, lograr recuperar la información geoespacial por su significado semántico y tomando como base las geo-ontologías.

Por otra parte una nueva generación de los SIG ha sido propuesta, se trata de los Sistemas de Información Geo-gráfica Gobernados por Ontologías (SIGGO) como nueva herramienta para recuperar la información espacial basándose en las geo-ontologías como forma de representación del conocimiento geográfico. Sin embargo todo esto aún constituye un tema abierto, en el cual todavía no se han logrado pasos significativos en la práctica, y sigue representando un gran reto para las

Comunidades de la Información Geoespacial de todo el mundo. En este trabajo nos centraremos específicamente en el tipo de datos espaciales raster y los métodos de recuperación y análisis. Veremos como la naturaleza de los datos geográficos no queda explícitamente representada en este modelo por lo que los métodos de recuperación y análisis actuales, sobre todo la superposición de mapas no favorecen la generación de nuevo conocimiento sobre la base de su valor semántico al codificar variables tanto cualitativas como cuantitativas utilizadas para describir los fenómenos u objetos que se están representando. El trabajo se estructura de la siguiente forma:

En la sección 2 de este trabajo se hace un análisis de las diferentes formas que existen en la actualidad para realizar la recuperación de datos espaciales raster, en la sección 3 se discuten los principales métodos de análisis raster y a continuación de ello, en la sección 4 se discuten los principales métodos de recuperación semántica de este tipo de información. En la sección 5 presentamos la propuesta para enfrentar el problema de la superposición

semántica de datos raster. En la sección 6 se expresan las conclusiones y los trabajos futuros.

2 Métodos de recuperación de datos espaciales raster

Basándonos en un estudio del estado actual de la recuperación de datos espaciales raster, se pudo representar una clasificación taxonómica de los principales métodos, así como las principales deficiencias detectadas en cada tipo de métodos. La Fig. 1 muestra como se pudo desglosar en 3 principales grupos de métodos, los Métodos basados en Consultas a Mapas, los Métodos basados en Extracción de Objetos en Mapas Topográficos y los Métodos de Análisis de Mapas

2.1 Métodos basados en consultas a mapas

Estos métodos han sido desarrollados tomando como punto de partida el conocido Lenguaje Estructurado de Consultas (*SQL, por sus siglas en inglés*), a partir del cual se han ido implementando nuevas variantes específicamente para el trabajo con los datos raster. Entre los principales problemas detectados en este tipo de métodos podemos mencionar que sólo se centran en las coordenadas y proyecciones cartográficas para las localizaciones de los objetos geoespaciales, es decir, si realizáramos una consulta a un mapa tendríamos que especificar las coordenadas geográficas exactas de la zona de la cual se quiere extraer la información. Por otra parte, realizan un procesamiento sintáctico de la consulta basándose exclusivamente en la sintaxis del lenguaje de consultas empleado, por lo que no se tiene en cuenta el valor semántico que se pudiera aprovechar de estos datos. O sea, se procesan los datos “por lo que contienen y no por lo que significan”. Además, es necesario resaltar que la mayor eficacia alcanzada en estos métodos va a estar condicionada por el grado de experticia en el dominio de la información geográfica por parte de los usuarios.

Dentro de este tipo de métodos tenemos dos variantes principales que se utilizan para recuperar datos espaciales raster, se trata del

Método Directo y del Indirecto. A continuación se profundiza en ellos.

2.1.1 Método directo

El Método Directo [Graeff y Carosio, 2001] se ejecuta directamente como un método de consulta tradicional, debido a que este no depende de un algoritmo o estrategias de segmentación de los objetos geográficos. Las estrategias de entrelazamiento de imágenes modificadas basadas en el conocimiento concierne a las características geométricas y topológicas de los objetos posibilitan consultar estructuras de datos raster de forma directa. De esta forma, no se necesita segmentar la imagen, y no se obtienen objetos desde la imagen, por lo que no se puede encuestar estructuras de datos raster a partes separadas de los objetos ni a más de un objeto segmentado.

En los documentos cartográficos se utilizan generalmente estrategias de entrelazamiento que proveen un extenso campo de algoritmos que se pueden usar para construir una herramienta de consulta. Su apariencia geométrica y topológica es bien conocida y declarada en leyendas de mapas que proveen un conocimiento base de patrones.

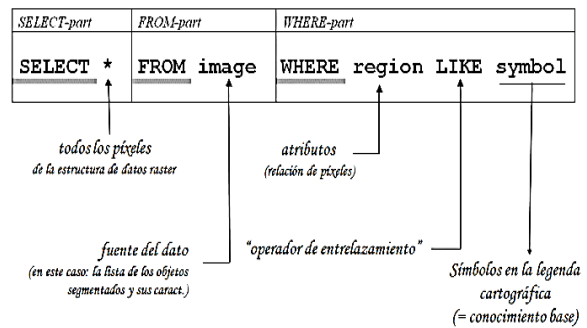


Fig. 2. Uso de las consultas en el método directo

En el Método Directo no se tiene una base de datos derivada de la imagen. La estructura de datos raster consultada es en sí el valor con que se cuenta. El atributo luego es llamado “región” debido a que la estructura de datos raster es concertada como regiones certeras de la imagen. Por otro lado, en vez de comparar exactamente los operadores matemáticos, se usan diferentes

operadores de entrelazamiento. La Fig. 2 muestra la estructura de consulta del Método Directo.

La elección del operador de entrelazamiento indicará el grado de ambigüedad que se mantenga. Entre los operadores de entrelazamiento, se han desarrollado cuatro operadores con varios grados de ambigüedad, especialmente varios límites de tamaños y orientaciones. La Tabla 1 muestra estos operadores así como los tamaños y las orientaciones posibles para cada operador.

Tabla 1. Operadores difusos del Método Directo [Graeff y Carosio, 2001]

Operador	Tarea de entrelazamiento Difuso
LIKE	patrones en un solo tamaño y una sola orientación
LIKE ROTATED	patrones en un solo tamaño, pero en todas las posibles orientaciones
SIMILAR	patrones en todos los tamaños posibles, pero sin rotar
SIMILIAR ROTATED	patrones en todos los tamaños y orientaciones posibles

Debido a que las estructuras de imágenes raster con el mismo significado no son representadas exactamente como el patrón de la leyenda, los operadores de entrelazamiento deben ser poco difusos. En los documentos cartográficos existen una gran cantidad de señalizaciones, las cuales están localizadas en una sola orientación y un sólo tamaño, y existen otras con diferentes tamaños y orientaciones. Entre las señalizaciones con diferentes tamaños y orientaciones, la orientación varía continuamente, pero los tamaños están meramente escalonados en pocos tamaños posibles [Graeff, 2002].

2.1.2 Método indirecto

El Método Indirecto [Graeff y Carosio, 2001] se basa en consultar las características de los objetos que son extraídos inicialmente de la imagen de un mapa topográfico escaneado. Después de un proceso de segmentación de la imagen y posterior generación de objetos por medio de características radiométricas, se puede

obtener una lista de objetos segmentados. A cada objeto segmentado se le puede asignar características como área, perímetro, densidad, así como muchas otras. Para extraer objetos de una clase de objetos es requerida una combinación ideal de características a identificar en la imagen. Luego de seleccionar por medio de consultas, las características relacionadas con el objeto que se encuentra en la base de datos, se puede visualizar el objeto seleccionado en la imagen. Este método es llamado indirecto debido a que la selección es realizada de las características relacionadas con el objeto que se encuentra en la base de datos, y no sobre la propia imagen.

Tanto en el método directo como en el indirecto se sigue la formulación conocida del lenguaje SQL utilizado para consultar bases de datos. De esta forma, las consultas comienzan por una declaración *SELECT*, seguida por la indicación del recurso en la parte *FROM* y la seleccionando la especificación de los parámetros a la parte *WHERE*. El Método Directo consume más tiempo de ejecución que el Método Indirecto debido a las herramientas de entrelazamiento implementadas en los algoritmos. Por lo tanto estas estrategias tienen que ser modificadas.

Para la implementación del Método Indirecto se emplea el lenguaje *RaQueL*, desarrollado por *Frischknecht* en el año 1999 [Graeff y Carosio, 2001] específicamente para la recuperación de datos raster, el cual se basa en la misma funcionalidad del lenguaje SQL. No obstante, después de ejecutadas las consultas, el resultado no es mostrado como una tabla de datos, sino visualizando directamente el resultado en un mapa y en donde todos los objetos seleccionados son mostrados con un color diferente al usado por el mapa, de manera que contraste. La fuente de datos está en la base de datos de las características de objetos, la cual hay que crear preliminarmente. Entre las características de los objetos, se usan atributos, en donde el lenguaje *RaQueL* calcula el área, perímetro, densidad, tamaño de los parámetros, orientación de los parámetros (ejes principales) y otra gran cantidad de características geométricas, las cuáles son independientes de la estructura raster.

Tabla 2. Tipos de análisis en un SIG Raster (Durand 2004)

Tipo de operación	Tipo de Localización		
	Píxel	Vecindad	Zona
Sólo opera con los atributos temáticos	valor típico	valor típico	valor típico
	valor no característico	valor no característico	valor no característico
	variedad	variedad	variedad
	comparación a un valor	comparación a un valor	Comparación a un valor
	reclasificación	reclasificación	reclasificación
	síntesis estadística	síntesis estadística	síntesis estadística
Sólo emplea los aspectos espaciales	posición geográfica	posición geográfica	posición geográfica
		variabilidad espacial	variabilidad espacial
		forma	forma
		tamaño	tamaño
		distancia y proximidad	
		dirección	
		vecinos a	

2.2 Métodos basados en la extracción de objetos en mapas topográficos

Estos métodos han sido desarrollados para la extracción de objetos en mapas topográficos digitalizados, obtenidos a partir de imágenes aéreas, satelitales, y de la rasterización de datos espaciales en formato vectorial. Estos métodos se centran en la extracción de determinados objetos contenidos en los mapas ya que los algoritmos son programados para que encuentren rasgos específicos dentro de dichos mapas. Ejemplo de esto son los trabajos desarrollado trabajos para la extracción de edificaciones [Miyoshi, 2004], extracción de líneas y caracteres [Chiang, 2006], extracción de intersecciones de carreteras [Chiang, 2009], entre otros. Estos métodos también realizan el proceso de extracción sin tener en cuenta la semántica asociada a los objetos espaciales, por lo que tampoco se tiene en cuenta el valor semántico que se pudiera aprovechar de estos datos.

2.3 Métodos de análisis de mapas

Debido a que el análisis espacial raster constituye la herramienta fundamental del proceso de toma de decisiones, nuestra trabajo se centrará en los

métodos de análisis de superposición de mapas y más adelante nos centraremos en explicar su esencia y principales deficiencias detectadas que actualmente presentan.

3 Análisis espacial raster

El proceso de análisis se considera como el conjunto de métodos que permiten al usuario del dato generar nueva información partiendo de la ya existente en función de un objetivo. El análisis espacial abarca un conjunto de técnicas y procedimientos para el estudio de las características de los datos espaciales [Bosque-Sendra, 2002]. Por otra parte este mismo autor plantea que en gran medida este tipo de análisis se ha ido desarrollando apoyándose en el uso de la metodología estadística y su adaptación en la integración con el estudio de los datos espaciales y que también se pueden incluir en él determinados procedimientos que analizan, de forma exclusiva, las características geométricas de los objetos geográficos sin considerar de forma simultánea, los valores temáticos asociados a dichas características.

En los datos espaciales raster los datos siempre se analizan teniendo en cuenta su representación espacial y los procesos de análisis se clasifican en aquellos que se realizan

sobre el conjunto de uno o varios mapas raster y los que se realizan sobre determinadas localizaciones especiales que es posible en él: el píxel, la vecindad y la zona, véase Tabla 2.

En las Fig. 3a, 3b y 3c se muestra el ejemplo de estos tipos de localización. Estos

procedimientos coinciden en gran medida con varios métodos empleados para la modelación cartográfica de sistemas vectoriales o, otros pocos casos, con el análisis de la componente espacial [Durand, 2004].

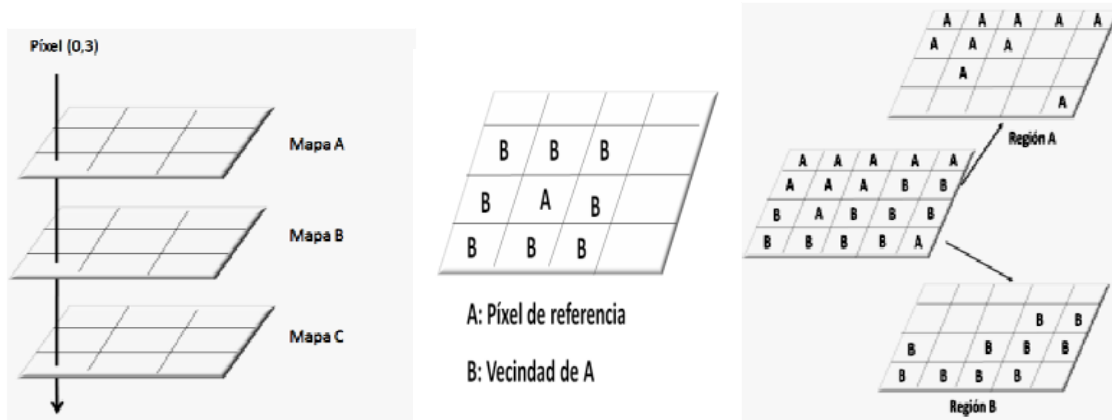


Fig. 3a. Localización de un “píxel”. **3b** Localización “vecindad” de un píxel. **3c** Localización “regiones o zonas”

Entre los métodos u operaciones analíticas empleados para analizar los datos espaciales raster que más se utilizan, se encuentran la superposición y la reclasificación de mapas [Durand, 2004]. En la subsección 3.1 y 3.2 se abordarán con mayor profundidad acerca de estos métodos analíticos.

3.1 Métodos de superposición de mapas

El método de superposición de mapas consiste en la combinación de dos o más capas o mapas en donde cada celda (píxel) de cada capa o mapa referencia la misma localización geográfica y con lo cual se genera una nueva capa o mapa que contiene la combinación de la información de las capas o mapas de entrada. En la Fig. 4, se muestra un ejemplo de cómo mediante la superposición de un mapa de relieve (orografía) que contiene los valores de alturas del terreno, con un mapa de recursos eólicos que contiene los valores de velocidades del viento, se pueden identificar posibles zonas factibles para la

ubicación de Parques Eólicos en el Municipio Especial Isla de la Juventud perteneciente a la República de Cuba.

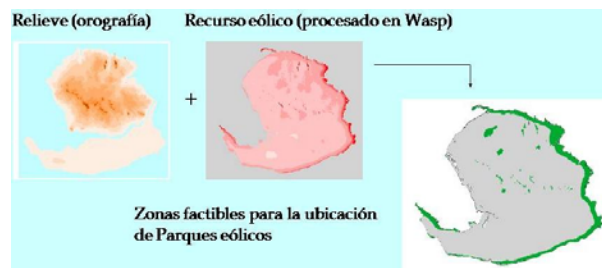


Fig. 4. Ejemplo de superposición de mapas. Se realiza una operación lógica (AND) entre los valores de las celdas de cada mapa combinando valores de altura del terreno y valores de velocidades del viento

Según Bosque-Sendra [Bosque-Sendra, 2002], la operación de análisis de superposición de mapas se puede realizar de dos formas:

- Análisis local píxel a píxel: Se realiza teniendo en cuenta de forma aislada píxel a píxel de una o varias capas de un mapa raster.
- Análisis en zonas o regiones: Se realiza teniendo en cuenta un conjunto de píxeles que se encuentren de forma contigua en una o varias capas de un mapa raster.

3.2 Reclasificación de mapas

Esta operación se realiza sobre un solo mapa. Al igual que las operaciones de superposición, constituye una de las técnicas más usadas en el análisis de los datos espaciales raster debido a que posibilitan la búsqueda y recuperación de forma selectiva de los datos raster almacenados en las bases de datos espaciales. La reclasificación es una técnica de generalización utilizada para reasignar valores en una capa de entrada raster como pueden ser la posición, el valor, la forma, el tamaño, la medida de los píxel, el grado de contigüidad y así crear una nueva capa de datos. La reclasificación cambia el valor de las celdas de entrada trabajando con una base "celda-a-celda" dentro del área de análisis. La Fig. 5 muestra un ejemplo de reclasificación de mapas en un Modelo Digital de Elevación (DEM por sus siglas en inglés) donde a partir de un mapa que contiene inicialmente 9 intervalos de alturas del terreno, se realiza una generalización de los intervalos con lo cual se reclasifican en 5 nuevos intervalos.

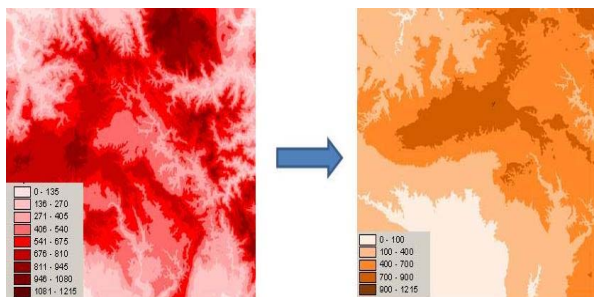


Fig. 5. Reclasificación de mapas. Reclasificación de un DEM en 5 intervalos de altura del terreno

Esta operación posibilita convertir los datos en escala de intervalos y de razón a una clasificación ordinal, para el modelado de la idoneidad de usos del suelo, utilizando el álgebra de mapas. Tiene el beneficio adicional de reducir el tamaño del archivo de las capas raster. La Fig. 6 muestra un ejemplo de un mapa que contiene n tipos de suelos y al que se le realiza una reclasificación binaria donde el nuevo mapa de salida contendrá sólo dos tipos de valores de suelos, valor 0 para los suelos no aptos y valor 1 para los suelos aptos. En este caso se buscan los suelos aptos para el Cultivo del Café en Cuba.

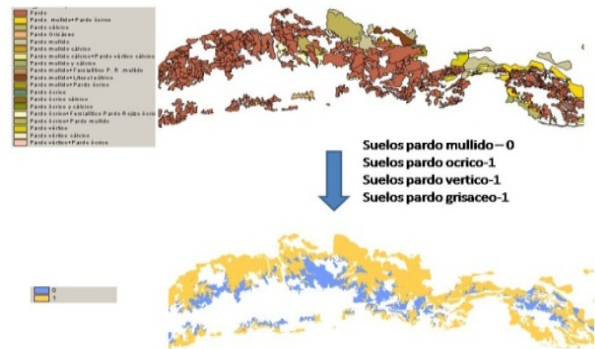


Fig. 6. Reclasificación de mapas. Mapa de suelos aptos para el cultivo del Café, se reclasifican con valores binarios

Los valores generalmente son números (enteros, flotantes) o expresiones Booleanas, o sea realizar la codificación de rasgos no numéricos es un procedimiento que puede considerarse inadecuado.

Por ejemplo: Un mapa raster de tipos de suelos que codifique el tipo Ferralítico Rojo = 1, Pardo = 2, Esquelético = 3. Se sabe que la media entre 1 y 3 es 2, pero para los edafólogos, pudiera ser difícil aceptar que en un área donde hay suelos ferralíticos y pardos que son muy fértiles, pueda catalogarse como un área, medianamente apta para cultivos varios. De este modo, los operadores para comparar atributos con valores son los mismos usados en la Matemática. Esta es la principal limitante tanto de la recuperación como del análisis de mapas

raster, pues como es sabida la naturaleza de los datos geográficos puede ser heterogénea, con variables tanto cualitativas como cuantitativas utilizadas para describir los fenómenos u objetos que se están representando.

4 Recuperación semántica de datos espaciales raster

La recuperación de la información espacial por su significado semántico contenida en los datos raster es aún un tema abierto, en el que se continúan realizando estudios en busca de dar pasos significativos. Las mayores aproximaciones hasta la actualidad en este sentido han llegado hasta la representación semántica de estos datos. Entre una de las mayores aproximaciones en este sentido lo constituye el trabajo realizado en [Oliva, 2009], en el cual se propone una metodología que presenta una arquitectura basada en estructuras de datos de búsqueda espacial (se escogió R-tree para datos vectoriales y Quadtree para datos raster) en dependencia del tipo de datos a anotar y como soporte para la persistencia de las anotaciones semánticas de datos geográficos tomando como bases sólidas las geo-ontologías. Este trabajo deja sentadas las bases para que se continúe trabajando en métodos eficientes para la recuperación de los datos una vez que se encuentren realizadas las anotaciones semánticas.

Entre otros de los trabajos que se han desarrollado y que se aproxima aún más a la recuperación semántica de los datos espaciales raster, está el realizado por [Quintero, 2009], que propone una metodología para realizar una representación semántica de datos espaciales raster, enfocándose en los modelos digitales de elevación como caso particular de los datos espaciales raster. Esta metodología se basa en tres etapas: Conceptualización, Síntesis y Descripción. La primera etapa que es la Conceptualización consta de tres partes fundamentales: la conceptualización del dominio geoespacial o de alto nivel, la conceptualización del dominio específico de un Conjunto de Datos Espaciales Raster (CDER) sobre el cual se vaya a operar, y la conceptualización de la aplicación a través de la cual se implementará la

representación de los datos. Esta primera etapa, mediante dos grandes ontologías; una ontología para el Dominio Geográfico y la otra para el Dominio de las Formas del Terreno; conceptualiza el dominio geográfico. Una segunda etapa, que es la Síntesis, en la cual mediante algoritmos, extraen las características del CDER, teniendo en cuenta al dominio en el que éstos son analizados. Y por último y no la menos importante, la tercera etapa, en la cual se determinan los extractos, que según este trabajo, son segmentos del CDER identificados con algún concepto en general, y los cuales se describen de acuerdo a la Conceptualización. Con esta metodología se logra describir que es lo que se encuentra contenido en un conjunto de datos espaciales raster, mediante un lenguaje muy cercano al natural utilizado por los humanos, pero a la vez en un formato legible computacionalmente que garantiza la integración y la interoperabilidad entre otros formatos semejantes. Sin embargo deja pendiente y abierta las posibles soluciones para recuperar y analizar la información una vez que se haya realizado la representación de los datos.

4.1 Ontologías de representación de datos para la integración semántica de datos geográficos

Larin-Fonseca y Garea-Llano [Larin y Llano, 2010] proponen un nuevo tipo de ontología, las Ontologías de Representación de Datos (ORD), la cual se genera automáticamente mediante técnicas de clasificación a partir de distancias y con el uso de medidas de disimilitud. Las ORD representan la abstracción semántica de los datos geográficos aportando un mayor grado de especialización y por tanto una mayor granularidad a partir del complemento entre la ORD y las Ontologías de Nivel Superior (ONS).

La ORD puede representarse como una especialización tanto de Ontologías de Aplicación como de las de Dominio o de Tarea. Esto depende del nivel del detalle conceptual con el que se estén manejando los datos. Esta Ontología constituye una capa intermedia entre los datos y las ONS, en la cual se representan de forma explícita tanto la semántica subyacente en

los datos como las relaciones existentes entre ellos.

La ORD es en esencia una ontología dinámica ya que su estructura, términos y relaciones siempre van a depender de los datos incluidos en el sistema. El uso de la ORD permite la integración de datos heterogéneos y proporciona una mayor riqueza semántica a partir del complemento que se genera entre la ORD y la(s) ONS utilizadas. Este complemento va dirigido en ambos sentidos, de la ONS a la ORD y viceversa. Si observamos la Fig. 7, en la zona 2 se muestra como a través de la capa de Interconexión se relacionan los conceptos de las ONS con los objetos representados en la ORD.

Esto se explica desde el hecho de que por un lado, la ONS puede contar con información precisa de los datos integrados al sistema lo que implica contar con mayores niveles de especialización que aumentan el nivel de precisión semántica en la ontología. Por el otro lado la ORD dispone de un contexto sobre el cual es posible procesar los datos, que es proporcionado por las ONS. Aquí se pone de manifiesto un proceso de sinergia, en el cual los resultados obtenidos a través del uso de estos dos tipos de Ontologías (ORD y ONS) son mejores que la suma de los resultados obtenidos por cada una de ellas por separado.

Basado en lo planteado anteriormente proponen que el uso de la ORD sea siempre en conjunto con al menos una ONS de manera que el usuario puede gestionar los datos geográficos incluidos en el sistema con la posibilidad de agregar nuevos datos, modificar datos y/o eliminar datos, este dinamismo de los datos incluidos en el sistema provoca cambios estructurales en la ORD ya que esta, como se ha mencionado anteriormente, es generada automáticamente a partir de estos datos. Cambios en estos datos implican cambios en los términos y relaciones existentes en la ORD, es por esta razón que plantean que esta es una ontología dinámica y cambia según cambian los datos. Otro aspecto a destacar es que todos estos cambios estructurales que puede sufrir la ORD no implican absolutamente ningún cambio en la(s) ONS manteniendo esta(s) su estructura original, cada uno de estos cambios ocurren por debajo de la capa de interconexión, véase Fig. 7.

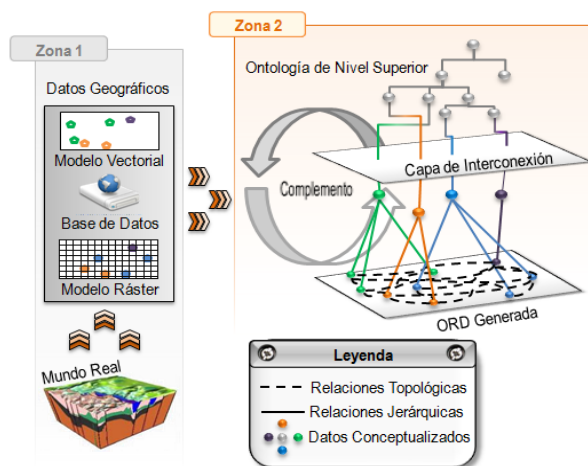


Fig. 7. Representación gráfica de los diferentes modelos en los que pueden ser almacenados los datos geográficos (Zona 1). La Zona 2 muestra la proyección en el Espacio Semántico de estos datos y las relaciones existentes entre ellos. También se muestran la ORD y la ONS complementadas entre sí (Larin y Garea 2010)

Esta ontología puede además contener nuevos conceptos que no existen de forma explícita ni en las ONS ni en los datos. Estos nuevos conceptos son aprendidos a partir del procesamiento realizado a los datos en el proceso de generación de la ORD aplicando técnicas de agrupamiento.

5 Propuesta para enfrentar el problema de la superposición semántica de datos raster (SSDR)

Sobre la base de las premisas analizadas y de las deficiencias detectadas es posible plantear una estrategia para el desarrollo de un método de superposición de mapas raster sobre una base semántica.

La estrategia se basa en la representación semántica de los datos raster mediante ORD [Larin y Llano, 2010] y su combinación con métodos desarrollados para el Reconocimiento Lógico Combinatorio de Patrones que trabajan con las descripciones mezcladas e incompletas [Ruiz-Shulcloper, Martínez-Trinidad y Guzmán-

Arenas, 1999], [Martínez-Trinidad y Guzmán-Arenas, 2001].

En esta nueva propuesta presentamos una modificación al trabajo descrito por Larin-Fonseca y Garea-Llano, en el que insertamos nuestra idea de solución, con lo cual se pretende modificar el análisis espacial raster al lograr integrar las variables tanto cuantitativas como cualitativas de los datos geográficos y dar origen a nueva información espacial la cual será representada a través de ontologías.

Al igual que el trabajo de Larin-Fonseca y Garea-Llano, partimos de una fuente de datos geográficos, que en nuestro caso particular serán mapas o capas del modelo de datos raster, de los cuales mediante un proceso de Síntesis propuesto por en el trabajo realizado por Quintero [Quintero, 2009] para los Modelos Digitales de Elevación (DEM), reutilizaremos los algoritmos de extracción implementados en dicho trabajo.

Basado en lo planteado anteriormente proponemos que las operaciones de superposición se realicen en el nivel semántico a través de la combinación de los Nodos de representación de datos (NRD), los cuales serán creados a partir de los objetos geográficos discretos y de las representaciones de variables continuas de distribución espacial que estén contenidos en los mapa raster y que serán representados en la ORD. Los NRD se presentan en dos variantes, NRD-O para objetos geográficos los cuales presentan las tres subestructuras (*ctemat*, *cespac* y *ctemp*) en el espacio semántico y NRD-C para clases abstractas de objetos geográficos, donde estas carecen de la subestructura *ctemp* para la componente espacial, ya que esta variante representa la abstracción semántica, en un nivel mayor, del conjunto de datos que ella agrupa, en este caso de los NRD-O. En las subsecciones 5.1 y 5.2 se explicaran en detalles las estructuras internas de los NRD para los objetos geográficos discretos y de los NRD para la representación de variables continuas de distribución espacial para el modelo de datos raster.

Una vez que en el nivel semántico se encuentren representados todos los NRD, en nuestra propuesta añadiremos una etapa posterior a la concebida por Larin-Fonseca y Garea-Llano. Procederemos a realizar los

métodos de SDR, los cuales a diferencia de los métodos convencionales de superposición de mapas, que realizan la operación entre los píxeles con la misma posición geográfica y las mismas dimensiones de los mapas o capas a procesar, esta SDR se realizará entre los NRD presentes en el espacio semántico. En la subsección 5.3 se mostrarán los pasos necesarios para llevar a cabo este método.

5.1 Estructura interna de los NRD para la representación de objetos geográficos discretos en el modelo raster

La estructura de los NRD correspondientes a los datos raster para el caso de la representación de objetos geográficos discretos, como puede ser un mapa aptitud de suelos, estaría basada en la representación de las características temáticas, espaciales y temporales de los objetos presentes en el mapa raster formado por conjuntos de píxeles o celdas, a través de tres subestructuras (*ctemat*, *cespac* y *ctemp*) en el espacio semántico. Estos NRD representan la abstracción semántica tanto de los objetos geográficos como de las clases aprendidas en la etapa de procesamiento de los datos. La Fig. 8 muestra un mapa con objetos geográficos discretos binarios donde sólo se toman valores de unos y ceros.

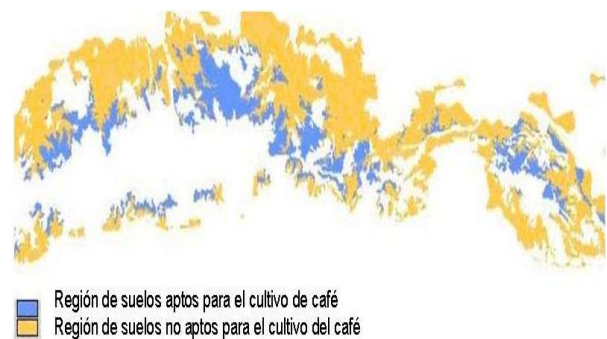


Fig. 8. Ejemplo de un mapa raster representando regiones de aptitud para el cultivo del café

Por tanto la estructura interna de los NRD de los datos raster discretos de manera general está determinada de la siguiente manera:

El NRD puede ser representado por la 5-tupla (id, rel, ctemp, ctemat, cespac) donde:

id: Es el identificador del NRD.

rel: Es el conjunto finito de relaciones.

ctemp: Representa las características temporales de la entidad geográfica representada.

ctemat: Representa las características temáticas de la entidad geográfica representada.

cespac: Representa las características espaciales de la entidad geográfica representada.

La estructura ctemp puede representarse por la 2-tupla (eventos, rel) donde:

eventos: Se define como un conjunto de acciones o hechos que afectan a una o varias propiedades de la misma entidad geográfica en el tiempo. (p.e: cambios de posición, forma etc.).

rel: Es el conjunto finito de relaciones temporales que puede tener la entidad geográfica con otra(s) entidad(es) geográfica(s).

La estructura eventos de la componente ctemp puede representarse por la 5-tupla (id, nombre, descrip, tiporeg, granul) donde:

id: Es el identificador del evento.

nombre: Nombre del evento.

descrip: Representa la descripción del evento.

tiporeg: Define el registro del tiempo de existencia del evento.

granul: Define el tipo de granularidad en el tiempo (Año, Mes, Día, etc.).

La estructura ctemat puede representarse por la 2-tupla (nombre, (...)) donde:

nombre: Nombre de la entidad geográfica representada.

(...): Propiedad Temática de la entidad geográfica representada que la definen semánticamente. Aquí es donde se realizará las operaciones de análisis lógico combinatorio de los datos de diferente naturaleza.

La estructura cespac puede representarse por la 3-tupla (ubic, repres, esca) donde:

ubic: Define la ubicación espacial de la entidad geográfica representada en el plano (coordenadas).

repres: Define el tipo de objeto geométrico en el que está representado la entidad (conjunto de celdas). Esta propiedad de la tupla es la que se tomará para saber sobre que NDR en la ORD se realiza la operación de superposición

esca: Define el tipo de escala de representación de la entidad geográfica a partir de la resolución de las celdas que conforman el mapa raster.

5.2 Estructura interna de los NRD para la representación de variables continuas de distribución espacial en el modelo raster

Para la estructura de los NRD correspondientes a los datos raster en el caso de la representación de variables continuas, como puede ser un mapa de valores de altura del terreno (véase Fig. 9), estaría basada en la representación de las características temáticas, espaciales y temporales de las celdas presentes en el mapa raster, a través de tres subestructuras ya descritas (*ctemat*, *cespac* y *ctemp*) en el espacio semántico.



Fig. 9. Ejemplo de un mapa raster representando regiones de aptitud para el cultivo del café

En este caso la estructura *cespac* estará representada por la 3-tupla (*ubic*, *repres*, *esca*) donde:

ubic: Define la ubicación espacial de la celda representada en el plano, (coordenadas).

repres: Define el tipo de objeto geométrico en el que está representado la entidad (celda). Esta propiedad de la tupla es la que se tomará

para saber sobre que NDR en la ORD se realiza la operación de superposición
esca: Define la resolución espacial de la celda representada en el NRD.

5.3 Superposición de los NRD

En esta operación entraría a jugar un papel preponderante los métodos de análisis lógico combinatorio los cuales darán la posibilidad de realizar una diversidad de operaciones matemáticas y lógicas entre las representaciones semánticas de los mapas consultados y permitirá trabajar con las descripciones mezcladas de manera simultánea, ya que frecuentemente y de forma errónea, la representación de los objetos es considerada como una secuencia de valores exclusivamente numéricos o categóricos. Sin embargo, existen muchos problemas reales de Reconocimiento de Patrones donde, en la descripción de cada objeto, ambos tipos de valores aparecen mezclados (se puede ver, por ejemplo, las bases de datos del Repositorio de la UCI [Merz, C. J. y Murphy P. M., 1998]).

En los intentos por resolver esta problemática, se han planteado 4 estrategias generales para tratar de trabajar con los datos mezclados:

Conversión de los datos numéricos a datos categóricos. En [Dougherty, Kohavi y Sahami, 1995] Se plantea que este proceso se realiza utilizando métodos de discretización. Estos están basados frecuentemente en histogramas, cálculo de la entropía, o alguna técnica de construcción de árboles de decisión. En este caso decidir el número de categorías puede ser difícil, y lo que es más importante, cuando las categorías son tratadas como nominales, la información del orden se pierde [Lumijarvi, Laurikkala y Juhola, 2004]. Podemos mencionar como ejemplo de esta estrategia el trabajo presentado por [Quintero, 2009], el cual propone una metodología para realizar una representación semántica de datos espaciales raster, enfocándose en los DEM como caso particular de los datos espaciales raster. En este trabajo se discretizan los valores continuos de las variables de alturas del terreno contenidas en los DEM, llevándolos a variables cualitativas que almacenarán los valores de elevación, llanura y depresión en dependencia de los intervalos de

altura que posea cada variable numérica para luego integrarlas con ontologías de dominio de formas del Terreno.

Convertir los datos categóricos a numéricos. Aunque no hemos encontrado trabajos dedicados especialmente a esta variante, si se encuentra en algunos artículos, donde a cada valor categórico se le asigna un código numérico, y después son aplicadas operaciones de suma y producto por un escalar. Por otra se puede decir que es de las más usadas a la hora de realizar operaciones de superposición y reclasificación de mapas raster. Esto se puede evidenciar si tomamos como ejemplo un mapa de tipos de suelos donde se muestren varios tipos de suelos y se requiera realizar una reclasificación de suelos con valores binarios donde sólo se necesite conocer de todos estos tipos de suelos cuales son factibles para un determinado objetivo, los cuales se reclasificaran con el valor de 1 y los demás no factibles con valor 0, o viceversa. Sin embargo, este enfoque es cuestionable, debido a que no tiene sentido para las operaciones aritméticas con valores no ordenados. Por ejemplo, supongamos que asignamos los códigos 0, 1, 2 y 3 para valores de "no posee", "izquierdo", "derecho" y "bilateral". En este caso, la distancia euclidiana de "derecho" a "no posee", como el doble de "derecho" a "izquierdo", lo que no es razonable desde la perspectiva del especialista [Lumijarvi, Laurikkala y Juhola, 2004].

Construir distancias heterogéneas, que trabajen directamente con datos mezclados. Existen varios resultados en esta dirección [Wilson y Martínez, 1997] y algunos estudios comparativos con distancias clásicas [Lumijarvi, Laurikkala y Juhola, 2004]. En este caso no se requieren transformaciones, por lo que no se pierde información. Sin embargo, varias de las funciones reportadas no cumplen con algunas de las propiedades de una distancia, y hay que tener cuidado cuando se utilizan en un algoritmo que lo requiera.

Trabajar con disimilitudes más generales y algoritmos que las permitan.

Actualmente ninguna de estas estrategias satisface las expectativas que se necesitan para lograr trabajar sobre las descripciones mezcladas e incompletas de los objetos geográficos en problemas de reconocimiento de patrones con un

nivel de veracidad muy confiable. La idea radica final que se persigue se basa en no discretizar las variables, no codificarlas o booleanizarlas, para ello nos basaremos en los métodos desarrollados en el campo del Reconocimiento Lógico Combinatorio de Patrones (LCPR por sus siglas en inglés).

El Enfoque Lógico Combinatorio, se origina a partir de los trabajos desarrollados en la antigua Unión Soviética continuados por un grupo de investigadores de Cuba y México encabezados por el Dr. José Ruiz Shulcloper. Este enfoque tiene su basamento en la idea de que la información debe de ser tratada de forma muy cuidadosa para evitar situaciones extremas, como por ejemplo el uso de números a manera de codificadores de descripciones en lugar de haberse utilizado algún otro método de codificación. Este tratamiento inadecuado de la información puede llevar a la obtención de resultados completamente erróneos que en algunos casos pueden ocasionar pérdidas económicas o propiciar situaciones catastróficas [Ruiz-Shulcloper, Martínez-Trinidad y Guzmán-Arenas, 1999]; [Martínez-Trinidad and Guzmán-Arenas, 2001].

En el caso de la información geográfica en el análisis raster actual, los objetos a tratar son n -arias, elementos del producto cartesiano simple sin ningún tipo de propiedades algebraicas, lógicas o topológicas que supone este espacio. Entonces, ¿cómo se seleccionan en estos casos las características más informativas? ¿Clasificar un nuevo objeto dado una muestra (de entrenamiento) de cada clase del problema o encontrar las relaciones entre todos los objetos en función de un cierto grado de similitud? Este es el principal problema resuelto por LCPR. Este enfoque es una respuesta metodológica necesaria para el hecho de que muchos problemas en función de selección y clasificación, que aparecen con frecuencia, las descripciones de los objetos son datos mezclados e incompletos.

Por ejemplo: la descripción de zonas geológicas para el establecimiento de algún pronóstico acerca de la presencia de yacimientos de minerales útiles. Para muchos geofísicos los problemas prácticos que deben tener en cuenta están no sólo en las medidas como los campos

magnéticos, la anomalía gravimétrica de Bouguer, la intensidad del campo magnético del aire, que son todas variables numéricas, deben tomar en cuenta también características como la serie genética de los suelos, la presencia o ausencia de defectos, en el nivel de desarrollo del suelo, y otros. Se puede decir entonces: bueno, vamos a procesar todas las características categóricas y tomar una decisión, mediante el procesamiento de todas las variables numéricas y hacer un segundo procesamiento mediante la codificación de las variables descriptivas. Por último, se puede llegar a una conclusión sobre las decisiones parciales hechas antes. Pero el problema es, como se puede ver, obviamente, que las variables categóricas no se deben analizar en conjunto con las variables numéricas. Para los geofísicos este es un gran problema. No es posible adoptar una solución basada en estas condiciones.

En la Fig. 10 se muestra gráficamente la estrategia de solución de nuestra propuesta, donde se puede observar cómo la ORD es generada a partir de los datos geográficos extraídos de los mapas raster. Dichos datos son representados en los NRD-O. Una vez creados estos nodos, se generan nuevos conceptos que pudieran estar subyacentes en los datos extraídos. Estos nuevos conceptos que se encuentran en un nivel superior de abstracción semántica serán representados en los NRD-C. Una vez representados los NRD, se procede a extraer las relaciones topológicas y/o espaciales entre los objetos geográficos raster, los cuales han sido ya representados en los NRD-O. Luego de que se hayan extraído las relaciones existentes, se tendrá generada la ORD. Posteriormente se procede a la integración semántica de la ORD generada con los conceptos existentes en la(s) ONS. Una vez que se tengan representados todos los objetos geográficos se puede comenzar a realizar los métodos de superposición semántica entre los NRD-O de la ORD. La ejecución de estos métodos devuelve como resultado, nuevos NRD-O con la información resultante de superponer la información de los nodos interactuantes en los métodos, los cuales vuelven a ser integrados con los conceptos de las ONS.

Este dinamismo de los datos incluidos en el sistema provocará la generación de nuevos mapas resultantes de la operación matemática y/o lógica y de nuevos NRD representados en la ORD ya que esta, como se ha mencionado anteriormente, es generada automáticamente a partir de estos datos. Esta actualización de la ORD provoca además el surgimiento de términos y relaciones existentes en la nueva información generada, de esta forma y tomando en cuenta lo planteado por Larin-Fonseca y Garea-Llano, el surgimiento de nuevos NRD en la ORD no implican absolutamente ningún cambio en la(s) ONS manteniendo esta(s) su estructura original, cada uno de estos cambios ocurren por debajo de la capa de interconexión.

Esta ontología puede además contener nuevos conceptos que no existen de forma explícita ni en las ONS ni en los datos. Estos nuevos conceptos son aprendidos a partir de la operación de superposición de los NRD de los datos raster realizado a los datos en el proceso de regeneración de la ORD.

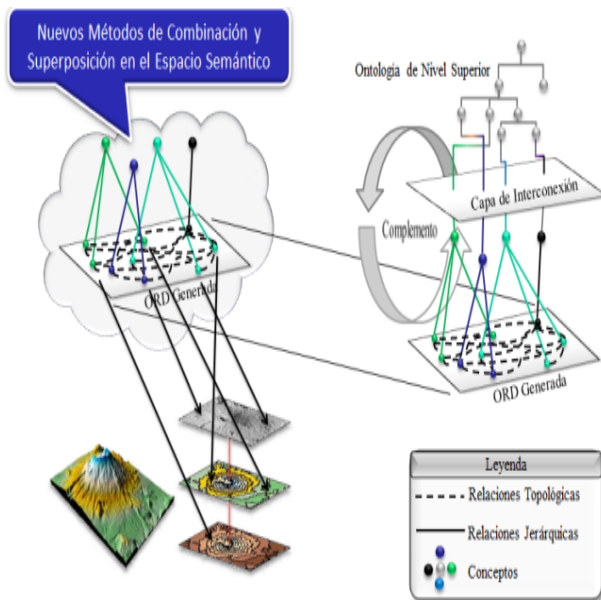


Fig. 10. Esquema de la estrategia propuesta para la Superposición de los Datos Espaciales Raster en el Nivel Semántico (modificado de Larin-Fonseca and Garea-Llano 2010)

A continuación se muestran los pasos necesarios para realizar el método de SSDR:

1. Definición de la estructura raster en la que están almacenados los datos:
Como ya se explicó esta estructura estarán conformadas por sus tres principales componentes, la espacial, la temporal y la temática.
2. Extracción de valores contenidos en los datos:
Para la extracción de los valores tanto temáticos como espaciales en los datos raster reutilizaremos los algoritmos de extracción implementados en el proceso de Síntesis propuesto en el trabajo realizado por Quintero [Quintero, 2009] para los DEM. Posteriormente se crearan los NRD-O con la información extraída.
3. Generación de nuevos conceptos a partir de los datos extraídos de los mapas:
Teniendo almacenada la información de la componente temática en los NRD-O, mediante el uso de algoritmos de agrupamiento jerárquico se pueden descubrir nuevos niveles de abstracción semántica de los objetos en clases, las cuales serán almacenadas en los NRD-C como ya se ha explicado anteriormente.
4. Extracción de las relaciones topológicas y/o jerárquicas que existen entre los objetos geográficos contenidos en los mapas raster:
Las relaciones existentes entre los datos raster pueden extraerse a partir de operaciones de análisis espacial (ver sección 3), utilizando la componente espacial almacenada en los NRD-O. Si observamos la Fig. 11, las zonas con diferentes colores representan diferentes municipios dentro de una provincia. Si nos enfocamos específicamente en el municipio A que se encuentra dentro del recuadro rojo, podremos apreciar como los otros municipios B, C, D, E y F tienen una relación de adyacencia con A. Esto además de apreciarse visualmente, si aplicaciones operaciones de análisis geoespacial sobre la componente espacial de estos municipios que serán representados como en NRD, podremos darnos cuenta de que su ubicación geográfica aparece de forma

contigua, por lo que tienen relaciones de adyacencia.

Luego de la ejecución de estos pasos se tendrá generada la ORD. Seguidamente se puede proseguir con la ejecución de los métodos que permiten realizar una superposición de datos raster, ya que como se ha mencionado anteriormente.

5. Ejecución de los métodos de SDR entre los NRD-O:

En estos métodos entran a jugar las componentes espaciales y temáticas. La componente espacial posibilitará seleccionar cuáles serán los NRD necesarios para ejecutar este método. Esto estará condicionado por las demandas de los resultados que se quieran obtener. Por otro lado la componente temática es la que contendrá los datos necesarios para poder realizar las operaciones de superposición. Aun cuando se han encontrado 4 estrategias para tratar el tema de los datos mezclados, todavía seguimos trabajando en buscar una mejor solución que garantice mejores resultados.

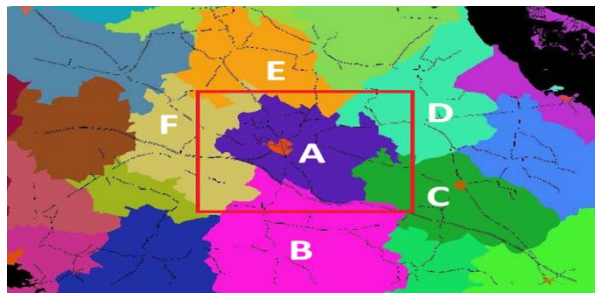


Fig. 11. Relaciones de adyacencia entre los objetos geográficos raster

6 Conclusiones y Trabajos Futuros

Una de las principales limitantes de las operaciones de recuperación y análisis raster radica en que los valores de las celdas son números (enteros, flotantes) o expresiones Booleanas, o sea realizan la codificación de rasgos no numéricos por lo que los operadores para comparar atributos con valores son los mismos usados en la matemática. La naturaleza

de los datos geográficos puede ser heterogénea, con variables tanto cualitativas como cuantitativas utilizadas para describir los fenómenos u objetos que se están representando. Por lo que se necesita encontrar métodos para que tomen en cuenta esta limitante y permitan trabajar con las descripciones mezcladas e incompletas de manera simultánea. El campo del Reconocimiento Lógico Combinatorio de Patrones puede ofrecer una serie de métodos y algoritmos en este sentido.

En este trabajo se ha realizado el esbozo de una estrategia de representación y análisis de datos geográficos raster, la misma se basa en su representación semántica mediante las ORD y a través de la combinación de los NRD extraídos de los objetos espaciales representados en el modelo raster en la ORD. Para el caso de los objetos geográficos discretos el NRD define el tipo de objeto geométrico en el que está representada la entidad, (conjunto de celdas) y para el caso de variables continuas de distribución espacial se toma como tipo de objeto geométrico la propia celda.

Como trabajos futuros nos proponemos continuar desarrollando esta estrategia de forma detallada, determinar, implementar y experimentar métodos de análisis de los datos a partir de sus descripciones mezcladas sobre la base de los métodos desarrollados en el campo del LCPR.

Referencias

1. **Bosque, J. (1994).** *Sistemas de Información Geográfica: prácticas con PC ARC/INFO e IDRISI*. Madrid: RA-MA.
2. **Dougherty, J., Kohavi R. & Sahami M. (1995).** Supervised and unsupervised discretization of continuous features. *12th International Conference on Machine Learning*, San Francisco, CA, USA, 194-202.
3. Geographic information systems in sustainable development. (June, 2004,). Retrieved from <http://www.fao.org/SD/eidirect/gis/eigis000.htm>.
4. **Graeff, B. (2001).** Querying Raster Data Structures – Probabilistic and non-probabilistic approaches on knowledge based template matching methods. *Vistas for Geodesy in the new Millennium, IAG 2001 Scientific Assembly*, Budapest, Hungary. 371-376.
5. **Graeff, B. & Carosio A. (2002).** Automatic interpretation of raster-based topographic maps by means of queries. *Spatial Information in Mapping and*

Cadastral Systems. FIG XXII International Congress, Washington D.C, USA, 1-12.

6. **Larin, R. & Garea E. (2010).** Automatic representation of semantic abstraction of geographical data by means of classification. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lectures Notes in Computer Science*, 6419, 557–568.
7. **Lumijarvi, J., Laurikkala J. & Juhola (2004).** A comparison of different heterogeneous proximity functions and Euclidean distance. In M. Fieschi, E. Coeira & Li, YC (Eds), *Medinfo 2004: Proceedings of the 11th World Congress On Medical Informatics*, (1362-1366). Amsterdam: IOS Press.
8. **Martínez, J. F. & Guzmán A. (2001).** The logical combinatorial approach to Pattern Recognition, An overview through selected works. *Pattern Recognition*, 34 (4), 741-751.
9. **Merz, C. J. & Murphy P. M. (1998).** *UCI Repository of Machine Learning Databases*. Department of Information and Computer Science, University of California, Irvine, USA.
10. **Oliva, R. (2009).** *Anotaciones Semánticas de Datos Geográficos*. Tesis de Maestría, Universidad de La Habana, La Habana, Cuba.
11. **Quintero, R., Torres M., Moreno M. & Guzmán G. (2009).** Towards a semantic representation of raster spatial data. *Geospatial Semantics. Lectures Notes in Computer Science*, 5892, 63–82.
12. **Ruiz, J., Guzmán A. & Martínez J. F. (1999).** *Enfoque Lógico-Combinatorio al Reconocimiento de Patrones, Selección de Variables y Clasificación Supervisada*. México D.F: Editorial IPN.
13. **Wilson, R.D. & Martínez T. R. (1997).** Improved Heterogeneous Distance Functions. *Journal of Artificial Intelligence Research*, 6, 1-34.



Eydel Jaime González

Obtuvo el título de Ingeniería en Ciencias Informáticas dado en la UCI (Universidad de las Ciencias Informáticas) en julio del 2007. En la actualidad se encuentra trabajando como investigador en el departamento de Reconocimiento de Patrones del Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV) en el área del Análisis Semántico de Datos Geoespaciales bajo la asesoría del Dr. Eduardo Garea Llano.



Rainer Larín Fonseca

Recibió el título de Ingeniería Informática y el título de Profesor de Ciencias Informáticas, dado en la CUJAE (Ciudad Universitaria José Antonio Echeverría) en julio del 2007. Actualmente se encuentra trabajando como investigador en el departamento de Reconocimiento de Patrones del Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV) en el área de la Representación Semántica de Datos Geoespaciales.



Eduardo Garea Llano

Dr. en Ciencias Técnicas, Investigador Auxiliar, obtuvo el título de Ingeniero Geólogo e Hidrogeólogo y de M.Sc. en Geología en el Instituto de Minas de San Petersburgo, Rusia en 1993 y el Grado Científico de Doctor en Ciencias Técnicas en el Instituto Politécnico Superior de La Habana “José Martí” en 2003. En la actualidad dirige el Departamento de Reconocimiento de Patrones del Centro de Aplicaciones de Tecnologías de Avanzada. También es el jefe del grupo de investigaciones en Semántica Geoespacial y Sistemas de Información Geográfica. Ha enfocado sus investigaciones hacia la interpretación compleja de datos espaciales utilizando herramientas de análisis de los Sistemas de Información Geográfica, el procesamiento Digital de Imágenes y video, la teledetección y la interpretación semántica de datos geoespaciales. Tiene publicados más de 20 artículos en revistas, libros y eventos nacionales e internacionales.