

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Nuevo tipo de ontología para la
representación semántica de objetos
geoespaciales**

Rainer Larín Fonseca

RT_053

octubre 2013





CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Nuevo tipo de ontología para la
representación semántica de
objetos geoespaciales**

Rainer Larín Fonseca

RT_053

octubre 2013



Ministerio de las Fuerzas Armadas Revolucionarias

Instituto Técnico Militar “*José Martí*”

Orden “*Antonio Maceo*”

Orden “*Carlos J. Finlay*”

GEOCUBA Investigación y Consultoría

Centro de Aplicaciones de Tecnologías de Avanzada

CENATAV

**NUEVO TIPO DE ONTOLOGÍA PARA LA
REPRESENTACIÓN SEMÁNTICA DE
OBJETOS GEOESPACIALES**

Tesis Presentada en Opción al Grado Científico de

Doctor en Ciencias Técnicas

Rainer Larín Fonseca

La Habana

2012

Ministerio de las Fuerzas Armadas Revolucionarias

Instituto Técnico Militar “*José Martí*”

Orden “*Antonio Maceo*”

Orden “*Carlos J. Finlay*”

GEOCUBA Investigación y Consultoría

Centro de Aplicaciones de Tecnologías de Avanzada

CENATAV

**NUEVO TIPO DE ONTOLOGÍA PARA LA
REPRESENTACIÓN SEMÁNTICA DE
OBJETOS GEOESPACIALES**

Tesis Presentada en Opción al Grado Científico de

Doctor en Ciencias Técnicas

Autor: Asp. Inv. Ing. Rainer Larín Fonseca

Tutor: Inv. Tit., Ing. Eduardo Garea Llano Dr.C.

La Habana

2012

Agradecimientos...

En este apartado quiero dejar plasmado un profundo agradecimiento a esas personas que me sirvieron de apoyo y guía a lo largo de este trabajo:

A mi querida Yení, por sus valiosas sugerencias y acertados consejos durante el desarrollo de este trabajo...

A Sandro, Diana, Oneysis y Dania, por brindarme la oportunidad de recurrir a su capacidad y experiencia científica en un marco de confianza, afecto y amistad, por su continuo y afectuoso aliento...

A Cynthia, por ser mi Fi en el trabajo con la librería Jena...

A mi tutor Eduardo Garea Llano, por su generosidad científica y valiosas críticas en las discusiones de los resultados de este trabajo...

A mis compañeros del Cenatav, por hacer más amenas las largas horas de trabajo...

Al grupo de Geo-Semántica, por su calidez y compañerismo al compartir inquietudes, éxitos y fracasos...

Al departamento de ICT, por sus valiosas búsquedas bibliográficas...

To Arturo, for helping me to learn English...

A Candy, por su acertada revisión ortográfica...

Dedicatorias...

Todo trabajo realizado con placer no es completo si no se ofrece en dedicatoria y esta va dirigida a todas aquellas personas que me han brindado parte importante de su tiempo, aportándome valiosas contribuciones que posibilitaron el desarrollo de este trabajo. A todas ellas muchas GRACIAS!!!!

Este trabajo va especialmente dedicado:

A mi querida hija Cristina, por ser mi principal fuente de motivación y para la cual tengo la obligación de ser cada día mejor...

A mis padres, por estar siempre presentes y por ser el modelo de personas que yo quiero ser, por inculcarme toda la ética y el rigor necesarios para la vida, por enseñarme que depende principalmente de mí el que pueda lograr mis metas...

A mi querida Yeny, por todo su amor, cariño y apoyo en todo momento...

A mis hermanos Muñeca, Daily, Andy y Dayan, a quienes quiero servir de ejemplo...

A mi prima Yulí, a quien siempre he considerado como mi hermana...

A mi querida Abuela, que no porque te hayas ido has dejado de estar presente...

A toda mi familia, que es muy grande y no quiero olvidarme de nadie, por ayudarme siempre...

A Sandro y a las queridísimas Diana, Oneysis, Malena y Dania, por su incondicional amistad, por ser mis fuentes de alegrías y dolores de cabeza...

A mi tutor Eduardo Garea Llano, por su visión y guía...

A mis profes de la CUJAE y especialmente a la profe Mayí por toda su ayuda y apoyo...

A todas mis amistades de siempre, a los del IF-x4 en la CUJAE, a los del CENATAV y a los del barrio...

SINTEISIS

En el presente trabajo de investigación es el resultado de las investigaciones desarrolladas por el autor. En la misma se propone un nuevo nivel de generalidad desde el punto de vista semántico dando lugar a un nuevo tipo de ontología para la representación semántica de objetos geoespaciales y las relaciones existentes entre estos.

Se introduce además una nueva teoría sobre el espacio semántico multidimensional para la representación de diferentes puntos de vista semánticos sobre los objetos geoespaciales así como también, se propone una modificación a la arquitectura de los Sistemas de Información Geográfica Gobernados por Ontologías para la generación automática de conocimiento a partir de fuentes de datos heterogéneas.

Los resultados alcanzados son válidos, ya que los métodos obtenidos permiten por un lado, representar y aprovechar la información semántica que subyace en los datos geoespaciales y por el otro, establecen las bases arquitectónicas para que los sistemas de información geográfica utilicen herramientas de procesamiento semántico de la información geoespacial.

ÍNDICE

ÍNDICE	1
INTRODUCCIÓN	1
Capítulo 1	11
1. Marco Teórico Referencial en el Procesamiento Semántico de Objetos Geoespaciales	11
1.1. Naturaleza Compleja de los Objetos Geoespaciales	11
1.2. Sistemas de Información Geográfica	12
1.3. Semántica e integración de Datos Geográficos.....	14
1.3.1. Redes Semánticas.....	18
1.3.2. Marcos Conceptuales	19
1.3.3. Ontologías	19
1.4. Representación Estática de las Ontologías Actuales.....	22
1.5. Ontologías como un componente más de los SIG.....	25
1.6. Conclusiones	28
Capítulo 2.....	30
2. Métodos para la Representación Semántica de Objetos Geoespaciales.....	31
2.1. Ontologías de Representación de Datos (ORD).....	31
2.1.1. Estructura Interna de la Ontología de Representación de Datos	35
2.1.2. Estructura Interna de los Nodos de Representación de Datos.....	36
2.1.3. Generación Automática de la ORD.....	37
2.1.3.1. Definición de la Estructura en la que está Almacenado el Dato	39
2.1.3.2. Extracción de Valores	40
2.1.3.3. Generación de Nuevos Conceptos.....	40
2.1.3.4. Extracción de Relaciones	42
2.2. Espacio Semántico Multidimensional.....	42

2.3.	Definición Contextual de los Objetos Geográficos.....	46
2.3.1.	Clasificación de Nodos de Representación de Datos Mediante la Regla del Vecino más Cercano.....	47
2.3.2.	Reducción del Espacio de Búsqueda para la Clasificación de los Nodos de Representación de Datos.....	50
2.3.3.	Medidas de Disimilitud para la clasificación de Nodos de Representación de Datos.....	53
2.4.	Ventajas del uso de la Ontología de Representación de Datos.....	57
2.4.1.	Rol de la Ontología de Representación de Datos en la Integración Semántica de Datos Geo-Espaciales Heterogéneos.....	57
2.4.2.	Sistemas Gestores de Bases de Datos Geoespaciales y ORD.....	59
2.4.2.1.	Sistemas Gestores de Bases de Datos Geoespaciales Semánticos.....	59
2.5.	Nueva Modificación a la Arquitectura SIGGO.....	61
2.6.	Conclusiones.....	63
Capítulo 3.....		64
3.	Resultados Experimentales Realizados para la Validación de los Métodos Propuestos.....	64
3.1.	Diseño Experimental.....	64
3.2.	Experimento 1: Representación Semántica Automática de Objetos Geoespaciales.....	66
3.3.	Experimento 2: Representación Automática de Objetos Geográficos con Contexto Multidimensional en el Espacio Semántico.....	71
3.4.	Experimento 3: Recuperación y análisis de la información para la toma de decisiones.....	75
3.5.	Experimento 4. Representación y Recuperación de objetos Geoespaciales en Imágenes de Teledetección.....	89
3.6.	Conclusiones.....	98
CONCLUSIONES Y RECOMENDACIONES.....		99
REFERENCIAS BIBLIOGRÁFICAS.....		101
BIBLIOGRAFÍA CONSULTADA.....		111

Anexos	113
ANEXO 1: Glosario de Términos.....	107
ANEXO 2: Glosario de Acrónimos	107
ANEXO 3: Estudio de Factibilidad del Prototipo de Sistema Gobernado por Ontologías para la Recuperación de Objetos Geospaciales en Imágenes Satelitales “Crudas”	108

INTRODUCCIÓN

Las herramientas para la captura y almacenamiento de los datos geoespaciales han ido evolucionando continuamente. Esto trae como consecuencia un aumento gradual en el volumen y la complejidad de este tipo de información. Estos además, son habitualmente generados por distintos productores de datos, lo que provoca una gran heterogeneidad en ellos, ya que son almacenados en formatos y estándares distintos [1]. A estos problemas, se le suma la naturaleza diversa de estos datos (*diversidad de escalas, en el grado de exactitud en la toma del dato, etc.*). Esto implica que los SIG no puedan procesarlos de forma directa, y por lo tanto, se requiera de un costoso proceso previo, denominado pre procesamiento, para la integración y compatibilización de los distintos datos a utilizar. Este pre procesamiento no siempre puede ser realizado, lo que induce a un manejo ineficiente de la información. La heterogeneidad de estos datos constituye una de las principales causas que afecta la interoperabilidad entre los diferentes sistemas, ya que el intercambio de información se ve limitado muchas veces por problemas de compatibilidad entre ellos. Importantes iniciativas para lograr un acceso uniforme a los datos lo constituyen consorcios como Open GIS [2] y W3C [3] que son asociaciones que buscan definir un conjunto de requisitos y normas para la estandarización de los datos en el ámbito geográfico, la cual va enfocada principalmente a resolver problemas derivados de la heterogeneidad sintáctica. Sin embargo, esta tampoco constituye toda la solución, ya que estos estándares también presentan diferencias entre ellos, por lo que el problema de la integración e interoperabilidad no queda totalmente resuelto.

Un camino que apunta a convertirse en una solución eficiente para el procesamiento de los datos geográficos es su tratamiento desde el punto de vista semántico. Prueba de ello es la tendencia que se manifiesta en numerosas investigaciones recientes que abordan esta problemática [4-10]. Esta solución se basa principalmente en el uso de ontologías como mecanismo de representación del conocimiento y se debe precisamente a que ellas combinan los paradigmas Orientado a Objeto [11] y de Entidad-Relación [12], los cuales resultan claves para la modelación de los diferentes fenómenos en el ámbito geográfico. La inclusión de las ontologías como una herramienta más en el procesamiento de los datos ha evolucionado los Sistemas de Información (SI), denominándose actualmente como Sistemas de Información Basados en Ontologías (SIBO) [13]. En el ámbito geográfico estos nuevos sistemas han sido denominados como Sistemas de Información Geográfica Gobernados por Ontologías (SIGGO) [9]. Los SIGGO incorporan las ontologías como un componente activo dentro de su arquitectura posibilitando gestionar los datos con mayor eficacia y robustez. De esta manera muchos de los problemas existentes con los datos geoespaciales pueden ser evitados.

El análisis realizado de la literatura sugiere que se hace necesario que, además de los esfuerzos actuales, se desarrollen nuevos mecanismos para el procesamiento de la información geoespacial sobre la base de su interpretación semántica. Del análisis surge también una contradicción la cual puede ser expresada como situación problemática a enfrentar en la presente investigación.

Situación Problemática

La arquitectura SIGGO según la concepción de sus autores [14] hace uso de los tipos de ontologías existentes en la literatura [15]. Estos tipos de ontologías son capaces de modelar

los diferentes dominios en el ámbito geográfico, así como las relaciones genéricas existentes entre las diferentes entidades geográficas, pero solo hasta cierto nivel de abstracción semántica. Esto viene dado principalmente porque la semántica que subyace en los datos geoespaciales y las relaciones existentes entre los objetos contenidos en ellos no pueden ser incluida en estos tipos de ontologías. Debido a esto, gran cantidad de información relevante no es considerada ni puede ser aprovechada por el SIGGO.

Por lo tanto, es necesario desarrollar mecanismos adecuados para representar esta información de forma que pueda resultar útil al SIGGO. Tomar en cuenta estos elementos dentro del análisis semántico, brindarían además, por un lado, la posibilidad de inferir nuevos conceptos a partir del agrupamiento de los objetos. De esta forma se aportan nuevos niveles de abstracción semántica en la(s) ontología(s) utilizadas lo que repercute en un enriquecimiento de la semántica del dominio representado. Por el otro lado, se toman en cuenta las relaciones existentes entre estos objetos (topológica, espacial, etc.), las cuales resultan clave para tareas comúnmente realizadas en el ámbito geoespacial como el análisis y la toma de decisiones.

A partir de lo planteado surge la necesidad de establecer primeramente, un nuevo nivel de abstracción que permita contener las definiciones necesarias para la representación de las relaciones y características que modelan y dan significado a los objetos en el ámbito geoespacial desde un punto de vista semántico y finalmente la modificación de la arquitectura SIGGO de forma que se habilite el uso de la semántica que subyace en los datos a través del uso de este nuevo nivel de abstracción. De esta manera se puede plantear el problema científico de esta investigación.

Problema Científico:

*Se hace necesario definir un **nuevo tipo de ontología** para la representación de la semántica que subyace en los datos geoespaciales y las relaciones que se generan entre los objetos contenidos en ellos de forma que pueda ser utilizado por un **SIGGO**.*

La disciplina en la que se define el **objeto de investigación** es la representación de las relaciones y características que dan significado a los objetos en el ámbito geoespacial desde un punto de vista semántico.

El **objetivo general** de esta investigación consiste en:

*Desarrollar métodos para la representación de la abstracción semántica de objetos geoespaciales integrados en un entorno **SIGGO**.*

Para lograr este objetivo se plantean los siguientes **objetivos específicos** en la investigación:

- 1. Definir un nuevo tipo de ontología de naturaleza dinámica para la representación de las relaciones y características que modelan y dan significado a objetos geoespaciales integrados en un entorno **SIGGO**.*
- 2. Diseñar una estructura para el tipo de ontología propuesta que sea capaz de contener la representación de la abstracción semántica de los objetos geoespaciales tomando en cuenta además, las relaciones existentes entre ellos.*
- 3. Proponer métodos para la representación automática de la abstracción semántica de los objetos geoespaciales integrados a un entorno **SIGGO** sobre la base de sus principales componentes: La Temática, la Espacial y la Temporal.*
- 4. Definir las características básicas que debe cumplir un método para ser utilizado en el descubrimiento automático de nuevos niveles de abstracción semántica que*

podieran estar subyacentes en los objetos a partir de las técnicas de agrupamiento jerárquico existentes de forma que pueda ser enriquecido automáticamente el dominio representado.

5. *Proponer un método para la interrelación automática del nuevo tipo de ontología propuesto con ontologías de niveles superiores (ontología de Tarea, de Dominio, etc.).*
6. *Modificar la arquitectura SIGGO para la conceptualización automática de los objetos geoespaciales y su uso con Ontologías de Nivel Superior.*
7. *Implementar una plataforma de experimentación en donde se integren todos los desarrollos alcanzados.*

Como **hipótesis** en la investigación planteamos que:

La representación semántica de objetos geoespaciales sobre la base de sus componentes fundamentales, incluyendo además, las relaciones existentes entre ellos en combinación con ontologías de niveles superiores, contribuye a mejorar (perfeccionar) las formas de análisis, explotación e integración realizadas por los SIG convencionales.

Para cumplir los objetivos y demostrar la hipótesis planteada se realizaron las siguientes

tareas:

1. *Estudiar el estado del arte en la representación semántica de objetos geoespaciales para detectar las limitaciones y ventajas de los enfoques existentes.*
2. *Estudiar el estado actual en las investigaciones sobre los Sistemas de Información Geográfica con relación al procesamiento y análisis de los objetos desde el punto de vista semántico.*

3. *Estudio, análisis y comparación de los diferentes mecanismos de representación de conocimientos que son más utilizados en la literatura haciendo énfasis en las Ontologías.*
4. *Estudio y análisis de los alcances y limitaciones de los diferentes tipos de ontologías propuestas en la literatura.*
5. *Proponer un nuevo tipo de ontología que, en combinación con ontologías de niveles superiores, proporcione nuevas y mejores formas de análisis y explotación de los objetos geográficos desde el punto de vista semántico.*
6. *Comparar la solución propuesta con las homólogas reportadas en la literatura consultada.*
7. *Estudio y análisis de la arquitectura de los SIG gobernados por ontologías propuestos en la literatura para su refinamiento a partir de la incorporación de los mecanismos de representación semántica multidimensional de objetos geoespaciales.*
8. *Implementación de una plataforma de experimentación donde se integren todos los desarrollos alcanzados.*
9. *Realización de experimentaciones para comprobar la validez de los métodos propuestos.*

Entre los **métodos de investigación** utilizados se destacan los siguientes:

- El método *dialéctico-materialista* para determinar las contradicciones existentes entre las teorías planteadas y lo que se evidencia en la práctica de forma que se generen nuevas soluciones a los problemas existentes.

- El método *hipotético-deductivo* para la elaboración de la hipótesis central de la investigación y deducir las consecuencias observables para la proposición de las nuevas líneas de trabajo a partir de los resultados parciales obtenidos.
- Los métodos *histórico-lógico* para el estudio crítico de trabajos previos y aprovechar de estos, los aspectos positivos para su utilización como puntos de referencia y su comparación con los resultados alcanzados en este trabajo.
- El método *analítico-sintético* para descomponer el problema de investigación en sus partes fundamentales y profundizar en cada uno de ellos por separado para luego integrarlos en la solución propuesta.
- El método de *modelado* para explicar las causas que hacen que el uso de la ontología que se propone en combinación con ontologías de nivel superior sea una buena solución para el problema.
- El método de *comparación-clasificación* para el análisis de los resultados obtenidos y su comparación con los resultados reportados en la literatura.
- El método *coloquial* para la presentación y discusión de los resultados en sesiones científicas.
- El método *experimental* para comprobar y fundamentar los estudios comparativos entre los métodos reportados en la literatura y el que se propone en este trabajo.

La **novedad científica** de la investigación está dada por el aporte que hace a la teoría y por su significación práctica y social. De forma detallada esta novedad se puede desglosar de la siguiente forma:

- Desde **el punto de vista teórico** se proponen aportes en el área de la *Geosemántica* contribuyendo en primer lugar con la especificación de un *nuevo tipo de ontología*

para la representación semántica de datos geoespaciales heterogéneos contenidos en SIG; la especificación de la teoría del *espacio semántico multidimensional* para la inclusión de diferentes puntos de vista acerca de los objetos representados para enriquecer semánticamente los procesos de análisis y toma de decisiones; finalmente la *modificación de la arquitectura de SIG gobernado por ontologías* para el aprovechamiento de la semántica que subyace es los datos geoespaciales manejados en el sistema.

- La **significación práctica** de este trabajo radica en la contribución que se hace al desarrollo de una de las herramientas más utilizadas en el procesamiento de la información geográfica, los SIG. Los métodos propuestos además, pueden ser implementados en distintas aplicaciones para resolver problemas de la vida real donde se hace necesario un análisis semántico de los datos geoespaciales. Dotar a los SIG de un procesamiento semántico de los datos aumenta el impacto en el desarrollo de la sociedad dado por la gran gama de aplicaciones en las que estos sistemas son utilizados. En nuestro país organizaciones, tales como *GEOCUBA*, *CITMA*, *MINFAR*, *MININT*, entre otros, utilizan los SIG en actividades como:
 - *La defensa y Orden Interior:* En esta área permiten el manejo catastral, ordenamiento del Teatro de Operaciones Militares, enmascaramiento de la técnica y monitoreo del desplazamiento del enemigo. Así como también el análisis de la distribución espacial del potencial delictivo y modus operandi.
 - *Desarrollo de Infraestructuras:* En esta área permiten el diseño, planeación, construcción y la valoración de proyectos de infraestructura como construcción de vías, redes de gas, petróleo, etc.

- *Medio Ambiente:* En este campo se utilizan en aspectos como el manejo de la tierra, su protección y conservación.
- La **actualidad** del tema abordado radica en que varias de las líneas abordadas se consideran problemas abiertos hoy en día y siguen destinándose grandes esfuerzos en este sentido. Prueba de ellos lo constituyen la gran cantidad de trabajos publicados en los últimos años en esta temática.

Alcances y Límites de la Investigación

Dado que el campo de la Geomática es bastante amplio es necesario aclarar que los métodos que se proponen en este trabajo han sido concebidos dentro del marco de los SIG y el procesamiento realizado de los datos geoespaciales por sistemas de este tipo sobre la base de sus principales características y limitaciones. Esto es debido a que se hace necesario que los SIG cuenten con herramientas para procesar robustamente no solo la semántica del dominio sino también la semántica que subyace en los datos geoespaciales. De esta forma pueden ser mejoradas tareas convencionales realizadas por estos sistemas como por ejemplo el análisis y la toma de decisiones.

Como límite fundamental de este trabajo de investigación tenemos que no existen implementaciones conocidas de SIG gobernados por ontologías con los que puedan realizarse experimentos de comparación.

Estructura de la tesis:

Capítulo 1: Aborda el estado del arte en la temática del procesamiento geoespacial. En este se exponen críticamente los principales problemas y enfoques existentes en el proceso de integración y uso de este tipo de información.

Capítulo 2: En este se reflejan los principales aportes y desarrollos obtenidos a lo largo de este trabajo de investigación para la representación semántica de objetos geoespaciales integrados en sistemas de nueva generación (SIGGO).

Capítulo 3: Se presentan las principales experimentaciones realizadas, las cuales contribuyen a la validación de los métodos propuestos.

La mayor parte del contenido de esta tesis es una recopilación de los métodos y resultados que aparecen publicados en los siguientes artículos del autor [16-33]. Estos artículos han sido publicados en eventos y revistas nacionales e internacionales.

Capítulo 1

**MARCO TEÓRICO REFERENCIAL EN EL
PROCESAMIENTO SEMÁNTICO DE LOS OBJETOS
GEOESPACIALES**

CAPÍTULO 1

Marco Teórico Referencial en el Procesamiento Semántico de los Objetos Geoespaciales

Este capítulo aborda el estado actual en el procesamiento de los objetos geoespaciales. Primeramente se presentan las características y complejidades de las distintas fuentes que contienen estos datos que dificultan su integración, uso y procesamiento. Seguidamente se describen brevemente los Sistemas de Información Geográfica (*SIG*) analizando críticamente las principales causas de los problemas existentes en estos sistemas que hacen necesaria la búsqueda de nuevas vías de solución en la integración de la geo información. A continuación se analizan los tipos de integración de la geo información existentes en la literatura y se define la óptima para el procesamiento de los objetos geoespaciales según las tendencias actuales. Posteriormente se analizan los principales mecanismos de representación de conocimiento profundizando en el que mayor aceptación ha tenido en estos últimos años. Seguidamente se exponen las últimas tendencias para el procesamiento semántico de la información geoespacial en los *SIG* y cómo esta unión (*Semántica + SIG*) da lugar a la creación de sistemas de nueva generación. Finalmente se exponen las principales conclusiones del capítulo.

1. Marco Teórico Referencial en el Procesamiento Semántico de Objetos Geoespaciales

1.1. Naturaleza Compleja de los Objetos Geoespaciales

Los objetos geográficos se representan a partir de tres componentes principales a partir de los cuales pueden ser discriminados. Estas componentes son: la *Componente Espacial*, la *Componente Temática* y la *Componente Temporal* [34]. La información Espacial es la referente a su localización sobre o cercana a la superficie de la Tierra. La Temática es acerca de los atributos de los objetos incluyendo su identificador y la Temporal es relativa a los aspectos cambiantes en el tiempo de las propiedades de estos objetos.

Resumiendo, podemos decir que estos datos constituyen la representación explícita y abstracta de objetos y/o fenómenos que poseen una ubicación con respecto a la superficie de la Tierra. Esta representación (*explícita y abstracta*) implica que estos sean definidos por medio de un sistema de clasificación que comprende las clases abstractas¹ en las que pueden ser agrupados. Esta clasificación da como resultado una taxonomía con diferentes niveles de abstracción donde los datos obtenidos de los objetos en el mundo real se definen como instancias de estas clases.

Los datos geoespaciales presentan características particulares que los diferencian de los datos de tipo no geoespacial y que dificultan considerablemente su integración y uso, por lo que muchas veces se hace necesario desencadenar complejos mecanismos de pre-procesamiento para la realización de las diferentes tareas que involucran este tipo de información. Estas características pueden resumirse principalmente en los siguientes aspectos (para mayor profundidad puede consultarse el trabajo de Leung [35]):

¹*Clase Abstracta*: Clase que definen las características y funcionalidades de los objetos en dependencia del nivel de abstracción en la taxonomía.

- *Gran Volumen de Almacenamiento:* Con el desarrollo de las nuevas tecnologías se hace posible la obtención de datos acerca de una mayor cantidad de objetos y/o fenómenos, igualmente el número de propiedades recogidas sobre estos pueden llegar a ser bastante numerosas.
- *Diversidad de Formatos:* Los datos geográficos pueden ser encontrados en formatos heterogéneos. Comúnmente se da el caso de que datos en diferentes formatos hacen referencia al mismo objeto en el mundo real.
- *Diversidad Semántica:* Los objetos geográficos pueden tener diferentes representaciones semánticas tomando en cuenta diferentes puntos de vista.

La naturaleza diversa y compleja de los datos geográficos ha traído como consecuencia grandes complicaciones en su procesamiento y uso. Esto se ha visto reflejado sobre todo en el intercambio de información entre los diferentes sistemas que manipulan este tipo de información, problemas que muchas veces son originados por las incompatibilidades existentes entre ellos.

1.2. Sistemas de Información Geográfica

Los Sistemas de Información Geográfica constituyen las herramientas por excelencia para el manejo de la información geoespacial. Estas herramientas juegan un papel fundamental en la representación, el procesamiento y el análisis de los datos geográficos, ya que proporcionan los mecanismos necesarios para la realización de estas tareas de manera mucho más rápida y eficiente en comparación con las formas de realización manual que existían antiguamente. Precisamente, la capacidad para realizar estas tareas es lo que los han situado a la cabeza de una gran gama de aplicaciones, así como también la gran cantidad de esferas en las que estos pueden ser utilizados.

La mayor utilidad de un SIG está en la capacidad que posee de construir nuevos modelos y/o representaciones del mundo real a partir de la integración de datos con la intervención de expertos. Los SIG además, pueden modelar y manipular amplias extensiones de terreno con enormes volúmenes de datos espaciales asociados a las diversas entidades en el ámbito geográfico y tienen la misión de facilitar soluciones que ayuden en diversos problemas.

En la literatura no existe una única definición de *SIG* totalmente consensuada, esto se debe a que definir el término *Sistema de Información Geográfica* es tan difícil como definir el término *Geografía*. Sin embargo sí existe unanimidad en cuanto a las principales características que debe tener un sistema de este tipo. Definiciones de SIG se pueden encontrar en [36-38], algunas de la cuales se presentan a continuación:

"Un sistema de hardware, software y procedimientos diseñado para realizar la captura, almacenamiento, manipulación, análisis, modelación y presentación de datos referenciados espacialmente para la resolución de problemas complejos de planificación y gestión".

NCGIA (1990)

"Un sistema de información geográfica integra hardware, software y datos para capturar, gestionar, analizar y presentar todo tipo de información geográficamente referenciada".

ESRI (2010)

Los SIG modelan el espacio geográfico a partir de un conjunto de capas temáticas de datos de la misma porción de territorio. Esta manera de tratar los datos geográficos permite, la selección independiente y organizada de la información del espacio geográfico de interés, prescindiendo del resto de los datos del modelo, para agilizar la obtención de los resultados. Las capas de información pueden ser representaciones vectoriales de los elementos de la realidad o bien, representaciones Ráster como las fotografías aéreas y las imágenes satelitales entre otros.

Un *SIG* por tanto, es un sistema capaz de almacenar, manipular y desplegar información referenciada geográficamente y provee al usuario las herramientas y operadores necesarios para modelar sobre el mundo real. Por esto se puede decir que para el manejo de datos espaciales, los sistemas de información geográfica son el paso adelante más importante desde la invención del mapa.

1.3. Semántica e integración de Datos Geográficos

La composición de los datos geospaciales es extremadamente heterogénea y constituye una de las principales causas que afectan la interoperabilidad entre los diferentes sistemas en el ámbito geográfico. Hoy en día son muchos los usuarios que afrontan dificultades para encontrar, acceder y manipular estos datos [39]. No obstante, en el caso de contar con la información a la mano surgen los problemas de compatibilidad entre los diferentes formatos de almacenamiento de estos datos. Pasos que resultan viables en la resolución de estos problemas se encuentran en la utilización de estándares. El empleo de estos contribuye a la solidez e interoperabilidad en la utilización uniforme de los datos geospaciales por partes de los diferentes usuarios y sistemas [40]. Una importante iniciativa para lograr un acceso uniforme a los datos lo constituyen organizaciones como la ISO [41] y consorcios como el *OpenGIS* [2] y el *W3C*[3], que son asociaciones que buscan definir un conjunto de especificaciones técnicas, requisitos y normas para la estandarización de los datos incluyendo además, el ámbito geográfico. En el caso de la *ISO*, la norma ISO TC-211[42] tiene como meta el desarrollo de una familia de normativas, entre otras, con el objetivo de brindar apoyo para comprensión y uso compartido de la información geográfica y el establecimiento de infraestructuras de datos espaciales a nivel local, regional y global, aumentar la disponibilidad, accesibilidad e integración de la

información geográfica que posibilite la interoperabilidad entre *SIG* heterogéneos y contribuir de modo eficaz y eficiente al desarrollo sostenible.

Por otro lado el Consorcio OpenGIS (*OGC*) se dedica fundamentalmente a la creación de especificaciones para la interoperabilidad entre sistemas y usuarios que tributen a un intercambio uniforme y ordenado de la información geográfica. El Consorcio World Wide Web (*W3C*) es una comunidad internacional que trabaja arduamente para desarrollar estándares Web y es liderado por el inventor de la Web Tim Berners-Lee [43]. La misión del *W3C* es guiar la Web hacia su máximo potencial con protocolos comunes que promuevan y aseguren su evolución e interoperabilidad. Varios de los estándares más utilizados se presentan en la Tabla 1:

Tabla 1. Principales estándares en los principales formatos de los Datos Geoespaciales[44]

Modelo Ráster	Modelo Vectorial	Modelo de Bases de Datos Espaciales
Web Map Service (<i>WMS</i>)	The Earth Science Markup Language (<i>ESML</i>)	Oracle Spatial (<i>ISO/TC 211 19107 2001</i>)
GeoTIFF	Geography Markup Language (<i>GML</i>)	PostGIS (<i>ISO/TC 211 19125-2 2001</i>)
Míster SID (<i>MrSID</i>)	Keyhole Markup Language (<i>KML</i>)	
Web Coverage Service (<i>WCS</i>)	ESRI Shapefile (<i>SHP</i>)	
	Web Feature Service (<i>WFS</i>)	

Sin embargo, la normalización de los datos geográficos no significa que se resuelvan todos los problemas de interoperabilidad, ya que, estos estándares también presentan diferencias entre ellos [45]. Para llegar a un consenso entre estas diferencias en 1999 la *ISO* y el *OGC* firman un acuerdo de cooperación técnica que implica revisiones mutuas en sus respectivos desarrollos.

El problema de tener un acceso uniforme a datos geográficos heterogéneos es conocido como *Integración*, el cual constituye un aspecto importante para lograr la interoperabilidad entre los diferentes SIG [46]. La integración de los geo-datos permite a los usuarios acceder a fuentes de datos heterogéneas, presentando una vista unificada de esos datos, de forma que no se llegue a percibir tal heterogeneidad [47]. En [48] se presenta una vista de los principales esquemas de integración existentes en la literatura, véase Fig. 1.

En la Fig. 1-A se presenta una panorámica de integración en la que se traducen las fuentes de información (FI) de un formato hacia el otro. Este tipo de integración tiene como deficiencias que no se asegura que la traducción sea posible para todas las entidades contenidas en las FI y que estas no siempre son biyectivas. El esquema de la Fig. 1-B representa la integración de las diferentes FI a través de la traducción de estas hacia un formato común a ambas fuentes, la cual también puede contener las deficiencias anteriormente mencionadas. Por último, en la Fig. 1-C se presenta un esquema basado en la integración a través del cotejo de conceptos abstractos generados a partir de las FI.

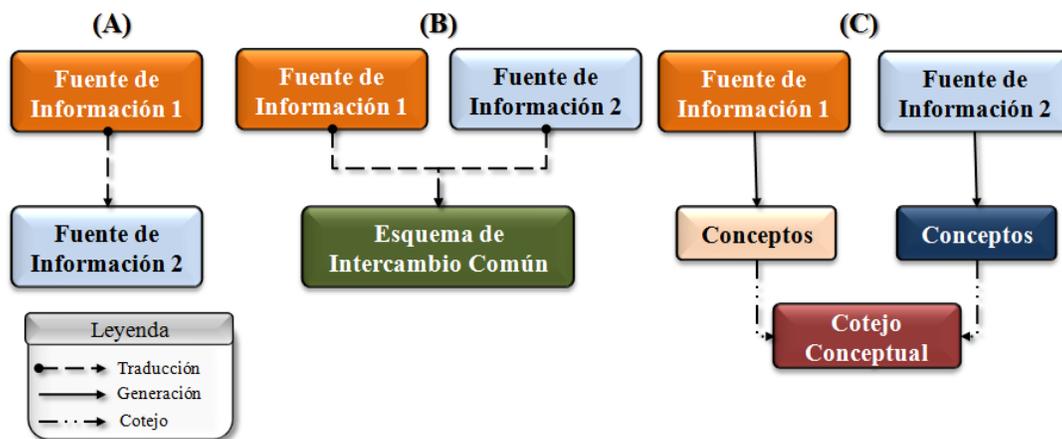


Fig. 1. Representación Gráfica de los principales esquemas de integración existentes en la literatura [48].

La integración de los geo-datos puede tratarse a varios niveles [49, 50], desde la integración física que aborda la *Integración Sintáctico-Estructural*, hasta llegar a la

denominada *Integración Semántica*. Los esfuerzos realizados en este sentido van enfocados principalmente a estos dos tipos de integración:

- *Integración Sintáctico-Estructural*: Plantea la existencia de una conexión técnica entre datos que pueden estar en diferentes sistemas de referencia o que se encuentren en diferentes formatos.
- *Integración Semántica*: Plantea la integración de la información proveniente de formatos heterogéneos basados en su semántica.

La integración *Sintáctico-Estructural* aunque ha sido en los últimos años la más utilizada presenta varias desventajas, lo que resulta que se estén buscando nuevas vías de integración. Como ventajas de la integración *Sintáctico-Estructural* podríamos decir que resulta fácil la conversión entre estándares que estén basados en los mismos formatos, como por ejemplo los que están basados en *XML*, pero por otro lado tienen como desventajas fundamentales como se mencionó anteriormente, que no siempre todas las transformaciones son biyectivas, además, se hace necesario que se definan los mismos modelos de datos para representar la información que se va a integrar (*idénticos nombres, estructuras, representaciones etc.*) y fundamentalmente que los datos no son tomados en cuenta de una manera consistente con su significado lo que puede provocar la existencia de errores.

La tendencia actual en la integración de la información geoespacial está haciendo uso como elemento fundamental la Semántica. La *Integración Semántica* facilita un procesamiento y análisis de los datos de manera más eficiente evitando los problemas de compatibilidad, los cuales la integración *Sintáctico-Estructural* no resuelve del todo. Estos tipos de integración han sido ampliamente abordados en la literatura [4, 5, 8, 9, 48, 51]. La gran mayoría de las

investigaciones en este sentido, en los últimos años, han sido sobre la base de la integración *Semántica*. Esto se debe principalmente por tratar los datos sobre la base de su conceptualización disminuyendo así la brecha semántica existente en el entendimiento *hombre-máquina*. Es así como surge la necesidad de integrar “*conocimiento*” con la información geoespacial y la premisa de gestionar este conocimiento de forma eficiente. Teniendo en cuenta la importancia que tiene el conocimiento subyacente en los datos es necesario considerar cuidadosamente cómo representarlo, de manera que se reduzcan las problemáticas existentes en la integración de la información geográfica.

Existen varias formas de representación del conocimiento y de manera general, están enfocadas a un dominio semántico específico. De las formas más usuales de representación de conocimiento existentes en la literatura se encuentran, entre otras, las *Redes Semánticas* y los *Marcos Conceptuales*. A finales de los '80 aparece un nuevo modelo que ha sido utilizado con más aceptación en estos últimos años, las *Ontologías*. A continuación en los apartados del 1.3.1 al 1.3.3 se describen brevemente estos mecanismos de representación de conocimiento.

1.3.1. Redes Semánticas

Una *Red Semántica* (creada en 1956 por Richens [52]) es una forma de representación de términos conceptuales interrelacionados entre sí, que determinan o hacen explícito el conocimiento sobre determinado dominio. Estas redes se visualizan como grafos orientados donde los nodos definen objetos o propiedades y los arcos representan las relaciones existentes entre ellos. Esta representación tiene como desventajas que no presenta una estructuración bien definida ya que los términos con diferentes niveles de abstracción pueden ser mezclados entre sí y que es el diseñador de la red quién debe decidir cuáles son

los mecanismos para la interrelación de los términos y para la resolución de conflictos, entre otras.

1.3.2. Marcos Conceptuales

Los Marcos Conceptuales, propuestos por Marvin Minsky [53], son una estructura para la representación estereotipada de una situación a través de la interrelación de términos. Esta es construida sobre la base de situaciones similares ocurridas anteriormente, permitiendo así, aplicar a situaciones nuevas los conocimientos de situaciones, eventos y conceptos previos. La representación de conocimiento basada en marcos le debe mucho a las redes semánticas. Sin embargo, los esquemas de representación basados en marcos insisten en una organización jerárquica de los conceptos utilizados, constituyendo esta, la principal diferencia con respecto a las redes semánticas, ya que las mismas no requieren tal organización.

1.3.3. Ontologías

Las ontologías han sido abordadas inicialmente desde el punto de vista de la filosofía [54-57] y en este ámbito tratan la naturaleza de la existencia y los tipos de las cosas que existen. Aunque toma su nombre de la Filosofía por analogía, el significado filosófico de esta palabra es similar a su significado dentro del campo de la Inteligencia Artificial (IA) en el sentido de la interpretación del filósofo Quine [58], el cual plantea que todo lo que puede ser cuantificado existe. Ya dentro del ámbito de la IA, en 1993 Gruber da una definición de Ontología que se convertiría en una de las más conocidas y citadas en la literatura:

"Una ontología es una especificación explícita de una conceptualización (...)"[59]

GRUBER (1993)

Para comprender claramente esta definición es necesario entender qué se quiere decir con conceptualización. Breis en [60] plantea que es una interpretación estructurada de una parte del mundo que usan los Seres Humanos para pensar y comunicar sobre ella. En este trabajo se entiende conceptualización como una forma de entendimiento consensuado sobre determinado objeto o dominio que puede ser representado y/o especificado; y por explícita se entiende que esta especificación y/o representación sea expresada de forma clara y precisa. Esta definición de ontología propuesta por Gruber más tarde en 1997 fue extendida por Borst incluyendo además que esta conceptualización debe ser consensuada:

“Una ontología es una especificación formal de una conceptualización compartida”[61]

BORST (1997)

Studer y demás autores en 1998 ofrecen una explicación de las definiciones de Gruber y Borst, en la que plantean lo siguiente:

“Conceptualización se refiere a un modelo abstracto de algún fenómeno en el mundo a través de la identificación de los conceptos relevantes de dicho fenómeno. Explícita significa que el tipo de conceptos y restricciones usados se definen explícitamente. Formal representa el hecho de que la ontología debería ser entendible por las máquinas. Compartida refleja la noción de que una ontología captura conocimiento consensual, esto es, que no es de un individuo, sino que es aceptado por un grupo”[62]

STUDER, BENJAMINS ET AL. (1998)

Cualquiera que sea el entorno de desarrollo de una ontología, es importante especificar que la base para su construcción es la conceptualización junto con un lenguaje para referirse a las entidades de un dominio en particular. Las ontologías precisan de los siguientes componentes para representar el conocimiento [63]:

- *Conceptos*: Son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, procesos de razonamiento, etc.
- *Relaciones*: Representan la interacción y enlace entre los conceptos, formando la taxonomía del dominio.
- *Instancias*: Se utilizan para representar objetos determinados de un concepto.
- *Axiomas*: Son teoremas que se declaran sobre las condiciones que deben cumplir los elementos de la Ontología.

En la literatura se definen varios niveles de generalidad que dan lugar a distintos tipos de ontologías [15], véase además la Fig. 2:

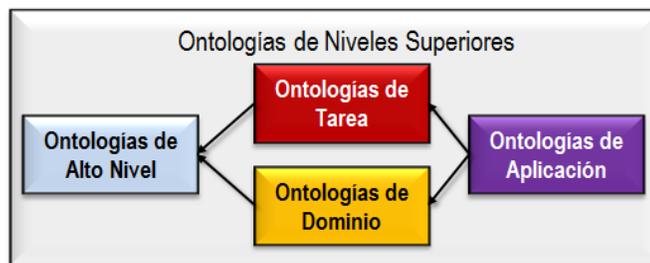


Fig. 2. Representación Gráfica de la jerarquía de los diferentes tipos de Ontologías propuesto en la literatura [15].

- *Ontologías de Alto Nivel*: Proporcionan términos genéricos reutilizables en diferentes dominios.
- *Ontologías de Dominio y de Tarea*: Expresan conceptos que son específicos de un dominio determinado (*Suelos*) o de determinada tarea (*Comercializar*). Estos conceptos usualmente son especializaciones de conceptos existentes en ontologías de Alto Nivel.
- *Ontologías de Aplicación*: Contienen todas las definiciones que son necesarias para modelar los conocimientos requeridos en un ámbito específico o aplicación particular en cierto dominio, por ejemplo la estimación del potencial agro

productivo de los suelos. Estos conceptos usualmente son especializaciones de conceptos existentes en ontologías de Dominio o de Tarea.

1.4. Representación Estática de las Ontologías Actuales

Hay que destacar que, los tipos de ontologías mencionados anteriormente proporcionan los elementos para la representación del conocimiento acerca de un dominio. Estos elementos siempre van a manifestarse en todo el ámbito del dominio representado, ya que representan el conocimiento adquirido y consensuado por expertos. El cambio de alguno de estos elementos siempre afecta la semántica del dominio representado, por lo que puede decirse que estas ontologías capturan el conocimiento de forma estática, o sea que sus componentes no cambian una vez definido y consensuado por los especialistas del dominio.

Por ejemplo, en el siguiente esquema, véase la Fig. 3, se representa un fragmento simplificado de una ontología de dominio de objetos urbanos.

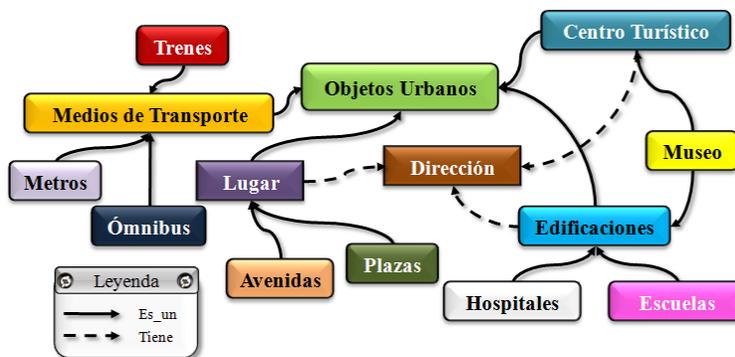


Fig. 3. Representación Gráfica Simplificada de una Ontología de Dominio de Objetos Urbanos.

El conocimiento representado en esta ontología siempre es el mismo y no cambia, porque representa la visión de estas entidades en el mundo real desde un punto de vista abstracto. Según esta, los *Trenes* siempre son *Medios de Transporte*, no otra cosa y los *Centros Turísticos* siempre tienen una *Dirección*. Cualquier cambio y/o modificación de los términos y/o relaciones en esta ontología podría implicar que cambie su semántica, o sea,

que su significado sea otro, porque estaría representando otro tipo de conocimiento y por tanto, su esencia no sería la misma, sería otra ontología.

La visión representada en estas ontologías no tiene en cuenta el conocimiento que subyace en los datos geoespaciales. Estos datos son obtenidos a partir de una abstracción realizada a objetos geográficos existentes en el mundo real, los cuales al ser combinados en *SIG*, pueden dar lugar a nuevos términos y nuevas relaciones entre ellos. Disponer de esta información aporta un nuevo conocimiento que dotaría de mayor robustez al procesamiento de los datos para la realización de tareas convencionales. Por ejemplo, al utilizar la ontología de la Fig. 3 no podría obtenerse información sobre que *Ómnibus* pasa cerca de determinada *Escuela* puesto que esa relación no se cumple para todos los objetos pertenecientes a estas clases, por lo tanto, no puede representarse en una ontología de este tipo. Sin embargo, este conocimiento puede ser extraído de los datos, lo que trae como consecuencia que aumente el nivel de especialización semántica de la ontología y se disponga de un conocimiento más preciso y cercano a la realidad que se modela.

Resumiendo, podemos plantear que con estos tipos de ontologías se pueden modelar dominios de conocimiento pero solo hasta un cierto nivel de abstracción, ya que:

- No tienen una relación directa con respecto a los objetos geoespaciales que están siendo modelados en el sistema, véase Fig. 4-A.
- No es posible modelar las características y relaciones existentes entre los objetos geoespaciales que subyacen en los datos, véase Fig. 4-B.

Esto es debido a que estas ontologías solo pueden modelar las diferentes relaciones que de manera general son siempre ciertas en el dominio, pero no pueden representar las nuevas relaciones que se generan a partir de la integración de datos en el *SIG*, por ejemplo:

Una relación topológica, como las que ha definido Egenhofer [64], del tipo adyacencia entre una instancia de la clase Lago (*Lago_A*) y una instancia de la clase Centro_Turístico (*Cent_Tur_C*), no es posible que pueda ser representada en estas ontologías ya que estas relaciones no siempre ocurren, estas dependen de los datos que se hayan cargado al SIG.

- No son capaces de descubrir nuevos niveles de abstracción subyacentes en los datos geoespaciales integrados al SIG, véase Fig. 4 (C).

Esto es debido a que es imposible representar nuevos niveles de abstracción semántica que pudieran estar subyacentes en los datos a partir de agrupaciones realizadas.

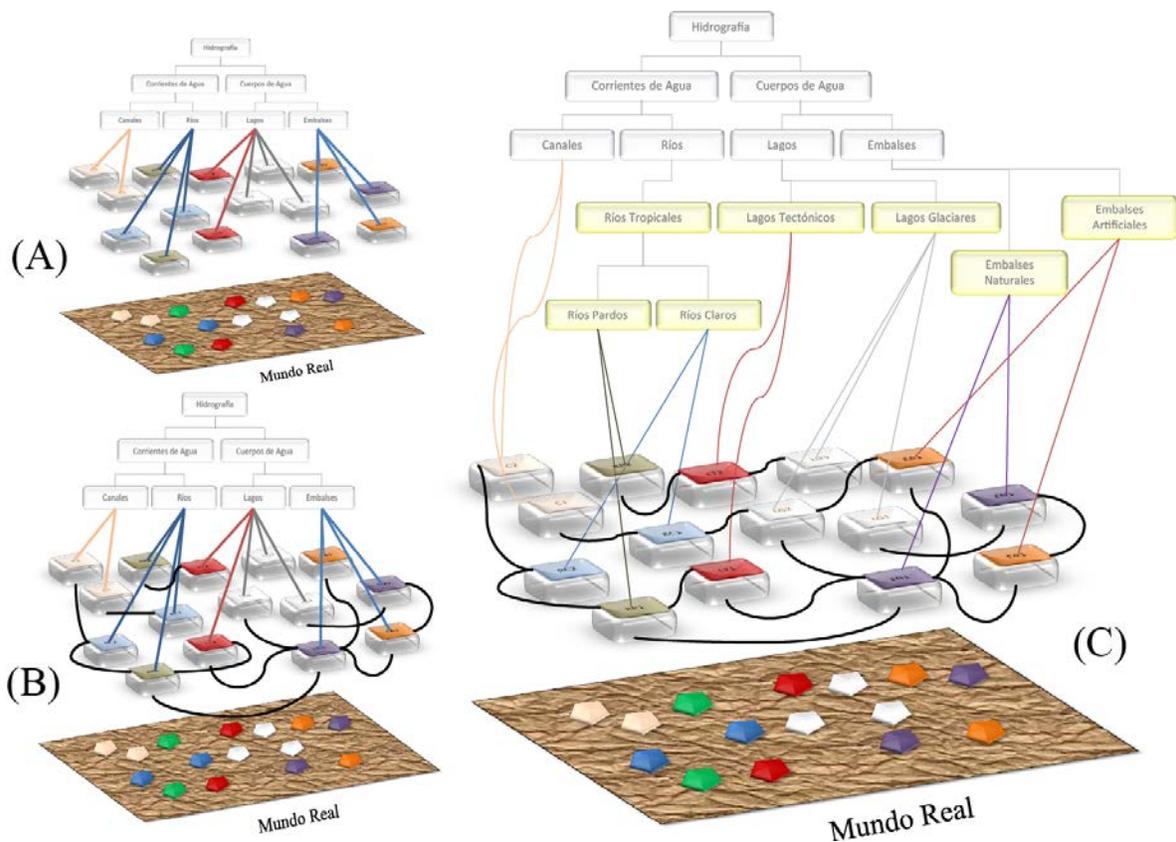


Fig. 4. Representación Gráfica de las carencias existentes en las Ontologías propuestas por Guarino con respecto a los datos integrados al SIG, (A) representación de la relación de la Ontología con los objetos integrados al SIG, (B) representación de las relaciones que se generan entre

estos objetos y (C) representación de las nuevas clases que pudieran descubrirse a partir de agrupamiento de objetos.

1.5. Ontologías como un componente más de los SIG

El uso de ontologías le permite al SIG centrarse en el significado de los datos en lugar de su sintaxis o estructura propiciando así una buena interoperabilidad entre los diferentes sistemas de este tipo. Además, permiten la integración de datos geográficos de una manera mucho más abstracta en la que el conocimiento juega un rol fundamental.

Todo esto ha llevado a pensar en SIG gobernados por ontologías (SIGGO), un concepto que ha sido introducido por Fonseca [9], abriendo nuevas maneras de pensar sobre las ontologías y los SIG. Los SIGGO pueden describirse como Sistemas de Información Geográfica en los que se incorpora el uso de ontologías como un componente activo dentro de su arquitectura. Estas constituyen una componente más, como lo es la base de datos temáticos o espaciales, que intervienen y cooperan de forma conjunta para alcanzar los objetivos para los cuales fue creado y sobre todo sientan la base para compartir información y colaborar, actividades que en la base, como cualquier proceso de sinergia, conduce a nuevas y mejores soluciones. La metodología para la construcción de SIGGO planteada por Fonseca en [9] es presentada a través de una arquitectura, la cual está centrada en dos aspectos fundamentales, la *generación* y *uso* del conocimiento.

La fase de *Generación del Conocimiento*, véase Fig. 5, según Fonseca, comprende la especificación de ontologías por parte de expertos a través del uso de un editor de Ontologías a partir de las cuales es posible definir estructuras de datos (*Clases*) que contengan todos los atributos y funciones que han de brindar el soporte a todas las funcionalidades del sistema. La principal ventaja en la utilización de estas Clases es que

estas van a contener todo el conocimiento subyacente en las ontologías a partir de las cuales fueron creadas.

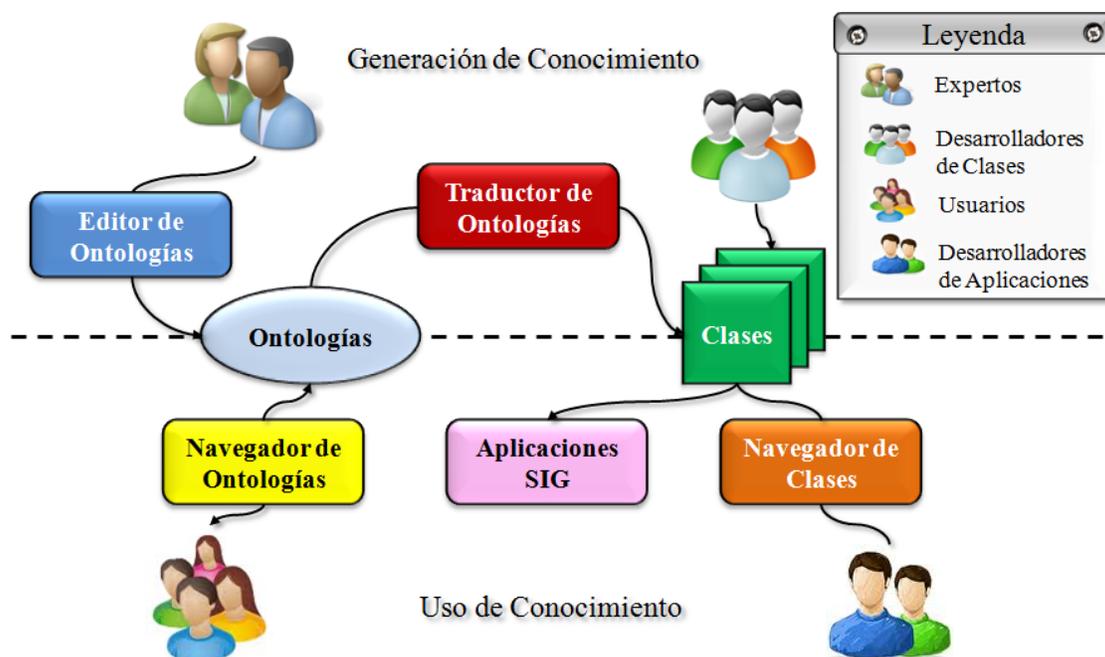


Fig. 5. Representación Gráfica de la arquitectura SIGGO propuesta por Fonseca [9].

La fase de *Uso del Conocimiento*, véase Fig. 5, se apoya en los productos obtenidos en la fase anterior (*Generación del Conocimiento*): una serie de ontologías especificadas en un lenguaje formal y una serie de clases derivadas de esas ontologías. Estas ontologías y Clases generadas se encuentran disponibles para ser navegadas y consultadas por los usuarios finales, como una vía de acceso al conocimiento incluido en ellas y también para la construcción de aplicaciones SIG.

Garea-Llano en [65], propone una modificación en la arquitectura de Fonseca, véase Fig. 6, en la cual incluye un módulo para la generación de conocimiento de forma automática desde fuentes de datos.

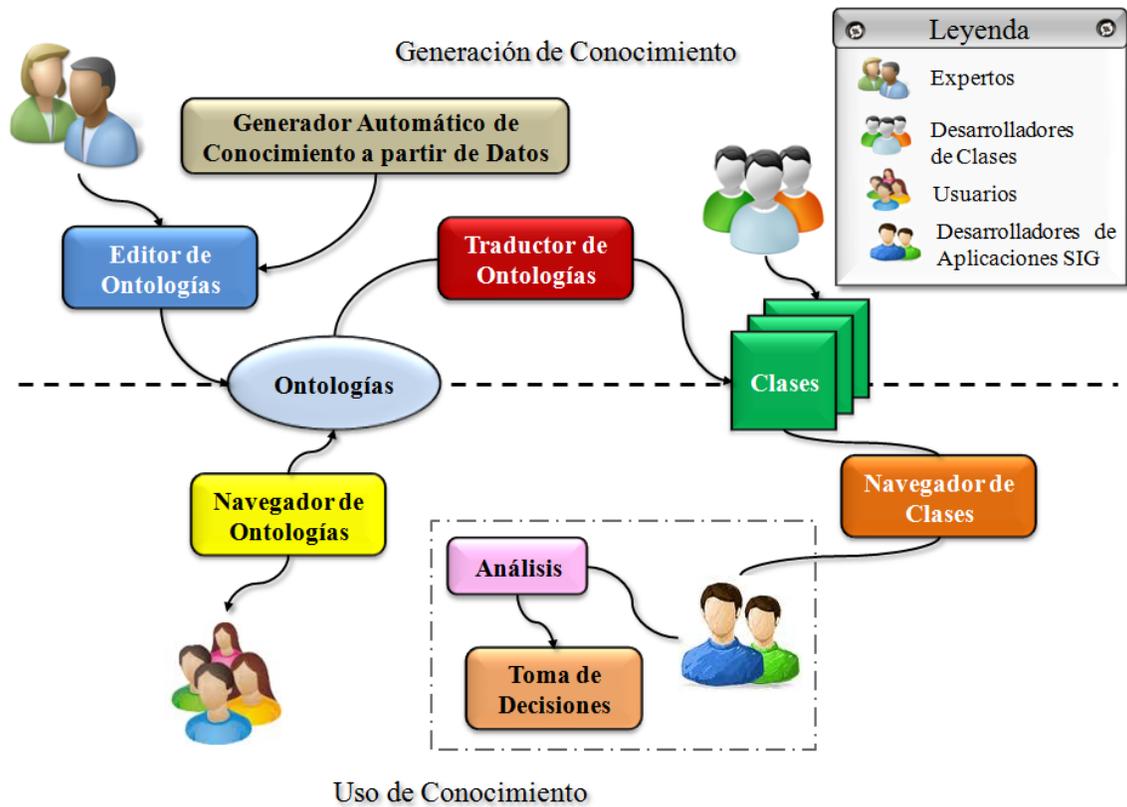


Fig. 6. Representación Gráfica de la modificación de la arquitectura SIGGO de Fonseca propuesta por Garea-Llano [65].

A criterio del autor, la lógica detrás de esta inclusión está fundamentada en que el esquema propuesto por Fonseca está basado en la suposición de un mutuo acuerdo alcanzado por un grupo de personas sobre el significado semántico de las entidades geográficas existentes en un determinado ámbito de análisis, las cuales son posteriormente conceptualizadas y representadas de forma explícita en una ontología o en la integración de varias de estas. Fonseca en este caso, no toma en cuenta la representación lógico-formal de esas entidades almacenadas en las diferentes estructuras de datos, como por ejemplo las bases de datos espaciales para la creación de las ontologías. Es principalmente por esta razón que Garea-Llano en la fase de *Generación de Conocimiento* incluye el módulo “*Generador Automático de Conocimiento a partir de Datos*”. En la fase de *Uso de Conocimiento* Garea-Llano también propone una ligera modificación en la arquitectura de Fonseca (véase

recuadro de la Fig. 6). Esta consiste en que los *Desarrolladores de Aplicaciones* en este caso desarrolladores de aplicaciones SIG, puedan crear aplicaciones capaces de utilizar el conocimiento subyacente en las Clases derivadas de las ontologías para mejorar tareas existentes en las aplicaciones SIG convencionales como el análisis y la toma de decisiones. Finalmente podemos plantear que los SIGGO, sin duda alguna, pueden constituir la base tecnológica a seguir para el logro de una verdadera interoperabilidad semántica entre las diferentes comunidades que utilizan los datos espaciales en tareas como la toma de decisiones y análisis de la información espacial en general. Aunque actualmente el desarrollo de estos se encuentra en sus primeras fases de estudio, en la literatura ya podemos encontrar varios trabajos enfocados al desarrollo y uso de ontologías en SIG [9, 16, 48, 66], lo que pone de manifiesto el interés científico a nivel mundial sobre esta rama. Por lo que puede plantearse que los SIGGO constituyen un nuevo paradigma en la forma de tratar la información geográfica.

1.6. Conclusiones

De lo analizado en el estudio realizado sobre el estado del arte y de lo presentado como marco teórico referencial de esta investigación, se concluye que existe una creciente necesidad de crear mecanismos para el procesamiento y la integración de datos geoespaciales desde un punto de vista semántico.

En la literatura se realizan algunas aproximaciones para tratar estos problemas. Las mejores vías de solución están basadas en el procesamiento de estos datos utilizando ontologías como mecanismo de representación de conocimiento. En este sentido, diversos tipos de ontologías han sido propuestas en la literatura. Estas van desde los niveles de abstracción más generales (*Ontologías de Alto Nivel*) hasta aquellos más específicos (*Ontologías de*

Aplicación), pero, a criterio del autor, existe todavía conocimiento que no está siendo aprovechando y es precisamente el que subyace en los mismos datos geoespaciales. Además, resulta necesario que los sistemas de información geográfica sean capaces de realizar un procesamiento semántico de la información geoespacial, por tales motivos en los próximos apartados se presentarán elementos que afrontan estos problemas y se mostraran los resultados alcanzados en este sentido.

Capítulo 2

**MÉTODOS PARA LA REPRESENTACIÓN
SEMÁNTICA DE OBJETOS GEOESPACIALES EN
EL ESPACIO SEMÁNTICO MULTIDIMENSIONAL**

CAPÍTULO 2

Métodos para la Representación Semántica de Objetos Geoespaciales en el Espacio Semántico Multidimensional

En este Capítulo se proponen métodos para la representación semántica de objetos geoespaciales en un espacio multidimensional. Primeramente se define un nuevo tipo de ontología, la *Ontología de Representación de Datos (ORD)*. La idea detrás de este concepto se basa en la definición de forma explícita de la estructura, relaciones y métodos necesarios para la representación del conocimiento subyacente en los datos geográficos. Primeramente se presenta el ámbito de aplicación y el alcance de esta nueva ontología enfocándose en como una ontología de este tipo puede contribuir de forma positiva a la integración semántica de fuentes heterogéneas de datos geográficos y a los problemas de interoperabilidad existentes entre los diferentes SIG. Seguidamente se continúa con la definición formal y la descripción detallada de la *ORD*. A continuación se introduce la definición de *Espacio Semántico Multidimensional* y cómo los objetos geoespaciales son representados en este espacio con la ayuda de la *ORD*. Posteriormente se propone una arquitectura en la que se contemplan estos elementos para su implementación dentro del módulo de *Generación Automática de Conocimiento a partir de Datos* (véase Fig. 6 del Capítulo 1). Finalmente se exponen las principales conclusiones del capítulo.

2. Métodos para la Representación Semántica de Objetos Geospaciales

Los diferentes tipos de ontologías propuestos en la literatura son capaces de modelar la semántica de los diferentes dominios de conocimiento a diferentes niveles de generalidad. Sin embargo, como se ha planteado anteriormente, existe un nivel de generalidad semántica que no puede ser representado en estos tipos de ontologías debido a la naturaleza del conocimiento que ellas representan. Este nivel de generalidad al que se hace referencia tiene un mayor grado de especialización con respecto a los demás niveles propuestos y el conocimiento que representa es dinámico. Este tipo de conocimiento no es provisto originalmente por los especialistas del dominio sino por la naturaleza y comportamiento de los datos que se integran en el SIG y las relaciones existentes entre los objetos geospaciales contenidos en estos datos.

Por tanto para representar este tipo de conocimiento, se hace necesaria la definición de un nuevo tipo de ontología. Este nuevo tipo de ontología resulta de gran utilidad por cuanto enriquece semánticamente al dominio mejorando los procesos de análisis de la información geoespacial. A continuación en la sección 2.1 se analizan estos aspectos con mayor profundidad.

2.1. Ontologías de Representación de Datos (ORD)

En el capítulo anterior se puso de manifiesto una serie de insuficiencias que presentaban los tipos de ontologías propuestos en la literatura para la representación semántica del conocimiento que subyace en los datos geospaciales integrados en los SIG. Para abordar estos problemas se propone la definición de un nuevo tipo de ontología, denominado como “*Ontología de Representación de Datos (ORD)*”. En este tipo de ontología se representa un

mayor grado de especialización semántica con respecto a los niveles representados en los tipos de ontologías existentes.

En la ORD se capturan los rasgos característicos de bajo nivel de los objetos geoespaciales almacenados en las fuentes de datos y las relaciones que se generan entre ellos al ser integrados en el SIG. De esta manera, puede representarse la naturaleza tanto semántica como dinámica de estos objetos independientemente del tipo y formato en el que se encuentren almacenados. En esencia, el objetivo de la ORD es representar la información de los objetos como conocimiento, alimentándose de los datos integrados al sistema. Este tipo de ontología contribuye conjuntamente con los tipos de ontologías existentes a que los objetos geoespaciales puedan no solo ser utilizados por el sistema sino también comprendidos.

De manera formal la *Ontología de Representación de Datos* se define como:

- ***Ontología de Representación de Datos:*** *Contiene las definiciones necesarias para la representación de los rasgos característicos y relaciones que modelan y dan significado a objetos pertenecientes a un dominio desde un punto de vista semántico.*

Los términos en este tipo de ontología pueden ser definidos como especializaciones de conceptos existentes en ontologías de Aplicación, de Dominio y/o de Tarea.

Basándonos en el esquema propuesto por Guarino, véase Fig. 2, incluimos la *Ontología de Representación de Datos* en el nivel de generalidad más bajo, por representar esta, el mayor grado de especialización con respecto a las demás ontologías, véase Fig. 7.

Sobre la base de la propuesta del nuevo tipo de ontología se nombrará los tipos de ontologías existentes como *Ontologías de Niveles Superiores (ONS)* a lo largo de este trabajo.

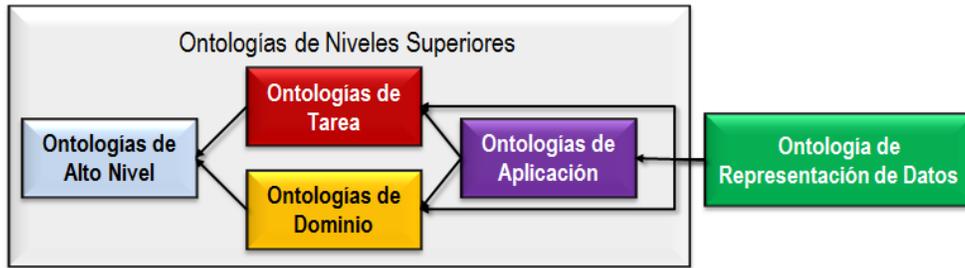


Fig. 7. Representación Gráfica de la jerarquía de los diferentes tipos de Ontologías propuesto por Guarino incluyendo la Ontología de Representación de Datos.

La ORD puede presentarse como una especialización tanto de ontologías de aplicación como de las de dominio o de tarea y esto depende del nivel de abstracción con el que se estén manejando los datos integrados al sistema. Este tipo de ontología define una ontología dinámica ya que la estructura, términos y relaciones de esta van a depender siempre de los objetos 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 empleadas por el usuario. Este complemento va dirigido en ambos sentidos, ($ONS \leftrightarrow ORD$), véase Fig. 8.

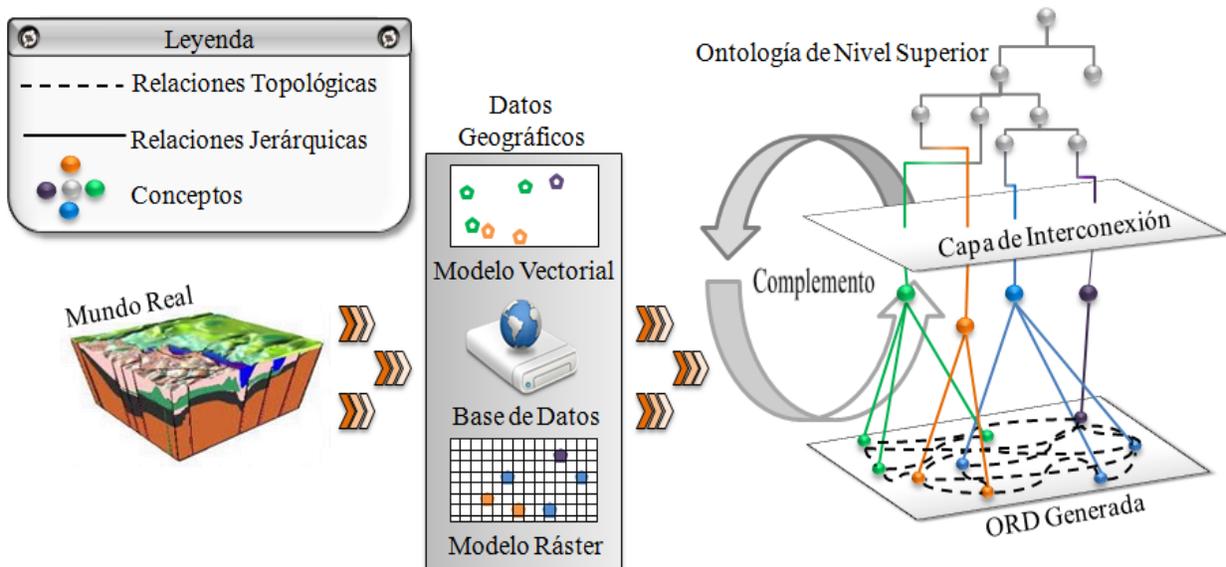


Fig. 8. Representación Gráfica del complemento generado entre la ORD y la OD en la Representación Semántica de los datos geográficos.

Esto se explica desde el punto de que por un lado, la ONS puede contar con información precisa de los objetos integrados al sistema, lo que implica contar con mayores niveles de especialización que aumentan el nivel de precisión semántica en el dominio especificado, por el otro lado la ORD dispone de un dominio que define la naturaleza semántica de esos objetos. De esta manera se pone de manifiesto un proceso de sinergia, en el que los resultados obtenidos a través de estos tipos de ontologías combinadas (ORD + ONS) son mejores que la suma de los resultados obtenidos por cada una de ellos por separado. Tomando en cuenta lo planteado anteriormente, se propone que el uso de la ORD sea siempre en conjunto con al menos una Ontología de Nivel Superior.

En este sentido puede destacarse que la gestión de los objetos geoespaciales incluidos en el sistema (*adición, modificación y/o eliminación*) provoca cambios estructurales en la ORD ya que esta es generada automáticamente a partir de estos datos. Es por esta razón que se plantea que este tipo de ontología define ontologías dinámicas ya que en ellas cambian los términos y relaciones según cambian los datos que se están representando, manteniendo la esencia semántica de estos en todo momento. Otro aspecto a destacar es que todos estos cambios estructurales no implican absolutamente ningún cambio en la(s) ONS manteniendo esta(s) su estructura original. Cada uno de estos cambios ocurre por debajo de la capa de interconexión véase Fig. 8.

Ontologías de este tipo pueden 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 análisis realizado a los datos en el proceso de generación de la ORD, aspecto que se explica con más detalle en el apartado 2.1.3.

2.1.1. Estructura Interna de la Ontología de Representación de Datos

Como se ha mencionado anteriormente la *ORD* constituye la proyección semántica de los objetos geospaciales. Estos objetos se describen sobre la base de una misma estructura por lo que se hace posible entonces un acceso uniforme a ellos independientemente del formato o estándar en el que han sido almacenados.

En la *ORD* los objetos geospaciales no son representados a través de conceptos sino a través de una estructura un poco más compleja, véase Fig. 9. Esta estructura es denominada como *Nodos de Representación de Datos (NRD)*, la cual es explicada en más detalles en la sección 2.1.2. Por tanto, la estructura de la *ORD* está compuesta por un conjunto de *NRD*, donde cada uno de estos representa la abstracción semántica de un objeto geoespacial en particular y aristas que interrelacionan estos *NRD* las cuales representan las relaciones existentes entre estos.

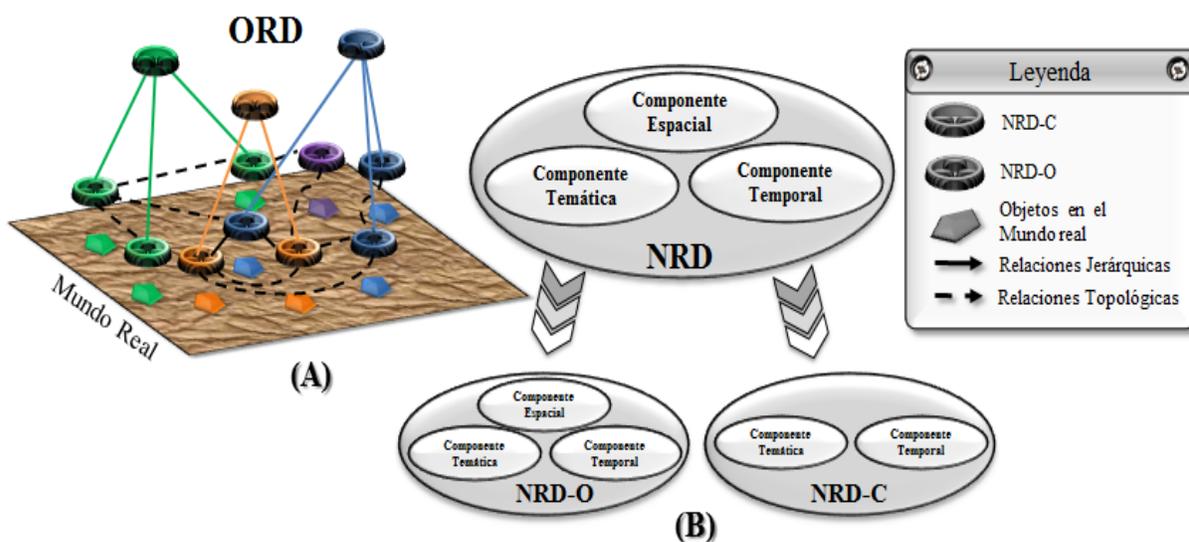


Fig. 9. (A) Representación Gráfica de la estructura de la Ontología de Representación de Datos. (B) Representación Gráfica del Nodo de Representación de Datos en sus variantes NRD-O para Objetos y NRD-C para Clases.

2.1.2. Estructura Interna de los Nodos de Representación de Datos

La estructura de los NRD está basada en la representación de las características temáticas, espaciales y temporales de los objetos geográficos, a través de tres subestructuras (*ctemat*, *cespac* y *ctemp*). Los NRD pueden presentarse en dos variantes, NRD-O para objetos geográficos los cuales presentan las tres subestructuras mencionadas anteriormente y los NRD-C para clases, las cuales carecen de la Componente Espacial, ya que esta variante representa la abstracción semántica del conjunto de objetos (NRD-O) que ella agrupa, véase Fig. 9-B. Por tanto, la estructura interna de los NRD de manera general está determinada de la siguiente manera:

- El NRD se representa 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* se representa 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* se representa 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 en la cual existe el evento.
 - **granul**: Define el tipo de granularidad en el tiempo (Año, Mes, Día etc.).
- La estructura *ctemat* se representa por la 2–tupla (*nombre*, (...)) donde:
 - **nombre**: Nombre de la entidad geográfica representada.
 - (...): Propiedades Temáticas de la entidad geográfica representada que la definen semánticamente.
- La estructura *cespac* se representa por la 3–tupla (*ubic*, *repres*, *esca*) donde:
 - **ubic**: Define la ubicación espacial de la entidad geográfica representada en el plano, (coordenadas).
 - **repres**: Define la geometría en la que está representada la entidad, p.e. *Punto*, *Línea*, *Polígono* en el caso de la representación vectorial.
 - **esca**: Define el tipo de escala de representación de la entidad geográfica.

2.1.3. Generación Automática de la ORD

La ORD se genera a partir de la información de los datos con los que se trabaja. Los mecanismos a utilizar para la generación de la ORD dependen del formato de

almacenamiento de los datos que se desea procesar. En este apartado se muestran los principales pasos a seguir para la generación automática de la ORD.

Los principales pasos a seguir para la generación automática de la ORD son:

1. Definición de la estructura en la que está almacenado el dato:
 - 1.1. Precisar cuáles son las estructuras que se refieren a sus tres principales componentes (*Temática, Espacial, Temporal*) y cómo acceder sus valores.
2. Extracción de valores:
 - 2.1. *Componente Temática*: Se extraen y normalizan los atributos temáticos del objeto analizado. Se almacenan en la estructura *ctemat* del *NRD-O*.
 - 2.2. *Componente Espacial*: Se extraen y normalizan los atributos vinculados con la ubicación en el espacio y la proyección cartográfica utilizada. Se almacenan en la estructura *cespac* del *NRD-O*.
 - 2.3. *Componente Temporal*: Se extraen y normalizan los atributos temporales del objeto analizado. Se almacenan en la estructura *ctemp* del *NRD-O*.
3. Generación de Nuevos Conceptos:
 - 3.1. Se descubren nuevos niveles de abstracción (conceptos) que pudieran estar subyacente en los datos. Se representan como *NRD-C*.
4. Extracción de Relaciones:
 - 4.1. Se extraen las relaciones existentes entre los objetos analizados. Se almacenan en la estructura *rel* del *NRD-O*.

Estos pasos son explicados con más detalles en los apartados siguientes 2.1.3.1, 2.1.3.2, 2.1.3.3, 2.1.3.4.

2.1.3.1. Definición de la Estructura en la que está Almacenado el Dato

Como se ha planteado anteriormente, en los datos geográficos existe una gran heterogeneidad de formatos y estándares. Por tanto, el acceso a las estructuras en que se encuentran las componentes *Temática*, *Espacial* y *Temporal* son variables.

Por ejemplo, para el caso de los datos en formato vectorial ESRI Shapefile tenemos que, su estructura de almacenamiento requiere como mínimo tres archivos, estos son:

- **.shp*: Es el archivo en el que se almacena la componente espacial.
- **.shx*: Es el archivo que almacena el índice de los objetos almacenados.
- **.dbf*: Es el archivo en el que se almacena la componente Temática.

Siguiendo con los datos en formato vectorial ahora presentamos el estándar KML[67]. Este estándar está basado en el lenguaje de marcado XML y ha sido desarrollado por la compañía Google para su utilización en Google Earth, aunque también es ampliamente utilizado por herramientas como Google Maps, NASA WorldWind, ESRI ArcGIS Explorer y AutoCAD entre otros. En su estructura, para el almacenamiento de los datos geográficos, tiene de manera resumida, el título y una descripción, elementos referidos a la componente Temática del objeto, además, presenta estructuras para representar las coordenadas del objeto referidas a la componente Espacial. Un ejemplo de un archivo KML podría ser, véase Fig. 10.

Como es posible observar estas dos variantes, aunque son formatos vectoriales, los mecanismos para el almacenamiento de las principales componentes de los datos geográficos varían drásticamente en ambos. Por lo que en este paso hay que determinar en qué formato y estándar se encuentran los datos y cómo acceder a estos valores. Una vez determinado esto se puede continuar al próximo paso.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <kml xmlns="http://www.opengis.net/kml/2.2">
3   <Placemark>
4     <name>La Habana</name>
5     <description>La Habana, capital de todos los Cubanos</description>
6     <Point>
7       <coordinates>23.143391,-82.388306,0</coordinates>
8     </Point>
9   </Placemark>
10 </kml>
```

Fig. 10. Representación gráfica de un archivo en formato KML.

2.1.3.2. Extracción de Valores

En este paso se extraen los elementos correspondientes a las tres principales componentes de los datos geográficos:

- *Componente Temática*: Son elementos generalmente referidos a las propiedades o atributos de los objetos geográficos que responden a la pregunta: *¿Qué es?*
- *Componente Espacial*: Son elementos generalmente referidos a la localización en el espacio que responden a la pregunta: *¿Dónde está ubicado el objeto?*
- *Componente Temporal*: Son elementos generalmente referidos al momento en que se manifiesta el dato y lo provee de propiedades que pueden ser medidas en el tiempo. Son atributos que responden a la pregunta: *¿Cuándo?*

Una vez determinados estos valores se crean cada uno de los NRD-O correspondientes a cada uno de los objetos, luego se pasa a determinar los nuevos términos/conceptos en los que estos pudieran ser agrupados para establecer nuevos niveles de abstracción y por último se determinan las relaciones existentes entre ellos.

2.1.3.3. Generación de Nuevos Conceptos

La tarea fundamental en este paso es el descubrimiento de nuevos niveles de abstracción en la taxonomía ontológica (conceptos) que pudieran estar subyacentes en los datos. Esto podría realizarse a partir de la agrupación de datos con características comunes utilizando

técnicas de agrupamiento jerárquico. El proceso de agrupamiento debe tener en cuenta al menos los siguientes elementos:

- Los datos son dinámicos (*referido a la adición, modificación y/o eliminación de datos*). Esto implica que pudieran ocurrir cambios estructurales en la ontología.
- Conservar las características que se tomaron en cuenta para la agrupación con el objetivo de generar el nuevo nivel de abstracción (*Representante del grupo*).
- La agrupación ha de ser jerárquica y cada nivel de la jerarquía representará diferentes niveles de abstracción semántica en los datos.
- La agrupación ha de ser incremental, o sea, cada nuevo objeto es colocado en un grupo existente o forma uno nuevo.
- Las agrupaciones realizadas no deben depender del orden de presentación de los datos. Esto significa que no importa el orden en que los datos son incluidos o eliminados del sistema, siempre se debe obtener la misma estructura ontológica.

El algoritmo *Jerárquico Incremental* propuesto por Pons [68] podría ser de utilidad. Este algoritmo es capaz de construir jerarquías de grupos con diferentes niveles de granularidad. En este caso, para el uso de este algoritmo en el descubrimiento de los nuevos conceptos es necesario tener en cuenta los siguientes aspectos:

- El uso de la medida de semejanza de Jaccard [69] ya que esta se ajusta a la naturaleza de los datos a procesar.
 - Esta medida ofrece un grado de similitud entre dos individuos basándose en las ocurrencias de características comunes entre ellos.
- El método de selección de los representantes del grupo (Centroide).

- Para el cálculo del centroide del grupo no se deben tomar en cuenta los métodos convencionales como la media, la mediana, el promedio o el elemento más similar a los restantes miembros del grupo, sino que se ha de generar un nuevo elemento que va a contener todas las características que son comunes en todos los miembros del grupo.

2.1.3.4. Extracción de Relaciones

En el dominio geoespacial las relaciones entre los objetos geográficos son tan importantes como los mismos objetos [70], por lo que en este paso se extraen todas las posibles relaciones que puedan existir entre los objetos analizados, como por ejemplo las relaciones *topológicas y/o espaciales*. Las formas de extraer estas relaciones varían también en dependencia del tipo y formato de datos a procesar. Estas son representadas a través de aristas que unen a los NRD.

2.2. Espacio Semántico Multidimensional

Como se ha planteado anteriormente la representación semántica de los objetos geoespaciales es realizada a través de la combinación de la ORD con las ONS. En esta representación los objetos geoespaciales son mapeados hacia un nuevo espacio, el cual ha sido denominado como “*Espacio Semántico*”. En este espacio las abstracciones de los diferentes objetos van a coexistir de forma conjunta, dando la posibilidad de que estos puedan ser accedidos de manera uniforme. De esta manera muchos de los problemas que trae consigo la integración desde el punto de vista *Sintáctico-Estructural* son evitados. Además, es bien conocido que los objetos geográficos tienden a tener diferentes significados para diferentes personas. En el ámbito geoespacial hay muchos ejemplos de objetos cuya semántica abarca diferentes dominios de conocimiento. Este hecho implica

que los diferentes especialistas asignen diferentes denominaciones para el mismo objeto. Un ejemplo podemos verlo con el objeto “*Ciénaga de Zapata*”, el cual puede ser denominado como “*Reserva Natural*”, sin embargo esta designación no tiene en cuenta toda la naturaleza semántica de este objeto, ya que también puede ser nombrado como “*Lugar Histórico*”. El hecho de que la semántica de un objeto abarque varios dominios disjuntos de conocimiento puede ser nombrado como “*Semántica Multidimensional*”.

La palabra “*Semántica*” hace referencia al significado del objeto, teniendo en cuenta todos los dominios de conocimiento que lo definen semánticamente. Tener en cuenta toda la semántica de un objeto es sumamente difícil, por lo que en general solo se toma en cuenta un subconjunto de esos dominios. De esta manera, se realiza una “*Abstracción Semántica*” del objeto analizado. Cuando la abstracción semántica realizada sobre un objeto abarca dos o más dominios disjuntos se define como “*Abstracción Semántica Multidimensional*”. Formalmente, podemos definir la semántica multidimensional de un objeto de la siguiente manera:

- Sea $D_o = (d_1, d_2, \dots, d_n)$ el conjunto de los dominios de conocimiento que definen la naturaleza semántica del objeto “ o ”, para lo cual se satisface que $d_i \cap d_j = \emptyset$ con $i \neq j$ y $n \geq 1$.

De esta manera podemos definir la abstracción semántica multidimensional de un objeto como un subconjunto de D_o con $n \geq 1$. La representación de la abstracción semántica multidimensional de los objetos geoespaciales trae grandes beneficios ya que permite que diferentes sistemas procesen estos objetos desde diferentes puntos de vista. Esta representación es realizada en el denominado “*Espacio Semántico*”. En este espacio, los diferentes objetos se representan sobre la base de su abstracción semántica la cual es

denominada como “Contexto”. El contexto define explícitamente el o los dominios sobre el cual los objetos geográficos son analizados. Es importante destacar que cada objeto en el Espacio Semántico tiene solamente un contexto y que cada contexto puede incluir uno o varios objetos, véase Fig. 11.

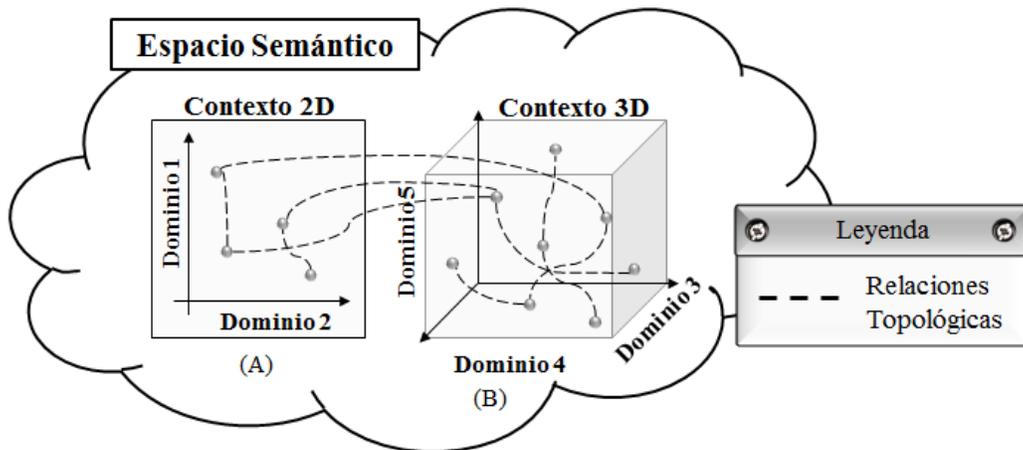


Fig. 11. Representación de dos contextos diferentes en el espacio semántico. En (A) se muestra un contexto 2D definido por dos dominios y en (B) se muestra un contexto 3D definidas por tres dominios. Objetos geográficos que pertenecen a diferentes contextos pueden ser vinculados a través de otras relaciones existentes entre ellos.

Para definir formalmente el Espacio semántico es necesario establecer que:

- Sea $A = \{a_1, a_2, \dots, a_k\}$ conjunto finito de atributos.
- Sea $B = \{b_1, b_2, \dots, b_h\}$ conjunto finito de axiomas.
- Sea \hat{a} el atributo que contiene el valor de georeferenciación.
- Sea $C = \{c_1, c_2, \dots, c_p\}$ conjunto finito de conceptos donde cada $c_i \subseteq A \forall c_i \in C$.
- Sea $X = \{x_1, x_2, \dots, x_n\}$ conjunto finito de conceptos que definen un objeto geoespacial donde se cumple que $x_j \subseteq A \cup \{\hat{a}\} \forall x_j \in X$.
- Sea $U = X \cup C$ conjunto finito de todos los conceptos.
- Sea $R = \{r_1, r_2, \dots, r_t\}$ conjunto finito de relaciones donde $r_s \subseteq U \times U \forall r_s \in R$.
- Sea $r' \subseteq X \times C$ relación existente entre un objeto y su clase (instanciación).

- Sea $\mathbb{O} = \{O_1, O_2, \dots, O_q\}$ conjunto finito de ontologías donde se cumple que $O_l = (\mathcal{U}_l, \mathcal{R}_l, \mathcal{B}_l)$ tal que $\mathcal{U}_l \subseteq U, \mathcal{R}_l \subseteq R, \mathcal{B}_l \subseteq B$.

Entonces el contexto t_j para un objeto geoespacial $x_j \in X$ se define como:

- Sea $t_j = (N, E)$ contexto para el j -ésimo objeto donde N es el conjunto finito de conceptos del dominio que tienen una relación de instanciación (r') con x_j y E es el conjunto finito de ontologías vinculadas con x_j , véase expresiones 1 y 2.
 - $N = \{c \in C : x_j r' c\}$. (1)
 - $E = \{O_t = (\mathcal{U}_t, \mathcal{R}_t, \mathcal{B}_t) \in \mathbb{O} : \exists c \in N \wedge c \in \mathcal{U}_t\}$. (2)

Por lo tanto el Espacio Semántico puede definirse como:

- Sea $T = \{t_1, t_2, \dots, t_n\}$ conjunto finito de contextos que se define como Espacio Semántico, donde t_j es el contexto del objeto x_j . Note que los objetos x_i, x_j con $i \neq j$ pueden tener el mismo contexto.

Sobre la base de la definición del Espacio Semántico se puede plantear que la ORD puede representarse en la forma de un grafo $G = (V, Z)$ donde por un lado, $V = \{v_1, v_2, \dots, v_n\}$ es el conjunto finito de NRD tal que para cada objeto x_i existe un NRD v_i y por el otro, Z es el conjunto finito de aristas tal que $z_{ij} \in Z$ existe entre los NRD v_i y v_j si y solo si existe entre los objetos x_i y x_j una relación $r \in R$. De manera similar se puede plantear que las ONS pueden definirse a través de la expresión E , véase expresión 2.

La principal ventaja de este enfoque es que permite representar diferentes puntos de vista semánticos de los objetos, fenómeno que Fonseca en [9] define como *Roles*. Los Roles son enfoques a través de los cuales los objetos pueden ser interpretados y ofrecen diferentes puntos de vistas acerca del mismo fenómeno. El uso de Roles ofrece una mayor flexibilidad en la interpretación de los objetos y amplía su horizonte conceptual. Estos tratados como

“*un todo*” definen la semántica en la que se va a desenvolver el objeto permitiendo que nuevos y mejores mecanismos sean desarrollados y utilizados para su procesamiento desde el punto de vista semántico. Todo esto se traduce en la posibilidad de contar con mayor información acerca de los objetos analizados, lo que implica la obtención de mejores resultados en tareas como la recuperación de información y análisis para la toma de decisiones.

2.3. Definición Contextual de los Objetos Geográficos

La ORD constituye la abstracción semántica de los objetos geoespaciales pero solo en el ámbito de estos objetos. Esto significa que se representa conocimiento acerca de qué son en sí pero, esto no incluye el conocimiento sobre su contexto. Esto trae como consecuencia que su semántica sea bastante pobre ya que, esta representa un conocimiento que no va más allá del entorno de los objetos y las relaciones existentes entre ellos. Por tanto, es necesaria la vinculación con su contexto (la(s) ONS).

Para contextualizar los objetos analizados es necesario vincular de forma explícita cada NRD, en la ORD, con los conceptos/términos en la(s) ONS. Una forma de llevarlo a cabo es a través de técnicas de clasificación supervisada sobre la base de criterios de semejanza semántica. Para lograr este objetivo de manera automática se proponen los siguientes pasos:

1. Para la(s) ONS (*Construyendo el conjunto de Entrenamiento*):

- 1.1. Extraer todas las propiedades presentes para todos los conceptos pertenecientes a la(s) ONS que son tomados en cuenta. Estos conceptos van a representar tanto las muestras como las clases que se van a utilizar en la clasificación.

- 1.2. Para cada uno de estos conceptos construir vectores de ocurrencias y ausencias a partir de las ocurrencias y ausencias de sus propiedades con respecto a las propiedades extraídas en el paso (1.1).
- 1.3. Etiquetar cada uno de estos vectores con respecto al concepto en la(s) *ONS* que estos representan.
2. Para los *NRD* (*Construyendo el conjunto de Prueba*):
 - 2.1. Para cada uno de los *NRD* construir vectores de ocurrencias y ausencias a partir de las ocurrencias y ausencias de sus propiedades con respecto a las propiedades extraídas en el paso (1.1).
3. Clasificar los *NRD* utilizando la clasificación a partir de distancias que se presenta en la próxima sección.

Dado que estos vectores de ocurrencias y ausencias representan la existencia o no de características comunes es muy conveniente representarlos con valores binarios donde se codifica con ceros y unos si existe o no la propiedad en el objeto.

2.3.1. Clasificación de Nodos de Representación de Datos Mediante la Regla del Vecino más Cercano

Existen varias técnicas para la clasificación de objetos, entre ellas se encuentran las técnicas de clasificación a partir de distancias. Una de estas es la regla vecino más cercano (1-NN) [71]. Esta ha constituido una de las formas más populares de predecir o clasificar nuevos datos en los últimos años, precisamente, por la sencillez de su funcionamiento y por los buenos resultados que se han obtenido con su utilización, además, de que no es necesario tener de antemano ningún conocimiento acerca de las distribuciones de los objetos a clasificar.

El 1-NN es un clasificador que se basa en distancias y puede incluso trabajar con disimilitudes que no cumplan propiedades de una métrica como la simetría o la desigualdad triangular. En esencia, cuando viene un nuevo objeto a clasificar, el método calcula las distancias con respecto a todas las clases, luego las ordena y le asigna al nuevo objeto la etiqueta de la clase que tuvo menor distancia al mismo. Esto significa en este caso que, mientras menor sea la distancia entre el objeto y la clase, mayor debe ser la correspondencia semántica entre ambos. En la clasificación de estos objetos, cada concepto en la(s) *ONS* constituye tanto la muestra como la Clase. Por lo tanto, solo se cuenta con un representante por clase, donde la *Muestra "i"* y la *Clase "i"* la constituye precisamente el *Concepto "i"* en la(s) *ONS*. Por este motivo, se utiliza el clasificador vecino más cercano 1-NN. Sin embargo, este clasificador siempre va a asignar una clase aunque la distancia al *NRD (abstracción del Objeto)* a clasificar sea muy grande, por lo que se hace conveniente establecer un umbral en el proceso de clasificación con el objetivo de no asignar clases incorrectas. De esta manera se asegura que no se asigne una clase al *NRD* si realmente no pertenece a esta. Para esto es utilizada la regla del vecino más cercano con rechazo, esta variante excluye de la clasificación a aquellos *NRD* para los que no se alcance el umbral predeterminado.

La clasificación de cada *NRD* con respecto a los conceptos en la(s) *ONS* se describe en el Pseudocódigo 1. Primeramente se tiene que:

- Sea cada $c_i \in C \forall i = 1 \dots p$ (definido anteriormente en el apartado 2.2) nombrado como m_i (*muestra*) para una mejor comprensión.
- Sea $M^* = \{m_1^*, m_2^*, \dots, m_p^*\}$ el conjunto de todas las muestras m_j^* que son entradas como parámetro en el proceso de clasificación, donde $M^* \subseteq C$.

- Sea cada $x_j \in X \forall j = 1 \dots n$ (definido anteriormente en el apartado 2.2) nombrado como nrd_j (*nodo de representación de datos*) para una mejor comprensión.
- Sea $\mathbb{N} = \{nrd_1, nrd_2, \dots, nrd_q\}$ conjunto finito de nrd que son entradas como parámetros en el proceso de clasificación.
- $\mathcal{D}(m_i^*, nrd_j)$: Función que determina la disimilitud entre la muestra m_j^* y el nrd_i .
- Sea $\mathbb{R} = \{\mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_k\}$ conjunto de pares de m_i y nrd_i con una relación de tipo r' (definida anteriormente en el apartado 2.2).
 - Donde $\mathbb{W} = \{(nrd_i, m_j)_r\} \forall r = 1 \dots k$.

El Pseudocódigo 1 muestra el algoritmo para la clasificación de cada nrd_j donde \mathcal{R} es el umbral de rechazo predeterminado:

Pseudocódigo 1. Pseudocódigo del algoritmo 1-NN con rechazo para la clasificación de los NRD con respecto a los conceptos en la(s) ONS.

Entradas: $M^*, \mathbb{N}, \mathcal{R}$

Salida: \mathbb{R}

```

1 mindist =  $\mathcal{D}(m_0^*, nrd_0)$ 
2  $\mathbb{W} = \{(m_0^*, nrd_0)\}$ 
3 for i = 1 to q do
4 |   for j = 1 to p do
5 |     |   if  $\mathcal{D}(m_i^*, nrd_j) < mindist$  and  $\mathcal{D}(m_0^*, nrd_0) < R$  then
6 |     |     |    $\mathbb{W} = \{(m_i^*, nrd_j)\}$ 
7 |     |     |   mindist =  $\mathcal{D}(m_i^*, nrd_j)$ 
8 |     |     |   end if
9 |     |   end for
10 |   Añadir  $\mathbb{W}$  a  $\mathbb{R}$ 
11 end for
```

Como este clasificador está relacionado con la noción de proximidad o similitud entre los individuos (*NRD* y *Muestras*), es necesario establecer una medida para calcular esas distancias o proximidades entre ellos. La medida que proponemos utilizar en este caso es la *Distancia de Semejanza Simple* [72, 73], la cual es explicada en el apartado 2.3.3 con más detalle.

2.3.2. Reducción del Espacio de Búsqueda para la Clasificación de los Nodos de Representación de Datos

El espacio de búsqueda del método anterior utilizando la regla del vecino más cercano exclusivamente abarca todos los conceptos existentes en la(s) *ONS*. Esto significa que para clasificar un nrd_j debe calcularse la distancia hacia cada una de las muestras existentes, véase Fig. 12-A.

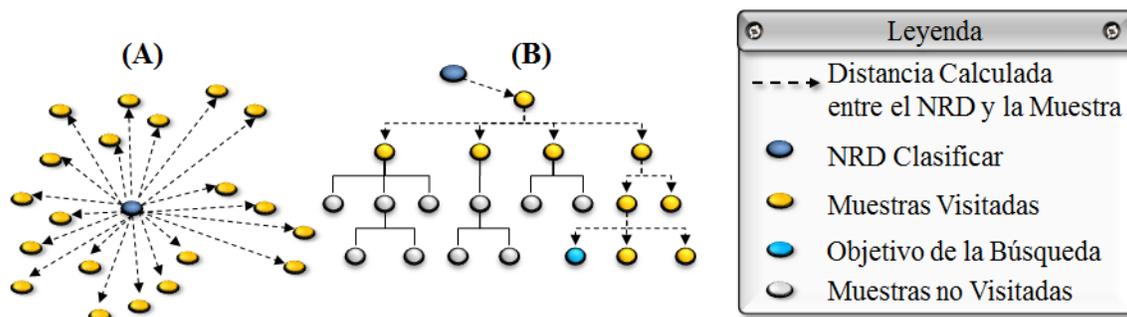


Fig. 12 Representación Gráfica que muestra en (A) el espacio de búsqueda para la clasificación de los NRD usando exclusivamente la regla del vecino más cercano y en (B) cuando esta búsqueda es basada en el paradigma de la búsqueda en árboles.

Dado que toda ontología lleva implícita una estructura taxonómica, es posible minimizar el espacio de búsqueda en el proceso de clasificación. Sobre la base de este hecho son utilizadas técnicas basadas en el paradigma de la búsqueda en árboles. Esta filosofía combinada con la de la regla del vecino más cercano posibilita disminuir considerablemente el espacio de búsqueda para el proceso de clasificación. Podemos definir entonces la taxonomía como un conjunto finito de conceptos tal que existe al menos uno, (c_1), que puede ser denominado como “Raíz” por constituir el término más general a partir del cual son definidos los demás conceptos. Los conceptos restantes (c_2, c_3, \dots, c_n) son descendientes de la Raíz, estos son llamados “*nodos-hijos*” para el caso de los que tienen una relación de tipo “*subclase_de*” con otros nodos incluyendo la Raíz. Estos nodos pueden ser divididos en (m_1, m_2, \dots, m_k) conjuntos disjuntos constituyendo a su vez “Sub-

taxonomías”. Además, aquellos conceptos/nodos que no tienen descendientes son definidos como “*Conceptos/Nodos Hojas o Externos*”.

A partir de esta abstracción el problema de la clasificación de un nrd_i puede ser abordada utilizando mecanismos de búsqueda en árboles donde la muestra más semejante a nrd_i puede ser encontrada sin tener que visitar todas las muestras existentes en la *ONS*, véase el esquema de la Fig. 12-B. Los pseudocódigos 2-A y 2-B muestran cómo se lleva a cabo lo planteado anteriormente.

Primeramente mostramos una modificación al *Pseudocódigo 1* para su utilización combinada con el método basado en la búsqueda en árboles, donde tenemos que:

- Sea $\mathbb{O} = \{m_1^*, m_2^*, \dots, m_q^*\}$ con $\mathbb{O} \subseteq \mathcal{C}$ conjunto de las conceptos raíces pertenecientes a los dominios del contexto del nrd_j .
- Sea $getSuccOf(m^*)$ función que devuelve en $\mathbb{M}^{*'}$ todos los conceptos los cuales son sucesores de m^* en la taxonomía.
- Sea $get1NN_R(nrd, \mathbb{M}^*, \mathcal{U}, \mathcal{R})$ algoritmo 1-NN modificado (Pseudocódigo 2-B).
 - Sea además $\mathbb{S} = (m^*, isCorrect, dist)$ salida del Pseudocódigo 2-B, donde:
 - m^* : Constituye la muestra cuya distancia al nrd es mínima.
 - $isCorrect$: Define si el nrd_j fue clasificado o no.
 - $dist$: Constituye la distancia calculada entre el nrd y la muestra m^* .
- Sea $getAntOf(m^*)$ función que obtiene el antecesor $m^{*'}$ de la muestra m^* en la taxonomía.
- Sea \mathcal{U} la distancia calculada del nrd a la muestra $m^{*'}$ cuyos descendientes están contenidos en \mathbb{M}^* . Además \mathcal{U} es tomado para determinar el criterio de parada.

El algoritmo descrito en el Pseudocódigo 2-B tiene como objetivo que en cada iteración se determine si el nrd_i clasifica o no con alguno de los conceptos m_j^* . Al término de las iteraciones el resultado S contiene aquella muestra m_j^* que mejor define al nrd_i y si la clasificación es válida o no.

Pseudocódigo 2. (A) Algoritmo para la clasificación de NRD basado en el paradigma de la búsqueda en árboles. (B) Modificación del algoritmo 1-NN con rechazo.

(A)	(B)
<pre> Entradas: $\mathcal{O}, N, \mathcal{R}$ Salida: \mathcal{R} 1 For $i=0$ to q do 2 $M^* = \text{getSuccOf}(\mathcal{O}_i)$ 3 For $j=0$ to p do 4 $U = \mathcal{D}(\mathcal{O}_i, N_j)$ 5 Repeat 6 $S(m^*, \text{isCorrect}, \text{dist}) = \text{get1NN}_R(N_j, M^*, U, \mathcal{R})$ 7 $U = S \rightarrow \text{dist}$ 8 If $S \rightarrow \text{isCorrect}$ then 9 $W = \{(N_j, S \rightarrow m^*)\}$ 10 Añadir W a \mathcal{R} 11 Else 12 $M^* = \text{getSuccOf}(S \rightarrow m^*)$ 13 End if 14 Until $S \rightarrow \text{isCorrect} \neq \text{true}$ 15 End for 16 End for </pre>	<pre> Entradas: nrd, M^*, U, \mathcal{R} Salida: S 1 $\text{mindist} = \mathcal{D}(m_0^*, nrd_0)$ 2 $S = \{m_0^*, \text{false}, \text{mindist}\}$ 3 for $i=0$ to p do 4 if $\mathcal{D}(m_i^*, nrd) < \text{mindist}$ then 5 $\text{mindist} = \mathcal{D}(m_i^*, nrd)$ 6 if $\text{mindist} < \mathcal{R}$ then 7 $S = \{m_i^*, \text{true}, \text{mindist}\}$ 8 else 9 $S = \{m_i^*, \text{false}, \text{mindist}\}$ 10 end if 11 end if 12 end for 13 if $S \rightarrow \text{mindist} > U$ then 14 $S \rightarrow m^* = \text{getAntOf}(S \rightarrow m^*)$ 15 end if 16 return S </pre>

De igual manera, como en el caso anterior (apartado 2.3.1), esta distancia d está relacionada con la noción de proximidad o similitud entre los individuos (NRD y $Muestras$). En este caso proponemos el uso de la distancia de Jaccard [69] la cual es explicada en el apartado 2.3.3 con más detalle.

La idea subyacente detrás de este método consiste en tomar el camino más corto que nos lleve a aquella muestra m^* (*concepto*) en la ONS que mejor defina al nrd que se está procesando. Con este objetivo, cada vez que una muestra en la taxonomía ontológica es expandida se determina cuál de sus descendientes (*nodos-hijos*) tiene menor distancia al nrd . Este método sigue la heurística de que la distancia entre el nrd y la muestra siempre disminuye a medida de que se va aproximando a aquella muestra que mejor lo define, o sea,

aquella muestra con la que el *nrd* clasifica y aumenta a medida de que se va alejando de ella. Basado en este criterio se determina la condición de parada para el caso de que la muestra con la que el *nrd* clasifica sea una sub-raíz en la taxonomía. Esta variante reduce el costo computacional de la clasificación ya que el algoritmo es capaz de determinar la profundidad de la búsqueda.

Este método se basa en dos aspectos fundamentales:

- La regla del vecino más cercano (1-NN) para la determinación de aquella muestra cuya distancia resulta ser mínima con respecto al *nrd*, la cual puede ser expandida si el *nrd* no clasifica con esta.
- El paradigma de la búsqueda en árboles el cual se apoya en la regla del vecino más cercano para determinar el camino hacia a aquella muestra que mejor define al *nrd* semánticamente, reduciendo así, el espacio de búsqueda y por lo tanto el costo computacional.

2.3.3. Medidas de Disimilitud para la clasificación de Nodos de

Representación de Datos

Uno de los aspectos claves para determinar la clasificación de los NRD lo constituye la elección de la medida que se desea utilizar para determinar la semejanza de estos con respecto a las muestras. De manera general es posible elegir entre un gran número de medidas para hallar semejanzas. Estas se diferencian básicamente por el tipo de datos para el que han sido diseñadas: *cuantitativos*, *categoricos*, *dicotómicos*.

Estas medidas también se diferencian por el tipo de distancia que evalúan:

- *Medidas de Similitud*: Las medidas de similitud evalúan el grado de parecido o proximidad entre elementos, donde los valores más bajos indican que los elementos son más diferentes.

- *Medidas de Disimilitud:* Las medidas de disimilitud hacen referencia a la lejanía entre objetos, donde los valores más bajos indican que los elementos son más semejantes.

En nuestro problema clasificar un nrd_i equivale a hallar la muestra (clase) cuya distancia hacia él sea mínima, por lo que se propone utilizar una medida basada en disimilitud.

Por lo general no existe una medida de disimilitud para todos los tipos de datos, esta debe ser escogida o adaptada en dependencia del tipo de dato y el problema específico que se ha de tratar. En el proceso de clasificación los NRD y las muestras son representados a través de vectores binarios que contienen la ocurrencia o no de las características medidas. En la literatura existen diferentes medidas para el cálculo de distancias entre dos objetos sobre la base de una tabla de contingencia, véase Tabla 2. En esencia estas medidas se diferencian en el valor que se le concede a cada una de las variables de la tabla (a, b, c, d) al calcular la semejanza.

Tabla 2. Tabla de contingencia para el cálculo de distancias entre los nrd y las muestras.

		<i>Muestra (m)</i>		
		1	0	
<i>NRD (nrd)</i>	1	<i>a</i>	<i>b</i>	<i>a + b</i>
	0	<i>c</i>	<i>d</i>	<i>c + d</i>
		<i>a + c</i>	<i>b + d</i>	<i>n</i>

Donde tenemos que:

- *a:* Es el número de características donde nrd y m tienen una ocurrencia.
- *b* y *c:* Número de características donde nrd y m tienen valores diferentes.
- *d:* Número de características donde nrd y m tienen una ausencia.
- *n:* Constituye el número total de casos.

Algunas medidas le conceden más importancia a las ocurrencias y ausencias conjuntas en las mismas características. Estas medidas consideran que dos objetos son más semejantes entre sí cuando presentan características comunes y además, carecen de las mismas características. Otras le conceden más importancia solo a las ocurrencias comunes o a las ausencias comunes de las mismas características, ya que estas no tienen por qué tener la misma importancia al valorar la semejanza.

Sobre la base de lo planteado anteriormente se propone utilizar en el apartado 2.3.1 para el cálculo de la semejanza entre las muestras y los NRD mediante el uso de la regla del vecino más cercano (1-NN) la de la Distancia de Semejanza Simple [72, 73]. Esta medida es conveniente utilizarla en tipos de datos en los que la existencia de ocurrencias y ausencias comunes en las mismas características tienen una contribución significativa en la clasificación (*ambos valores son igualmente importantes*). En este caso podemos determinar para cualquier par (nrd,m) las propiedades comunes en las que tienen una ocurrencia o una ausencia por lo que la Distancia de Igualdad Simple es una opción tentadora para determinar la semejanza entre ellos. Esta distancia es mostrada en la expresión (3).

$$\mathcal{D}_{IS(nrd,m)} = 1 - S_{CIS(A,B)} \quad (3)$$

Donde $S_{CIS(m,nrd)}$ es el Coeficiente de Igualdad Simple determinado por la expresión (4).

$$S_{CIS(A,B)} = \frac{|A \cap B| + |\mathcal{G} - (A \cup B)|}{|A \cap B| + |B \cap (\mathcal{G} - A)| + |A \cap (\mathcal{G} - B)| + |\mathcal{G} - (A \cup B)|} \quad (4)$$

Dónde:

- A : Conjunto de características de la muestra.
- B : Conjunto de características del nrd.
- \mathcal{G} : Conjunto de todas las características (Universo).

Esta medida también puede plantearse como el cociente entre la suma de las ocurrencias y ausencias comunes y el número total de características, véase la expresión (5) y Tabla 2.

$$S_{CIS(m,nrd)} = \frac{a+d}{a+b+c+d} \quad (5)$$

Para el caso de la clasificación basada en el paradigma de la búsqueda en árboles solo se cuenta con las propiedades del NRD y la muestra que se está analizando, por lo que resulta imposible saber si estos carecen de propiedades comunes entre ellos. Por tanto, en este caso proponemos utilizar una medida que solo toma en cuenta las ocurrencias comunes en ambos elementos para valorar la semejanza entre ellos. Esta medida es la Distancia de Jaccard [69], véase expresión (6).

$$D_{J(m,nrd)} = 1 - S_{j(m,nrd)} \quad (6)$$

Donde $S_{j(m,nrd)}$ es el Coeficiente de Jaccard determinado por la expresión (7).

$$S_{j(A,B)} = \frac{|A \cap B|}{|A \cup B|} \quad (7)$$

Dónde:

A: Conjunto de características de la muestra.

B: Conjunto de características del *nrd*.

Esta medida también puede plantearse como el cociente entre el número de concordancias y el número total de características, véase la expresión (8) y Tabla 2.

$$S_{J(m,nrd)} = \frac{a}{a+b+c} \quad (8)$$

De manera general dado que cada una de las medidas existentes para el cálculo de la semejanza entre dos individuos utilizando valores binarios enfatiza en un aspecto concreto de la tabla de contingencia mostrada anteriormente (Tabla 2), la decisión sobre qué medida

es más conviene utilizar en cada caso va a depender de la naturaleza del problema a resolver.

2.4. Ventajas del uso de la Ontología de Representación de Datos

El uso de la ORD supone un nuevo enfoque para el procesamiento semántico de los datos geoespaciales por parte de los SIG. Este enfoque puede ser aplicado sin grandes contratiempos puesto que no supone grandes restricciones para su utilización a diferencia de enfoques semejantes como es el caso de las bases de datos semánticas, las cuales están siendo ampliamente promovidas por las compañías que las producen. En los apartados siguientes se ponen de manifiesto algunas de las ventajas más notables del uso de este nuevo tipo de ontología.

2.4.1. Rol de la Ontología de Representación de Datos en la Integración

Semántica de Datos Geo-Espaciales Heterogéneos

Como se ha planteado anteriormente con respecto a la integración de fuentes de datos heterogéneas desde el punto de vista semántico, son muchos los autores que consideran el uso de Ontologías como el mecanismo más adecuado para realizar esta tarea. Los principales enfoques son [46]: *Enfoque de Ontología Única*, *Enfoque de Múltiples Ontologías* y *Enfoque Híbrido de Ontologías*. Estos enfoques no incluyen la semántica subyacente en los datos mismos, por lo que gran cantidad de información no es tomada en cuenta. El uso de la ORD combinado con estos enfoques ayuda a enriquecer la semántica con la que los objetos son analizados, ya que permite la representación explícita de estos objetos en el espacio semántico, así como también las relaciones generadas entre ellos. Por esta razón se propone un nuevo enfoque en la integración semántica de los datos geoespaciales sobre la base del uso de la ORD, véase Fig. 13.

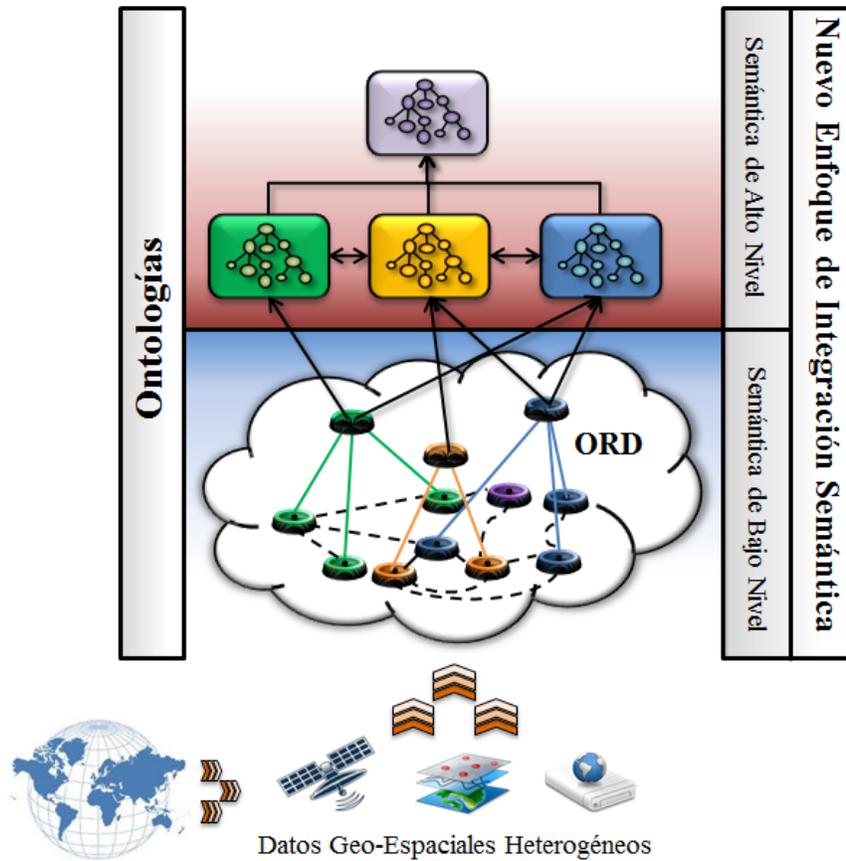


Fig. 13 Nuevo enfoque de integración semántica de datos geo-espaciales.

Entre las principales ventajas del uso de este nuevo enfoque podemos mencionar:

- Se habilita la integración tomando en cuenta la semántica de bajo nivel.
 - La incorporación del conocimiento subyacente en los geo-datos y las relaciones entre ellos en el proceso de integración.
- Permite el uso de diferentes enfoques de integración semántica a alto nivel incluso simultáneamente.
- Habilita la integración basada en la representación semántica multidimensional de los objetos.
- Habilita el trabajo con fuentes de datos dinámicas y Heterogéneas.
- Habilita la generación de nuevos niveles de abstracción.

2.4.2. Sistemas Gestores de Bases de Datos Geoespaciales y ORD

Aunque este trabajo se centra en campo de los SIG y dentro de este específicamente en los SIG gobernados por ontologías, en el ámbito geoespacial en general existe también una marcada tendencia al uso de tecnologías semánticas en los sistemas gestores de bases de datos geoespaciales (SG-BDG). Por la importancia que tienen estos dentro de este ámbito a continuación se presenta brevemente un análisis crítico sobre algunas limitantes en el procesamiento semántico de los datos geoespaciales detectadas en estos sistemas y cómo puede utilizarse la ORD para afrontar las mismas.

2.4.2.1. Sistemas Gestores de Bases de Datos Geoespaciales Semánticos

El modelo relacional es, indiscutiblemente, el modelo de base de datos más utilizado hoy en día, sin embargo, no logra captar adecuadamente la semántica de los datos que modela. Actualmente los SG-BDG se enfrentan al reto de proveer modelaciones de datos que soporten tipos de relaciones y restricciones mucho más complejas para dar respuesta a la demanda de los usuarios. Por estas razones Saurabh en [74] plantea que el modelo relacional no es suficiente para satisfacer las nuevas demandas de gestión de bases de datos. Para afrontar estos problemas, en la actualidad se están utilizando técnicas de procesamiento semántico de datos como es el caso de la generación automática de ontologías desde el propio SG-BDG.

Los SG-BDG más populares como **Oracle[®] Database Semantic Technologies**, **Microsoft[®] SQL Server** y **PostgreSQL[®]** entre otros, para proveer semántica, se basan en la modelación ontológica (en RDF²) conocida informalmente como “*Tabla-a-Clase, Columna-a-Predicado*”. Esta modelación está basada principalmente en los trabajos

² Marco de Descripción de Recursos (del inglés Resource Description Framework, RDF). <http://www.w3.org/RDF/>

propuestos en la literatura para la generación de conocimiento de forma automática desde fuentes de datos (en este caso no heterogéneas) generalmente utilizando bases de datos. Estos trabajos se centran en la generación de los esquemas semánticos a partir del esquema relacional de la BD.

En el caso del trabajo de Man Li y demás autores [75] un conjunto de reglas son definidas para la generación de ontologías a partir de una base de datos relacional. La principal limitación que presenta este trabajo es que la base de datos tiene que estar normalizada en al menos tercera forma normal (3FN). Otro aspecto a destacar con respecto a este método es que la ontología generada está basada solamente en el esquema, estructura y relaciones existentes entre las tablas de la base de datos sin tomar en cuenta la información contenida en los datos en sí para la generación de la ontología. El trabajo de Miriam Baglioni y demás autores [76] constituye en sí mismo una extensión del trabajo de Man Li y demás autores para el ámbito geográfico. En este se propone la inclusión de dos nuevas reglas para la construcción de una ontología de aplicación con el objetivo de representar la georeferenciación de los objetos geográficos almacenados en la base de datos relacional. Además de las limitaciones del trabajo de Man Li y demás autores, este método presenta que, la interrelación con la ontología de dominio está enfocada en la comparación sintáctica entre los términos con el apoyo de tesauros. Esto repercute negativamente en el hecho de que solo las interrelaciones entre los términos generados desde la base de datos y los de la ontología que se tomarán en cuenta serán aquellas que se correspondan desde el punto de vista sintáctico. En la Fig. 14 se muestra el esquema del método propuesto.

El método comienza construyendo una ontología de aplicación a partir de la base de datos, por medio del módulo de extracción, esta ontología está compuesta por un conjunto de conceptos y las relaciones existentes entre estos conceptos, los cuales representan el

esquema de la base de datos, luego esta ontología es enriquecida con una ontología de dominio para proveer mayores niveles de abstracción. Esta ontología enriquecida representa entonces la vista conceptual de los datos, los cuales son mapeados para posibilitar consultas espaciales semánticas.

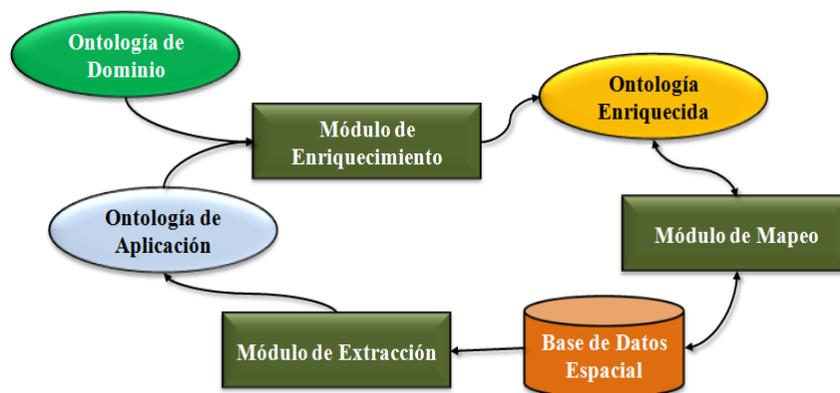


Fig. 14. Representación Gráfica del método propuesto por Baglioni y demás autores [76].

Como puede observarse los sistemas gestores de bases de datos geoespaciales semánticos utilizan en su gran mayoría ontologías de aplicación para su modelado semántico, las cuales como se ha comentado anteriormente, presentan muchas limitantes para representar la semántica que subyace en los datos almacenados, esto sumado a las propias limitantes del modelado semántico en las SG-BDE. Además, la integración de fuentes heterogéneas de datos utilizando estas modelaciones semánticas resulta bastante difícil. Por lo que el uso de la ORD para enfrentar estos problemas parece ser una opción tentadora de investigación en este campo.

2.5. Nueva Modificación a la Arquitectura SIGGO

En este apartado se presenta una propuesta para la estructura interna del módulo de generación automática de conocimiento a partir de datos propuesta por Garea-Llano [65] como una modificación a la arquitectura SIGGO propuesta por Fonseca [9], la cual fue presentada en el Capítulo 1 sección 1.5 y mostrado en la Fig. 6. Este módulo contiene los

mecanismos que permiten al sistema extraer la semántica subyacente en los datos. Esta posteriormente, puede ser refinada por expertos a través del módulo de Edición de Ontologías.

Los métodos propuestos en este capítulo están contenidos en este módulo, pero estos también utilizan las ontologías disponibles en el módulo de ontologías. Por esta razón resulta necesario realizar un nuevo cambio a la arquitectura de Fonseca incluyendo un nuevo elemento en la modificación propuesta por Garea-Llano. Esta nueva modificación va dirigida en el sentido de enlazar el módulo de generación automática de conocimiento a partir de datos con el módulo de las Ontologías, véase Fig. 15.

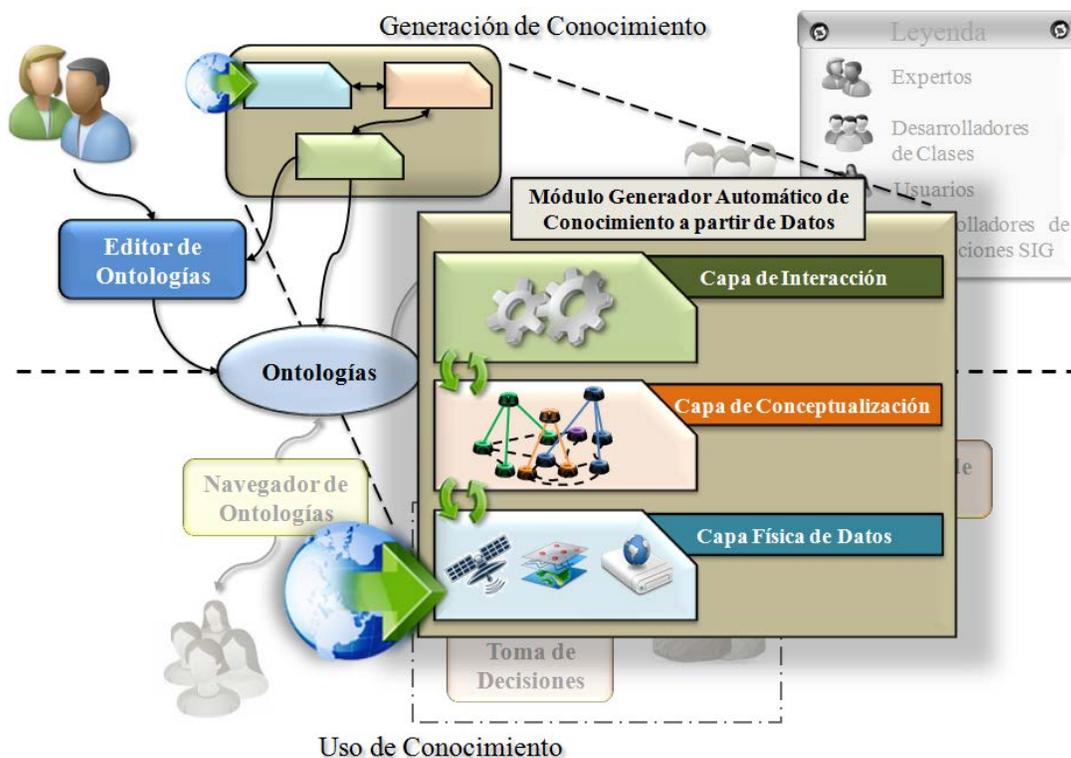


Fig. 15. Representación Gráfica de la modificación a la propuesta de Garea-Llano.

Además, se propone una estructura interna para este módulo, la cual se presenta a través de una vista arquitectónica definida en varias capas de abstracción interrelacionadas entre sí, véase Fig. 15. La primera capa, denominada “*Capa Física de Datos*”, es la encargada de

contener la información geográfica en su formato original. A partir de esta capa es que se toman los elementos necesarios para la generación de la Ontología de Representación de Datos en la llamada “*Capa de Conceptualización*”. Es entonces a través de la llamada “*Capa de Interacción*” que este módulo se relaciona con el módulo de edición y el de ontologías para el establecer el vínculo entre la ORD y la(s) ONS, así como también para su refinamiento en el módulo de edición de ontologías.

Esta estructura muestra la organización de los métodos propuestos en un entorno SIGGO. El principal objetivo de este módulo es establecer las bases para dotar a estos sistemas de las herramientas necesarias para la representación semántica de objetos geoespaciales y la integración de estos independientemente del formato de almacenamiento.

2.6. Conclusiones

De lo tratado en el capítulo puede concluirse que los nuevos conceptos y métodos propuestos de representación semántica de objetos geoespaciales en un entorno SIGGO buscan sentar las bases para que sistemas de este tipo puedan realizar un procesamiento semántico de la información geoespacial.

Con la propuesta de representación multidimensional semántica se hace posible representar con mayor fidelidad la naturaleza semántica de los diferentes objetos, la cual está generalmente dispersa en diferentes ontologías y este enfoque permite utilizarlas sin tener que interrelacionarlas o realizar algún proceso de alineamiento de ontologías. La multi-representación semántica permite la representación de una mayor cantidad de información acerca de estos objetos. Todos los elementos tratados en el capítulo contribuyen significativamente a la mejora de tareas realizadas por los SIG convencionales como son por ejemplo, la recuperación de información y el análisis para la toma de decisiones.

Capítulo 3

**RESULTADOS EXPERIMENTALES REALIZADOS
PARA LA VALIDACIÓN DE LOS MÉTODOS
PROPUESTOS**

CAPÍTULO 3

Resultados Experimentales Realizados para la Validación de los Métodos Propuestos

3. Resultados Experimentales Realizados para la Validación de los Métodos Propuestos

En este capítulo se presentan los principales resultados obtenidos en las experimentaciones realizadas para la validación de la hipótesis planteada a través de los métodos propuestos.

3.1. Diseño Experimental

A partir de la hipótesis trazada para esta investigación es posible establecer las variables dependientes a comprobar y las independientes que deben ser manipuladas para comprobar su efecto sobre las variables dependientes.

Variables Dependientes:

- **MS:** Mejorar (perfeccionar) las formas de análisis, explotación e integración desde el punto de vista semántico de la información geoespacial.
- **MAI:** Mejorar tareas existentes en los SIG convencionales (p.e: La recuperación y/o el análisis para la toma de decisiones).

Variables Independientes:

- **RS:** La utilización del nuevo tipo de ontología (ORD) para la representación semántica de objetos geoespaciales sobre la base de sus componentes fundamentales y las relaciones existentes entre ellos.

- **COS:** La combinación de la ORD con ontologías de niveles superiores.
- **MASIGGO:** Modificación de la arquitectura SIGGO.

Tomando en cuenta lo anterior es posible plantear la hipótesis de investigación funcionalmente como:

$$MS = f(RS, COS) \quad MAI = f(RS, COS, MASIGGO)$$

Para la comprobación se realizó el siguiente diseño de experimento, véase Tabla 3.

Tabla 3. Diseño del experimento. Variable Independiente (VI), Variable Dependiente (VD).

Nº Exp.	VI experimentada	VD a Comprobar Objetivo del Experimento.
1	RS	Comprobar la validez de los métodos propuestos para la utilización del nuevo tipo de ontología en la representación semántica automática de objetos geoespaciales.
2	RS	Comprobar la utilidad del uso del nuevo tipo de ontología propuesto para una representación semántica de objetos geoespaciales más cercana a la realidad tomando en cuenta que la semántica de estos se encuentra en un contexto que comprende más de un dominio de conocimiento.
3	COS, MASIGGO	Comprobar la utilidad del enfoque del nuevo tipo de ontología propuesto, combinado con las ontologías de niveles semánticos superiores para mejorar las formas de análisis, explotación e integración de los objetos geoespaciales, las cuales repercuten en el mejoramiento de las tareas existentes en los SIG.
4	COS, MASIGGO	Comprobar la validez de los métodos para: <ul style="list-style-type: none"> • La combinación automática de las ORD con la ONS y su validez para mejorar las tareas de recuperación de la información mediante su

aplicación en la detección y recuperación de objetos geoespaciales en imágenes de teledetección.

3.2. Experimento 1: *Representación Semántica Automática de Objetos Geoespaciales*

Objetivo del Experimento:

Este experimento tiene como objetivo comprobar la validez de los métodos propuestos para el uso del nuevo tipo de ontología propuesto en la representación automática de la abstracción semántica de objetos geoespaciales.

Materiales y Métodos:

Para llevar a cabo este objetivo se han tomado los datos procedentes de los mapas de Suelo y Geología, de los Institutos de Suelos y Geología de Cuba [77, 78]. Para determinar la precisión en el proceso de clasificación se ha utilizado una ontología de dominio de Coberturas Terrestres (*OD-CT*), donde, además de otros conceptos, se encuentran los de Suelos y Geología. Se han escogido los datos de Suelo y Geología para ilustrar este experimento, precisamente porque estos datos son semejantes desde el punto de vista de representación geoespacial pero, no desde el punto de vista semántico, véase Fig. 16. Cada una de las muestras de las clases a partir de las cuales son clasificados estos datos, como se ha planteado anteriormente, están representadas por cada uno de los conceptos presentes en la OD-CT, por ejemplo, el concepto *Suelos* en la OD-CT representa la clase *Suelos* y es a su vez la única muestra existente de esa clase. Estas muestras en el caso de las clases Suelos y Geología contienen las siguientes características:

- Clase Suelos: *Id, Nombre, Grupo, Tipo, Textura, Erosión, Acidez y Salinidad.*

- Clase Geología: *Id, Código, Nombre, Descripción, Edad, Textura y Dureza*.

Estas clases contienen algunas características comunes ya que es totalmente posible que esto ocurra.

Siguiendo los pasos descritos anteriormente en la sección 2.1.3 para el caso específico de los datos almacenados en el formato vectorial *ESRI ShapeFile* tenemos que:

1. *Definición de la estructura en la que está almacenado el dato:*

El Shapefile es un formato vectorial que almacena la ubicación de los objetos geográficos y los atributos asociados a ellos pero carece de la capacidad para el almacenamiento de la información topológica de estos. De manera general un Shapefile está compuesto por varios archivos, pero deben existir al menos los tres principales.

Estos tienen las siguientes extensiones:

- *.shp: Este es el archivo principal, en él se almacena las entidades geométricas de los objetos (*Componente Espacial*).
- *.dbf: Este es el archivo que almacena la información de los atributos de los objetos (*Componente Temática*).
- *.shx: Este es el archivo que almacena el índice de las entidades geométricas actuando como vínculo entre la componente espacial y la temática para cada uno de los datos almacenados.

2. *Extracción de valores:*

Para el acceso y extracción de los valores tanto temáticos como espaciales, en este caso, se ha utilizado la función “*shaperead*” existente en la herramienta MatLab R2009a [79].

3. *Generación de Nuevos Conceptos:*

Para la generación de nuevos conceptos nos basamos en la componente temática de los datos, la cual se encuentra almacenada en el fichero *.dbf. A partir de la información temática extraída sobre cada objeto se pueden utilizar distintos algoritmos de agrupamiento jerárquico que preferiblemente cumplan con las condiciones mencionadas en la sección 2.3.3 para el descubrimiento de los nuevos niveles de abstracción. Dado que estamos tratando solo con dos tipos de datos pertenecientes a las capas de datos de Geología y Suelos se hace innecesaria la generación de nuevos conceptos en este caso particular. Esto se debe a que estos datos desde un punto de vista semántico no son semejantes por lo que nuevos niveles de abstracción semántica no pueden ser generados.

4. *Extracción de Relaciones:*

Las relaciones existentes entre los objetos almacenados en estos datos, como el caso del ShapeFile (*SHP*) que carece de este tipo de información, pueden establecerse a partir de operaciones geoespaciales y análisis basados en la componente espacial. Esta componente en el formato *SHP* se encuentra almacenada en el archivo *.shp.

Una vez procesado los datos y generada la ORD se procede a la vinculación de esta con la(s) ONS. Este vínculo complementa y enriquece, al verse como un todo (ORD + ONS) la información semántica existente en ambos tipos de Ontologías. Por tanto, es necesario extraer los vectores de características comunes para cada dato y cada clase.

El vector extraído por clases contiene las ocurrencias y/o ausencias comunes de las características que se han tomado en cuenta. Con estas se ha construido la siguiente matriz de características por clase, véase Tabla 4:

Tabla 4. Vectores de ocurrencias y ausencias de características comunes por Clases.

Caract. Muest.	Id	Código	Nombre	Descripción	...	Edad	Textura	Salinidad
Geología	1	1	1	1	..	1	1	0
Suelos	1	0	1	0	..	0	1	1

Cada uno de los objetos analizados en el proceso de clasificación contiene un subconjunto del total de estas características. Cada fila de la Tabla 5 representa una capa diferente de datos sobre Suelos y Geología en la que los objetos han sido almacenados.

Tabla 5. Cantidad de objetos contenidos en las capas de Suelos y Geología.

Capa Geología		Capa Suelos	
Capa	Nº de Objetos	Capa	Nº de Objetos
geoLayer_1	928	sueLayer_1	307
geoLayer_2	1145	sueLayer_2	385
geoLayer_3	712	sueLayer_3	524
geoLayer_4	1098	sueLayer_4	233
geoLayer_5	977	sueLayer_5	189

Cada capa registra las propiedades existentes de los objetos que ella contiene, véase Tabla 6, por lo que se hace posible la clasificación de estos clasificando la capa a la cual ellos pertenecen. De esta manera una buena clasificación de la capa representa una buena clasificación de los objetos contenidos en ella.

Tabla 6. Vectores de ocurrencias y ausencias de características comunes por capa de datos.

Caract. Muest.	Id	Código	Nombre	Descripción	...	Edad	Textura	Salinidad
geoDat_1	1	0	1	0	...	1	1	0
geoDat_2	1	0	1	1	...	1	0	0
geoDat_3	0	1	1	0	...	1	1	0
geoDat_4	1	1	0	0	...	0	1	0
geoDat_5	0	1	1	1	...	1	1	0
sueDat_1	1	0	1	0	...	0	0	1
sueDat_2	0	0	1	0	...	0	1	1
sueDat_3	1	0	1	0	...	0	1	0
sueDat_4	0	0	1	0	...	0	1	0
sueDat_5	1	0	0	0	...	0	1	1

En la Fig. 16 se muestra la representación visual de estos dos tipos de objetos por lo que puede constatarse que aunque diferentes desde el punto de vista semántico visualmente son muy semejantes.



Fig. 16. Representación Gráfica de las capas de datos de Geología (A) y Suelos (B), los cuales tienen una representación geo-espacial similar pero desde el punto de vista semántico no lo son.

Resultados y Discusión:

La clasificación de las capas de datos fue realizada y todos los objetos pertenecientes a cada una de las capas temáticas fueron correctamente clasificados de forma automática utilizando el método de la *Regla del Vecino más Cercano* con rechazo presentado en la sección 2.3.1, la *Distancia de Igualdad Simple* del toolbox Distools y las estructuras "dataset" de la *PRTools* [80]. En la Fig. 17 se muestran los vectores de disimilitud de cada capa de datos con respecto a las clases Geología y Suelo, lo cual da la medida de que las capas pertenecientes a una misma clase se encuentran agrupadas.

Este experimento demuestra cómo es posible realizar la representación semántica de objetos geoespaciales automáticamente a través del uso de técnicas de clasificación y medidas de distancias.

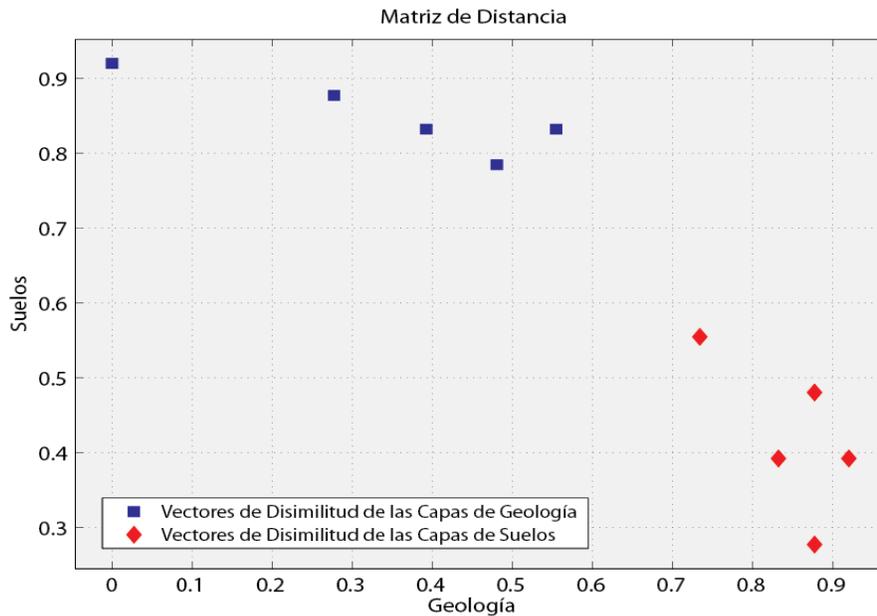


Fig. 17. Visualización de los vectores de disimilitud de cada dato con respecto a las clases Suelo y Geología.

Conclusiones del Experimento

En esta experimentación se muestra cómo es realizada la representación automática de la abstracción semántica de objetos geoespaciales. En este además, se presenta el uso de los pasos propuestos para la generación de la ORD descritos en la sección 2.1.3 para el caso específico de los datos almacenados en el formato vectorial *ESRI ShapeFile*. Los resultados ponen de manifiesto la validez de este tipo de representación utilizando técnicas de clasificación y medidas de semejanza semántica. De esta manera son contextualizados los objetos geoespaciales en los diferentes dominios dados por las Ontologías de Niveles Superiores a través del uso de la ORD.

3.3. Experimento 2: *Representación Automática de Objetos Geográficos con Contexto Multidimensional en el Espacio Semántico*

Objetivo del Experimento:

En este experimento el objetivo que se persigue es comprobar la utilidad del uso del nuevo tipo de ontología que se propone en este trabajo de investigación para representar objetos

geoespaciales cuya abstracción semántica comprende más de un dominio de conocimiento, siendo estos dominios disjuntos. Donde además, cada uno de estos dominios está definido por una ontología diferente.

Materiales y Métodos:

En este experimento se ha conformado con la ayuda de expertos una base de datos de objetos geográficos cuya semántica abarca diferentes dominios. Cada uno de estos dominios es representado por una Ontología de Nivel Superior diferente y en su conjunto definen un contexto multidimensional.

Los dominios en los que se representan los objetos son: *Reservas Naturales (RN)*, *Atracciones Turísticas (AT)* y *Lugares Históricos (LH)*. Estos constituyen las clases a partir de las cuales se determinan los valores de pertenencias de los distintos objetos. Los objetos geoespaciales tomados pueden pertenecer a uno o varios de estos dominios, estos son: *Playa Girón (PG)*, *el Castillo de los Tres Reyes del Morro (CM)*, *el Parque Baconao (PB)*, *Sierra del Rosario (SR)*, *Ciénaga de Zapata (CZ)*, *Sierra Maestra (SM)*, *Cayo Coco (CC)*, *Catedral de La Habana (CLH)* y *el Escambray (Es)*.

Cada uno de estos objetos contiene características/propiedades que son tomadas en cuenta para el análisis a través de los diferentes puntos de vista. En la Tabla 7 podemos observar los vectores de características obtenidos a partir de estos objetos y las clases utilizadas.

Tabla 7. Vectores de ocurrencias y ausencias conjuntas utilizados en el proceso de clasificación.

Dominios	Vectores binarios de ocurrencias y ausencias para las clases																
NR	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	
TA	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	
HP	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
Objetos	Vectores binarios de ocurrencias y ausencias para los objetos																“Ground Truth”³
PG	1	1	1	0	1	1	1	0	1	0	1	1	0	1	1	1	NR, TA, HP
CM	1	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	-, TA, HP

³ http://en.wikipedia.org/wiki/Ground_truth

PB	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	NR, TA, HP
SR	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	NR, -, -
CZ	1	1	1	1	0	1	0	1	0	0	1	0	1	1	0	0	0	0	NR, TA, -
SM	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1	1	0	0	NR, -, HP
CC	1	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	-, TA, -
CLH	1	1	0	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	-, TA, HP
SE	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	NR, -, HP
CHC	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	-, -, HP

Resultados y Discusión

La clasificación de la variante propuesta fue realizada en MatLab [81] utilizando la Distancia de Jaccard implementada en el toolbox Distools y las estructuras "dataset" de la PRTools [80]. En la Fig. 18 se muestran los objetos geoespaciales representados en un contexto n-dimensional ($n = 3$) en el espacio semántico. Los ejes indican el dominio semántico y los objetos se representan sobre la base del valor de pertenencia a cada uno de estos dominios.

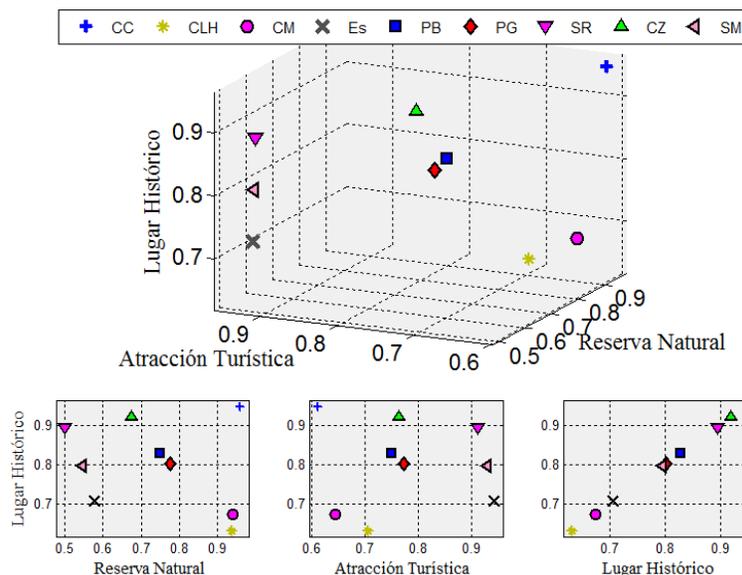


Fig. 18. Representación gráfica de los valores de pertenencia de los diferentes objetos con respecto a las clases.

En la Fig. 18 puede observarse cómo se separan los diferentes objetos en el espacio semántico tomando en cuenta los valores de disimilitud calculados a partir de los vectores de características entre los objetos y las clases. Para cada uno de los objetos es calculada la

disimilitud a cada una de las clases siendo la(s) menor(es) aquella(s) clase(s) que en un mayor grado represente(n) la naturaleza del objeto analizado.

La mayor importancia de este tipo de representación radica en la posibilidad de representar una mayor cantidad de información sobre el objeto conceptualizado. Esta información no tiene necesariamente que estar contenida en una sola Ontología de Nivel Superior sino que también puede estar contenida en varias de estas ontologías. De esta manera es posible representar con mayor flexibilidad toda la naturaleza semántica de los objetos geoespaciales. La representación semántica multidimensional de objetos geoespaciales mejora muchas tareas existentes hoy en día tanto en los SIG convencionales como en los SIG de nueva generación (*SIGGO*). Incluso brinda un importante aporte en actividades como la toma de decisiones a partir de la mayor cantidad de información que proveen los contextos multidimensionales en los que se pueden representar estos objetos.

Conclusiones del Experimento

El experimento realizado en esta sección tuvo como objetivo la representación, también de forma automática, de objetos geoespaciales con contexto multidimensional en el espacio semántico. De esta manera se hace posible la representación de objetos desde diferentes puntos de vistas y así ofrecer una mayor riqueza semántica. Esto pone a disposición una mayor cantidad información sobre los objetos la cual es de gran utilidad en tareas de análisis y tomas de decisiones. Los resultados mostraron la validez de realizar este tipo de representación de forma automática tomando en cuenta que los diferentes dominios pueden estar definidos por diferentes ontologías y de su implementación tanto en aplicaciones SIG convencionales como en aplicaciones con arquitectura *SIGGO*.

3.4. Experimento 3: Recuperación y análisis de la información para la toma de decisiones.

Objetivo del Experimento:

Comprobar la utilidad del enfoque del nuevo tipo de ontología propuesto, combinado con las ontologías de niveles semánticos superiores para mejorar las formas de análisis, explotación e integración de los objetos geoespaciales y el mejoramiento de las tareas existentes en los SIG (p.e: La recuperación y/o el análisis para la toma de decisiones).

Materiales y Métodos Utilizados:

Para la ejecución del experimento se seleccionaron dos casos reales de aplicación de las herramientas de recuperación y análisis de la información en SIG. Estos casos de estudio corresponden a aplicaciones desarrolladas para la toma de decisiones en dos áreas en las que el manejo de la información geoespacial es el elemento fundamental para llegar a una posible solución o variantes de solución. Ellos son:

- Estimación de riesgos ante fenómenos atmosféricos para la ubicación de instalaciones eólicas generadoras de electricidad en zonas con potencial eólico de la Isla de la Juventud.
- La zonificación de suelos con vistas a la aplicación de fertilizantes naturales.

Procedimiento y procesos realizados:

Se desarrolló una plataforma de experimentación, la cual se nombró como ViSem y cuyo objetivo fundamental es la implementación y prueba de los métodos automáticos desarrollados. Esta plataforma presenta las funcionalidades básicas existentes en los SIG convencionales accesibles a través de la barra de herramientas, véase Fig. 19, estas son:

- Carga y visualización de capas de datos.

- Interacción con el mapa (*Paneo, Zoom, Restablecer*).
- Consulta a los datos.

La plataforma está compuesta por un panel de capas. En esta se visualizan las capas de datos cargadas así como la descripción temática de estas. Con respecto a estas capas la plataforma también posee un visor de mapas en el cual se visualizan los datos cargados a través de su componente espacial, véase Fig. 19-A. Esta plataforma además, contiene un visor de ontologías que permite la carga, la visualización y la salva tanto de las ontologías contenidas en el sistema como de la ORD generada, véase Fig. 19-B.

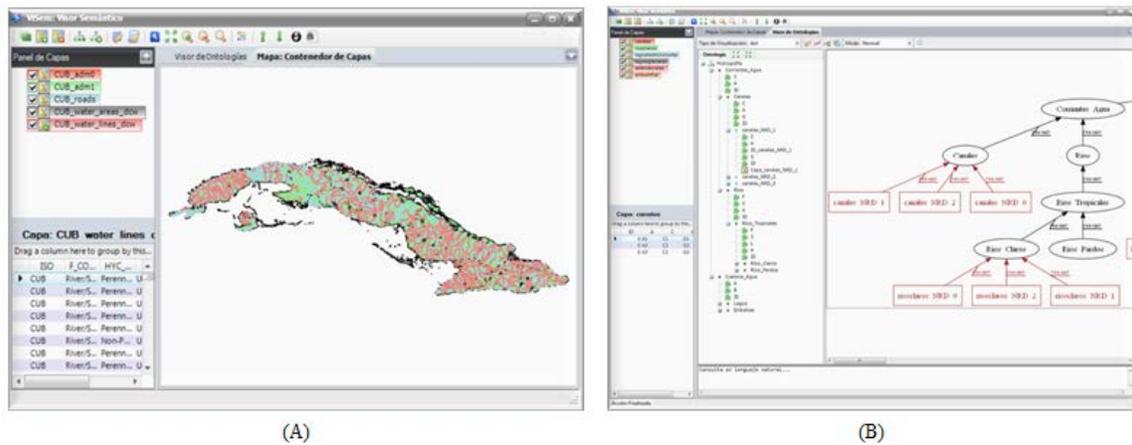


Fig. 19. Representación gráfica de la plataforma de experimentación. (A) Visor de Mapas. (B) Visor de ontologías.

En la plataforma se implementaron además las siguientes funcionalidades:

- Conceptualización de fuentes datos heterogéneas (Generación automática de la ORD a partir de datos geográficos provenientes de fuentes heterogéneas).
- Vinculación de forma automática de la ORD generada con las ontologías existentes en el módulo de ontologías.
- Persistencia de los objetos geospaciales procesados y las ontologías generadas. Esta funcionalidad tiene como objetivo hacer persistente en fichero los objetos

procesados. Esta persistencia en fichero se realiza en formato OWL (*.owl), por lo que es posible que sea utilizada por otros sistemas.

- Consulta y visualización de los objetos según el nivel de abstracción semántica especificado. Lo importante de las consultas basadas en la abstracción semántica de los datos es que habilita el uso de términos no existentes en las fuentes de datos y brinda una mayor flexibilidad en las consultas realizadas.

Esta plataforma ha sido diseñada e implementada utilizando el entorno de desarrollo *Microsoft .NET* [82] a través de su lenguaje insignia *C#* [83]. Para la gestión de ontologías se ha utilizado la biblioteca *Jena* [84]. Esta es una plataforma de código abierto desarrollada por los laboratorios de *Hewlett Packard* para la programación de la web semántica que permite la gestión de ontologías. Dado que esta plataforma se encuentra desarrollada en lenguaje *Java* [85] se ha utilizado el *IKVM.NET* [86]. Este es una implementación de la plataforma de *Java* para *Mono* [87] y *Microsoft .NET Framework* [82]. Para la visualización de las ontologías se ha utilizado la biblioteca de código abierto *Graphviz* [88], la cual provee herramientas para la visualización de objetos con vínculos entre ellos (gráfos).

Estimación de Riesgos

Las instalaciones generadoras de energía (*campos eólicos, paneles solares, etc.*), por sus características están principalmente localizadas en zonas de difícil acceso. Estas zonas están caracterizadas por diferentes niveles de complejidad en términos de la influencia de fenómenos naturales que pueden afectar su integridad, elementos que tienen que ser tomados en cuenta cuando se realiza la estimación de su factibilidad y los costos de instalación. El caso de estudio que se presenta se basa en la estimación automática de las regiones y los niveles de riesgo que los fenómenos atmosféricos representan para posibles

instalaciones eólicas a implantar en el territorio de la Isla de la Juventud tomando en cuenta las estimaciones de su potencial eólico [89] y el criterio especializado de los expertos, véase Tabla 8.

Tabla 8. Criterio especializado de los expertos.

<i>Criterios</i> <i>Tipo de riesgo</i>	Velocidad del viento (Km/h)	Altura (m)	Precipitaciones (mm/h)	Temperatura (C ⁰)
Muy Alto	>100	>0	> 600	>30
Alto	>100	200-400	400-600	> 30
Medio	30-99	>0	> 200	> 25
Bajo	< 30	>0	0- 600	0- 25

Para la evaluación de este caso de estudio se ha realizado una serie de acciones que se describen a continuación.

- Se determinó el objetivo del análisis que consistía en determinar y clasificar las zonas por su grado de riesgo ante fenómenos atmosféricos a partir de la localización previa de regiones con potencial eólico, para ello se tomaron los criterios de expertos plasmados en la Tabla 8.
- Se integraron a la plataforma los datos correspondientes a las temáticas de velocidades medias del viento y su potencial generador (*base de datos georreferenciada, en formato ACCESS*); relieve (*modelo digital del terreno en formato raster, escala 1:100 000, en formato ASCII*); precipitaciones (*mapa vectorial de isoyetas, escala 1:10 000, en shapefile*), temperatura (*mapa vectorial poligonal de las regiones de temperatura, escala 1:25 000, en shapefile*) correspondientes a la Isla de la Juventud [90, 91].
- A partir de los criterios de dos expertos del Centro de Gestión de la Información y Desarrollo de la Energía (CUBAENERGIA) del CITMA se desarrolló mediante la metodología METHONTOLOGY [92] una ontología de aplicación de criterios de expertos (*OA-CE*), un esquema simplificado de esta ontología se muestra en la Fig. 20.

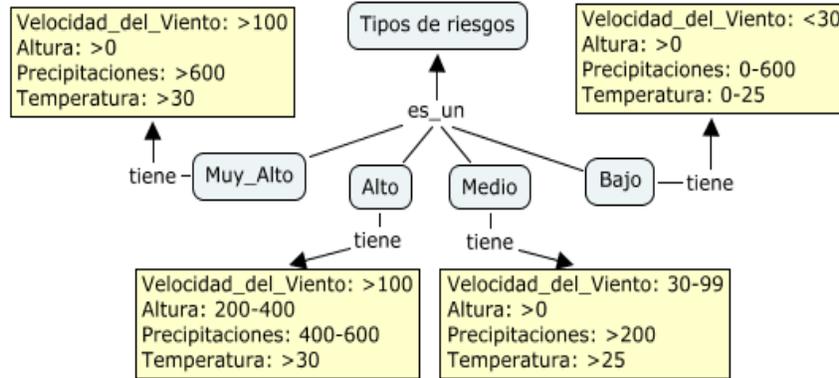


Fig. 20. Esquema simplificado de la Ontología de Aplicación utilizada como apoyo en el proceso de estimación de riesgos.

Se desarrolló además, por la misma metodología una ontología de dominio de estimación de riesgos (*OD-ER*). Esta comprende la representación de las relaciones de cada una de estas temáticas desde el punto de vista semántico y la información necesaria para la determinación del conocimiento requerido sobre los diferentes conceptos tratados en este caso de estudio. La Fig. 21 muestra un esquema simplificado de esta ontología.

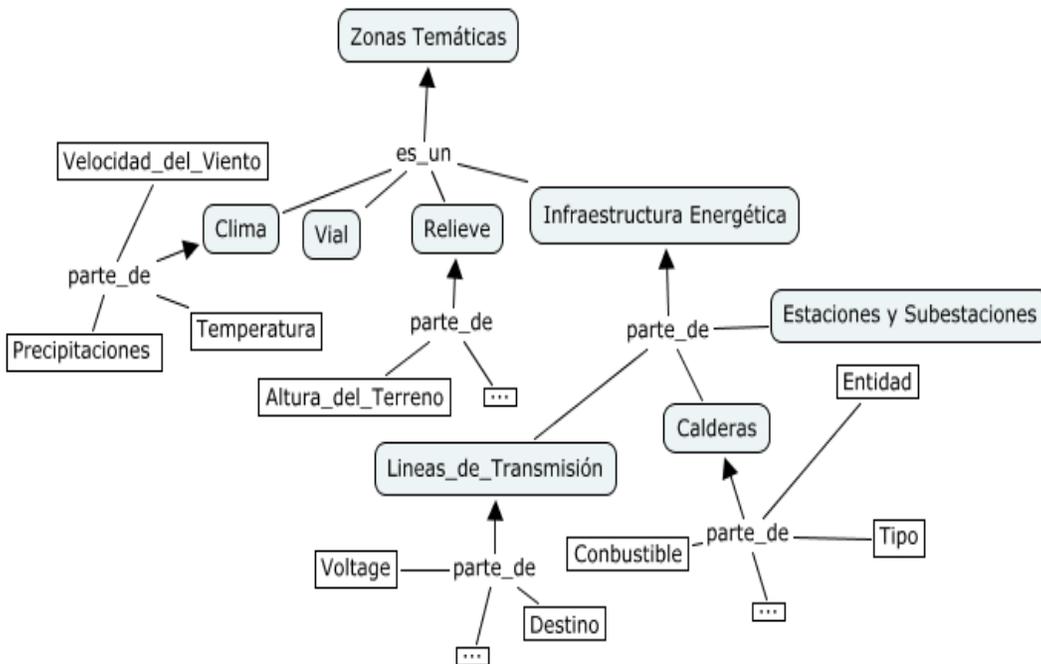


Fig. 21. Esquema simplificado de la ontología de dominio para la estimación de riesgos de las instalaciones eólicas.

Para determinar los niveles de riesgo se utilizan la ORD y las ONS, que en este caso, son las ontologías de aplicación OA-CE y la de dominio OD-ER. La ORD es generada a partir de los datos contenidos en las capas temáticas anteriormente mencionadas y posteriormente es vinculada con la OD-ER. La Tabla 9 muestra un subconjunto del conjunto de las características tomadas en cuenta para la construcción de los vectores binarios para cada muestra m en la OD-ER y cada nodo nrd en la ORD (*Conjunto de Entrenamiento*).

Tabla 9. Características y Vectores de ocurrencias y ausencias por muestras utilizados en el proceso de clasificación (Conjunto de Entrenamiento).

Caract. Muest.	ID	TIPO	VOLTAGE	COMBUST	...	ALTURA	PRECIP
Viales	1	1	0	0	..	0	0
Relieve	1	0	0	0	..	1	0
...
Clima	1	0	0	0	..	0	1
Inf_Energ	1	0	1	1	..	0	0

Los diferentes datos integrados son los utilizados en el proceso de clasificación y contienen un subconjunto de estas características y se encuentran contenidos en diferentes capas. Por lo tanto, clasificar correctamente cada una de estas capas implica una buena clasificación de cada uno de los objetos contenidos en ella reduciendo así el costo computacional del proceso. Para estas capas se construyeron los vectores de ocurrencias y ausencias de las características que se tomaron en cuenta para el proceso de clasificación. La clasificación usando el clasificador 1-NN con rechazo basado en el paradigma de la búsqueda en árboles, la Distancia de Jaccard del toolbox Distools y las estructuras "dataset" de la PRTools [80] fue correcta para todas las capas analizadas. En las Fig. 22 (Vista todos contra todos) y Fig. 23 (vista 3D) se muestra la clasificación para cada una de las capas de datos con respecto a las distintas clases.

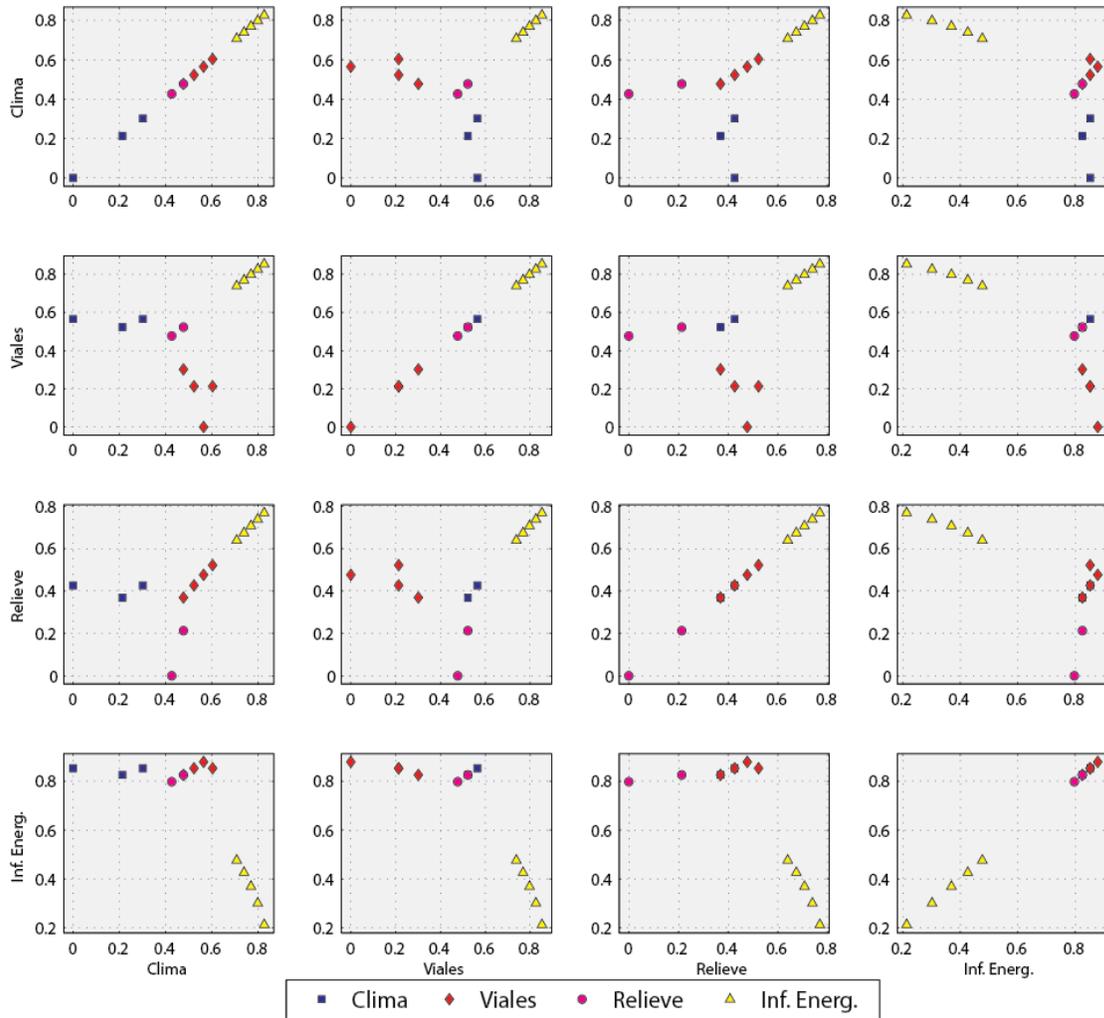


Fig. 22. Representación gráfica de la clasificación realizada utilizando 1-NN con rechazo.

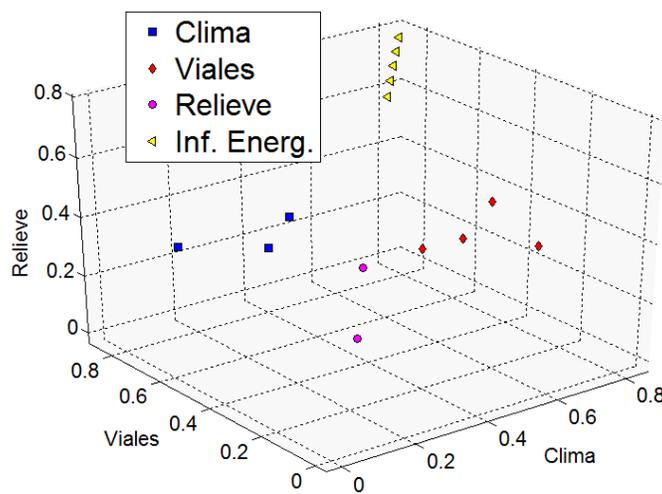


Fig. 23. Representación gráfica de la clasificación realizada utilizando 1-NN con rechazo. Presentación en una vista 3D.

De esta manera se muestra la validez de representar de manera automática la abstracción semántica de los datos geográficos utilizando ORD y ONS. Otro aspecto a tener en cuenta para la determinación de los niveles de riesgo en este caso son las relaciones topológicas existentes entre los objetos geográficos representados en estos datos, ya que, aquellas zonas de riesgo serán las zonas donde estén presentes a la vez y superpuestas cada una de las zonas determinadas por las variables que están siendo tomadas en cuenta, o sea, las zonas (*regiones*) de riesgos son aquellas regiones dadas por la intersección de todas y absolutamente todas las regiones de interés definidas por los especialistas.

Entre las relaciones topológicas que podemos tomar en cuenta se encuentran las definidas por Egenhofer [64, 93, 94] denominado como el *Modelo de las 9-intersecciones*, las cuales están definidas de la siguiente forma:

Sea:

- A, B : Regiones pertenecientes a los objetos A y B respectivamente.
- $R(A, B)$: Define las relaciones existentes entre las regiones A y B dada por las siguientes expresiones:
 - A^0 : Referente al interior de la región del objeto A.
 - ∂A : Referente al contorno de la región del objeto A.
 - A^- : Referente al exterior de la región del objeto A.
 - B^0 : Referente al interior de la región del objeto B.
 - ∂B : Referente al contorno de la región del objeto B.
 - B^- : Referente al exterior de la región del objeto B.

Entonces es posible modelar las relaciones espaciales entre A y B a través de la expresión (9) y/o la Fig. 24, donde cada una de intersecciones puede tomar un valor de vacío (\emptyset) o no vacío ($\sim\emptyset$).

$$R(A, B) = \left\{ \begin{array}{lll} A^0 \cap B^0 = (\emptyset, \sim\emptyset) & A^0 \cap \partial B = (\emptyset, \sim\emptyset) & A^0 \cap B^- = (\emptyset, \sim\emptyset) \\ \partial A \cap B^0 = (\emptyset, \sim\emptyset) & \partial A \cap \partial B = (\emptyset, \sim\emptyset) & \partial A \cap B^- = (\emptyset, \sim\emptyset) \\ A^- \cap B^0 = (\emptyset, \sim\emptyset) & A^- \cap \partial B = (\emptyset, \sim\emptyset) & A^- \cap B^- = (\emptyset, \sim\emptyset) \end{array} \right\} \quad (9)$$

Las relaciones tomadas en cuenta son las del tipo *contiene_a*, *contenido_en*, *iguales*, *cubre_a*, *cubierto_por* y *superposición*, véase Fig. 24.

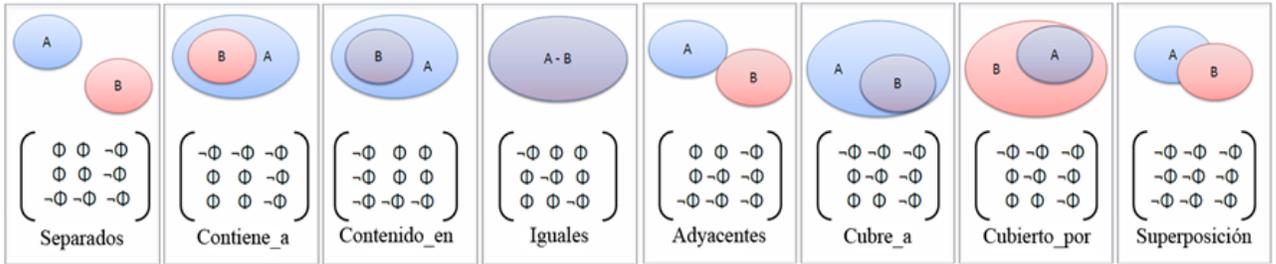


Fig. 24. Relaciones topológicas entre los objetos A y B.

Una vez generada la ORD y vinculada con la OD-ER apoyados también tanto en la OA-CE como en las relaciones topológicas existentes entre los objetos podemos determinar los niveles de riesgo a partir de las siguientes expresiones:

Sea:

- $\text{getDat}(R_i)$: Función que obtiene los nrd la relación R_i .
- $\text{getClas}(\text{nrd}_j)$: Función que obtiene la Clase que define a nrd_j .
- n: Número máximo de relaciones.
- m: Número máximo de nrd.
- $P = (\text{Veloc. Viento, Altura, Temp, Precip, LineasTrans, EstySubEst, Calderas})$
- $R = (\text{Contiene}_a, \text{Contenido}_{en}, \text{Iguales, Cubre}_a, \text{Cubierto}_{por}, \text{Superpos}) \quad (10)$

Entonces dada la expresión 11:

$$I_1 = \text{getDat}(R_1), I_2 = \text{getDat}(R_2), \dots, I_n = \text{getDat}(R_n) \quad (11)$$

Donde $I_i \forall i = 1..n$ constituye el conjunto formado por todos los *nrd* que contengan las relaciones $R_i \forall i = 1..n$, entonces las áreas de categorías de riesgo son aquellas zonas Z dada por la expresión 12:

$$Z = (I_1 \cap I_2 \cap \dots \cap I_n) \quad (12)$$

Siempre y cuando se cumpla con las siguientes restricciones:

- $R_i \in R \forall i = 1 \dots n$
- $\text{getClas}(\text{nrd}_j) \in P \forall j = 1 \dots m$
- $\bigcup_{j=1}^m \{\text{getClas}(\text{nrd}_j)\} = P$

Una vez determinadas estas zonas se procede a determinar las categorías de riesgo de cada una de ellas sobre la base de los criterios de los especialistas conceptualizados en la OA-CE.

- Se realizaron diferentes operaciones de búsqueda y análisis de la información espacial integrada en la plataforma de experimentación⁴, mediante la navegación por la ontología generada y se comparó la eficacia de los resultados con los obtenidos utilizando los métodos de superposición de mapas y recuperación basada en el lenguaje SQL, mediante el Error de Contorno calculado por la expresión 13, véase Tabla 10:

$$EC = \frac{\sum |(RA-DA)|}{RA.100} \quad (13)$$

Dónde:

- RA- es el área real de la región a detectar por la consulta (*Ground Truth*)
 - DA- es el área de la región detectada por la consulta
- Se compararon los tiempos de cada operación y sus costos.

⁴ Esta plataforma de experimentación es tratada con más detalle en la sección 3.4.

Zonificación de suelos

El suelo es el elemento fundamental de la superficie de la Tierra que sirve de sustento a la vida, su uso racional y sostenible en la agricultura es una cuestión de importancia estratégica por lo que su zonificación agroecológica constituye una herramienta de suma utilidad para la planificación adecuada de las labores agrícolas a realizar. La aplicación de fertilizantes requiere de este instrumento para una correcta aplicación de las dosis que realmente necesita el suelo sin detrimento de sus propiedades físicas y químicas a la par de lograr el necesario aumento de su fertilidad. El caso de estudio que se presenta se basa en la obtención del mapa de zonificación agroecológica de la región montañosa Nipe-Sagua-Baracoa ubicada en la región más oriental de Cuba [102].

Este proceso de zonificación se basa en la combinación de diferentes métodos de análisis, como son, la superposición de mapas y la reclasificación para la obtención de mapas de aplicación de los fertilizantes naturales a partir de la combinación de datos espaciales y alfanuméricos integrados al SIG, véase Fig. 25.

Para la evaluación de este caso de estudio se siguieron los siguientes pasos:

- Se determinó el objetivo del análisis que consistía en determinar la zonificación de los suelos con vistas a la aplicación de fertilizantes naturales en la región montañosa de Nipe Sagua Baracoa, para ello se tomaron los criterios de expertos plasmados en [102].
- Se integraron a la plataforma ViSem los datos correspondientes a las temáticas de suelos (*mapa vectorial poligonal de los suelos, escala 1:100 000, en shapefile e*); relieve (*modelo digital del terreno en formato raster, escala 1:100 000, en formato ASCII*); precipitaciones (*mapa vectorial de isoyetas, escala 1:10 000, en shapefile*), temperatura (*mapa vectorial poligonal de las regiones de temperatura, escala 1:25 000, en shapefile*) correspondientes al macizo montañoso Nipe-Sagua-Baracoa [102].

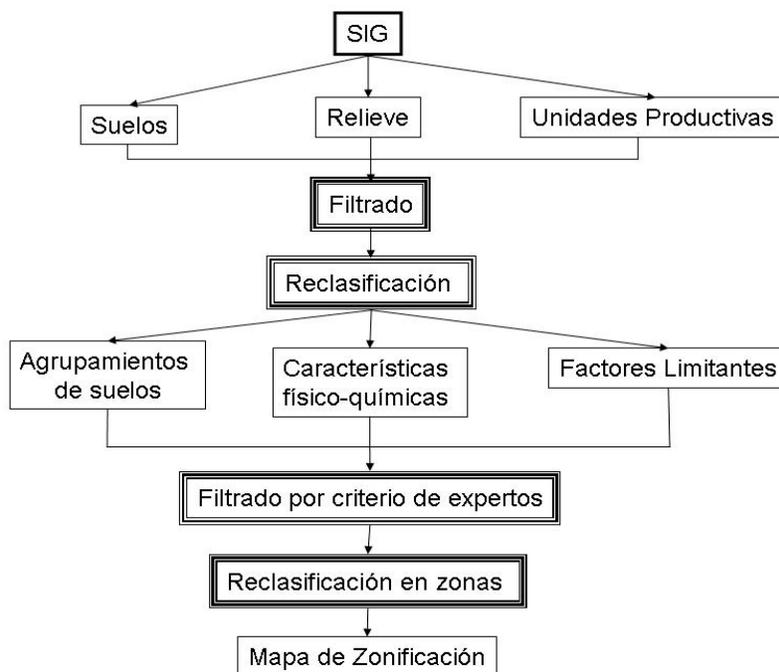


Fig. 25 Esquema metodológico empleado para la zonificación de los suelos con vistas a la aplicación de fertilizantes naturales.

- A partir de los criterios de un experto del Instituto de Suelos y otro del Centro de Desarrollo de la Montaña de Guantánamo se desarrolló mediante la metodología METHONTOLOGY [92] una ontología de aplicación de criterios de expertos (*OA-CES*), un esquema simplificado de esta ontología se muestra en la Fig. 26. Se desarrolló además por la misma metodología una ontología de dominio de zonas agroecológicas (*OD-ZA*). Esta comprende la representación de las relaciones de cada una de estas temáticas desde el punto de vista semántico y la información necesaria para la determinación del conocimiento requerido sobre los diferentes conceptos tratados en este caso de estudio.

Para determinar las zonas agroecológicas se utilizan la ORD y las ONS (*OA-CES* y la *OD-ZA*). La ORD es generada a partir de los datos contenidos en las capas temáticas anteriormente mencionadas y posteriormente es vinculada con la *OD-ZA*.

De la misma forma que en el caso de estudio anterior, otro aspecto a tener en cuenta para la determinación de las zonas agroecológicas son las relaciones topológicas existentes entre los objetos geográficos representados en estos datos, ya que, la aptitud para la aplicación de fertilizantes de cada zona agroecológica estará dada por la presencia y superposición a la vez de cada una de las regiones determinadas por las variables que están siendo tomadas en cuenta, o sea, las zonas agroecológicas son aquellas regiones dadas por la intersección de todas y absolutamente todas las regiones de interés definidas por los especialistas. Una vez determinadas estas zonas se procede a determinar las categorías de aptitud de cada una de ellas sobre la base de los criterios de los especialistas conceptualizados en la OA-CES.

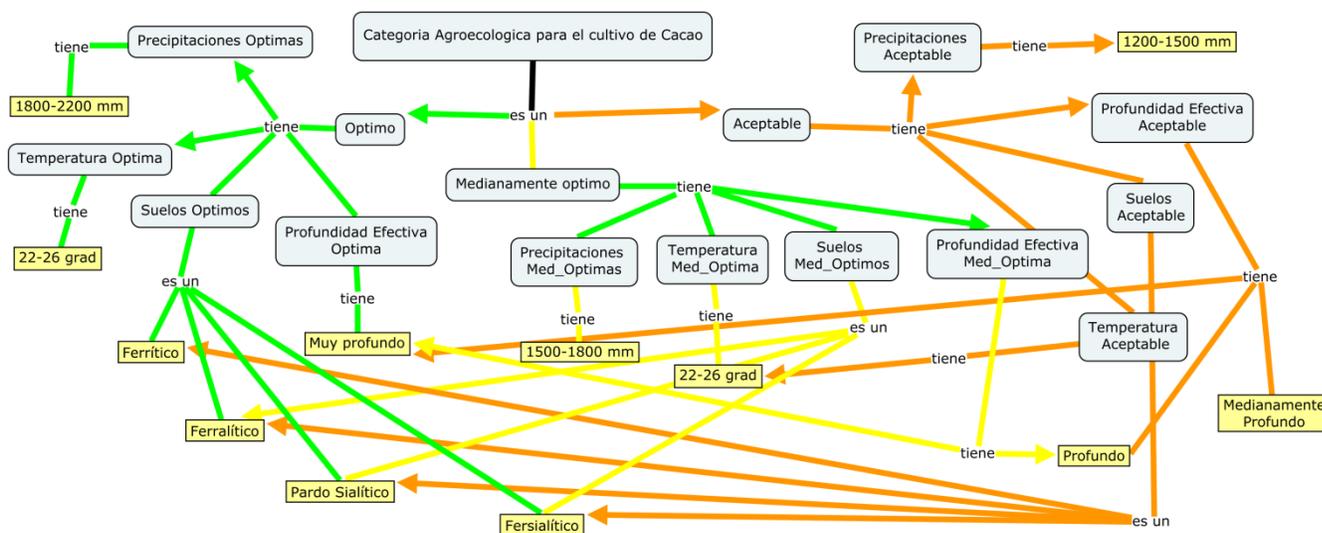


Fig. 26. Esquema simplificado de la Ontología de Aplicación utilizada como apoyo en el proceso de zonificación agroecológica.

- Se realizaron diferentes operaciones de búsqueda y análisis de la información espacial integrada en la plataforma de experimentación⁵ mediante la navegación por la ontología generada y se comparó la eficacia de los resultados con los obtenidos, utilizando los

⁵ Esta plataforma de experimentación es tratada con más detalle en la sección 3.4.

métodos de superposición de mapas y recuperación basada en el lenguaje SQL, mediante el Error de Contorno calculado por la expresión 13, véase Tabla 13.

- Se compararon los tiempos de cada operación y sus costos, véase Tabla 10.

Resultados y Discusión

En las tablas Tabla 10, Tabla 11, Tabla 12 y Tabla 13 se presentan los resultados comparativos en cuanto a eficiencia y eficacia de los métodos propuestos respecto a sus equivalentes de superposición de mapas y recuperación basados en el lenguaje SQL.

Estos resultados ponen de manifiesto cómo las relaciones existentes entre los objetos, en este caso las relaciones topológicas, resultan útiles en tareas como la determinación automática de los niveles de riesgo o la zonificación agroecológica de los suelos. Otro aspecto a destacar es que las tareas de recuperación de datos utilizando la estructura y los métodos propuestos resultan ser más simples y rápidas que los métodos convencionales para el caso de usuarios no expertos en aplicaciones SIG, ya que posibilita que estos puedan navegar por la ontología y realizar distintos tipos de análisis a partir de las recuperaciones sin la necesidad de conocer las herramientas de análisis de los SIG. Los resultados demuestran que la recuperación de la información mediante el método propuesto aumenta la calidad de los análisis pues reduce los posibles errores en las operaciones que deben realizar, por ejemplo, al crear la consulta y ofrecer un resultado de mayor exactitud, se eliminan operaciones de rasterización y superposición que dependen de la resolución y la escala de la celda.

Además, se logra un ahorro de tiempo en la realización de los análisis espaciales. Esto es debido a que se suprime una serie de pasos que conllevan a un gasto de tiempo y en los cuales es posible la comisión de errores, tales como:

- La búsqueda de las capas de datos con las que se desea trabajar ya que estas se encuentran conceptualizadas en la ontología generada.
- La creación de una consulta en lenguaje SQL, ya que el usuario navega libremente por dicha ontología, además, podrían implementarse métodos de búsqueda basados en lenguaje natural.

Ejemplos de consultas en que el sistema automáticamente no podría dar respuestas utilizando los métodos convencionales, serían las siguientes:

¿Cuáles son las zonas de alto riesgo?

¿Cuáles son las regiones de suelos aptas para la aplicación de fertilizantes?

Este tipo de consulta no puede ser obtenido a través de lenguaje *SQL* ya que, tanto en la base de datos como en el archivo *ShapeFile*, no existe ninguna tabla nombrada con los términos utilizados en la consulta. Por tanto, a través del uso de la *ORD* conjuntamente con las *ONS* (*OD-ER* y *OA-CE*), el sistema puede entender fácilmente lo que se desea recuperar. Esto es posible ya que a partir de la *OA-CE* se pueden conocer qué zonas temáticas y cuáles valores de estas, son necesarios para responder la consulta, véase la Fig. 20. Luego, tomando en cuenta la información que provee la *OA-CE*, se buscan en la *OD-ER* complementada con la *ORD* aquellas zonas que cumplen con estos requerimientos dando así la respuesta esperada. Hay que destacar el importante papel que juegan las relaciones topológicas existentes entre los distintos objetos almacenados en los datos, ya que cada una de las zonas de riesgos o de suelos a recuperar está compuesta por la superposición de un conjunto de objetos que contengan las relaciones descritas en la expresión 10.

Además, las recuperaciones fueron realizadas sin la necesidad de la construcción de una consulta en lenguaje SQL por lo que es posible expandir su utilización a usuarios no expertos.

Tabla 10. Resultados de las comparaciones de la eficacia de las operaciones de búsqueda y análisis de la información espacial utilizando el método propuesto (SIGGO) y métodos de superposición de mapas y recuperación basada en el lenguaje SQL. (Estimación de Riesgos)

Operación	Resultado en el SIG			Resultado en el SIGGO	
	RA (ha)	DA (ha)	EC	DA (ha)	EC
Selección de Zonas de Riesgos Muy Altos	11263	11272.048	0.08	11266.96	0.03
Selección de Zonas de Riesgos Altos	41044	41053.500	0.02	41045.85	0.004
Selección de Zonas de Riesgos Medios	166331	166533.75	0.12	166330.49	0.0003

Tabla 11. Resultados de las comparaciones entre tiempos y costos de la realización de las operaciones de búsqueda y análisis de la información espacial utilizando el método propuesto (SIGGO) y métodos de superposición de mapas y recuperación basada en el lenguaje SQL. (Estimación de Riesgos).

Pasos Metodológicos	Operación en el SIG	Tiempo Prom. (min)	Costo material (CUP)	Costo salario (CUP)	Costo Total (CUP)	Operación en la plataforma SIGGO	Tiempo Prom. (min)	Costo material (CUP)	Costo salario (CUP)	Costo Total (CUP)
• Selección de regiones por velocidades del viento	<ul style="list-style-type: none"> • Selección manual de la capa (mediciones). • Conformación manual de la Sentencia SQL • Selección automática de los objetos 	3.0	0.006	0.17		<ul style="list-style-type: none"> • Selección manual del criterio en la ontología. • Selección automática de las instancias 	0.5	0.001	0.03	0.031
• Selección de los intervalos de relieve para cada zona	<ul style="list-style-type: none"> • Selección manual de la capa (relieve). • Reclasificación manual por intervalos de relieve. • Conformación manual de la Sentencia SQL • Selección automática de los objetos 	3.0	0.006	0.17	0.175	<ul style="list-style-type: none"> • Selección manual del criterio en la ontología. • Selección automática de las instancias 	0.5	0.001	0.03	0.031
• Selección de los intervalos de	<ul style="list-style-type: none"> • Selección manual de la capa (clima) • Conformación manual de la Sentencia SQL 	3.0	0.006	0.17	0.175	<ul style="list-style-type: none"> • Selección manual del criterio en la ontología. • Selección automática 	0.5	0.001	0.03	0.031

precipitaciones para cada zona	• Selección automática de los objetos.					de las instancias				
• Selección de los intervalos de temperatura para cada zona	• Selección manual de las capas (clima) • Conformación manual de la Sentencia SQL • Selección automática de los objetos.	3.0	0.006	0.17	0.175	• Selección manual del criterio en la ontología. • Selección automática de las instancias	0.5	0.001	0.03	0.031
• Obtención de las zonas de riesgos	• Conversión a formato raster y reclasificación de cada uno de los resultados de las consultas • Superposición de los mapas raster y reclasificación del mapa resultante	20	0.04	1.12	1.17	• Selección manual del criterio en la ontología de aplicación • Selección automática de las instancias	0.5	0.001	0.03	
Totales		32	0.06	1.80	1.87		2.5	0.005	0.15	0.155

Tabla 12. Resultados de las comparaciones de la eficacia de las operaciones de búsqueda y análisis de la información espacial utilizando el método propuesto (SIGGO) y métodos de superposición de mapas y recuperación basada en el lenguaje SQL.

Operación	Resultado en el SIG			Resultado en el SIGGO	
	RA (ha)	DA (ha)	EC	DA (ha)	EC
Selección de áreas con la clasificación “optima”	2269.80	2279.75	0.43	2269.87	0.003
Selección de áreas con la clasificación “medianamente óptima”	6620.00	6634,75	0.21	6615.42	0.069
Selección de áreas con la clasificación “aceptable”	37698.00	37737.00	0.1	37699.75	0.004
Totales	46587.8	46651.5	0.74	46585.04	0.076

Tabla 13. Resultados de las comparaciones entre tiempos y costos de la realización de las operaciones de búsqueda y análisis de la información espacial utilizando el método propuesto (SIGGO) y métodos de superposición de mapas y recuperación basada en el lenguaje SQL. (Suelos)

Pasos Metodológicos	Operación en el SIG	Tiempo Prom. (min)	Costo material (CUP)	Costo salario (CUP)	Costo Total (CUP)	Operación en la plataforma SIGGO	Tiempo Prom. (min)	Costo material (CUP)	Costo salario (CUP)	Costo Total (CUP)
• Selección de agrupamientos de suelos para cada zona	<ul style="list-style-type: none"> • Selección manual de la capa (suelos). • Conformación manual de la Sentencia SQL • Selección automática de los objetos 	3	0.006	0.17	0.176	<ul style="list-style-type: none"> • Selección manual del criterio en la ontología. • Selección automática de las instancias 	0.5	0.001	0.02	0.021
• Selección de los intervalos de relieve para cada zona	<ul style="list-style-type: none"> • Selección manual de la capa (relieve). • Reclasificación manual por intervalos de relieve. • Conformación manual de la Sentencia SQL • Selección automática de los objetos 	8	0.017	0.45	0.467	<ul style="list-style-type: none"> • Selección manual del criterio en la ontología. • Selección automática de las instancias 	0.5	0.001	0.02	0.021
• Selección de los tipos de suelos por factores limitantes	<ul style="list-style-type: none"> • Selección manual de las capas(suelos y factores limitantes) • Conformación manual de la Sentencia SQL combinando suelos y factores limitantes • Selección automática de los objetos. 	16	0.034	0.90	0.934	<ul style="list-style-type: none"> • Selección manual del criterio en la ontología. • Selección automática de las instancias 	0.7	0.0015	0.03	0.0315
Obtención de las regiones agroecológicas	<ul style="list-style-type: none"> • Conversión a formato raster y reclasificación de cada uno de los resultados de las consultas. • Superposición de los mapas raster y reclasificación del mapa resultante 	50	0.108	2.82	2.928	<ul style="list-style-type: none"> • Selección manual del criterio en la ontología de aplicación. • Selección automática de las instancias 	0.8	0.0017	0.04	0.0417
Totales		77	0.16	4.34	4.511		2.5	0.005	0.11	0.115

Conclusiones del Experimento

Se demuestra que con el uso del enfoque del nuevo tipo de ontología y los métodos propuestos, la eficacia y la eficiencia de las formas de análisis, explotación e integración de los objetos geográficos y las tareas existentes en los SIG (p.e: La recuperación y/o el análisis para la toma de decisiones), son mejorados.

Estos resultados muestran que el uso de la ORD complementada con la OD-ER como una capa intermedia entre el sistema de recuperación y la base de datos espacial, permite el manejo de los objetos a través de su representación conceptual (*abstracción semántica*), dando la posibilidad de que el SIG disponga de herramientas para el procesamiento semántico de los datos geoespaciales. Esto permite que los diferentes tipos de análisis realizados sobre estos tipos de datos puedan ser realizados tanto por especialistas como por usuarios no expertos en el uso de SIG.

3.5. Experimento 4. *Representación y Recuperación de objetos Geoespaciales en Imágenes de Teledetección.*

Objetivo del Experimento:

Comprobar la validez de los métodos propuestos para demostrar cómo el uso del tipo de ontología propuesto combinado con técnicas de procesamiento de imágenes permite enfrentar las tareas de detección y recuperación de la información mediante su aplicación en la detección y recuperación de objetos geoespaciales en imágenes de teledetección de forma automática, reduciendo así errores y el desgaste que conlleva este tipo de trabajo.

Materiales y métodos empleados:

Con el desarrollo de las nuevas tecnologías el volumen de imágenes de teledetección ha aumentado considerablemente, así como la extensión territorial que estas cubren. Buscar y

recuperar objetos contenidos en estas imágenes son tareas caras y complejas desde el punto de vista computacional y humano.

Datos de partida

Se cuenta con dos bases de datos de imágenes SPOT pancromáticas con una resolución espacial de 20 m. Esta contiene dos tipos diferentes de objetos a detectar, objetos artificiales (*construidos por el Hombre*) y objetos naturales. La segunda contiene imágenes de teledetección que pueden o no contener los objetos de interés. La Tabla 14 y la Tabla 15 presentan descripciones de las bases de datos utilizadas.

Tabla 14. Descripciones Generales de las bases de datos utilizadas

Base de Datos de Obj. Geospaciales		Base de datos de imágenes satelitales
Nº de Objects		Nº de imágenes
Artificiales	Naturales	
21	13	166

Tabla 15. Nº de imágenes en las cuales los objetos están contenidos.

Objetos Geográficos	Nº de imágenes
Objeto 1	6 Imágenes
Objeto 2	3 Imágenes
Objeto 3	9 Imágenes
Objeto 4	12 Imágenes
Objeto 5	2 Imágenes
...	...
Objeto 29	2 Imágenes
Objeto 30	7 Imágenes
Objeto 31	5 Imágenes
Objeto 32	3 Imágenes
Objeto 33	3 Imágenes
Objeto 34	2 Imágenes

La Fig. 27 muestra una representación gráfica de un subconjunto de las imágenes contenidas en estas bases de datos. Las imágenes contenidas en ambas bases de datos

pueden tener tanto problemas de rotación como de escala incluyendo que no están georeferenciadas.

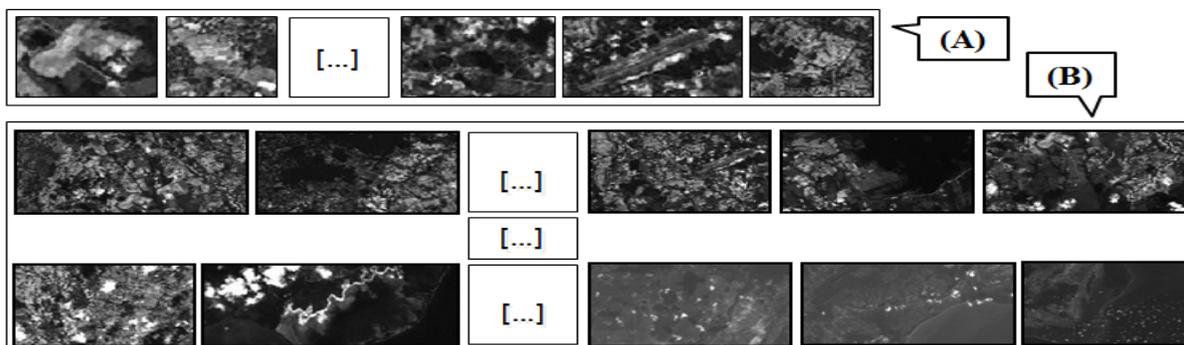


Fig. 27. Representación gráfica de un subconjunto de las imágenes contenidas en las bases de datos utilizadas. (A) Base de datos de los objetos geospaciales de interés. (B) Base de datos de las imágenes de teledetección.

Se asume además, que, los objetos abarcan más de un pixel en la imagen y pueden tener problemas de escala y rotación. Otro aspecto a tener en cuenta es el número insuficiente de instancias de los objetos por lo que los miles de ejemplos de entrenamiento requeridos por muchos de los algoritmos de clasificación no están disponibles.

Procedimiento y procesos realizados

En este acápite se hace una propuesta de un framework basado en ontologías con el objetivo de enfrentar el problema de la detección y recuperación de objetos en imágenes satelitales. Este framework tiene una arquitectura gobernada por ontologías en la que se hace uso del nuevo tipo de ontología propuesto, así como también ontologías de niveles semánticos superiores. La propuesta presentada utiliza tres ontologías, ellas son: la Ontología de Representación de Datos (ORD) y dos Ontologías de Nivel Superior (*una ontología de tarea y una ontología de dominio*). En este caso cabe destacar que en la estructura *ctemat* de los nodos (NRD) que conforman la ORD, el campo “(...)” de la 2-tupla (*nombre, (...)*) va a contener los rasgos de bajo nivel invariantes a la escala y rotación

del objeto representado en el NRD. La ontología de tarea es utilizada en el proceso de recuperación y la de dominio para representar la abstracción semántica del dominio en cuestión. El framework presentado contiene tres módulos principales, ellos son:

- *Módulo de Pre-procesamiento Semántico de Objetos*: En este módulo los objetos a recuperar son pre-procesados. El esquema del pre-procesamiento es mostrado a continuación:

1. Por cada objeto se extraen y representan en la ORD descriptors invariantes a la escala y la rotación. El vector descriptor es generado con la combinación de los algoritmos SURF(*Speeded Up Robust Features*) [95] y SIFT (*Scale-Invariant Features Transform*) [96]. Entonces, cada Nodo de Representación de Datos (NRD) en la ORD contiene los rasgos semánticos y de bajo nivel del objeto representado.
2. Una vez que la ORD es generada los objetos son contextualizados desde el punto de vista semántico. Esto se realiza mediante el enlace de los NRD con la ontología de dominio.

La fase de pre-procesamiento se presenta en la Fig. 28.

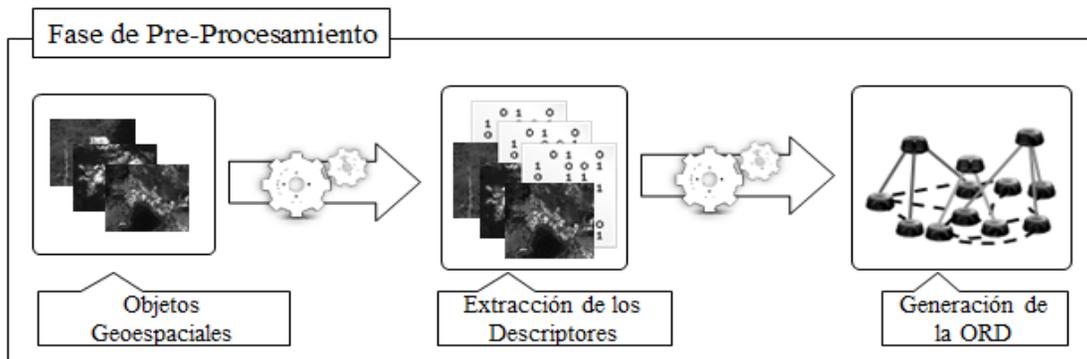


Fig. 28. Diagrama de la fase de pre-procesamiento de los objetos geoespaciales.

- *Módulo de Procesamiento de la Consulta de Usuario:* Este módulo procesa las consultas del usuario para la recuperación de los objetos. La consulta es realizada en lenguaje natural. El esquema del procesamiento es mostrado a continuación.
 1. Primeramente se realiza un proceso de desglose de la consulta en la que los conceptos claves son identificados para determinar los objetos a recuperar. Esto se realiza mediante la eliminación de palabras que no aportan nada en el proceso de recuperación (*stop words* (conjunciones, artículos, pronombres, etc.)) apoyados con la herramienta Freeling [97]. También se utiliza un diccionario de sinónimos.
 2. Se reformula la consulta con los términos de interés resultantes de la fase previa.
 3. Finalmente, los términos de interés son buscados en la ontología y sus rasgos descriptivos son obtenidos. Este resultado es utilizado como parámetro de entrada en la fase de detección de objetos.

El diagrama de la fase de procesamiento de la consulta se muestra en la Fig. 29.

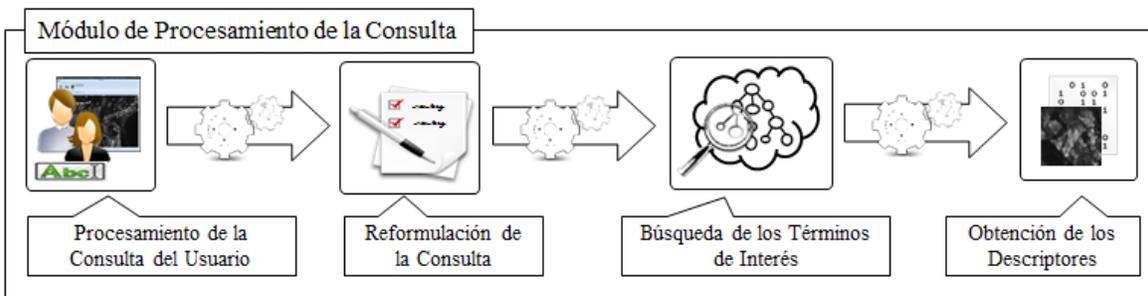


Fig. 29. Diagrama de la fase de procesamiento de la consulta

- *Módulo de Detección de Objetos:* Este módulo procesa imágenes de entrada para el reconocimiento de objetos. El esquema del procesamiento es mostrado a continuación:
 1. El primer paso en esta fase es la extracción de los rasgos de bajo nivel invariantes a escala y rotación en las imágenes de entrada, las cuales dado su gran extensión espacial, son procesadas por regiones. Los algoritmos SURF [95]

y SIFT [96] son utilizados para describir estas imágenes dando como resultado un vector descriptor combinado.

- El siguiente paso es el proceso de comparación de descriptores para identificar el objeto buscado en la imagen. El clasificador de la regla del vecino más cercano con rechazo es utilizado para determinar qué descriptores de la imagen corresponden con los descriptores del objeto. Se utilizó el umbral de rechazo (0.8) propuesto por Lowe [96] ya que da buenos resultados. Finalmente, se obtiene la localización del objeto.

El diagrama de la fase de detección se muestra en la Fig. 30.

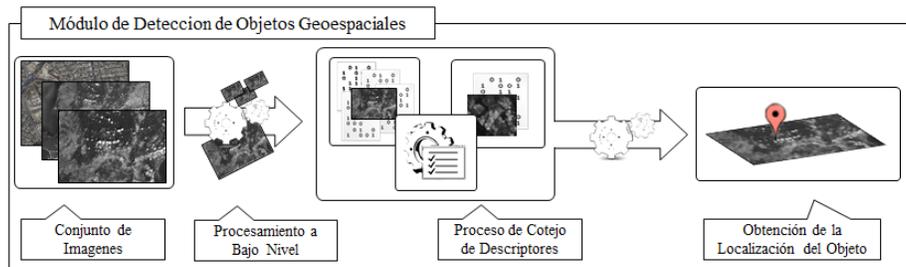


Fig. 30. Diagrama de la fase de detección de los objetos.

. La implementación realizada contiene tres secciones, ellas son:

- *Sección Management*: En esta sección se realizan las configuraciones.
- *Sección Pre-Processing*: Esta sección implementa el módulo de pre-procesamiento semántico de los objetos, ver Fig. 31– A.

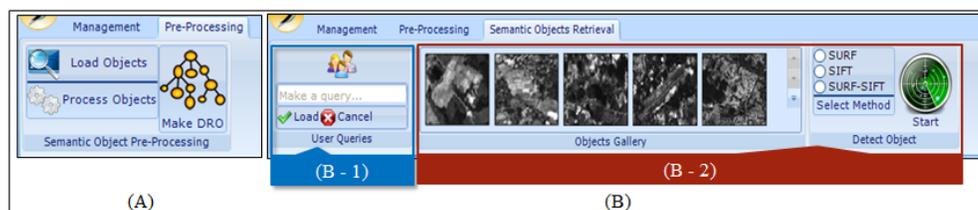


Fig. 31. Representación Gráfica de dos secciones de la plataforma. (A) Sección de Pre-procesamiento Semántico de los Objetos. (B) Sección de Recuperación Semántica de Objetos.

- *Sección Semantic Objects Retrieval*: En esta sección se implementan los módulos de *Procesamiento de la Consulta de Usuario* (ver Fig. 31 – B1) y de *Detección de*

Objetos (ver Fig. 31 – B2). En el panel de procesamiento de la consulta pueden realizarse consultas en lenguaje natural. Como resultado de la consulta son cargados los objetos semánticamente representados en la ORD en la *galería de objetos* (ver Fig. 31 – B2).

Los pasos que se desarrollaron en el experimento fueron los siguientes:

1. Se realizó el pre-procesamiento de la base de datos que contiene los objetos a detectar en el módulo de *Pre-procesamiento Semántico de Objetos* utilizando los descriptores SIFT y SURF (Tabla 15).
2. Se realizó el procesamiento de dos consultas a la base de datos de imágenes satelitales para la recuperación de objetos naturales y objetos artificiales en el módulo de *Procesamiento de la Consulta de Usuario*.
3. Se realizó la detección de los objetos contenidos en ambas consultas en el módulo de *Módulo de Detección de Objetos*.
4. Se realizó la medición del parámetro Precisión (*Pr*) que se define como la razón entre número de objetos recuperados y el número total de objetos a recuperar [98].
5. Se midieron los tiempos de procesamiento y se estimaron costos correspondientes a la recuperación de los objetos de interés para los procesos 1, 2 y 3.
6. Se realizó la selección, extracción y salva en ficheros de los descriptores correspondientes a los objetos a detectar utilizando los algoritmos SIFT y SURF en la plataforma de procesamiento digital de imágenes TN Estudio [99].
7. Se realizó la detección de los objetos en la base de datos de imágenes satelitales mediante la clasificación supervisada de las mismas, utilizando el clasificador paramétrico de Máxima Probabilidad ya que este hace uso de una figura geométrica

para ubicar dentro de sus contornos los puntos pertenecientes a una clase. El término “paramétrico”, entonces, se refiere a los parámetros que describen la figura geométrica [99].

8. Se realizó la medición del parámetro Pr para este proceso.
9. Se midieron los tiempos de procesamiento y se estimaron los costos correspondientes a la recuperación de los objetos de interés para los procesos 6 y 7.
10. Se compararon los resultados, en cuanto a eficacia medida por el parámetro Pr y eficiencia mediante los tiempos y costos del proceso de recuperación.

Resultados y Discusión

En la Tabla 16 se presenta la comparación del desempeño del método propuesto con el desempeño del enfoque de clasificación supervisada [99] para la recuperación de objetos geospaciales utilizando tres descriptores de rasgos para la base datos de prueba (*Base de datos de imágenes satelitales*). Como puede observarse en todos los casos la precisión (Pr) alcanzada en la recuperación mediante el método propuesto es mayor que la alcanzada por el método de clasificación supervisada, lo que pone de manifiesto que el proceso de recuperación realizado en el nivel semántico, posibilita una mayor eficacia que realizarlo mediante la comparación de los rasgos de los pixeles seleccionados como patrones contra los pixeles de las imágenes a clasificar.

En la Fig. 32, se presentan algunos ejemplos de la detección de objetos. Estos ejemplos fueron tomados utilizando diferentes descriptores de rasgos. Por otro lado como resultado adicional del experimento, se comprobó que el uso combinado de los descriptores provistos por los algoritmos SURF y el SIFT rechazan eficientemente sub-aéreas similares que no contienen el objeto de interés por lo que muestran mejores resultados que con cada uno de ellos de forma individual. En la Fig. 32-D se puede ver un ejemplo de ello.

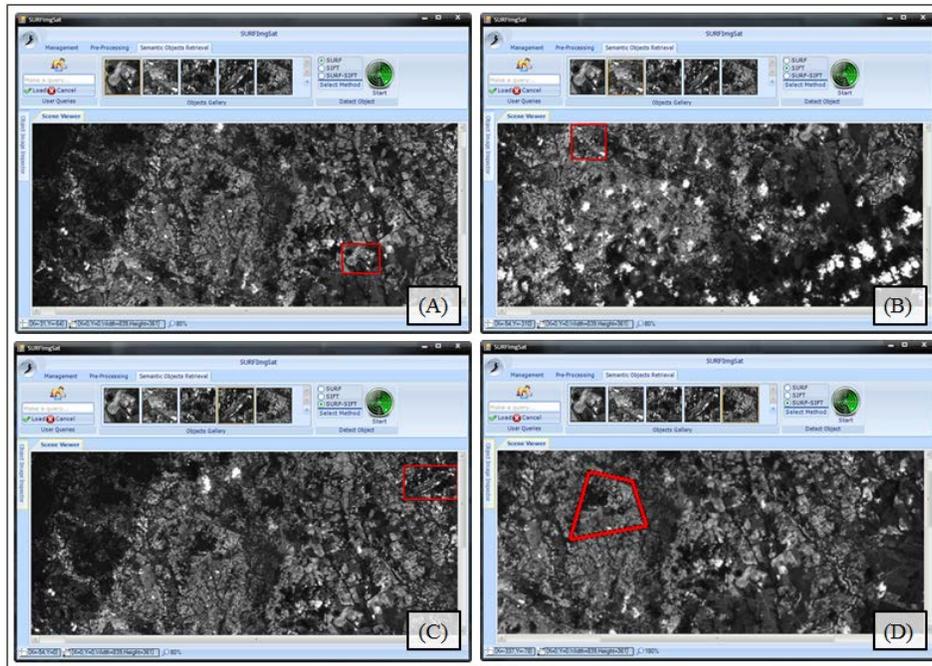


Fig. 32. Representación Gráfica de los resultados de la detección de objetos en imágenes satelitales.

En la Tabla 17 se presenta la estimación de la eficiencia en el proceso de recuperación de los objetos mediante la comparación de los tiempos y costos de procesamiento, como puede observarse en el caso del método propuesto se produce una disminución sustancial del tiempo y costo empleado en el procesamiento, lo que pone de manifiesto su mayor eficiencia respecto al proceso de recuperación de los objetos mediante la clasificación supervisada de las imágenes.

En la Tabla 18 se presenta el cálculo del costo económico utilizando el Análisis de Puntos de Función [100-102] para la implementación de un sistema como el propuesto y su comparación con algunos sistemas comerciales disponibles que poseen funcionalidades comparables en el sentido de su resultado final. El anexo 3 muestra en detalle los cálculos realizados para llegar al costo total del proyecto. Como puede verse el costo de implementación es significativamente inferior que el costo de la licencia de algunos de los productos comerciales que incluyen funcionalidades cuyos resultados pueden ser

comparados en cuanto a eficacia, estos productos no incluyen un procesamiento semántico de la información por lo que la detección y recuperación de objetos en las imágenes se basa en técnicas de segmentación y clasificación.

Conclusiones del Experimento

Se demuestra que el uso del tipo de ontología propuesto combinado con técnicas de procesamiento de imágenes permite enfrentar las tareas de detección y recuperación de objetos geoespaciales en imágenes de teledetección de forma automática aumentando la eficacia y la eficiencia de este proceso. Queda demostrado además, que el uso de la ORD permite la combinación de rasgos de bajo nivel con abstracciones semánticas, disminuyendo así la brecha semántica existente entre Hombre-Máquina. Posibilitando también que, la recuperación de los objetos de interés sea sobre la base del procesamiento de consultas realizadas en lenguaje natural habilitando el procesamiento semántico de objetos en las aplicaciones de este tipo.

3.6. Conclusiones

Las experimentaciones realizadas muestran la validez de la hipótesis de investigación trazada a partir de la prueba de los métodos propuestos para la representación semántica automática de objetos geoespaciales utilizando técnicas de clasificación y medidas de semejanza. Se demuestra además, cómo es posible la representación semántica automática de objetos geo-espaciales en un contexto que comprende más de un dominio de conocimiento haciendo posible la representación de objetos desde diferentes puntos de vistas. Se muestra cómo estos métodos pueden ser implementados en un entorno SIGGO, así como también su utilidad en la mejora de las formas de análisis, explotación e integración mediante tres casos de uso.

Tabla 16. Comparación del desempeño utilizando diferentes vectores de descripción

Objetos Geográficos	Clasificación Supervisada			Método Propuesto		
	Pr			Pr		
	SURF	SIFT	SURF-SIFT	SURF	SIFT	SURF-SIFT
Objetos Artificiales						
Objeto 1	0.6	0.6	0.7	0.8	0.8	0.9
Objeto 2	0.7	0.7	0.8	0.9	0.9	0.9
Objeto 3	0.5	0.7	0.7	0.8	0.9	1
...
Objeto 19	0.7	0.7	0.8	0.9	0.9	1
Objeto 20	0.6	0.7	0.8	0.8	0.8	0.9
Objeto 21	0.8	0.8	0.9	0.9	1	1
Objetos Naturales						
Objeto 1	0.4	0.6	0.7	0.7	0.8	0.9
Objeto 2	0.7	0.7	0.8	0.8	0.8	0.9
Objeto 3	0.4	0.5	0.6	0.6	0.8	0.8
...
Objeto 11	0.6	0.7	0.7	0.8	0.9	0.9
Objeto 12	0.7	0.7	0.8	0.9	0.9	1
Objeto 13	0.6	0.6	0.8	0.8	0.8	0.9

Tabla 17. Resultados de las comparaciones entre tiempos y costos de la realización de las operaciones utilizando el método propuesto y los estimados para método de clasificación supervisada de las imágenes.

Clasificación supervisada						Método propuesto					
Operación	Tiempo operación (min)	Tiempo total (min)	Costo Material (CUP)	Costo Salario (CUP)	Costo Total (CUP)	Operación	Tiempo operación (min)	Tiempo total (min)	Costo Material (CUP)	Costo Salario (CUP)	Costo Total (CUP)
Selección, extracción y salva en ficheros de los patrones correspondientes a los objetos a detectar	10	120	0,26	6,77	7,03	Pre-procesamiento de la base de datos que contiene los objetos a detectar en el módulo de <i>Pre-procesamiento Semántico de Objetos</i>	0,5	6	0,01	0,34	0,35
Detección de los objetos en la base de datos de imágenes satelitales mediante la clasificación supervisada	3	48	0,10	2,71	2,81	Procesamiento de las consultas en el módulo de <i>Procesamiento de la Consulta de Usuario</i>	0,001	0,002	0,000004	0,00011	0,000117
						Detección de los objetos contenidos en ambas consultas en el módulo de <i>Módulo de Detección de Objetos.</i>	0,5	6	0,01	0,34	0,35
Totales	13	168	0,35728	9,48	9,84		1,00	12,00	0,03	0,68	0,70

Tabla 18. Costo económico de un sistema como el propuesto y su comparación con algunos sistemas comerciales disponibles

Conceptos	Valor (CUP)	Software para el procesamiento de imágenes de teledetección	Valor (USD) (El valor de una licencia, sin soporte técnico)
Factor de peso de los actores (UAW).	3	ERDAS IMAGINE	5400
Factor de peso de los Casos de Uso	240	ER MAPPER	3600
Puntos de Casos de Uso (UUCP).	243	SCANMAGIC	2500
Factor de complejidad técnica (TCF).	1.04	ENVI	4232
Factor de ambiente (EF)	0.47		
Puntos de Casos de Uso Ajustados (UCP).	1.187.784		
Factor de conversión (CF).	28		
Esfuerzo(E)	33.257.952		
Tiempo de desarrollo	1662.8976 hrs. (8.7 meses)		
Costo Total en CUP	13.511,04		
Costo Total en USD (1 USD=25CUP)	540,44		

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

De manera general, al culminar este trabajo de tesis, podemos plantear que con los métodos propuestos para la representación semántica de objetos geoespaciales integrados en un entorno de SIGGO, la hipótesis de investigación trazada ha sido validada. Estos métodos fueron desarrollados sobre la base de la propuesta de un nuevo tipo de ontología, que representa la naturaleza semántica y dinámica de los objetos geoespaciales incluyendo las relaciones existentes entre ellos.

De manera específica podemos plantear las siguientes conclusiones de la investigación:

- Se presentó un nuevo tipo de ontología para la representación semántica de objetos geoespaciales integrados en sistemas con arquitectura *SIGGO*.
- Se definió una estructura ontológica para la representación semántica de objetos geoespaciales tomando en cuenta además, las relaciones existentes entre ellos.
- Se obtuvo un método general para la extracción automática de las características de bajo nivel que definen a los objetos geoespaciales.
- Se definieron las características básicas que debe cumplir un algoritmo de agrupamiento jerárquico para ser utilizado en el descubrimiento automático de nuevos niveles de abstracción semántica subyacentes en datos geoespaciales, proponiéndose además, un candidato de los existentes en la literatura.
- Se obtuvieron dos métodos para la interrelación automática de la representación semántica de los datos geoespaciales con ontologías de nivel superior.

- Se obtuvo una plataforma de experimentación en la cual se implementaron los métodos desarrollados a lo largo de este trabajo de investigación en un entorno de aplicaciones SIGGO.
- Se obtuvo un prototipo de sistema para la detección y recuperación de objetos geoespaciales en imágenes de teledetección guiado por ontologías. Este es un ejemplo de aplicación de los métodos propuestos en un sistema con arquitectura de SIGGO.

Recomendaciones

Con los resultados obtenidos a lo largo de esta investigación no se da por concluido el trabajo en esta temática. Es principalmente por esta razón que, a continuación se plantean algunos aspectos que pueden seguir siendo desarrollados a modo de trabajos futuros, estos son:

- Profundizar en la teoría de la representación de objetos geoespaciales en el espacio semántico y su extensión a otros tipos de objetos.
- Investigar sobre métodos de recuperación y análisis de objetos geoespaciales dada su representación en el espacio semántico con contexto multidimensional.

REFERENCIAS BIBLIOGRÁFICAS

- [1] M. Lutz and E. Klien, "Ontology-Based Retrieval of Geographic Information," *International Journal of Geographical Information Science (IJGIS)* 20(3): 233-260, 2008.
- [2] Internet. (2010, Enero). *Open Geospatial Consortium, Inc.® (OGC)*. Available: <http://www.opengeospatial.org/>
- [3] Internet. (2010, Marzo). *World Wide Web Consortium, (W3C)*. Available: <http://www.w3.org/>
- [4] U. Visser, "Intelligent Information Integration for the Semantic Web," *Lecture Notes in Computer Science. Springer Berlin - Heidelberg*, vol. volume 3159, 2004.
- [5] M. Kavouras, M. Kokla, and E. Tomai, "Comparing categories among geographic ontologies," *Computers & Geosciences, Special Issue, Geospatial Research in Europe: AGILE 2003*.
- [6] A. Schwering and M. Raubal, "Spatial relations for semantic similarity measurement," *Lecture Notes in Computer Science, editor, Proceedings of the ER'05 Perspectives in Conceptual Modeling 3770, pages 259-269. Springer Berlin/Heidelberg*, 2005.
- [7] A. Sotnykova, C. Vangenot, N. Cullot, N. Bennacer, and M. Aufaure, "Semantic mappings in description logics for spatio-temporal database schema integration," *Journal on Data Semantics III, pages 143-167.*, 2005.
- [8] F. Hakimpour, "Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata," *PhD thesis Zurich University*, 2003.
- [9] F. Fonseca and M. Egenhofer, "Ontology-Driven Geographic Information Systems," *Proc. 7th ACM Symposium on Advances in Geographic Information Systems 1999*.

- [10] L. Stoimenov, A. Stanimirovic, and S. Djordjevic-Kajan, "Discovering mappings between ontologies in semantic integration process," *Proceedings of the AGILE'06: 9th Conference on Geographic Information Science*, pages 213-219, Visegrád, Hungary, 2006.
- [11] I. Graham, *Métodos Orientados a Objetos*: DIAZ DE SANTOS, S.A. EDICIONES, 1996.
- [12] R. Barker, *Entity Relationship Modelling*: Addison-Wesley Publishing Company. Inc., 1990.
- [13] G. Barchini, M. Álvarez, and S. Herrera, "Sistemas de Información: Nuevos Escenarios Basados en Ontologías," *Revista de Gestão da Tecnologia e Sistemas de Informação Journal of Information Systems and Technology Management*, 2006.
- [14] F. Fonseca, M. Egenhofer, and P. Agouris, "Using Ontologies for Integrated Geographic Information System," *Transaction in GIS* 6(3), 2002.
- [15] N. Guarino. (1998). *Formal Ontology and Information Systems*.
- [16] R. Larin-Fonseca and E. Garea-Llano, "Integración semántica de datos espaciales con sistemas de información geográfica," *Reportes Técnicos Publicados del Centro de Aplicaciones de Tecnología de Avanzada*, ISSN 2072-6287, 2009.
- [17] R. Larin-Fonseca and E. Garea-Llano, "Automatic Generation of Geospatial Application Ontology " *VII Congreso Nacional de Reconocimiento de Patrones, RECPAT*, ISBN: 978-959-155-048-6, 2009.
- [18] R. Larin-Fonseca and E. Garea-Llano, "Combination of Spatial Analysis Methods and Semantics Extraction in a SDI Environment for Agro-ecological Zonation," *Proceedings of The 12th AGILE International Conference on GIScience en Hannover, Germany*. ISSN 2073-8013, 2009.
- [19] R. Larin-Fonseca and E. Garea-Llano, "Integración Semántica de Datos Espaciales con Sistemas de Información Geográfica," *Informática 2009, XIII Convención y Feria Internacional, Memorias*, ISBN: 978-959-286-010-0., 2009.

- [20] R. Larin-Fonseca and E. Garea-Llano, "Representación Semántica Automática de Datos Geoespaciales," *Segundo Taller Cubano Eureka 2010. Red Iberoamericana de Descubrimiento de Conocimientos (EUREKA IBEROAMERICA)* 2010.
- [21] R. Larin-Fonseca and E. Garea-Llano, "Topological Relations as Rule for Automatic Generation of Geospatial Application Ontology," ed: Fco Macias y Juan C.Monllor Eds. "Desarrollo de grandes aplicaciones de Red" JDARE 2010, Editorial OFP Velografic, Universidad de Alicante España, ISBN 978-84-614-3720-7, 2010, pp. 27-38.
- [22] R. Larin-Fonseca and E. Garea-Llano, "Automatic Representation of Semantic Abstraction of Geographical Data by Means of Classification," *Springer-Verlag Berlin Heidelberg 2010*, vol. CIARP 2010, LNCS 6419, pp. 557–568, 2010.
- [23] R. Larin-Fonseca and E. Garea-Llano, "Automatic Representation of Geographical Data from Semantic Point of View throughout a New Ontology and Classification Techniques " *Transaction in GIS. Blackwell Publishing Ltd, ISSN: 1467-9671*, vol. Vol 15, 2011.
- [24] R. Larin-Fonseca and E. Garea-Llano, "Automatic Generation of the Data-Representation Ontology for Semantic Integration of Heterogeneous Geographical Data," *VII Congreso Internacional GEOMATICA 2011, Memorias ISBN:978-959-7213-01-7*, 2011.
- [25] E. Garea-Llano and R. Larin-Fonseca, "Reglas Para La Inclusión De Relaciones Topológicas En Ontologías De Aplicaciones SIG, Caso De Estudio, Evaluación De Factibilidad," *VII Congreso Internacional GEOMATICA 2011, Memorias, ISBN: 978-959-7213-01-7*, 2011.
- [26] E. J. González, R. Larin-Fonseca, and E. Garea-Llano, "Hacia métodos de Análisis de Datos Espaciales Raster en el nivel Semántico," *VIII Congreso Nacional de Reconocimiento de Patrones, RECPAT, ISBN: 978-959-16-1287-8*, 2010.
- [27] R. Larin-Fonseca and E. Garea-Llano, "Representación Semántica N-Dimensional de Datos Geográficos en un Entorno Siggo," *VIII Congreso Nacional de Reconocimiento de Patrones, RECPAT, ISBN: 978-959-16-1287-8*, 2010.

- [28] E. Garera-Llano, R. Larin-Fonseca, R. Oliva-Santos, and F. Vera-Voronisky, "Sistemas de Información Geográfica como Herramienta para la Integración e interpretación semántica de la información espacial," in *Semántica Espacial y Descubrimiento de Conocimiento para desarrollo sostenible*, ed: en Tatiana Delgado, Jose Capote, eds. Parte I, Cap I.2. Ed CUJAE. ISBN 978-959-261-282-2. La Habana, Cuba, 2009, pp. 147- 165.
- [29] E. J. González, R. L. Fonseca, and E. G. Llano, "Hacia Métodos de Análisis de Datos Espaciales Raster en el Nivel Semántico (ext.)," *Computación y Sistemas*, Vol. 15, No. 1 pp. 91-106. ISSN 1405-5546, 2011.
- [30] R. Larin-Fonseca and E. Garea-Llano, "Relaciones Topológicas para la Generación Automática de Ontologías de Aplicación en Sistemas de Información Geográfica," *Revista Cienc. Tierra y Esp., II Época*, No. 12 / pp. 47-57 ISSN 1729-3790., 2011
- [31] R. Larin-Fonseca and E. Garea-Llano, "Hacia la Construcción de Sistemas de Información Geográfica Gobernados por Ontologías (SIGGO)," ed. GrupoM, Universidad de Alicante: Desarrollo de Grandes Aplicaciones de Red. VIII Jornadas, JDARE 2011. Alicante, España, noviembre 24-25, 2011. Actas, ISSN: 1889-7819, ISBN: 978-84-615-4940-5, 2011, pp. pp. 115 - 126.
- [32] R. Larin-Fonseca and E. Garea-Llano, "Hacia la Construcción de Sistemas de Información Geográfica de Nueva Generación," *IX Congreso Nacional de Reconocimiento de Patrones RECPAT en el Evento Internacional COMPUMAT 2011*. ISBN: 978-959-250-658-9, 2011.
- [33] R. Larin-Fonseca and E. Garea-Llano, "Semantic Representation of Geospatial Objects Using Multiples Knowledge Domains," *Lecture Notes in Computer Science, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 7441, pp. 187-195, 2012.
- [34] F. Escobar, G. Hunter, I. Bishop, and A. Zerger, "Introduction to GIS," *Department of Geomatics, The University of Melbourne*, 2000.
- [35] Y. Leung. (2010). *Knowledge Discovery in Spatial Data*.

- [36] Internet. (2010). *Grupo para el Desarrollo de los Sistemas de Información Geográfica*. Available: <http://www.gdsig.com.ar/?q=definiciones>
- [37] Internet. (2010). *The guide to Geographic Information Systems*. Available: <http://www.gis.com/content/what-gis>
- [38] NCGIA, "Introduction to GIS," *National Center for Geographic Information and Analysis. University of California*, vol. Vol. 1, 1990.
- [39] J. Goizueta, "Bases de datos geográficos," *ECAS Técnicos Asociados S.A.*, 2000.
- [40] A. Levinsohn, "La Interoperabilidad Geoespacial: El Santo Grial del Campo SIG," *Revista Internacional de Ciencias de la Tierra*, 2001.
- [41] Internet. (2010, Julio). *International Organization for Standardization*. Available: <http://www.iso.org/iso/home.htm>
- [42] Internet. (2010, Julio). *ISO/TC-211 Geographic Information/Geomatics*. Available: <http://www.isotc211.org/>
- [43] T. Berners-Lee. (2010, Julio). *Tim Berners-Lee Home Page*. Available: <http://www.w3.org/People/Berners-Lee/>
- [44] Internet, "OGC® Standards and Supporting Documents," *Open Geospatial Consortium. OpenGIS® and OGC®*. <http://www.opengeospatial.org/standards>. Acceso en Feb 2012, 2012.
- [45] D. Anaya, O. Cantán, J. Lacasta, J. Noguerras, and F. J.Zarazaga, "Interoperabilidad entre estándares de meta-datos geográficos," *Depto. Informática e Ingeniería de Sistemas Universidad de Zaragoza*, 2002.
- [46] H. Wache, T. V. Gele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. H. bner, "Ontology-based integration of information—a survey of existing approaches," *Proceedings of the IJCAI'01: 17th International Joint Conferences on Artificial Intelligence, Seattle, WA, pp.108–117*, 2001.

- [47] C. Batini and M. Scannapieco, "Data Quality: Concepts, Methodologies and Techniques," *Berlin, Springer-Verlag Berlin Heidelberg.*, 2006.
- [48] H. Stuckenschmidt, U. Visser, G. Schuster, and T. Vögele, "Ontologies for Geographic Information Integration," *Pergamon Press, Inc. Tarrytown, NY, USA*, 2007.
- [49] R. Lemmens, "Semantic Interoperability of Distributed Geo-services," *Delft, Netherlands. NCG*, 2006.
- [50] P. A. Power, "Estándares y Recomendaciones del OGC y CSG. Capítulo B Interoperabilidad y Ejemplos de Interoperabilidad," *Instituto Geográfico Nacional*, 2006.
- [51] G. N. Hess and C. Iochpe, " Ontology-driven resolution of semantic heterogeneities in gdb conceptual schemas," *Proceedings of the GEOINFO'04: VI Brazilian Symposium on GeoInformatics*, 2004.
- [52] F. Lehmann. (1992). *Semantic Networks in Artificial Intelligence. Vol. 23*.
- [53] M. Minsky, "A framework for representing knowledge," *P.H. Winston (ed.), The Psychology of Computer Vision, McGraw-Hill*, 1975.
- [54] T. Honderich and C. G. Trevijano, "Enciclopedia Oxford de filosofía," *Oxford University Press*, 2001.
- [55] F. J. G. Peñalvo, "Web Semántica y Ontologías," *Tendencias en el Desarrollo de Aplicaciones Web*, 2005.
- [56] G. W. Leibniz, "Opuscles et fragments inédits de Leibniz," in *extraits des manuscrits de la Bibliothèque royale de Hanovre.*, ed. Alcan, Paris, France: Reprinted Hildesheim, Georg Olms, 1961, 1903.
- [57] I. Kant, "Lectures on metaphysics - Part III Metaphysik L2," *Cambridge University Press.*, 2001.
- [58] W. O. Quine, "From a Logical Point of View, Nine Logico-Philosophical Essays," *Cambridge, Massachusetts: Harvard University Press*, 2006.

- [59] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications.," *Knowledge Acquisition*, vol. vol 5, pp. 199-220., 1993.
- [60] J. T. F. Breis, "Un Entorno de Integración de Ontologías para el Desarrollo de Sistemas de Gestión de Conocimiento," TESIS DOCTORAL, Departamento de Ingeniería de la Información y las Comunicaciones, UNIVERSIDAD DE MURCIA, 2003.
- [61] W. N. Borst, "Construction of Engineering Ontologies for Knowledge Sharing and Reuse," CTIT Ph.D-thesis series No.97-14., University of Twente, Enschede, The Netherlands, 1997.
- [62] S. Studer, R. Benjamins, and D. Fensel, "Knowledge Engineering: Principles and Methods," *Data and Knowledge Engineering*, 1998.
- [63] T. Gruber, "Ontolingua: A mechanism to support portable ontologies.," 1992.
- [64] M. J. Egenhofer, "A model for detailed binary topological relationships," *Geomatica* vol. Vo. 47, No. 3 & 4, 1993.
- [65] E. Garea-Llano, "Estado Actual de la Interpretación Semántica de Datos Espaciales," *Reporte Técnico Reconocimiento de Patrones CENATAV. ISSN 2072-6287*, 2007.
- [66] M. VIINIKKALA, "Ontology in Information Systems," *Disponible en:* <http://www.cs.tut.fi/~kk/webstuff/Ontology.pdf>, 2003.
- [67] KML. (2010, August). *Keyhole Markup Language*. Available: <http://code.google.com/apis/kml/documentation>
- [68] A. Pons-Porrata, R. Berlanga-Llavori, and J. Ruiz-Shulcloper, "Building a hierarchy of events and topics for newspaper digital libraries," *Lectures Notes on Computer Sciences 2633*, Springer-Verlag, Berlin Heidelberg, F. Sebastiani (Ed.), 2003.
- [69] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin de la Société Vaudoise des Sciences Naturelles* 37, 547-579., 1901.
- [70] E. Klien and M. Lutz, "The Role of Spatial Relations in Automating the Semantic Annotation of Geodata," *Cohn, A. and Mark, D (eds.). Proceedings of the Conference of*

- Spatial Information Theory (COSIT'05), Ellicottville, NY, USA. Lecturer Notes in Computer Science, Vol. 3693, pp. 133-148., 2005.*
- [71] E. Fix and J. L. Hodges, "Discriminatory analysis, nonparametric discrimination: Consistency properties," *Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas*, 1951.
- [72] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, "Multivariate analysis methods. An application-oriented introduction," *Springer, Berlin*, 2000.
- [73] M. E. R. Salazar, S. Á. Hernández, and E. B. Nuñez, *Coefficientes de Asociación: Plaza y Valdés*, 2001.
- [74] S. Boyed and P. Cannata, "A Semantic Database Management System: SIM," *Computer Sciences Honors Thesis, Department of Computer Science The University of Texas at Austin*, 2003.
- [75] M. Li, X.-Y. Du, and S. Wang, "Learning Ontology from Relational Database," *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou*, 2005.
- [76] M. Baglioni, M. V. Masserotti, C. Renso, and L. Spinsanti, "Building Geospatial Ontologies from Geographical Databases," *LNCS 4853, Proceedings of the Second International Conference on Geospatial Semantics, Mexico City*, 2007.
- [77] J. Tremols and A. Hernandez, "Mapa Genetico Digital de Suelos de Cuba, en escala 1:250 000," *Instituto de Suelos, MINAG, La Habana, Cuba*, 2001.
- [78] IGP, "Mapa Geologico de Cuba en formato digital, escala 1:100 000," *Instituto de Geologia y Paleontologia, IGP, La Habana, Cuba*, 2003.
- [79] MatLab, "MatLab R2009a," *The MathWorks. Inc.*, 2009.
- [80] R. P. W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. d. Ridder, D. M. J. Tax, and S. Verzakov, "PRTools4 A Matlab Toolbox for Pattern Recognition Version 4.1.5," *Delft*

Pattern Recognition Research, Faculty EWI - ICT, Delft University of Technology, The Netherlands, 2009, Online: <http://prtools.org> 2009.

- [81] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc. , 2010.
- [82] Internet. (2010, October). *Microsoft .NET Framework*. Available: <http://www.microsoft.com/net/>
- [83] Internet. (2010, October). *Centro de Desarrolladores de Visual C Sharp*. Available: <http://msdn.microsoft.com/es-es/vcsharp/default.aspx>
- [84] Internet. (2010, October). *Jena – A Semantic Web Framework for Java*. Available: <http://jena.sourceforge.net/>
- [85] Internet. (2010, October). *Java*. Available: <http://www.java.com/es/>
- [86] Internet. (2010, October). *IKVM.NET*. Available: <http://www.ikvm.net/>
- [87] Internet. (2010, October). *Mono*. Available: http://www.mono-project.com/Main_Page
- [88] Internet. (2010, October). *Graphviz - Graph Visualization Software*. Available: <http://www.graphviz.org/>
- [89] B. Garea, A. Curbelo, O. Jiménez, E. Garea, and G. Martin, " Informe de los resultados del proyecto Internacional SWERA en Cuba," *SWERA. Cuba Progress Report*, 2003.
- [90] Internet. (2010, Enero). *ESRI Home Page*. Available: <http://www.esri.com/>
- [91] Microsoft. (2010, September). *Microsoft Access*. Available: <http://office.microsoft.com/en-us/access/>
- [92] M. F. Lopez, A. G. Perez, and N. Juristo, "METHONTOLOGY: from Ontological Art towards Ontological Engineering," *Proceedings of the AAAI97 Spring Symposium*, pp. 33-40, 1997.
- [93] M. J. Egenhofer, "On the equivalence of topological relation," *National Center for Geographic Information and Analysis and Department of Surveying Engineering, Department of Computer Science, University of Maine*, 1994.

- [94] M. J. Egenhofer and J. R. Herring, "Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases," *Department of Surveying Engineering, University of Maine*, 1991.
- [95] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *In ECCV*, pp. pages. 404 - 417, 2006.
- [96] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. Volume 60 Issue 2, 2004.
- [97] J. Atserias, B. Casas, E. Comelles, M. González, L. Padró, and M. Padró., "FreeLing 1.3: Syntactic and semantic services in an open-source NLP library," *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, 2006.
- [98] D. Y. Liu, "A survey of content-based image retrieval with high-level semantics. ," *Pattern Recognition*, vol. 40, pp. 262- 282, 2007.
- [99] J. L. Gil, E. B. García, D. R. Ponvert–Delisles, R. Sánchez, and M. B. Vega, "Enfoques para la Clasificación Digital de Imágenes Mono y Multiespectrales y su Implementación en el Software Cubano Tn Estudio V2.0.," *III Congreso Internacional de Geomática, en Informática 2002, Ciudad de La Habana*, 2002.
- [100] L. Probasco, "Dear Dr. Use Case: What About Function Points and Use Cases?," *RationalEdge*, 2002.
- [101] IFPUG, "Frequently Asked Questions," *International Function Point Users Group*, 2000.
- [102] U. K. S. M. A, "MKII Function Points Analysis Counting Practices Manual, Versión 1.3.1," *United Kingdom Software Metrics Association*, 1998.
- [103] Internet. (2010). *El Diccionario de la Lengua Española (vigésima segunda edición) - RAE_(Abstraer)*. Available:
http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=abstraer

BIBLIOGRAFÍA CONSULTADA

Además de las referencias bibliográficas tenemos:

Aerts, K., Maesen, K., Van Rompaey, A., 2006. A practical example of semantic interoperability of large-scale topographic databases using semantic web technologies. In: Proceedings of the AGILE'06: 9th Conference on Geographic Information Science, Visegrád, Hungary, pp. 35–42.

Cali', A.,Calvanese,D.,Giacomo,G.D.,Lenzerini,M.,2002.Onthe expressive power of data integration systems. In: Proceedings of the ER'02: 21st International Conference on Conceptual Modeling. Lecture Notes in Computer Science, vol. 2503. Springer, Berlin, Heidelberg, pp.338–350.

Janowicz, K., 2005. Extending semantic similarity measurement thematic roles. In: Proceedings of the GeoS'05: First International Conference on GeoSpatial Semantics. Springer, Mexico City, Mexico, pp. 137–152.

Shariff, A., Egenhofer, M., Mark, D., 1998. Natural-language spatial relations between linear and areal objects: the topology and metric of english-language terms. *International Journal of Geographical Information Science* 12 (3), 215–245.

Cabral L., Domingue J., Motta E., Payne T.R., and Hakimpour F. 2004 Approaches to Semantic Web Services: An Overview and Comparisons. In Proceedings of the 1st European Semantic Web Symposium (ESWS2004). Heraklion, Crete, Greece.

Klien E. 2007 A rule-based strategy for the semantic annotation of geodata. *Transactions in GIS, Special Issue on the Geospatial Semantic Web*, forthcoming.

Lutz M. 2006 Ontology-Based Descriptions for Semantic Discovery and Composition of Geoprocessing Services. *GeoInformatica*, forthcoming.

Lutz M., and Klien E. 2006 Ontology-Based Retrieval of Geographic Information. *International Journal of Geographical Information Science (IJGIS)* 20(3): 233-260.

- Haarslev, V., Lutz, C., Moller, R., 1998. Foundation of spatiotemporal reasoning with description logics. In: Proceedings of the KR'98: 6th International Conference on Principles of Knowledge Representation and Reasoning. Morgan Kaufmann, Trento, Italy, pp. 112–123.
- Hill, L.L., 2000. Core elements of digital gazetteers: placenames, categories, and footprints. In: Proceedings of the ECDL '00: 4th European Conference on Research and Advanced Technology for Digital Libraries. Springer, London, UK, pp. 280–290.
- Holt, A., 2000. Understanding environmental and geographical complexities through similarity matching. *Complexity International* 7, 1–16.
- Horrocks, I., 1998. The fact system. In: Proceedings of the TABLEAUX'98: Automated Reasoning with Analytic Tableaux and Related Methods. Lecture Notes in Computer Science, vol. 1397. Springer, Berlin, pp. 307–312.
- Janowicz, K., 2005. Extending semantic similarity measurement thematic roles. In: Proceedings of the GeoS'05: First International Conference on GeoSpatial Semantics. Springer, Mexico City, Mexico, pp. 137–152.

Anexos

ANEXO 1: Glosario de Términos

<i>Abstracción</i>	Se entiende en este trabajo como la consideración aislada de las cualidades esenciales de un objeto excluyendo los detalles que no son de interés. El Diccionario de la Lengua Española (vigésima segunda edición) lo define como acción de separar por medio de una operación intelectual las cualidades de un objeto para considerarlas aisladamente o para considerar el mismo objeto en su pura esencia o noción [103].
<i>Abstracción semántica</i>	Se entiende en este trabajo como el proceso que implica reducir un fenómeno a sus componentes fundamentales para conservar sus rasgos más relevantes con el objetivo de extrapolarlos hacia un espacio semántico en el cual este esté definido según su significado.
<i>Ontologías de nivel superior (ONS)</i>	Se entiende en este trabajo como aquellas Ontologías de mayor grado de generalización, por ejemplo, una Ontología de Dominio es una Ontología de nivel superior con respecto a una Ontología de Aplicación ya que contiene términos más generales.

ANEXO 2: Glosario de Acrónimos

<i>ONS</i>	Ontologías de Niveles Superiores
------------	---

ANEXO 3: Estudio de Factibilidad del Prototipo de Sistema Gobernado por Ontologías para la Recuperación de Objetos Geoespaciales en Imágenes Satelitales “Crudas”

Para la estimación del tamaño de un sistema a partir de sus requerimientos, una de las técnicas más difundidas es el *Análisis de Puntos de Función* [100-102]. Esta técnica permite cuantificar el tamaño de un sistema en unidades independientes del lenguaje de programación, las metodologías, plataformas y/o tecnologías utilizadas.

En etapas tempranas del desarrollo de un sistema se identifican los actores y los casos de uso, y se documenta cada uno de ellos mediante una breve descripción. Aplicando el Análisis de Puntos de Función a estos casos de uso, se podrá obtener una estimación inicial del tamaño y esfuerzo en el desarrollo del software. Es importante destacar que esta estimación será imprecisa debido principalmente a la escasa información que se tiene sobre el software al principio de un proyecto, pero, permitirá obtener una idea del esfuerzo necesario para llevar adelante el mismo.

Cálculos realizados

Primeramente se procede a evaluar la complejidad de los actores con los que tendrá que interactuar el sistema. Este puntaje se calcula determinando si cada actor es una persona u otro sistema, además, evalúa la forma en la que este interactúa con el caso de uso, y la cantidad de actores de cada tipo.

- **Justificación 1.** Cálculo del Factor de peso de los actores (*UAW*).

$$UAW = \sum (\text{Cantidad de actores de un tipo} * \text{Factor})$$

$$UAW = 1 * 3$$

$$UAW = 3$$

Tabla 19. Complejidad de los actores de acuerdo a su naturaleza.

Actores	Factor	Complejidad
Especialista	3	Complejo

- **Justificación 2.** Cálculo del Factor de peso de los Casos de Uso (*UUCW*).

Tabla 20. Complejidad de los casos de uso de acuerdo al número de transacciones.

Paquete	Nº de CU	Factor de peso / CU	Complejidad
Módulo de Pre-procesamiento			
Semántico de Objetos	9	15	Complejo
Módulo de Procesamiento de			
la Consulta de Usuario	4	15	Complejo
Módulo de Detección de			
Objetos	3	15	Complejo

$$UUCW = 16 * 15$$

$$UUCW = 240$$

- **Justificación 3.** Cálculo de Puntos de Casos de Uso (*UUCP*).

$$UUCP = UAW + UUCW$$

$$UUCP = 3 + 240$$

$$UUCP = 243$$

- **Justificación 4.** Cálculo del Factor de complejidad técnica (TCF).

Este se compone de 13 puntos que evalúan la complejidad de los módulos del sistema, cada uno de estos factores tienen un peso definido (0 - 5). Para una mejor comprensión, a continuación se muestra una tabla con los ítems:

Tabla 21. Cuantificar factor de complejidad técnica.

Factor	Descripción	Peso	Valor
T_1	Sistema distribuido	2	0
T_2	Tiempo de respuesta	1	5
T_3	Usuario final	1	5
T_4	Procesamiento interno	1	5
T_5	Reutilización	1	5

T_6	Facilidad de instalación	0.5	5
T_7	Facilidad de uso	0.5	5
T_8	Portabilidad	2	5
T_9	Facilidad de cambio	1	3
T_{10}	Concurrencia	1	4
T_{11}	Objetivos especiales de seguridad	1	0
T_{12}	Acceso directo a terceras partes	1	0
T_{13}	Facilidades especiales de entrenamiento	1	2

$$TFactor = \sum_{i=1}^{13} Valor_i * Peso_i$$

$$TFactor = 44$$

$$TCF = 0.6 + (0.01 * TFactor)$$

$$TCF = 1,04$$

- **Justificación 5.** Cálculo del Factor de ambiente (EF).

Los factores sobre los cuales se realiza la evaluación son ocho puntos, que están relacionados con las habilidades y experiencia del grupo de personas involucradas con el desarrollo del proyecto. Estos factores se muestran a continuación:

Tabla 22. Cuantificar factor de ambiente según criterio de relevancia.

Factor	Descripción	Peso	Valor
E_1	Familiaridad con el modelo de proyecto utilizado	1.5	5
E_2	Experiencia con la aplicación	0.5	2
E_3	Experiencia en orientación a objetos	1	5
E_4	Capacidad del analista líder	0.5	5
E_5	Motivación	1	5
E_6	Estabilidad de los requerimientos	2	3
E_7	Personal a tiempo medio	-1	4
E_8	Dificultad del lenguaje de programación	-1	2

$$EF = 1.4 + \left(-0.03 * \sum_{i=1}^8 Valor_i * Peso_i \right)$$

$$EF = 0.47$$

- **Justificación 6.** Cálculo de Puntos de Casos de Uso Ajustados (UCP).

$$UCP = UUCP * TCF * EF$$

$$UCP = 243 * 1.04 * 0.47$$

$$UCP = 118.7784$$

- **Justificación 7.** Cálculo del Factor de conversión (CF).

Según Karner cada Punto de Casos de Uso requiere 20 horas – hombre.

Posteriormente esta propuesta se ha refinado tomando el siguiente criterio:

- Se contabiliza cantidad de factores ambientales con una puntuación por debajo de 3, para los factores de E_1 a E_6 .
- Se contabiliza cantidad de factores ambientales con una puntuación por encima de 3, para los factores E_7 y E_8 .
- Se evalúa el resultado o la cantidad total según la siguiente tabla:

Tabla 23. Cantidad de horas-persona según el valor.

Horas-Persona (CF)	Descripción
20	Si el valor es ≤ 2
28	Si el valor es ≤ 4
36	Si el valor es ≥ 5

Efectuando los planteamientos antes descritos se considera un Factor de conversión de **CF = 28 horas – hombre.**

- **Justificación 8.** Cálculo del Esfuerzo (E).

$$E = UCP * CF$$

$$E = 118.7784 * 28$$

$$E = 3325.7952$$

- **Justificación 9.** Cálculo del Tiempo de Desarrollo en Horas (CTD).

- **CH:** Cantidad de hombres desarrollando el sistema.

$$CTD = E / CH$$

$$CTD = 3325.7952 / 2$$

$$CTD = 1662.8976 \text{ Hrs.}$$

- **Justificación 10.** Cálculo del Costo total del proyecto (CTP).

- **THM:** Total de horas mensuales en el desarrollo del proyecto.

Suponiendo que se trabajen 24 días al mes y 8 horas diarias

- $THM = 24 * 8$

- $THM = 192$

- **SP:** Salario promedio de las personas que trabajan en el proyecto.

Suponiendo que el salario estándar de un trabajador es \$390.00

- **SP = (390.00) * 2/2 = \$390.00**
- **THP:** Tarifa horaria promedio.
 - $THP = SP / THM$
 - $THP = 390.00 / 192$
 - **THP = 2.03125 \$/Hrs.**
- **CTP:** Costo total del proyecto.
 - **K:** Coeficiente que tiene en cuenta los costos indirectos, puede tomar valores de 1.5 ó 2.0, para este caso, suponemos un valor de 2.0.
 - $CTP = E * K * THP$
 - $CTP = 3325.7952 * 2 * 2.03125$
 - **CTP = \$13511.043**

Tabla 24. Cálculos Efectuados.

Cálculos	Valor	Justificación
UAW	3	Ver Justificación 1
UUCW	240	Ver Justificación 2
UUCP	243	Ver Justificación 3
TCF	1.04	Ver Justificación 4
EF	0.47	Ver Justificación 5
UCP	118.7784	Ver Justificación 6
CF	28	Ver Justificación 7
Esfuerzo(E)	3325.7952	Ver Justificación 8
	1662.8976 hrs.	
Tiempo de desarrollo	(8.7 meses)	Ver Justificación 9
Costo	\$13511.043	Ver Justificación 10

RT_053, octubre 2013

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2013

Editor: Lic. Lucía González Bayona

Diseño de Portada: Di. Alejandro Pérez Abraham

RNPS No. 2142

ISSN 2072-6287

Indicaciones para los Autores:

Seguir la plantilla que aparece en www.cenatav.co.cu

C E N A T A V

7ma. A No. 21406 e/214 y 216, Rpto. Siboney, Playa;

La Habana. Cuba. C.P. 12200

Impreso en Cuba

