

Línea de Investigación: Integración semántica de Información Geográfica

†María Mercedes Vitturini, †Pablo Fillottrani*

†Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (LISSI)
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Av. Alem 1253 - (8000) Bahía Blanca - Argentina
Tel/Fax: (+54) 291-4595135/6
E-mail: mvitturi@cs.uns.edu.ar
E-mail: prf@cs.uns.edu.ar

Resumen

Muchas aplicaciones se ven beneficiadas con el uso de Información Geográfica (IG). En algunos casos, sus propias bases de IG cuentan con datos acumulados y con distinto formato estructural y significado semántico. Esta información es una fuente de conocimiento que interesa compartir y relacionar. En un sentido estricto, cualquier Sistema de Información Geográfica (SIG) puede definir su ontología de aplicación O_A desde el momento que le asigna un significado a los símbolos que usa. El problema está en que existen muchas O_A 's independientes, para dominios superpuestos y con distinta semántica. Una propuesta que permite integrar IG heterogénea es relacionar a través de asociaciones semánticas a los conceptos de las distintas O_A 's y con los conceptos de una ontología de dominio O_D compartida.

Palabras clave: Integración de Información Geográfica, Ontologías, SIG

Contexto

En este trabajo se presenta una línea de investigación del proyecto *Integración de Información y Servicios en la Web* bajo la dirección del Dr. Pablo R. Fillottrani. El mismo se encuadra en las actividades de investigación del Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (LISSI) del Departamento de Ciencias e Ingeniería de la Computación y está financiado por la Universidad Nacional del Sur y aprobado por la Secretaría

de Ciencia y Tecnología.

Motivación

Las aplicaciones SIG comienzan a utilizarse en la década del '70, con representaciones heterogéneas para la IG. En la década del '90 impulsadas por las exigencias de las tecnologías emergentes: Infraestructuras de Datos Espaciales (IDE)¹ y Sistemas Móviles Basados en Ubicación (LBMS)², organizaciones internacionales comienzan a trabajar fuertemente en la definición de estándares para sistemas geográficos [2, 1, 4]. Finalmente, la tendencia actual es enriquecer semánticamente la información y analizar su impacto en cuanto a capacidad de integración a través del uso de ontologías [10, 11, 18, 17], en este caso particular, relacionadas con el ámbito de aplicaciones de IG. La figura 1 esquematiza la relación entre la definición de estándares y el uso de ontologías para SIG como mediadores en la integración de IG.

Las calidades "interoperabilidad" e "integración" en varias oportunidades se utilizan indistintamente, aunque en un sentido estricto se las puede distinguir[8]. En el área de Ingeniería de Software, la calidad de *interoperabilidad* se relaciona con la capacidad de un sistema para integrar componentes de software que provienen de diferentes desarrolladores. Esto permite a un nuevo producto ingresar al mercado a competir con otros y ofrecer software integrable con otros módulos. Desde el modelo de datos, la calidad de *integración* se refiere a la posibilidad de compartir información, esto es datos y su significado. Cuando

*Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC).

¹En inglés Spatial Data Infrastructure (SDI)

²En inglés Local Based Mobile System (LBMS)

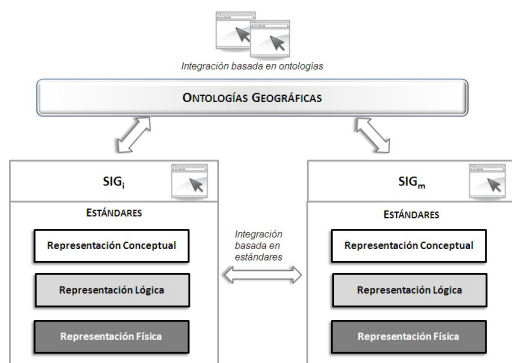


Figura 1: Integración de IG

la información es producida y gestionada localmente, en general, resulta precisa en el ámbito de su comunidad de uso. Sin embargo, cuando la misma ha sido producida por otros esto deja de ser cierto, y se requiere de conocimiento y de reglas de transformación para integrar información “heterogénea” de fuentes no relacionadas.

Interoperabilidad e integración son calidades que ganan importancia en el campo de los productos tecnológicos con manejo de IG. En la última década ha aumentado la exigencia por interoperabilidad e integración en distintos ámbitos de aplicación: SIG, IDE y LBMS. La integración para las aplicaciones que trabajan con IG presenta una dificultad adicional: al momento de intercambiar información no solo aparecen diferencias a nivel de modelos del dominio, sino también en los modelos lógicos y físicos.

Los problemas de *heterogeneidad en IG* se clasifican en niveles o vistas que impactan de distinta forma en cuanto a las dificultades de integración. La vista *heterogeneidad sintáctica* considera que distintas soluciones pueden ofrecer diferentes formatos y modelos de datos (vector y mosaico) y hasta distintos sistemas de coordenadas para representar fenómenos idénticos. Se considera como nivel de *heterogeneidad estructural* en IG cuando los modelos de distintas soluciones representan los mismos fenómenos con distintos atributos (estructuras distintas). El nivel de *heterogeneidad semántica* en IG se establece cuando distintas soluciones interpretan significados distintos para los mismos fenómenos del mundo a modelar. El cuadro 1 ejemplifica tales niveles de heterogeneidad. Las dificultades de heterogeneidad sintáctica son más simples de resolver usando sistemas de transformación. Una solución parcial a los problemas de heterogeneidad estructural y semántica es contar con tesauros que normalicen los términos del dominio de aplicación. La integración

Ejemplos		
Heterogeneidad	Sistema S_1	Sistema S_2
<i>sintáctica</i>	S_1 representa zonas por densidad de población de Bahía Blanca bajo el modelo vector.	S_2 representa zonas por densidad de población de Bahía Blanca bajo el modelo mosaico.
<i>estructural</i>	S_1 representa zonas por densidad de población de Bahía Blanca, incluyendo detalle de los servicios públicos.	S_2 representa zonas por densidad de población de Bahía Blanca, incluyendo detalles sobre los tipos de construcción (edificios, casas, barrios cerrados, etc.).
<i>semántica</i>	S_1 representa densidad de población de Bahía Blanca considerando número de habitantes.	S_2 representa densidad de población de Bahía Blanca considerando unidades familiares.

Cuadro 1: Ejemplos de heterogeneidad entre SIG

semántica de IG motiva importantes investigaciones en el área de Ciencias de la Computación. Las soluciones a problemas de integración semántica generalmente se vinculan con la definición de ontologías que incluyan especificaciones formales para los modelos de representación de cada solución.

Líneas de Investigación y Desarrollo

La presente línea da continuidad a la investigación realizada en el ámbito de formación de posgrado para obtener el título de Magister en Ciencias de la Computación y como parte de la actividad desarrollada en el LISSI.

Ontologías para Información Geográfica

Las ontologías se encargan de la especificación formal de *símbolos* o *términos* de forma tal que dicha in-

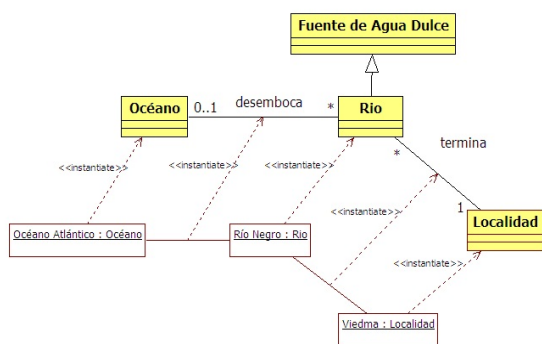


Figura 2: Ejemplo de Modelado de Conceptos y Relaciones

interpretación sea única. Esta afirmación es válida tanto entre personas así como también si participan en la comunicación agentes automáticos. En la figura 2 se presenta un ejemplo simple de especificación usando UML[16]. Los conceptos están representados por rectángulos. Entre los conceptos se distinguen las *clases* (en amarillo), que definen las propiedades comunes a los elementos del mismo tipo y las *instancias* (en blanco) que representan a un concepto concreto, ejemplar o “instancia”. Se puede observar una relación de jerarquía entre las clases FUENTE DE AGUA DULCE y RÍO por la asociación distinguida *es_un*, que define a RÍO como un tipo de FUENTE DE AGUA DULCE. El uso de la asociación *es_un* refleja la relación entre un concepto especializado y su padre. Otras asociaciones definidas en el modelo del ejemplo son *desemboca* que especifica que un RIO *desemboca* en el OCEÁNO y *termina* para representar que RÍO *termina* en una LOCALIDAD. Las instancias de clases y asociaciones están representados por la dependencia estereotipada «*instanciate*». Con esta simple definición es posible obtener algunas conclusiones básicas, como por ejemplo:

- El Río Negro es un RÍO.
- El Atlántico es un OCEANO.
- Viedma es una LOCALIDAD.
- El Río Negro *desemboca* en el Atlántico.
- El Río Negro *termina* Viedma.

Y otras conclusiones implícitas más elaboradas:

- El Río Negro es *una* FUENTE DE AGUA DULCE.
- Viedma *está proxima* al Atlántico.

- Viedma *cuenta con una* FUENTE DE AGUA DULCE.

Lo atractivo de esta representación de conocimiento es que las mismas conclusiones podría obtenerlas un agente automático. Existen diversas ontologías y para diferentes dominios. Entre ellas difieren en el nivel de detalle que expresan. Un indicador de la complejidad de una ontología es el conjunto de relaciones conceptuales que admite.

Partiendo de que una ontología es una especificación sobre una conceptualización compartida, expertos del dominio, usuarios y diseñadores acuerdan el conocimiento especificado en la ontología para que la misma sea *usable*, al menos por los miembros de la comunidad. Naturalmente este acuerdo es difícil de lograr. Una forma de abordar a una solución es dividir el conocimiento compartido en distintas ontologías, según su alcance o nivel de generalidad. Distintos autores proponen formas levemente diferente en cuanto a la forma de nombrar y clasificar ontologías según el nivel de generalidad [13, 11, 17]. En general se puede afirmar que distinguen los siguientes niveles de generalidad:

- *Ontologías genéricas (Top-Level)*, capturan conocimiento de propósito general, en general independiente del dominio de aplicación, por ejemplo espacio, el tiempo, evento, acción, etc. En teoría se espera que estas ontologías se adopten por una amplia comunidad de usuarios.
- *Ontologías de dominio y ontologías de tarea*, definen el conocimiento usado por un dominio (medicina, geografía) o por una actividad específica (comercio) describiendo su vocabulario, mediante la especialización de los términos introducidos en la ontología de alto nivel.
- *Ontologías de aplicación*, capturan el conocimiento necesario de un sistema o aplicación específica. Describen conceptos que dependen tanto del dominio como de la actividad.

Trabajos Futuros

De las investigaciones previas se tiene conocimiento sobre los modelos de representación de IG, características y requisitos de los SIG, tecnologías existentes, arquitecturas y tipos de heterogeneidades relacionadas con IG que deben enfrentarse como parte del proceso de integración [19, 3]. La construcción de *ontologías de dominio* necesita considerar la definición

de un esquema conceptual general para uno o varios dominios de uso. Durante el diseño de ontologías de dominios, las relaciones o roles entre conceptos tienen un significado relevante y deberían ser usados en la descripción semántica siempre que fuera posible. Esto es, se deberían incluir vínculos taxonómicos y no taxonómicos entre conceptos. De esta forma se pueden proveer significados conceptuales más flexibles y precisos que los que tienen con definiciones jerárquicas estáticas.

La propuesta de integración con *ontologías híbridas* [18] trabaja definiendo al menos dos niveles de ontologías: ontología de aplicación (O_A) y ontología de dominio (O_D). El nivel O_D provee la caracterización formal de las entidades de dominio acordadas. La ontología O_A define el significado del esquema de aplicación de una solución. Finalmente, se establece la asociación o *anotación semántica* entre los conceptos de O_D y O_A . De esta forma pueden coexistir distintas aplicaciones, cada una con su propia ontología O_{A_i} y anotaciones semánticas hacia la ontología O_D compartida que sirve de mediadora (figura 3). La integración basada en ontologías híbridas define asociaciones semánticas entre la ontología O_D y los fenómenos de una cada solución SIG particular. Esta organización construye semánticas complejas, sin perder flexibilidad si se desean incorporar nuevas fuentes de información.

El siguiente ejemplo muestra una aplicación de esta propuesta. Supongamos una ontología O_D que define los conceptos $RÍO_D$, $ESPECIE_D$ y $REGIÓN_D$ y establece entre ellos las relaciones “ $RÍO$ se crían $ESPECIE$ ” y “ $RÍO$ pertenece $REGIÓN$ ”. Sea A_1 una aplicación que tiene información sobre pesca en Argentina y cuenta con el conocimiento “*en el Río Paraná se práctica pesca deportiva del Dorado*”. Por otro lado, sea la aplicación A_2 con información turística de Argentina que posee el conocimiento “*El Litoral es una región turística de Argentina que está próxima al Río Paraná*”. Definiendo las siguientes relaciones:

- En A_1 *Río Paraná* como instancia del concepto de aplicación $RÍO_{A1}$.
- En A_1 *Dorado* como instancia del concepto de aplicación PEZ_{A1} .
- En A_1 la relación $RÍO_{A1}$ *pesca deportiva de* PEZ_{A1} .
- En A_2 *Litoral* como instancia del concepto de aplicación $REGIÓN_{TURISTICA_{A2}}$.

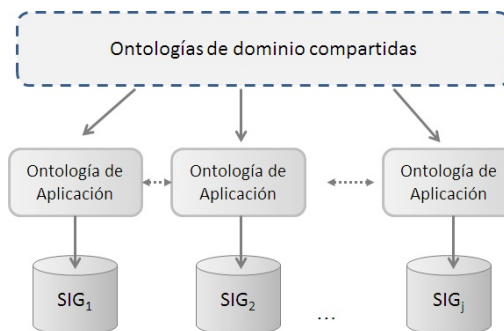


Figura 3: Integración entre aplicaciones usando ontología híbrida

- En A_2 *Río Paraná* como instancia del concepto de aplicación $RÍO_{A2}$.
- En A_2 la relación $RÍO_{A2}$ *próximo a* $REGIÓN_{TURISTICA_{A2}}$.
- Entre A_1 y O_D la anotación semántica $RÍO_{A1}$ *corresponde con* $RÍO_{OD}$.
- Entre A_1 y O_D la anotación semántica PEZ_{A1} *corresponde con* $ESPECIE_{OD}$.
- Entre A_2 y O_D la anotación semántica $REGIÓN_{TURISTICA_{A2}}$ *corresponde con* $REGIÓN_{OD}$.
- Entre A_2 y O_D la anotación semántica $RÍO_{A2}$ *corresponde con* $RÍO_{OD}$.

De esta integración es posible deducir el conocimiento “*En la región turística del Litoral se práctica pesca deportiva del Dorado*”. Este conocimiento actualmente no está en las aplicaciones A_1 , A_2 y sin embargo se creó a partir de la integración. Además al mismo razonamiento podría llegar un agente automático si cuenta con la información recientemente enumerada.

El resultado planteado en el ejemplo es interesante, pero para llegar al estado de información ideal quedan temas por investigar. Por un lado, ontologías de dominio geográfico existentes. El dominio geográfico en sí mismo es amplio y es posible que los conceptos de una aplicación no estén restringidos a una única ontología de dominio. Esto plantea nuevos desafíos: qué restricciones se deben satisfacer al considerar más de una ontología de dominio para no dar lugar a contradicciones, terminos sobredefinidos y relaciones de un conceptos a más de una O_D [7, 15, 14].

La integración basada en ontología híbrida considera vincular semánticamente los conceptos de la aplicación con los conceptos de la ontología de dominio.

En Ingeniería de Software la conceptualización juega un rol primordial en las etapas iniciales del desarrollo de un sistema y plantea las bases para la reutilización e integración. Así, aunque la propuesta de ontologías híbridas[18] no determina el momento para establecer la asociación entre la aplicación SIG y la ontología de dominio, empíricamente afirmamos que cuanto antes se definan estas relaciones semánticas (por ejemplo en la actividad de *análisis del problema*), menor será el costo y mejor será el resultado. Queda propuesto como trabajo de investigación definir un modelo de proceso para SIG que considere la actividad de asociación semántica con la ontología de dominio como parte del proceso de desarrollo [6, 12].

Transversal a los puntos anteriores está la selección de la/s ontología/s O_D semánticamente cohesivas con el problema. Se planteó una técnica empírica para medir la interrelación entre una aplicación y una ontología de dominio. Esta técnica está basada en experiencias en desarrollo y migración de sistemas de información. Un desafío es formalizar la propuesta para contar con un mecanismo que permita valorar nominalmente la calidad de cohesión entre la ontología de dominio y la ontología de aplicación [5, 9].

Formación de Recursos Humanos

Los temas de investigación planteados forman parte del trabajo de tesis para obtener el título de Magister en Ciencias de la Computación. Temas relacionados con este proyecto se trabajan en una tesis de licenciatura.

Referencias

- [1] ISO/TC 211 Geographic information/Geomatics. Website, 1994. <http://www.isotc211.org/>.
- [2] Open Geospatial Consortium, Inc. (OGC). Website, 1994. <http://www.opengeospatial.org/>.
- [3] P. Burrough and R. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, 1998.
- [4] INSPIRE CEN. Cen tc 287 geographic information. inspire, 2007.
- [5] Mariano Fernández-López and Asunción Gómez-Pérez. Overview and analysis of methodologies for building ontologies. *Knowl. Eng. Rev.*, 17:129–156, June 2002.
- [6] Mariano Fernandez-Lopez, Asuncion Gomez-Perez, and Natalia Juristo. Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium*, pages 33–40, Stanford, USA, March 1997.
- [7] F. Fonseca, Max J. Egenhofer, P. Agouris, C. Cmara, Frederico Fonseca, Max J. Egenhofer, Peggy Agouris, and Gilberto Câmara. Using ontologies for integrated geographic information systems, 2002.
- [8] Carlo Ghezzi, Mehdi Jazayeri, and Dino Mandrioli. *Fundamentals of Software Engineering*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 2002.
- [9] Asunción Gomez-Perez. Some ideas and examples to evaluate ontologies, 1994.
- [10] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993.
- [11] Nicola Guarino. Formal ontology and information systems. pages 3–15, 1998.
- [12] Wolfgang Hesse. Ontologies in the software engineering process. In *In (R. Lenz et al., Ed.): EAI 2005 - Proceedings of the Workshop on Enterprise Application Integration*, 2005.
- [13] Vakgroep Informatica, En Toegepaste Informatica, R.A. Meersman, and R. A. Meersman. Semantic ontology tools in is design, 1999.
- [14] Margarita Kokla and Marinos Kavouras. Fusion of top-level and geographical domain ontologies based on context formation and complementarity. *International Journal of Geographical Information Science*, 15(7):679–687, 2001.
- [15] Hardy Pundt and Yaser Bishr. Domain ontologies for data sharing-an example from environmental monitoring using field gis. *Comput. Geosci.*, 28:95–102, February 2002.
- [16] James Rumbaugh, Ivar Jacobson, and Grady Booch. *Unified Modeling Language Reference Manual, The (2nd Edition)*. Pearson Higher Education, 2004.
- [17] Mike Uschold. Knowledge level modelling: concepts and terminology. *Knowl. Eng. Rev.*, 13(1):5–29, 1998.
- [18] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner. Ontology-based integration of information - a survey of existing approaches. pages 108–117, 2001.
- [19] M. Zeiler. *Modeling our World. The ESRI Guide to Geodatabase Design*. Environmental Systems Research Institute, Inc, 1999.