

Integración de Sistemas de Información Geográfica

Agustina Buccella¹, Alejandra Cechich¹, and Pablo Fillottrani²

¹ GIISCO Research Group
Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Neuquen, Argentina

² Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Bahía Blanca, Argentina

Resumen Hoy en día gran cantidad de entidades en el mundo están utilizando la información geográfica como base para sus sistemas de información. Este gran crecimiento trajo consigo la necesidad de compartir ésta información que se encuentra tanto en sistemas privados como en públicos (por ejemplo, Internet). Pero la integración de estos sistemas no es una tarea trivial, hay varios aspectos que se deben analizar de manera de conocer los posibles problemas que la heterogeneidad provoca. En este trabajo describimos el marco conceptual en el cual se ubican estos sistemas de información geográfica describiendo los aspectos fundamentales que se deben tener en cuenta al iniciar un proceso de integración.

1. Introducción

El avance de la tecnología en cuanto a dispositivos de colección de datos más sofisticados y el desarrollo de gran cantidad de nuevos sistemas de información espaciales han generado un alto volumen de datos sobre nuestra tierra. Por ejemplo, hoy en día el uso de dispositivos GPS es tan común que están disponibles incluso en los relojes de pulsera.

Toda esa información se captura y almacena dentro de diferentes sistemas de información con diferentes niveles de detalle. A su vez, con el crecimiento de la tecnología que incluye al área de las redes de comunicación, los mismos son también distribuidos. Este escenario de distribución y heterogeneidad nos lleva a pensar que para una buena explotación de todos o de gran cantidad de los datos geográficos existentes necesitaremos realizar esfuerzos en lo que se refiere a la *integración*.

La *integración de los datos*, sin importar el tipo de los sistemas a integrar presenta una serie de problemas que en general son complejos de resolver. El concepto de *integración* en sí mismo involucra una serie de decisiones que se deben tomar en forma correcta para lograr un resultado consistente. Cuando hablamos de consistencia nos referimos a que la respuesta que se le brinda a un usuario, luego de efectuar una consulta, debe ser coherente y por lo tanto

satisfacer su necesidad. Para esto, el sistema debe recuperar la información de las fuentes de información relacionadas con la consulta que están disponibles en ese momento.

Aproximadamente desde hace una década se ha introducido el concepto de ontología [5] como herramienta para solucionar los problemas principales de la integración. Las ontologías fueron introducidas por Gruber [6] como una “especificación formal y explícita de una conceptualización compartida”. El término *conceptualización* se refiere a un modelo abstracto de cómo el ser humano piensa comúnmente cosas del mundo real, como por ejemplo, *una mesa*. El término *especificación explícita* se refiere a que se ha dado un nombre y una definición a los conceptos y relaciones creados en el modelo abstracto. *Formal* se refiere al hecho de que la ontología debe ser entendida por una computadora (machine readable) [3,21]. Y *compartida* refleja la noción de que una ontología captura conocimiento consensuado, es decir, es aceptada por un grupo o comunidad.

Así surgieron una gran cantidad de propuestas diferentes con mayor o menor nivel de complejidad. Por ejemplo, las propuestas de [16,15,18,20] definen posibles enfoques de integración de GIS basados en ontologías. En general, los problemas de integración que surgen en los GIS son los mismos que en sistemas convencionales además de aquellos que sólo se aplican a este tipo de sistemas debido a la naturaleza de los mismos. Es decir, debido a la naturaleza de la información geográfica. Por ejemplo, uno de los problemas más comunes es el de la heterogeneidad, la cual genera diferentes incompatibilidades como por ejemplo:

- *Heterogeneidad en el modelo conceptual*: Una calle se representa como una clase objeto en un sistema y como una relación en otro.
- *Heterogeneidad en el modelo espacial*: Las calles pueden representarse como polígonos (o un segmento de pixels) en un sistema y como líneas en otro.
- *Heterogeneidad de estructura o esquema*: Dos sistemas poseen el nombre de la calle pero uno posee información sobre la vereda y el otro no.
- *Heterogeneidad semántica*: Un sistema considera la parte pavimentada como la calle mientras que otro considera la calle pavimentada mas la vereda como la calle. Uno puede definir una calle como una vía pavimentada donde transitan autos con una vereda mientras que otro sistema define la calle como cualquier tipo de via usada por autos dentro de un área residencial.

Cada una de estas heterogeneidades tiene un tratamiento diferente y existen varias propuestas enfocadas en uno o más de ellos.

En la sección siguiente se describen los aspectos principales que se deben analizar cuando se inicia un proceso de integración de información geográfica.

2. Integración de Información Geográfica

De los trabajos recientes y mas referenciadas en la literatura [20,12,17,18,8,9,4,11,19,1] que han propuesto la integración de información geográfica utilizando ontologías podemos extraer una serie de aspectos que se deben tener en cuenta a la hora de iniciar un proceso de estas características:

- *Representación de la información geográfica:* La representación de la información geográfica es uno de los principales aspectos a tener en cuenta. Como ejemplo, podemos citar la hipótesis analizada por Mark et al. [14] la cual expresa que “las entidades geográficas y no geográficas son diferentes en un número de maneras”. Sus experimentos tienen como objetivo comprobar el grado de acuerdo para el cual las personas codifican el dominio geográfico a nivel conceptual. Como conclusión de su estudio, encontraron que existe un conjunto de términos geográficos que poseen una mayor frecuencia, es decir, son términos recurrentes. De esta manera, éstos términos podrían ser utilizados para simplificar el proceso de integración y a su vez para brindar mayor expresividad a las ontologías a nivel conceptual.
- *Representación formal de las ontologías:* Las ontologías consisten de *axiomas lógicos* que expresan el significado de los términos para una comunidad en particular. Es la lógica precisamente la que se utiliza para formalizar y manipular la semántica y las ontologías. Los dos formalismos lógicos más usados son las Lógicas para la Descripción (Description Logics DL) [2] y las Lógicas basadas en Marcos (Frame-based Logic FLogic) [13]. Una comparación exhaustiva respecto a estos dos lenguajes se encuentra en [8]. Como conclusión, las DL ofrecen mayor poder expresivo para formalizar relaciones y conceptos, en cambio FLogic se centra en la definición de estos últimos, en vez de relaciones. Como ventaja principal del uso de la lógica en la formalización (ya sea utilizando cualquiera de los dos lenguajes mencionados anteriormente) es el uso de razonadores tal como FACT [10] y RACER [7]. Estos poseen la capacidad de realizar inferencias para determinar por ejemplo relaciones de generalización/especialización entre conceptos y relaciones, reconocimiento de individuos, detección de inconsistencias (en algunos casos) y fortalecimiento de restricciones de cardinalidad [8]. De esta manera el proceso de integración de ontologías puede realizarse de forma semi-automática aprovechando éstas ventajas que provee el uso de las mismas.
- *El proceso de integración:* En general, los procesos de integración propuestos en la literatura proveen un conjunto de pasos para encontrar correspondencias útiles entre las fuentes de información. En particular cuando poseemos fuentes de información geográfica estos procesos varían ya que se deben considerar nuevos conceptos como objetos geográficos, granularidad, campos, relaciones cuantitativas y cualitativas, variaciones espacio-temporales, escalas, etc. Por ejemplo, si no se aplican mecanismos para analizar la granularidad [4] dos ontologías no poseerán elementos en común. Así, los procesos de integración deben adicionar técnicas que implementen soluciones para estas nuevas características.

Todos estos aspectos se deben analizar previamente de manera de que el proceso de integración que se inicie pueda evitar problemas comunes y realizar una definición de las ontologías correcta.

3. Conclusiones

En este trabajo se describieron los aspectos principales que deben considerarse al iniciar un proceso de integración de sistemas de información geográfica. Entre ellos se destaca el uso de ontologías formales como herramienta fundamental para mejorar y acelerar tanto el proceso en si mismo, como también la consistencia y veracidad de los resultados. A su vez, se analizaron las razones por las cuales la información geográfica necesita de una atención especial y de como las propuestas deben proveer nuevos mecanismos para hacerles frente. Igualmente todos estos aspectos dependerán en última instancia de las ontologías fuente. En la literatura existen miles de trabajos que utilizan las ontologías para diferentes problemas. Pero, aspectos como la evaluación de si están representando realmente el mundo real, o aspectos como calidad o completitud no son debidamente tenidos en cuenta. Un proceso de integración fallará o no proveerá resultados útiles si las ontologías en las cuales estan basados son de baja calidad.

Como trabajo futuro se propone continuar con este análisis de los aspectos principales a considerar antes de iniciar un proceso de integración de información geográfica. Es decir, se estudiarán las diferentes arquitecturas y procesos de integración utilizando ontologías propuestos en la literatura. Una vez que este estudio finalice, se propondrá la contrucción de una arquitectura que explote las ventajas y beneficios del uso de las ontologías y a su vez tenga en cuenta las particularidades de los sistemas de información geográficos. Luego partiendo de la misma se propondrá un proceso con sus métodos y técnicas que realicen la integración. Por último se deberá validar la viabilidad de la propuesta.

Referencias

1. K. Aerts, K. Maesen, and A. van Rompaey. 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*, pages 35–42, Visegrád, Hungary, 2006.
2. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook - Theory, Implementation and Applications*. Cambridge University Press, 2003.
3. D. Fensel. *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, 2nd edition edition, Berlin 2003.
4. F. Fonseca. *Ontology-driven Geographic Information Systems*. PhD thesis, University of Maine, 2001.
5. T. Gruber. Ontolingua: A mechanism to support portable ontologies. Technical Report KSL 91-66, Knowledge Systems Laboratory, Stanford University, Stanford, 1992.
6. T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
7. V. Haarslev and R. Moller. Racer system description. In P. Lambrix, A. Borgida, M. Lenzerini, R. Moller, and P. Patel-Schneider, editors, *Proceedings of the CEUR-WS International Workshop on Description Logics*, number 22, Linköping, Sweden, August 1999.

8. F. Hakimpour. *Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata*. PhD thesis, Zurich University, 2003.
9. G.Ñ. Hess and C. Iochpe. Ontology-driven resolution of semantic heterogeneities in gdb conceptual schemas. In *Proceedings of the GEOINFO'04: VI Brazilian Symposium on GeoInformatics*, pages 247–263, Campos do Jordão, 2004.
10. I. Horrocks. The fact system. In H. de Swart, editor, *Automated Reasoning with Analytic Tableaux and Related Methods: International Conference Tableaux*, number 1397, pages 307–312, Berlin, May 1998. Lecture Notes in Artificial Intelligence, Springer-Verlag.
11. K. Janowicz. Extending semantic similarity measurement by thematic roles. In *Proceedings of the GeoS'05: First International Conference on GeoSpatial Semantics*, pages 137–152, Mexico City, Mexico, 2005. Springer Verlag.
12. M. Kavouras, M. Kokla M., and Tomai E. Comparing categories among geographic ontologies. *Computers & Geosciences, Special Issue, Geospatial Research in Europe: AGILE 2003*, 31(2):145–154, March 2003.
13. Michael Kifer, Georg Lausen, and James Wu. Logical foundations of object-oriented and frame-based languages. *J. ACM*, 42(4):741–843, 1995.
14. Mark D. M., A. Skupin, and B. Smith. Features, objects, and other things: Ontological distinctions in the geographic domain. In Lecture Notes in Computer Science, editor, *Spatial Information Theory: Foundations of Geographic Information Science*, 2205, pages 488–502, Berlin, 2001. Berlin: Springer-Verlag.
15. M.A. Rodríguez and M.J. Egenhofer. Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2):442–456, 2003.
16. M.A. Rodríguez and M.J. Egenhofer. Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science*, 18(3):229–256, 2004.
17. A. Schwering and M. Raubal. Spatial relations for semantic similarity measurement. In Lecture Notes in Computer Science, editor, *Proceedings of the ER'05 Perspectives in Conceptual Modeling*, 3770, pages 259–269. Springer Berlin/Heidelberg, 2005.
18. 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.
19. L. Stoimenov, A. Stanimirovic, and S. Djordjevic-Kajan. Discovering mappings between ontologies in semantic integration process. In *Proceedings of the AGILE'06: 9th Conference on Geographic Information Science*, pages 213–219, Visegrád, Hungary, 2006.
20. U. Visser. *Intelligent Information Integration for the Semantic Web*, volume 3159 of *Lecture Notes in Computer Science*. Springer Berlin - Heidelberg, 2004.
21. U. Visser and C. Schlieder. Modelling with ontologies. In *Proceedings of the Ontology and Modeling of Real Estate Transactions in European Juristictions*, Ashgate, 2002.