

Explotación de Información Geográfica Basada en Integración de Ambientes de Trabajo

García-Martínez, R., Merlino, H., Rodríguez, D., Martins, S., Baldizzoni, E., Diez, E.,
Amatriain, H., Ribeiro, F., Segura, A., Santamaría, P., Mieres, F., Aguirre, D.

Grupo de Investigación en Sistemas de Información (UNLa GISI)
Licenciatura en Sistemas. Universidad Nacional de Lanús, Argentina
<http://www.unla.edu.ar/sistemas/gisi/>
rgm1960@yahoo.com, hmerlino@gmail.com

Resumen. Los sistemas de información geográfica son susceptibles de ser tratados como bases de datos que contiene información geográfica de utilidad vinculada a un determinado territorio de interés. Desde esta visión, pueden ser sujetos de búsqueda de patrones y regularidades significativas utilizando ingeniería de explotación de información. En esta comunicación, se plantea un ambiente de trabajo integrado que permite desarrollar procesos de explotación de información sobre bases de datos georeferenciados.

1. Introducción

La Explotación de Información es la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Mobasher et al., 1999; Srivastava et al., 2000; Abraham, 2003; Coley, 2003]. Ha sido definido como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información [Fayad et al., 1996; Grossman et al., 1998].

Un Proceso de Explotación de Información se define, como un grupo de tareas relacionadas lógicamente [Curtis et al., 1992] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del Aprendizaje Automático [García-Martínez, 1997; García-Martínez et al., 2003], susceptibles de ser utilizadas en cada uno de estos procesos.

Las técnicas de minería de datos del área de sistemas inteligentes más utilizadas son: los algoritmos TDIDT, los mapas auto organizados (SOM) y las redes bayesianas. Los algoritmos TDIDT permiten el desarrollo de descripciones simbólicas de los datos para diferenciar entre distintas clases [Quinlan, 1990]. Los mapas auto organizados pueden ser aplicados a la construcción de particiones de grandes masas de información, tienen la ventaja de ser tolerantes al ruido y la capacidad de extender la generalización al momento de necesitar manipular datos nuevos [Kohonen, 1995].

Las redes bayesianas pueden ser aplicadas para identificar atributos discriminantes en grandes masas de información, detectar patrones de comportamiento en análisis de series temporales. [Heckerman et al., 1995].

Un Sistema de Información Geográfica (SIG o GIS, en su acrónimo inglés Geographic Information System) es una integración organizada de hardware, software y datos geográficos [Moreno et al., 2012; Wikipedia, 2013] diseñada para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión geográfica. También puede definirse como un modelo de una parte de la realidad referido a un sistema de coordenadas terrestre y construido para satisfacer unas necesidades concretas de información. En el sentido más estricto, es cualquier sistema de información capaz de integrar, almacenar, editar, analizar, compartir y mostrar la información geográficamente referenciada. En un sentido más genérico, los SIG son herramientas que permiten a los usuarios crear consultas interactivas, analizar la información espacial, editar datos, mapas y presentar los resultados de todas estas operaciones.

Se han utilizado técnicas de minería de datos para la identificación de patrones en diversas aplicaciones de los sistemas de información geográfica que van desde: la determinación de la salinidad de suelos [Eklund, 1998], el análisis de zonas con alto índice de criminalidad [Estivill-Castro y Lee, 2001], el desarrollo urbano y ocupación territorial [Mennis y Liu, 2005], la predicción de incendios en bosques [Stojanova et al., 2006]; hasta el estudio de migraciones humanas [Spielman y Thill, 2008].

Referentes de la comunidad académica han señalado la necesidad de disponer de procesos de explotación de información georeferenciada [Verma y Nashine, 2012] que permitan obtener conocimiento a partir de información digital disponible [Roddick y Lees, 2001], su caracterización [Shekhar et al., 2003] y tecnologías involucradas [Ng y Han, 2002].

En este trabajo se reseñan los procesos de explotación de información (sección 2), como problema se identifica la necesidad de disponer de un ambiente de trabajo para explotación de información georeferenciada (sección 3), se propone una solución al problema planteado (sección 4), y se describe un caso de estudio donde se utiliza el ambiente integrado construido (sección 5).

2. Procesos de Explotación de Información

Britos [2008] propuso los siguientes procesos de explotación de información: descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos, descubrimiento de reglas de pertenencia a grupos y ponderación de reglas de comportamiento o de pertenencia. Estos procesos fueron validados durante un periodo de cinco años en casos de aplicación y reportados a la comunidad internacional en [García-Martínez, et al., 2013].

El proceso de descubrimiento de reglas de comportamiento aplica cuando se requiere identificar cuales son las condiciones para obtener determinado resultado en el dominio del problema. Para el descubrimiento de reglas de comportamiento definidos a partir de atributos clases en un dominio de problema que representa la masa de

información disponible, se propone la utilización de algoritmos de inducción TDIDT [Quinlan, 1986].

El proceso de descubrimiento de grupos aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema. Para el descubrimiento de grupos a partir de masas de información del dominio de problema sobre las que no se dispone ningún criterio de agrupamiento “a priori” se propone la utilización de Mapas Auto Organizados de Kohonen o SOM por su sigla en inglés [Kohonen, 1995].

El proceso de ponderación de interdependencia de atributos aplica cuando se requiere identificar cuales son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema. Para ponderar en que medida la variación de los valores de un atributo incide sobre la variación del valor de un atributo clase se propone la utilización de Redes Bayesianas [Jensen, 1996].

El proceso de descubrimiento de reglas de pertenencia a grupos aplica cuando se requiere identificar cuales son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema. Para descubrir este tipo de reglas se propone la utilización de Mapas Auto Organizados de Kohonen y tomando como clases a los grupos descubiertos se propone utilizar algoritmos de inducción TDIDT.

El proceso de ponderación de reglas de comportamiento o de la pertenencia a grupos aplica cuando se requiere identificar cuales son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo. Para ponderar en que medida la variación de los valores de un atributo antecedente en las reglas descubiertas incide sobre la variación del valor de un atributo clase se propone la utilización de Redes Bayesianas.

3. Identificación del Problema

Los sistemas de información geográfica son susceptibles de ser tratados como bases de datos que contiene información georeferenciada de utilidad vinculada a un determinado territorio de interés. Por esta razón, pueden ser considerados como objetos de estudio y experimentación en el marco de la ingeniería de explotación de información [García-Martínez et al., 2011]. Se asume factible poder extender los procesos de explotación de información usuales, o generar variaciones de estos, para su aplicación a la información contenida en la base de datos georeferenciada y en consecuencia, poder extraer de manera ingenieril y sistémica, patrones ó piezas de conocimiento de interés sobre el territorio asociado. Sin embargo, no existen ambientes de trabajo de acceso libre que permitan desarrollar procesos de explotación de información sobre bases de datos georeferenciadas.

En este contexto, se plantean las siguientes preguntas de investigación: [a] ¿Se pueden integrar ambientes de trabajo abiertos de explotación de información y de gestión de de información geográfica?, y de ser posible resolver la integración [b] ¿Se puede

validar el funcionamiento del ambiente integrado de trabajo a través de casos de estudio?

4. Solución Propuesta

En esta sección se describen: los ambientes de trabajo considerados como candidatos al proceso de integración (sección 4.1) y el sistema de integración desarrollado (sección 4.2).

4.1. Ambientes de Trabajo Considerados para la Solución

En este apartado se describen en la Sección 3.1 las funcionalidades del ambiente de trabajo para explotación de información (TANAGRA) y en la Sección 3.2 las funcionalidades del ambiente de trabajo para información geográfica (gvSIG); elegidos para la experiencia de integración de ambientes.

4.1.1. Ambiente de Trabajo para Explotación de Información

El ambiente de trabajo para explotación de información TANAGRA es un paquete de software de uso libre de aprendizaje automático desarrollado con fines académicos y de investigación por Ricco Rakotomalala [2005a; 2005b] en la Universidad Lumière Lyon 2 (Francia). Tanagra brinda una amplia variedad de métodos para la implementación de proyectos de minería de datos, desde aquellos destinados al análisis exploratorio de los datos, al aprendizaje estadístico, al aprendizaje automático y así como herramientas pertenecientes al área de base de datos.

Los algoritmos de minería de datos destinados a extraer patrones de conocimiento ocultos en grandes masas de datos, pueden clasificarse en dos categorías: aprendizaje supervisado y aprendizaje no supervisado.

Tanagra provee una amplia variedad de algoritmos pertenecientes a ambas categorías. Entre los algoritmos de aprendizaje supervisado se encuentran: Binary logistic regression, k-Nearest Neighbor (k-NN), Multi-layer perceptron, Prototype-NN, ID3, Linear Discriminant Analysis y Naive Bayes. Entre los algoritmos de agrupamiento se encuentran: K-Means, Kohonen's SOM, LVQ y HAC.

Los formatos de archivos que Tanagra reconoce para la lectura de la base de datos son TXT y XLS, mientras que para el desarrollo del proyecto, utiliza los formatos BDM o TDM.

4.1.2. Ambiente de Trabajo para Gestión de Información Geográfica

El ambiente de trabajo para información geográfica gvSIG [Anguix et al., 2008] es un paquete de software de uso libre para desarrollo de Sistemas de Información Geográfica que incluye las aplicaciones gvSIG Desktop y gvSIG Mobile. La aplicación gvSIG Desktop es un programa informático para el manejo de información geográfica con precisión cartográfica que se distribuye bajo licencia GNU GPL v2. Permite acceder a información vectorial y rasterizada así como a servidores de mapas que cumplan las especificaciones del OGC. Esta es una de las principales

características de gvSIG respecto a otros Sistema de Información Geográfica, la importante implementación de servicios OGC: WMS (Web Map Service), WFS (Web Feature Service), WCS (Web Coverage Service), Servicio de Catálogo y Servicio de Nomenclador.

Está desarrollado en lenguaje de programación Java, funcionando con los sistemas operativos Microsoft Windows, Linux y Mac OS X, y utiliza bibliotecas estándar de GIS reconocidas, como Geotools o Java Topology Suite (JTS). Asimismo, gvSIG posee un lenguaje de scripting basado en Jython y también se pueden crear extensiones en Java utilizando las clases de gvSIG.

Entre los formatos gráficos de fichero más habituales cuenta entre otros con acceso a formatos vectoriales GML, SHP, DXF, DWG, DGN, KML y formatos de imagen rasterizada como MrSID, GeoTIFF, ENVI o ECW.

4.2. Sistema de Integración Desarrollado

La solución desarrollada para el problema consiste en la implementación de un Sistema de Integración de los ambientes de Información Geográfica y de Explotación de Información. Se describen: los requisitos del sistema (sección 4.2.1), los requisitos funcionales (sección 4.2.2), los requisitos no funcionales (sección 4.2.3), los requisitos en negativo (sección 4.2.4), y aspectos operativos de la implementación del sistema (sección 4.2.5).

4.2.1. Requisitos del Sistema

El sistema se integra como extensión a gvSIG logrando acceder a los datos en formato dBase (.dbf) incluidos en los archivos Shape que forman parte de las capas utilizadas por la herramienta. Una vez extraídos los datos, el sistema los traduce a un formato que la herramienta TANAGRA pueda analizar, con el fin de permitir al usuario utilizar todas las técnicas de explotación de información que esta herramienta ofrece.

4.2.2. Requisitos Funcionales

Se enumera la lista de Requisitos Funcionales del Ambiente Integrado de Explotación de Información Geográfica que indican las funcionalidades que formarán parte del Sistema que lo soporta:

- El sistema permitirá a los usuarios utilizar todas las herramientas de explotación de información disponibles en TANAGRA extrayendo los datos de los archivos Shape manejados por gvSIG de forma automática.
- El sistema facilitará al usuario la conversión de información contenida en los archivos Shape al formato soportado por TANAGRA.
- El software formará parte de la aplicación gvSIG quedando disponible en el menú contextual de la capa seleccionada.

4.2.3. Requisitos No Funcionales

Se enumera la lista de Requisitos No Funcionales del sistema que soporta el Ambiente Integrado de Explotación de Información Geográfica:

- La integración entre las herramientas se dará de forma transparente para el usuario, esto es, el usuario no verá ni participará de los procesos internos de conversión de datos ni llamados a funciones de herramientas externas.
- El sistema será una aplicación de escritorio.
- El sistema deberá implementarse sobre plataformas Windows XP o Windows 7, 32bits.
- El sistema sólo trabajará con archivos en formato Shape.
- La integración funcionará sobre la versión 1.12 de gvSIG y la versión 1.4 de Tanagra.
- La interfaz del sistema se desarrollará tomando como base la interfaz de gvSIG 1.12.
- El sistema se desarrollará en Java.

4.2.4. Requisitos en Negativo

Se enumera la lista de Requisitos en Negativo del sistema que soporta el Ambiente Integrado de Explotación de Información Geográfica:

- El sistema no realizará ninguna función sobre archivos Shape que no estén correctamente formados. Los mismos deberán contener las tres partes necesarias en un Shapefile: .shp, .shx, .dbf
- El desarrollo del sistema no contemplará la corrección de bugs encontrados en las herramientas integradas.

4.2.5. Implementación del Sistema

La implementación del Sistema que soporta el Ambiente Integrado de Explotación de Información Geográfica fue realizada por Alumnos de la Cátedra de Proyecto de Software de la Licenciatura en Sistemas de La Universidad Nacional de Lanus durante el segundo cuatrimestre del 2013. La documentación del sistema puede verse en [Segura et al, 2013a] y una visualización (demo) del funcionamiento del sistema puede encontrarse en [Segura et al, 2013b].

5. Caso de Estudio

El caso de estudio se centra en el análisis de los datos geográficos y descriptivos sobre los operativos realizados por el Departamento de Defensa Civil en la Ciudad de Buenos Aires. La naturaleza de los procesos de análisis que proveen los Sistemas de Información Geográfica debe ser potenciada con procesos de explotación de información. En la Figura 1 se presenta una visualización de dicha información utilizando un Sistema de Información Geográfica.

Utilizando el Proceso de Descubrimiento de Reglas de Pertenencia a Grupos se observa:

- En la Figura 2 se agrupan los incidentes por Lluvias, Emanaciones, Seguimiento de Sustancias Peligrosas e Inundación, que están esparcidos ampliamente por los

barrios San Nicolás, Saavedra, Barracas, Almagro, Palermo, Monte Castro, Mataderos, Puerto Madero, Caballito, Constitución.

- En la Figura 3 se agrupan los incidentes por escapes de gas y residuos patológicos, que se encuentran bien centralizados en los barrios San Cristóbal, Recoleta y Villa Santa Rita.

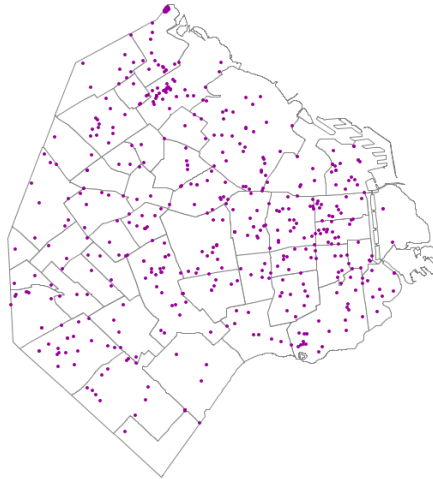


Fig. 1. Visualización SIG de operativos de defensa civil en CABA

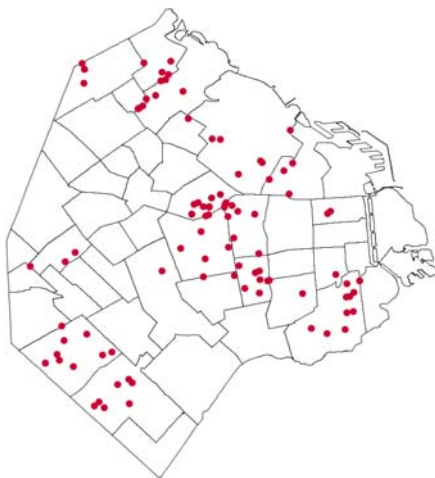


Fig. 2. Incidentes por lluvias, emanaciones, seguimiento de sustancias peligrosas e inundación

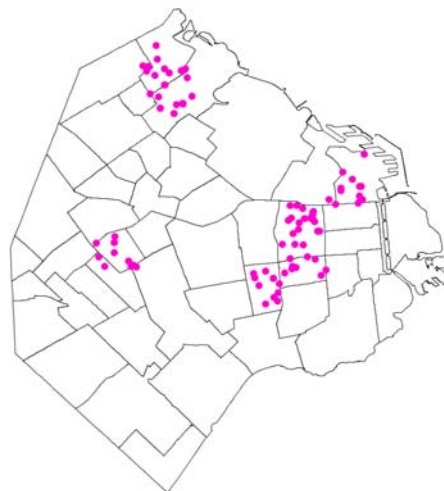


Fig. 3. Incidentes por escapes de gas y residuos patológicos

- En la Figura 4 se agrupan los incidentes por incendio, que se encuentran bien centralizados en los barrios Balbanera, San Cristobal, Boedo, Retiro y La Boca.

- En la Figura 5 se agrupan los incidentes por árboles caídos, que se encuentran en los barrios Caballito, Balbanera, Mataderos, Retiro, San Cristobal y Boca.

Utilizando el Proceso de Descubrimiento de Interdependencia de Atributos se observa en la Figura 6 que el barrio en el que se registraron más operativos en el año 2012 fue Palermo, mientras que el barrio en el que se registraron menos operativos fue Escollera Exterior. También se podría analizar en qué Barrio hay más posibilidades que se registre un desastre específico. Por ejemplo, los operativos relacionados con Inundaciones, según el análisis, son más frecuente en el barrio de Villa Devoto.

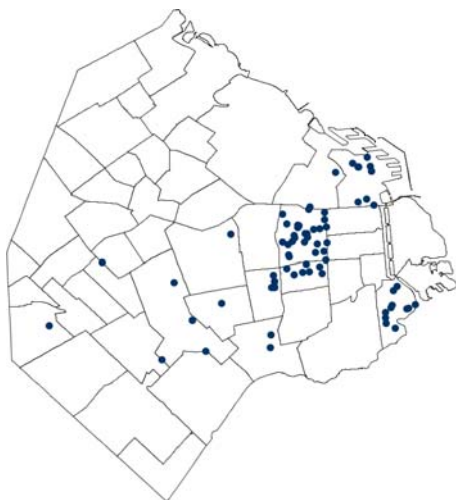


Fig. 4. Incidentes por incendios

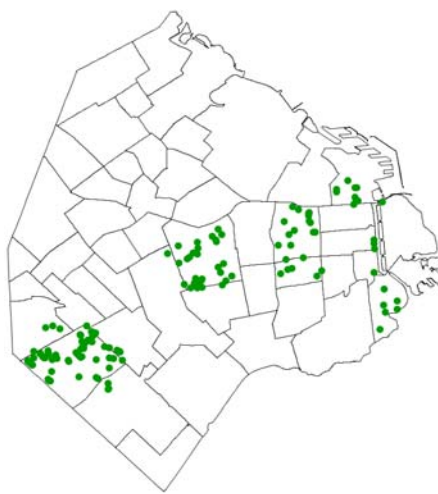


Fig. 5. Incidentes por árboles caídos

6. Conclusiones

En resultado presentado en este trabajo se enmarca en los proyectos 33A167 “Instrumentos para la Gestión de Proyectos de Explotación de Información” y 33B133 “Procesos de Explotación de Información Georeferenciada” de la Universidad Nacional de Lanús que buscan explorar la aplicación de la ingeniería de explotación de información en el dominio de sistemas de información geográfica.

Se ha presentado un ambiente de trabajo para realizar explotación de información en bases de datos geográficos, resultante de la integración de dos ambientes abiertos, uno para explotación de información (TANAGRA) y el otro para procesamiento de información geográfica (gvSIG).

Se ha desarrollado un caso de estudio del proceso de integración sobre una base de datos geográficos sobre operativos realizados por el Departamento de Defensa Civil en la Ciudad de Buenos Aires.

Entre las futuras líneas de trabajo se han establecido las siguientes:

- [i] Caracterizar los procesos de explotación de información georeferenciada.
- [ii] Definir, por asimilación con los procesos de explotación de información estándares, los procesos: “Descubrimiento de Reglas de Asociación en Espacios

- Georeferenciados”, “Descubrimientos de Grupos en Espacios Georeferenciados”, y “Descubrimiento de Interdependencias en Espacios Georeferenciados”-
- [iii] Desarrollar un esquema de proceso de cada uno de los procesos de explotación de información georeferenciada definidos.
 - [iv] Identificar los algoritmos de minería de datos aplicables a cada proceso de explotación de información georeferenciada definido.

7. Financiamiento

Las investigaciones que se reportan en este artículo han sido financiadas parcialmente por los Proyectos de Investigación 33A167 y 33B133 de la Secretaria de Ciencia y Técnica, y por la Cátedra de Proyecto de Software de la Licenciatura en Sistemas de la Universidad Nacional de Lanús (Argentina).

8. Referencias

- Abraham, A. 2003. Business Intelligence from Web Usage Mining. *Journal of Information & Knowledge Management*, 24: 375-390.
- Anguix, A., Díaz, L., & Carrera, M. (2008). gvSIG: A GIS desktop solution for an open SDI. *Journal of Geography and Regional Planning*, 1(3): 41-48.
- Cooley, R. 2003. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. *ACM Transactions on Internet Technology*, 32: 93-116.
- Curtis, B., Kellner, M., Over, J. 1992. Process Modelling. *Communications of the ACM*, 359: 75-90.
- Eklund, P. W. (1998). Data mining and soil salinity analysis. *International Journal of Geographical Information Science*, 12(3): 247-268.
- Estivill-Castro, V., & Lee, I. (2001, September). Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data. *En Proceedings 6th International Conference on Geocomputation* (pp. 24-26).
- Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. 1996. *Advances in Knowledge Discovery and Data Mining*. AAAI Press.
- Ferreira, J., Takai, O., Pu, C. 2005. Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services. *Proceedings Seventh IEEE International Conference on E-Commerce Technology*. Pág. 471-474.
- García Martínez, R. 1997. *Sistemas Autónomos. Aprendizaje Automático*. Editorial Nueva Librería. ISBN 950-9088-84-6.
- García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- García Martínez, R., Servente, M. y Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011. Towards an Information Mining Engineering. *En Software Engineering, Methods, Modeling and Teaching*. Páginas 83-99. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6.
- Grossman, R., Kasif, S., Moore, R., Rocke, D., Ullman, J. 1998. *Data Mining Research: Opportunities and Challenges*. <http://www.csl.mtu.edu/cs5811/common/2002-fall-reading-group/2002-10-25/grossman98.ps>. Ultimo acceso 17 de Abril del 2008.

- Heckerman, D., Chickering, M., Geiger, D. (1995). Learning bayesian networks, the combination of knowledge and statistical data. *Machine learning* 20: 197-243.
- Jensen, F. 1996. An introduction to Bayesian networks (Vol. 210). London: UCL press.
- Kanungo, S. 2005. Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. Pág. 231-240.
- Kohonen, T. (1995). *Self-Organizing Maps*. Springer Verlag Publishers.
- Kohonen, T. 1995. *Self-Organizing Maps*. Springer Verlag Publishers.
- Langseth, J., Vivatrat, N. (2003). Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound. *Intelligent Enterprise* 5(18): 34-41.
- Mennis, J., & Liu, J. W. (2005). Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change. *Transactions in GIS*, 9(1): 5-17.
- Mobasher, B, R Cooley and J Srivastava 1999. Creating adaptive web sites through usage-based clustering of URLs. *Proceedings Workshop on Knowledge and Data Engineering Exchange*, Pág. 19-25.
- Moreno, A., Buzai, G., Fuenzalida, M., Colso, A. (2012). *Sistemas de Información Geográfica*. Editorial RA-MA. ISBN-13: 978-8499641317
- Ng, R. y Han, J. (2002). CLARANS: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering*, 14(5): 1003-1016.
- Quinlan, J. (1990). Learning Logic Definitions from Relations. *Machine Learning*, 5:239-266
- Quinlan, J. 1986. Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Rakotomalala, R. (2005a). TANAGRA, une plate-forme d'expérimentation pour la fouille de données, *Revue MODULAD*, 32: 70-85., 2005. (Journal)
- Rakotomalala, R. (2005b). TANAGRA: Un logiciel gratuit pour l'enseignement et la recherche. *Actes de 5èmes Journées d'Extraction et Gestion des Connaissances, EGC-2005, RNTI-E-3*, vol. 2, pág. 697-702.
- Roddick, J. F., & Lees, B. G. (2001). Paradigms for spatial and spatio-temporal data mining. *Geographic data mining and knowledge discovery*, 33-50.
- Segura, A., Santamaria, P., Mieres, F., Aguirre, D. (2013a). *Ambiente de Explotación de Información para Sistemas de Información Geográfica*. Informe. Cátedra de Proyecto de Software. Licenciatura en Sistemas. Universidad Nacional de Lanús. <http://www.unla.edu.ar/sistemas/sls/ls-3-Proyecto-de-Software/pdf/2013-TF-PdS-EI-SIG.pdf>. Pagina vigente al 11/05/2014.
- Segura, A., Santamaria, P., Mieres, F., Aguirre, D. (2013b). *Demo Ambiente de Explotación de Información para Sistemas de Información Geográfica*. Demo. Cátedra de Proyecto de Software. Licenciatura en Sistemas. Universidad Nacional de Lanús. <http://www.youtube.com/watch?v=T3i17KJgxUc&feature=youtu.be> Pagina vigente al 11/05/2014.
- Shekhar, S., Zhang, P., Huang, Y., & Vatsavai, R. R. (2003). Trends in spatial data mining. *Data mining: Next generation challenges and future directions*, 357-380.
- Spielman, S. E., & Thill, J. C. (2008). Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems*, 32(2): 110-122.
- Srivastava, J., Cooley, R., Deshpande, M., Tan, P. 2000. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 12: 12-23.
- Verma, D. y Nashine, R. (2012). Data Mining: Next Generation Challenges and Future Directions. *International Journal of Modeling and Optimization*, 2(5): 603-608.
- Wikipedia. (2013). *Sistemas de Información Geográfica*. http://es.wikipedia.org/wiki/Sistema_de_Informaci3n_Geogr3fica. Pagina vigente al 29/08/2013.