

# INTEGRACIÓN DE SISTEMAS BASADOS EN CONOCIMIENTO Y DE DESCUBRIMIENTO. APLICACIÓN A LA TOMA DE DECISIONES ESTRATÉGICAS OPERACIONALES

Claudio Rancan<sup>1,2</sup>, Patricia Pesado<sup>1,3</sup>, Ramón García-Martínez<sup>1,4</sup>

<sup>1</sup> Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP

<sup>2</sup> Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

<sup>3</sup> Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CIC

<sup>4</sup> Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

crancan@itba.edu.ar, ppesado@lidi.info.unlp.edu.ar, rgarciamar@fi.uba.ar

## CONTEXTO

El Proyecto articula líneas de trabajo de los siguientes Proyectos: [a] “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

## RESUMEN

Este proyecto se centra en definir una arquitectura de integración entre sistemas de descubrimiento de conocimiento (adquisición automática) y sistemas basados en conocimiento (sistemas expertos). Se busca estimar la viabilidad de implementación de dicha arquitectura en función de la madurez de las tecnologías involucradas.

**Palabras clave:** Descubrimiento de conocimiento. Sistemas basados en conocimiento. Arquitecturas de Sistemas

## 1. INTRODUCCION

Los sistemas basados en conocimientos (SBC) ó sistemas expertos emulan el comportamiento humano experto en un área de conocimiento determinada. Constituyen sistemas de ayuda a la toma de decisiones en áreas tan diversas como la selección de estrategias instruccionales [Sierra *et al.*, 2004], el control de variables ambientales [Sierra *et al.*, 2005], la configuración de ventiladores en neonatología [Bermejo *et al.*, 2002] la individualización y acuerdos para penas en procesos judiciales [Gómez *et al.*, 2001] o la generación asistida del mapa de actividades de proyectos de desarrollo de software [Diez *et al.*, 2003].

La base de conocimiento de un sistema experto encapsula en algún formalismo de representación (reglas, marcos, redes semánticas entre otras), el conocimiento del dominio que debe ser puesto en juego por el sistema para resolver un problema dado.

Las metodologías de construcción de bases de conocimiento se han consolidado en los últimos 15 años [Debenham, 1990; 1998; Gomez *et al.*, 1997; García-Martínez y Britos, 2004].

Los sistemas inteligentes constituyen el campo de la Informática en el que se estudian y desarrollan algoritmos que implementan los distintos modelos de aprendizaje y su aplicación a la resolución de problemas prácticos [Michalski, 1991; Dejong & Money 1986]. Entre los problemas abordados en este campo, está el de descubrir conocimientos [Michalski *et al.*, 1983; 1986; 1998; Michalski y Tecuci, 1994; Mitchel, 1996; Michie, 1988].

El descubrimiento de conocimiento (KD Knowledge Discovery) es la búsqueda de patrones interesantes y de regularidades importantes en grandes bases de información [Fayad *et al.*, 1996; Grossman *et al.*, 1999]. Al hablar de descubrimiento de conocimiento basado en sistemas inteligentes o Minería de Datos/Información Inteligente [Evangelos *et al.*, 1996; Michalski *et al.*, 1998] nos referimos específicamente a la aplicación de métodos de aprendizaje automático u otros métodos similares, para descubrir y enumerar patrones presentes en dicha información.

En este contexto, trabajos recientes sobre sistemas de ayuda a la toma de decisiones estratégicas operacionales (SATD-EO) basados en SBC [Sierra *et al.*, 2006] en áreas como el control aéreo [Ierache y García-Martínez, 2004] ó el alistamiento de unidades navales [Rancán, 2004] han puesto de manifiesto que es un problema abierto la definición de cómo se puede integrar a este tipo de sistemas, procesos de descubrimiento de conocimiento basados en aprendizaje automático [García-Martínez y Borrajo, 1997; 2000; Grosser *et al.*, 2005; Felgaer *et al.*, 2006, Cogliati *et al.*, 2006] que les permitan mejorar “en línea” la calidad de la base de conocimiento utilizada para la ayuda a la toma de decisiones.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

Las líneas de investigación de este proyecto son:

- [a] El paradigma de arquitecturas integradas de planificación y aprendizaje basado en construcción de teorías y su potencial aplicación al descubrimiento de conocimiento estratégico operacional [Fritz, *et al.*, 1989; García-Martínez y Borrajo, 1996;1997; 1998; 2000; Sierra *et al.*, 2006].
- [b] La arquitectura subyacente a sistemas basados en conocimiento de ayuda a la toma de decisiones conocidos [García Martínez y Britos, 2004; Britos, 2001; Rizzi, 2001; Ierache y García-Martínez, 2004; Bermejo *et al.*, 2002; Gomez *et al.*, 2001; Hostian 2003].
- [c] Los paradigmas de descubrimiento de conocimiento con énfasis en la evaluación de conocimiento [Jensen, 2002], sus estructuras [Utgof *et al.*, 2000; Jensen y Neville, 2002; Neville y Jensen, 2002], los procesos de adquisición distribuidos [Jensen *et al.*, 1999] y las tecnologías de sistemas inteligentes asociadas al descubrimiento de conocimiento [Britos *et al.*, 2005].

### 3. RESULTADOS OBTENIDOS/ESPERADOS

A la fecha se ha definido la arquitectura de integración de sistemas de descubrimiento de conocimiento y sistemas basados en conocimiento [Rancan *et al.*, 2007a] y se han realizado experimentos exploratorios sobre la calidad de los resultados obtenibles [Rancan *et al.*, 2007b]. El año pasado (2008) se han identificado casos reales de aplicación año se prevé la prueba de conceptos sobre casos reales.

### 4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y tres tesis de ingeniería informática (dos de ellas defendidas y otra en curso). Esta línea vincula al LIDI de la Facultad de Informática de la Universidad Nacional de La Plata y al Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

### 5. BIBLIOGRAFIA

Bermejo, F., Britos, P., Rossi, B y García Martínez, R. 2002. *Sistema de Asistencia para la Configuración de Ventiladores OAF en Neonatología*. Revista del Instituto Tecnológico de Buenos Aires. 28: 24-68. ISSN: 0326-1840. 2002.

Britos, P. 2001. *Sistema de Ayuda sobre Legislación Argentina en Riesgos de Trabajo*. Tesis de Magister en Ingeniería del Conocimiento. Facultad de Informática de la Universidad Politécnica de Madrid.

Britos, P., Hossian, A., García Martínez, R. y Sierra, E. 2005. *Minería de Datos Basada en Sistemas Inteligentes*. 876 páginas. Editorial Nueva Librería. ISBN 987-1104-30-8.

Cogliati, M., Britos, P. y García-Martínez, R. (2006). *Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT*. Springer IFIP Series. Aceptado.

De Giusti E., Mollo Brisco G., La Battaglia, J., Pasini, A. Dirección: Pesado, P. 2004. *Sistema de Simulación de Escenarios y Decisiones Empresarias*. Proceedings de las XII Jornadas de Jóvenes Investigadores AUGM 2004. Universidad Federal Do Paraná - Curitiba, Brasil.

Debenham, J. 1990. *Knowledge Systems Design*. Prentice Hall.

Debenham, J. 1998. *Knowledge Engineering: Unifying Knowledge Base and Database Design*. Springer-Verlag.

DeJong, G.F., Mooney, R.J. 1986. *Explanation-Based Learning: An Alternative View*, Machine Learning, 1: 145-176.

Diez, E., Britos, P., Rossi, By García-Martínez, R. 2003. *Generación Asistida del Mapa de Actividades de Proyectos de Desarrollo de Software*. Reportes Técnicos en Ingeniería del Software. 5(1):13-18. ISSN 1667-5002.

Evangelos, S., Han, J, (eds). 1996. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, EE.UU.

Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. (eds). 1996. *Advances in Knowledge Discovery and Data Mining*, San Mateo, AAAI Press.

Felgaer, P., Britos, P. and García-Martínez, R. 2006. *Prediction in Health Domain Using Bayesian Network Optimization Based on Induction Learning Techniques*. International Journal of Modern Physics C 17(3): 447-455.

Fritz, W., García Martínez, R., Rama, A., Blanqué, J., Adobatti, R, y Sarno, M. 1989. *The Autonomous Intelligent System*. Robotics and Autonomous Systems, 5(2): 109-125.

García Martínez, R. & Borrajo Millán, D. 1996. *Unsupervised Machine Learning Embedded in Autonomous Intelligent Systems*. Proceedings of the XIV International Conference on Applied Informatics. Páginas 71-73. Innsbruck. Austria.

García Martínez, R. & Borrajo Millán, D. 1998. *Learning in Unknown Environments by Knowledge Sharing*. Proceedings of the Seventh European Workshop on Learning Robots. Páginas 22-32. Editado University of Edinburg Press.

García Martínez, R. y Borrajo Millán, D. 1997. *Planning, Learning and Executing in Autonomous Systems*. Lecture Notes in Artificial Intelligence. 1348:208-210.

García Martínez, R. y Borrajo Millán, D. 2000. *An Integrated Approach of Learning, Planning and Executing*. Journal of Intelligent and Robotic Systems 29(1):47-78.

García-Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería.

- Gomez, A., Juristo, N., Montes, C. y Pazos, J. 1997. *Ingeniería del Conocimiento*. Editorial R. Areces. Madrid.
- Gómez, S., Perichinsky, G. y Garcia Martinez, R. 2001. *Un Sistema Experto Legal para la Individualización y Acuerdos para Penas*. Proceedings del Simposio Argentino de Informática y Derecho. Pág. 23-33. Septiembre. Editado por la SADIO.
- Grosser, H., Britos, P. y García-Martínez, R. 2005. *Detecting Fraud in Mobile Telephony Using Neural Networks*. Lecture Notes in Artificial Intelligence 3533: 613-615.
- Grossman, R., Kasif, S., Moore, R., Rocke, D. and Ullman, J. 1999. *Data Mining Research: Opportunities and Challenges*, A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data, January 1999, Chicago
- Hossian, A. 2003. *Sistema de Asistencia para la Selección de Estrategias y Actividades Instruccionales*. Tesis de Magister en Ingeniería del Software. ITBA.
- Ierache, J. y García-Martínez, R. 2004. *Sistema Experto Aplicado al Control del Espacio Aéreo*. Proceedings del IX Congreso Argentino de Ciencias de la Computación.
- Jensen D. 2002. *Knowledge Evaluation*. Handbook of Data Mining and Knowledge Discovery. Kloege, W. and J. Zytkow (Eds.). Oxford: Oxford University Press
- Jensen D. and J. Neville 2002. *Schemas and models*. Proceedings of the Multi-Relational Data Mining Workshop, 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Jensen D., Y. Dong, B. Lerner, E. McCall, L. Osterweil, S. Sutton Jr., and A. Wise 1999. *Coordinating agent activities in knowledge discovery processes*. In Proceedings of the International Joint Conference on Work Activities Coordination and Collaboration. pp. 137-146.
- La Battaglia, J., Rodriguez, I., Thomas, P., Bertone, R. Dirección: Pesao P. 2003. *Tecnología aplicada a gestión distribuida*. Proceedings de las XI Jornadas de Jóvenes Investigadores AUGM 2003. Universidad Nacional de La Plata – Argentina.
- Miatón, I., Pesado, P., Bertone, R. y De Giusti. 2003. *Agentes Basados en Sistemas Distribuidos*. Proceedings V WICC.
- Michalski, R. 1993. *A Theory and Methodology of Inductive Learning*, Artificial Intelligence, 20:111-161.
- Michalski, R. Bratko, I. Kubat, M (eds.) 1998. *Machine Learning and Data Mining, Methods and Applications*, John Wiley & Sons Ltd, West Sussex, England
- Michalski, R. S. 1991. *Toward an Unified Theory of Learning: An Outline of Basic Ideas*, Proceedings of the 3rd World Conference on the Fundamentals of Artificial Intelligence, Paris.
- Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (eds.), 1986. *Machine Learning: An Artificial Intelligence Approach, Vol. II*, Morgan-Kauffman
- Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (eds.). 1983. *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Morgan-Kauffman
- Michalski, R. S., Tecuci, G. (eds) 1994. *Machine Learning: A Multistrategy Approach, Vol. III*, Morgan Kauffman
- Michie, D. 1988. *Machine Learning in the next five years*, EWSL-88, 3rd European Working Session on Learning, Glasgow, Londres, Pitman.
- Mitchell, T. M. 1996. *Machine Learning*, McGraw-Hill.
- Neville J. and D. Jensen 2002. *Supporting relational knowledge discovery: Lessons in architecture and algorithm design*. Proceedings of the Data Mining Lessons Learned Workshop, 19th International Conference on Machine Learning.
- Pesado, P. Feierherd G. y Pasini, A. 2005. *Requirement Specifications for Electronic Voting System*. Journal of Computer Science & Technology, 5(4): 312-319 . ISSN: 1666-6038
- Rancán, C. 2004. *Arquitectura de Sistema Híbrido de Evaluación del Alistamiento de Unidades Navales Auxiliares*. Reportes Técnicos en Ingeniería del Software. 6(1): 45-54. ISSN 1667-5002.
- Rancan, C., Kogan, A., Pesado, P. and García-Martínez, R. (2007b). *Knowledge Discovery for Knowledge Based Systems. Some Experimental Results*. Research in Computing Science Journal, 27: 3-13.
- Rancán, C., Pesado, P. y García-Martínez, R. (2007a). *Toward Integration of Knowledge Based Systems and Knowledge Discovery Systems*. Journal of Computer Science & Technology, 7(1): 91-97.
- Rizzi, M. 2001. *Sistema Experto Asistente de Requerimientos*. Tesis de Magister en Ingeniería del Software. ITBA.
- Sierra, E., García-Martínez, R., Hossian, A., Britos, P. y Balbuena, E. 2006. *Providing Intelligent User-Adapted Control Strategies in Building Environments*. Research in Computing Science Journal, 19: 235-241.
- Sierra, E., Hossian, A. y García-Martínez, R. 2003. *Sistemas Expertos que Recomiendan Estrategias de Instrucción. Un Modelo para su Desarrollo*. Revista Latinoamericana de Tecnología Educativa. 1(1): 19-30. ISSN: 1695-288X. 2003.
- Sierra, E., Hossian, A., García-Martínez, R. y Marino, P. 2005. *Sistema Experto para Control Inteligente de las Variables Ambientales de un Edificio Energéticamente Eficiente*. Proc. XI RTPIC. Pág. 446-452.
- Utgoff P., V. Lesser, and D. Jensen 2000. *Inferring task structure from data*. University of Massachusetts, Department of Computer Science. Technical Report UM-CS-2000-054.