

Técnicas de explotación de información aplicadas al análisis del rendimiento académico de alumnos de la FACENA-UNNE

Porcel, Eduardo; Dapozo, Gladys; López, María V.; Caputo, Liliana N.
Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste. 9 de Julio N° 1449. CP 3400. Corrientes. Argentina.
TE: (03783) 423126 - (03783) 473930 Fax
{gndapozo, eporcel, mvlopez, lcaputo}@exa.unne.edu.ar

CONTEXTO

Este trabajo sintetiza los avances logrados y las líneas de trabajo examinadas en el proyecto F008-2008: "Rendimiento Académico de alumnos de la FACENA – UNNE: Su análisis mediante métodos cuantitativos", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste. El mismo se desarrolla en el Grupo de Matemática Aplicada a la Investigación Educativa, en el Departamento de Matemática de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste (FACENA-UNNE).

RESUMEN

Las actividades enmarcadas en el proyecto "Rendimiento Académico de alumnos de la FACENA – UNNE: Su análisis mediante métodos cuantitativos", tienen como objetivo fundamental la identificación de factores que afectan el rendimiento académico de los alumnos de las carreras de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste (FACENA-UNNE). En este trabajo se describen las principales metodologías y técnicas de explotación de información empleadas para la construcción de modelos orientados al análisis de la información recopilada para satisfacer los objetivos del proyecto, y se sintetizan los resultados obtenidos, los trabajos en curso y la integración y formación de recursos humanos que intervienen en el mismo. Los resultados de este proyecto aportarán a un mayor conocimiento de los posibles factores que inciden en el desempeño de los alumnos y permitirán que, desde la gestión institucional, se aborden mecanismos superadores que contribuyan, por un lado, al mejoramiento de los índices de permanencia y culminación exitosa de los estudios universitarios y, por otro lado, aporten información significativa a los procesos de evaluación y acreditación de carreras, en consonancia con las políticas educativas nacionales que priorizan la formación en Informática, Ingeniería y Ciencias Básicas, por considerarlas estratégicas para el desarrollo del país.

Palabras clave: *Rendimiento académico. Educación Superior. Métodos estadísticos clásicos. Métodos inteligentes.*

1. INTRODUCCION

En nuestro país, en la década del '90, el Estado Nacional incluye en su agenda de política educativa la evaluación de la calidad de la formación universitaria, y la mayoría de las universidades nacionales inician procesos de evaluación institucional.

En 1996, se conocen los primeros resultados referidos al rendimiento académico de los estudiantes de las trece carreras que por entonces podían cursarse en la FACENA-UNNE. Dicha información hace referencia a elevados índices de desgranamiento en todos los años de estudio pero, fundamentalmente, al término del primer cuatrimestre del primer año de estudios. Asimismo, da cuenta de que el retraso promedio en el egreso de todas las carreras alcanza al 50% de la duración teórica de las mismas, llegando en algunas a superarlo (Informe de Autoevaluación de FACENA-UNNE, 1996).

El rendimiento académico es un claro indicador del avance exitoso en la carrera de estudios de algún alumno en un momento particular, y a su vez también es un pronosticador de la posibilidad de completar exitosamente dicha carrera (Toer, 2000).

Debe tenerse en cuenta que se trata de un constructo teórico complejo y multidimensional, atravesado y determinado por múltiples factores socioeconómicos (situación laboral, estado civil, nivel educativo del grupo familiar, etc.), institucionales (diseño curricular, práctica docente, valores y concepciones institucionales, etc.) e individuales de los estudiantes (formación previa, hábitos de estudios, etc.).

Por tal motivo, el rendimiento académico ha sido representado de diferentes maneras en los diversos estudios que han abordado el tema. En algunos, está representado sólo por el número de materias aprobadas por un alumno en una carrera, en otros por los resultados de tests específicamente diseñados o el promedio de notas de las asignaturas cursadas. Esta variedad de interpretaciones del concepto de rendimiento académico está ligada a las particularidades de las investigaciones en cuestión, principalmente al momento histórico en que se realiza la investigación y las concepciones de quienes llevan a cabo y financian la misma. Restringir el concepto a uno solo de estos indicadores, supone una postura ingenua y hasta simplista de lo que el acto educativo significa, pues

equivale a descontextualizarlo, aislándolo de la realidad social e histórica de la que forma parte. Dado que el objetivo de este proyecto es la construcción de modelos de predicción del rendimiento académico de los estudiantes, se utilizó para ello el caudal de información que la institución universitaria dispone en sus sistemas informáticos de administración de alumnos, la cual incluye los datos socioeconómicos y educativos aportados por el mismo alumno en el momento de su inscripción, y datos del desempeño académico en el período 2001-2009 de los estudiantes de las diversas carreras de FACENA, lo cual comprende información histórica del desarrollo de las distintas actividades académicas previstas en el proceso formativo (cursos de ingreso, inscripción a materias, calificaciones obtenidas en los exámenes, etc.), que permiten analizar cómo se realiza el tránsito de los alumnos hacia la titulación. Esto requiere el análisis de bases de datos para lo cual se utilizaron técnicas de Explotación de Información.

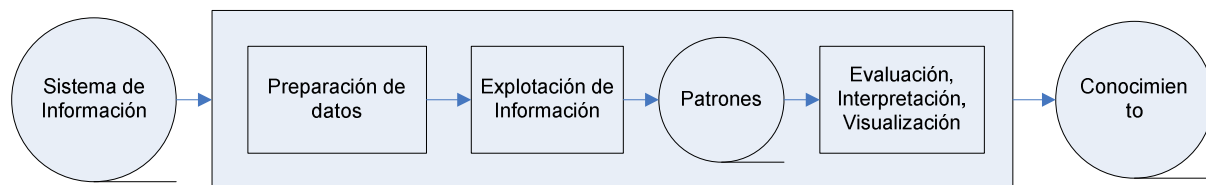


Figura 1

La Explotación de Información toma del análisis estadístico (García, 2004 en Britos, 2008) las siguientes técnicas: análisis de varianza, análisis de regresión, prueba Chi-cuadrado, análisis de agrupamiento, análisis de componentes, análisis discriminante, series de tiempo.

De los sistemas inteligentes (García Martínez *et al.*, 2003 en Britos, 2008) usa las siguientes tecnologías: algoritmos genéticos, algoritmos (TDIDT - Top Down Induction Decision Trees) que generan árboles de decisión, redes neuronales, redes bayesianas.

Además, en el marco de este proyecto, se han obtenido datos complementarios mediante instrumentos diseñados especialmente para recabar información de aspectos específicos, no sistematizados, como por ejemplo el test diagnóstico de conocimientos matemáticos previos realizado a los ingresantes en el año 2009.

Esta información se analiza empleando métodos del análisis estadístico (regresión logística, análisis de conglomerados) y de los sistemas inteligentes (Árboles de Decisión y Redes Neuronales), como así también métodos de programación por camino crítico (CPM), utilizando duraciones determinísticas y borrosas.

La Regresión Logística es un método lineal que intenta modelizar la probabilidad de ocurrencia de un evento de interés. La variable dependiente es

La explotación de información engloba un conjunto de técnicas encaminadas a la extracción de conocimiento procesable pudiendo abordarse la solución a problemas de predicción, clasificación y segmentación en la extracción de conocimiento no trivial que reside de manera implícita en los datos disponibles en distintas fuentes de información (Schiefer *et al.* (2004) en Britos (2008)).

La Explotación de Información puede definirse como un elemento fundamental de un proceso más amplio, el cual, como lo muestra la Figura 1, tiene una primera etapa de preparación de datos, luego el proceso de explotación de información, la obtención de patrones de comportamiento, y la evaluación e interpretación de los patrones descubiertos (Fayyad *et al.*, 1996; Britos *et al.*, 2005). Este conjunto de técnicas pueden aplicarse a la solución a problemas de predicción, clasificación y segmentación. Las bases de estas técnicas se encuentran en el análisis estadístico y en los sistemas inteligentes (Umapathy, 2007 en Britos, 2008).

categorica dicotómica o policotómica, a los efectos de facilitar la interpretación (Britos, 2005).

Las Redes Neuronales son modelos computacionales inspirados en las características neurofisiológicas del cerebro humano y están formadas por un gran número de neuronas dispuestas en varias capas e interconectadas entre sí mediante conexiones con pesos. Las redes neuronales constituyen herramientas analíticas que permiten examinar los datos con el objeto de descubrir y modelar las relaciones funcionales existentes entre las variables. Pueden comportarse como técnicas de aproximación o de clasificación universales (Castillo *et al.*, 1999).

Como antecedentes de aplicación de la técnica de redes neuronales en el ámbito de educación pueden mencionarse los trabajos de González (1999), Salgueiro *et al.* (2006), Borracci y Arribalzaga (2005).

Los árboles de decisión son una serie de reglas o condiciones organizadas de forma jerárquica, a modo de árbol. Son muy útiles para encontrar estructuras en espacios de alta dimensionalidad y en problemas que mezclan datos categóricos y numéricos.

Como antecedente de aplicación de la técnica de árboles de regresión aplicado el rendimiento de alumnos universitarios puede mencionarse a Bacallao Gallestey *et al.* (2004).

El CPM es un método de planificación, evaluación y control de proyectos formados por un gran número de actividades que guardan entre sí una relación de orden o precedencia. (Pérez Mackeprang *et al*, 1998). Considerando que el plan de estudios es un proyecto constituido por actividades (la aprobación de las asignaturas que lo forman), el CPM resulta una herramienta apropiada para el seguimiento de los estudios de los alumnos y para detectar cuáles actividades son las causantes de los largos periodos de permanencia de éstos en el sistema. El desempeño de cada alumno es una ejecución particular del proyecto y resulta pues interesante construir - para cada carrera - categorías de alumnos según las similitudes de dichas ejecuciones y/o de los tiempos requeridos para la ejecución de cada actividad. Para la construcción de estas tipologías se utiliza análisis de conglomerados, el cual tiene como propósito clasificar una población de sujetos u objetos en función de ciertas variables (cualitativas o cuantitativas), con el propósito de que los elementos de una misma clase o conglomerado sean lo más semejantes posible, pero que elementos de conglomerados distintos sean lo más disímiles posible. (Levy Mangin y Varela Mallou.; 2003).

2. LINEAS DE INVESTIGACION y DESARROLLO

En el marco de este proyecto, se encuentran en ejecución las siguientes líneas o ejes temáticos:

2.1. Metodologías de integración de datos y técnicas de preprocesamiento con el objetivo de asegurar la calidad de la información que constituye el soporte de las actividades del proyecto.

2.2. Modelado de datos y predicción del rendimiento académico en el primer cuatrimestre del primer año de estudios en función del nivel de conocimientos matemáticos previos de los ingresantes a la FACENA, empleando métodos estadísticos clásicos e inteligentes.

2.3. Formulación y ajuste de un modelo de predicción del avance en las carreras en función de los retrasos en las tareas de regularizar y aprobar las materias de primer año, empleando técnicas estadísticas clásicas e inteligentes.

2.4. Formulación y ajuste de modelos para predecir el rendimiento académico de los alumnos en función de las características socioeconómicas de los mismos, empleando métodos estadísticos clásicos e inteligentes.

2.5. Construcción de categorías de alumnos de las carreras Licenciatura en Sistemas de Información, Agrimensura, Ingenierías (Eléctrica y Electrónica) y Bioquímica, en función del análisis de sus trayectos de formación,

mediante CPM y análisis de conglomerados (cluster).

3. RESULTADOS OBTENIDOS/ESPERADOS

Dentro de las líneas de trabajo mencionadas en la sección 2, se han obtenido los siguientes resultados:

En relación con la línea 2.1: Se han estudiado técnicas de preprocesado de datos para mejorar la calidad de la información obtenida desde los sistemas de información existentes, y para mantener actualizado un repositorio con toda la información sistematizada existente en la unidad académica respecto del desempeño de los alumnos.

En esta misma línea, se encuentra en curso el desarrollo de modelos de información de Inteligencia de Negocios para la gestión académica universitaria, cuyo objetivo es ampliar las posibilidades de obtención de información del sistema actual mediante la utilización de herramientas software de tipo open source y facilitar el acceso a la información pertinente y relevante a un amplio abanico de usuarios, entre ellos los docentes-investigadores.

En relación con la línea 2.2: Se ha dado inicio a esta línea, mediante la realización de un test de diagnóstico de conocimientos matemáticos previos a los ingresantes de las cohortes 2009 de las trece carreras de la FACENA (Porcel, Ramírez Arballo y Mata, 2009). Los datos del diagnóstico se utilizarán como variables explicativas en un modelo que permita predecir el rendimiento de los alumnos al finalizar el primer cuatrimestre, empleando métodos estadísticos e inteligentes.

En la línea 2.3, bajo el supuesto de que la dificultad de aprobar las distintas asignaturas incide en los retrasos en los estudios, se consideró necesario estimar la probabilidad de aprobación de cada asignatura, en función de los intentos previos de aprobación, realizando el ajuste a una distribución de probabilidad de tiempo de espera discreto (distribución geométrica), así como también analizar el rendimiento académico de los alumnos de las trece carreras de grado de la FACENA, utilizando indicadores basados en la relación entre el número de exámenes rendidos y el número de asignaturas aprobadas por los mismos, estimados mediante regresión lineal paramétrica y no paramétrica (Porcel, López, Dapozo y Caputo, 2009a).

Con respecto a la estimación de las probabilidades de aprobación de las asignaturas, se comunicaron y/o publicaron los resultados obtenidos para las carreras de Ingeniería Eléctrica, Ingeniería en Electrónica y Profesorado en Matemática de la FACENA (UNNE) (Porcel, López, Dapozo y Caputo, 2009b) (Porcel, López, Dapozo y Caputo, 2009c).

En el futuro se desarrollarán modelos de predicción del avance en las carreras utilizando técnicas de árboles de decisión, entre otras.

En relación con la línea 2.4: Se ha analizado el perfil socioeconómico y educativo de los alumnos ingresantes de la FACENA y su relación con su rendimiento académico, medido en términos de su desempeño en la primera asignatura de Matemática, en el primer año de carrera universitaria, utilizando modelos ajustados mediante técnicas estadísticas clásicas (regresión logística) e inteligentes (redes neuronales). Se ha realizado además un estudio comparativo entre ambas metodologías, con el objeto de contrastar el desempeño y la eficiencia de las mismas en el problema de la predicción del rendimiento académico de los estudiantes. Los resultados obtenidos han permitido determinar la vinculación que existe entre el rendimiento académico de los alumnos del primer año de todas las carreras con las condiciones socioeconómicas de los mismos (Porcel, Dapozo y López, 2008) (Porcel, Dapozo y López., 2009) (Porcel, López y Dapozo, 2010).

En relación con la línea 2.5: Se analizó el trayecto de formación de los egresados de Licenciatura en Sistemas de Información, en el periodo 2001-2008, a través de CPM y en base a redes de actividades estocásticas, obteniéndose las duraciones media, mediana y borrosas de la carrera (Cáceres y Caputo, 2009), así como también las probabilidades de ejecución de los distintos trayectos conformados por las diversas opciones de las asignaturas optativas. Se clasificó la población de egresados según la similitud de sus trayectos de formación, utilizando análisis de conglomerados (cluster).

Estos resultados, y los que se obtendrán hasta la culminación del proyecto, aportarán a un mayor conocimiento de los posibles factores que inciden en el desempeño de los alumnos. Esta información permitirá que, desde la gestión institucional, se aborden mecanismos correctivos o superadores que contribuyan al mejoramiento de los índices de desgranamiento, abandono, bajo rendimiento académico y prolongación excesiva de la duración de la carrera, comunes en el ambiente universitario nacional, principalmente notorios en la actuación de los alumnos en el primer año de carrera.

El conocimiento más profundo respecto al rendimiento estudiantil que pueda generarse a partir de este proyecto, representa un aporte para los procesos de evaluación y acreditación de títulos, los cuales se basan en la construcción de indicadores para detectar fortalezas y debilidades de la formación universitaria, por lo que “reflexionar sobre todos los elementos que la evaluación del rendimiento del alumnado proporciona se convierte en un mecanismo claro para la mejora de la calidad del proceso educativo” (Muñoz, 2005).

Asimismo, el conocimiento de las causas que determinan el nivel del rendimiento académico de los estudiantes universitarios, permitirá profundizar la política educativa del gobierno nacional que, mediante la dotación de recursos especiales (como becas, planes de tutorías, entre otros) prioriza la formación en Informática, Ingeniería y Ciencias Básicas, por considerarla fundamental para el desarrollo y crecimiento de la Nación.

4. FORMACION DE RECURSOS HUMANOS

En este proyecto se enfatiza el enfoque interdisciplinario dado que sus integrantes proceden de distintas disciplinas: Matemática, Estadística e Informática. Esto permite un abordaje sistémico de los problemas de la investigación, a la vez que se complementan y enriquecen las distintas miradas disciplinares.

Los alumnos en proceso de formación de la Licenciatura en Matemática y la Licenciatura en Sistemas de Información que se suman al proyecto en calidad de becarios o para la realización del Trabajo Final de Aplicación, requisito académico de la carrera de Sistemas, tienen la oportunidad de aplicar y ampliar sus conocimientos en estas aplicaciones interdisciplinarias. En tal sentido, se pueden mencionar los siguientes Trabajos Finales de Aplicación, dirigidos por integrantes de este grupo de investigadores:

- “Modelos de información de Inteligencia de Negocios para la Gestión Académica Universitaria mediante herramientas open source”.
- “Técnicas de Simulación aplicadas al análisis de la duración real de carreras universitarias”.

Además, se han incorporado al proyecto los siguientes recursos humanos:

- Un becario de Pre-grado de la Secretaría General de Ciencia y Técnica (UNNE). Tema: “Descripción de la trayectoria académica de los estudiantes de la Licenciatura en Sistemas de Información de FACENA - UNNE, mediante redes de actividades estocásticas y programación por camino crítico”. Período: 2009-2010.
- Un becario de Pre-grado de la Secretaría General de Ciencia y Técnica (UNNE). Tema: “Desarrollo de modelos de RNA para la predicción del rendimiento académico de los alumnos de la FACENA (UNNE)”. Período: 2010-2011.

5. BIBLIOGRAFIA

BACALLAO GALLESTEY C., PARAPAR DE LA RISTRA, J., ROQUE GIL M., BACALLOA GUERRA J. “Árboles de regresión y otras opciones metodológicas aplicadas a la predicción del rendimiento académico”. Revista Cubana de Educación Médica Superior, vol.18,

- Nº3. 2004. Disponible en: http://bvs.sld.cu/revistas/ems/vol18_3_04/ems02304.htm
- BORRACCI, R. A., ARRIBALZAGA, E. B. "Aplicación de análisis de conglomerados y redes neuronales artificiales para la clasificación y selección de candidatos a residencias médicas". Educación Médica, Vol 8, Nº 1. ISSN 1575-1813. Barcelona. 2005.
- BRITOS, P. "Procesos de explotación de información basados en sistemas inteligentes". Tesis Doctoral Universidad Nacional de la Plata. La Plata. Buenos Aires. Argentina. 2008.
- BRITOS, P. V. "Minería de Datos". Buenos Aires: Nueva Librería. 2005.
- BRITOS, P., HOSSIAN, A., GARCÍA MARTINEZ, R., SIERRA, E. "Minería de Datos Basada en Sistemas Inteligentes". Nueva Librería. 2005.
- CACERES, B.; CAPUTO, L. "Duración borrosa de la Licenciatura en Sistemas de Información de la FACENA – UNNE". IX Encuentro Regional de Docentes de Matemática. FACENA – UNNE. 2009.
- CASTILLO, E., COBO, A., GUTIÉRREZ, J. M., PRUNEDA, R. E. "Introducción a las Redes Funcionales con Aplicaciones. Un Nuevo Paradigma Neuronal". Editorial Paraninfo S.A. Madrid. España. 1999.
- FAYYAD, U. M., PIATETSKY-SHAPIRO, G., SMYTH, P., UTHURUSAMY, R. "Advances in Knowledge Discovery and Data Mining". AAAI/MIT Press. 1996.
- GARCÍA MARTÍNEZ, R., SERVENTE, M., PASQUINI, D. "Sistemas Inteligentes". Editorial Nueva Librería. Buenos Aires. 2003. En: Britos, 2008.
- GARCÍA, R. "Inferencia Estadística y Diseño de Experimentos". EUDEBA. Buenos Aires. 2004. En: Britos, 2008.
- GONZÁLEZ, D.S. "Detección de alumnos de riesgo y medición de la eficiencia de centros escolares mediante redes neuronales". Biblioteca de Económicas y Empresariales. Servicios de Internet. Universidad Complutense de Madrid. 1999.
- INFORME DE AUTOEVALUACION DE FACENA (UNNE). Facultad de Ciencias Exactas y Naturales y Agrimensura. Universidad Nacional del Nordeste. 1996.
- LEVY MANGIN, J.P.; VARELA MALLOU, J. "Análisis Multivariable para las Ciencias Sociales". Editorial Pearson Prentice Hall. 862 pp. 2003
- MUÑOZ, S. "Indicadores de rendimiento académico del alumnado de la universidad de La Laguna. Jornadas sobre Políticas de Calidad en la Universidad de La Laguna", 2005.
- PÉREZ MACKEPFRANG, C; ALBERTO, C.; CARIGNAGNO, C. Y CASTRO, S. "Programación por Camino Crítico; Introducción al Método: Actividades en los Vértices". Revista de la Escuela de Perfeccionamiento en Investigación Operativa Nº 14. pp. 15 – 28. 1998
- PORCEL E. A., DAPOZO G. N., LÓPEZ M. "Técnicas clásicas de minería de datos aplicadas al estudio del rendimiento académico de alumnos de primer año de carreras de la FACENA". Comunicaciones Científicas y Tecnológicas de la UNNE 2008.
- PORCEL, E. A.; DAPOZO, G. N., LÓPEZ, M. V. "Predicción del rendimiento académico de alumnos de primer año de la FACENA (UNNE) en función de su caracterización socioeducativa". Inédito. 2009.
- PORCEL, E. A.; LÓPEZ, M. V.; DAPOZO, G. N. "Predicción del rendimiento académico de alumnos de primer año de universidad mediante redes neuronales". Inédito. 2010.
- PORCEL, E. A.; LÓPEZ, M. V.; DAPOZO, G. N.; CAPUTO, L. N. "Relación entre el número de exámenes rendidos y el número de asignaturas aprobadas como indicador del rendimiento académico de alumnos universitarios". XX Jornadas de la Escuela de Perfeccionamiento en Investigación Operativa (EPIO). Buenos Aires. Argentina. ISBN: 978-950-42-0113-7. 2009a.
- PORCEL, E. A.; LÓPEZ, M. V.; DAPOZO, G. N.; CAPUTO, L. N. "Probabilidad de aprobar las asignaturas de la carrera de Profesorado en Matemática de la FACENA (UNNE)". Reunión de Comunicaciones Científicas y Tecnológicas 2009 de la UNNE. 2009b.
- PORCEL, E. A.; LÓPEZ, M. V.; DAPOZO, G. N.; CAPUTO, L. N. "Probabilidad de aprobar las asignaturas de las carreras de Ingeniería Eléctrica e Ingeniería en Electrónica de la FACENA (UNNE)". Revista "Investigación Operativa". UTN Facultad Regional Buenos Aires. ISSN: 0329-7322. Nº 31. 2009c.
- PORCEL, E. A.; RAMÍREZ ARBALLO, M. G.; MATA, L. "Comparación del nivel de conocimientos matemáticos de alumnos ingresantes a la carreras de Ingeniería Electricista Electrónica de FaCENA en 2001 y 2009". Reunión de Comunicaciones Científicas y Tecnológicas 2009 de la UNNE. 2009.
- SALGUEIRO, F., COSTA, G., CÁNEPA, S., LAGE, F., KRAUS, G., FIGUEROA, N., CATALDI, Z. "Redes Neuronales para predecir la aptitud del alumno y sugerir acciones". VIII Workshop de Investigadores en Ciencias de la Computación (WICC 2006
- SCHIEFER, J., JENG, J., KAPOOR, S., CHOWDHARY, P. "Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence". Proceedings 2004 IEEE International

Conference on E-Commerce Technology. Pág. 162-169. 2004. En: Britos, 2008.

STEFANOVIC, N., MAJSTOROVIC, V., STEFANOVIC, D. "Supply Chain Business Intelligence Model". Proceedings 13th International Conference on Life Cycle Engineering. Pág. 613-618. 2006. En: Britos, 2008.

TOER, M. "El caso de los ingresantes de 1998 al Ciclo Básico Común de la Universidad de Buenos Aires". CBC, Universidad de Buenos Aires. 2000. Disponible en <http://caraya.cbc.uba.ar/dat/sbc/perfil/perfil.html#1>.

UMAPATHY, K. "Towards Co-Design of Business Processes and Information Systems Using Web Services". Proceedings 40th Annual Hawaii International Conference on System Sciences. Pág. 172-181. 2007. En: Britos, 2008.