

Las Motivaciones para Estudiar y su Relación con el Rendimiento Académico: El Caso de Sistemas Operativos de la LSI de la UNNE

David L. la Red Martínez¹, Julio C. Acosta¹, Valeria E. Uribe¹, Federico Agostini¹,
Alice R. Rambo¹

¹ Grupo de Sistemas Operativos y TICs, Dpto. de Informática,
FACENA, UNNE, 9 de julio 1449,
(3400) Corrientes, Argentina
lrmdavid@exa.unne.edu.ar

Abstract. The relatively low % of students promoted and regularized (academic success) in Operating Systems Cathedra of the Bachelor's Degree in Information Systems (LSI) of the Faculty of Sciences and Natural Surveying (FACENA) of the Northeast National University (UNNE), has motivated an project research which this work is a small part, whose objective is to determine the variables that affect the academic performance, whereas the final status of the student according to the Res. N° 185/03 CD (scheme for evaluation and promotion): promoted, regular or free¹. The variables considered are: status of the student, educational level of parents, secondary education, social and economic level, age, nature, if it works and the attitude towards study and ICTs. This work especially considers the reasons to study expressed by students. Data Warehouse (DW) and Data Mining (DM) techniques were used, to search for profiles of the students and identify situations of success or failure academic.

Keywords: ICTs, Data Warehouse, Data Mining, Clustering, Demographic Cluster, Academic Performance, Student's Profiles, Operating Systems.

Resumen. El relativamente bajo % de alumnos promocionados y regularizados (éxito académico) en Sistemas Operativos de la Licenciatura en Sistemas de Información (LSI) de la Facultad de Ciencias Exactas y Naturales y Agrimensura (FACENA) de la Universidad Nacional del Nordeste (UNNE), motivó un proyecto de investigación del cual este trabajo es una pequeña parte, cuyo objetivo es determinar las variables que inciden en el rendimiento académico, considerando la situación final del alumno según la Res. N° 185/03 CD (régimen de evaluación y promoción): promocionado, regular o libre. Las variables consideradas son: situación del alumno, nivel educacional de los

¹“Promoted” refers to students that to be exempt from the final exam. “Regularized” refers to students that approve the partial examinations; they test theoretical concepts in the final exam. A student is “Free” when he reprovess the partial examinations and he should make again the course or must to test out of the course.

padres, educación secundaria, nivel socio-económico, edad, género, si trabaja y la actitud hacia el estudio y las TICs. En este trabajo se considera especialmente el aspecto relacionado con las razones para estudiar expresadas por los alumnos. Se utilizaron técnicas de Almacén de Datos (Data Warehouses: DW) y de Minería de Datos (Data Mining: DM), para buscar perfiles de los alumnos y determinar situaciones de éxito o de fracaso académico.

Palabras Clave: TICs, Almacén de Datos, Minería de Datos, Clustering, Cluster Demográfico, Rendimiento Académico, Perfiles de Alumnos, Sistemas Operativos.

1 Introducción

Tomando como referencia la información oficial de la UNNE², la carrera de Sistemas de la FACENA ha registrado una matrícula considerable de alumnos de la UNNE (2005: 4,42%; 2006: 3,93%; 2007: 3,82%; 2008: 3,53%; 2009: 3,34% (aspirantes); 2010: 2,79% (aspirantes)); ha sido la carrera de mayor número de alumnos en la FACENA (2005: 37,64%; 2006: 34,77%; 2007: 33,23%; 2008: 30,32%; 2009: 26,47% (aspirantes); 2010: 24,80% (aspirantes)), la de mayor cantidad de nuevos inscriptos en la FACENA (2005: 33,92%; 2006: 29,89%; 2007: 29,71%; 2008: 38,74%; 2009: 23,70%; 2010: 23,30%) y la que más egresados ha producido en la FACENA (2004: 56,05%; 2005: 41,99%; 2006: 44,02%; 2007: 54,30%; 2008: 46,63%). Estos datos demuestran de manera elocuente la importancia de la carrera de Sistemas (Licenciatura en Sistemas de Información: LSI) de la FACENA de la UNNE.

Un análisis más detallado permite observar los relativamente bajos porcentajes de egresados respecto de nuevos inscriptos en la LSI; estos porcentajes varían si se considera solamente el título terminal de grado (Licenciado en Sistemas de Información) o si además se considera el título intermedio (Programador Universitario de Aplicaciones):

- Sin considerar el título intermedio: 2005: 4,81%, 2006: 5,27%, 2007: 9,49%, 2008: 5,42%.
- Considerando el título intermedio: 2005: 21,81%, 2006: 20,22%, 2007: 18,98%, 2008: 15,51%.

Estos porcentajes relativamente bajos en la relación egresados respecto de nuevos inscriptos se observan también considerando a la FACENA y a la UNNE en su totalidad:

- FACENA: 2005: 17,62%, 2006: 13,73%, 2007: 10,39%, 2008: 12,89%.
- UNNE: 2005: 22,57%, 2006: 22,11%, 2007: 20,49%, 2008: 22,36%.

² Fuente: Secretaría General Académica UNNE: <http://www.unne.edu.ar/institucional/>, última consulta: 15/07/10.

Lo señalado en los párrafos precedentes permite afirmar que la relación entre egresados y nuevos inscriptos es en general relativamente baja, y especialmente baja si se considera a la LSI sin los egresados con título intermedio.

Las relativamente bajas tasas de egresados respecto de nuevos inscriptos mencionadas en el apartado anterior, que podríamos considerar el “rendimiento académico global” de una Carrera, Facultad o Universidad, se observan también en numerosas asignaturas de la LSI, considerando “rendimiento académico particular” o simplemente “rendimiento académico”, a los resultados de las evaluaciones de los alumnos durante el cursado de una asignatura, y la condición final lograda por los mismos en el marco de la Res. N° 185/03 CD (régimen de evaluación y promoción): promocionado, regular o libre.

Para la asignatura Sistemas Operativos los valores de los últimos años son los siguientes: alumnos promocionados y regularizados respecto de los que rindieron algún examen parcial: 2006: 21,05%, 2007: 32,89%, 2008: 34,86%, 2009: 13,51%, 2010: 38,39%.

Además se ha observado que un considerable porcentaje de alumnos se inscriben para cursar la asignatura, pero luego no completan el cursado (73,39% en el 2010).

Asimismo, reiteradamente se ha observado una escasa concurrencia de los alumnos a las tutorías establecidas para brindarles apoyo en la realización de las diferentes actividades planificadas (teorías, prácticos, laboratorios).

Ante la situación mencionada se consideró de gran importancia efectuar una investigación que permita determinar las variables que inciden en el relativamente bajo rendimiento académico de los alumnos de Sistemas Operativos de la LSI de la FACENA de la UNNE, identificar los perfiles de alumnos exitosos (los que promocionan o regularizan la asignatura), como así también los perfiles de alumnos que no lo logran (los que quedan en la condición de libres). Una vez determinados los perfiles de alumnos con bajo rendimiento académico, se podrán encarar acciones tendientes a evitar potenciales fracasos académicos. Para la determinación de los perfiles de alumnos se consideró apropiado utilizar técnicas de DW y DM.

Este artículo se ha estructurado de la siguiente manera: primeramente se planteará el objetivo principal del mismo, luego se hará una muy breve revisión de los principales conceptos involucrados en cuanto a DW y DM, seguidamente se indicará brevemente el software utilizado, para continuar con la metodología seguida y la presentación de algunos resultados obtenidos relacionados con las razones para estudiar expresadas por los alumnos, para finalizar con las conclusiones y líneas futuras de acción, los reconocimientos y las referencias.

2 Objetivo Principal

El objetivo principal de este trabajo es encontrar perfiles de alumnos por medio de la aplicación de técnicas de DM a un DW con datos académicos, socio económicos y demográficos correspondientes a alumnos de SO de la Licenciatura en Sistemas de Información (LSI) de la FACENA de la UNNE, especialmente desde la perspectiva de las razones para estudiar expresadas por los alumnos.

3 Revisión Conceptual

Un DW es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales [1], [2]. Es también un conjunto de datos integrados orientados a una materia, que varían con el tiempo, y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración [3], [4], [5], [6].

La DM es la etapa de descubrimiento en el proceso de KDD (Knowledge Discovery from Databases), es el paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos preprocesados [7], [8], [9], [10].

Es también un mecanismo de explotación, consistente en la búsqueda de información valiosa en grandes volúmenes de datos. Está muy ligada a los DW ya que los mismos proporcionan la información histórica con la cual los algoritmos de minería obtienen la información necesaria para la toma de decisiones [11], [12].

La DM es un conjunto de técnica de análisis de datos que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y extraer patrones y tendencias para predecir comportamientos futuros [5], [13], [14], [15].

4 Software Utilizado

Se ha utilizado el IBM Data Warehouse Edition (DWE) V.9.5, que incluye al DB2 Enterprise Server Edition (DB2 ESE), al Design Studio (DS) y al Intelligent Miner (IM), que se han obtenido de la empresa IBM Argentina S.A. en el marco de la Iniciativa Académica de dicha empresa y de los Acuerdos realizados entre la misma y la FACENA de la UNNE (Acuerdo del 18/06/04 D, Res. N° 1417/04 D, Res. N° 858/06 CD).

5 Metodología Seguida

El presente estudio se realizó sobre datos obtenidos mediante encuestas realizadas al alumnado de SO, considerando además los resultados de las distintas instancias de evaluación previstas durante el cursado de dicha asignatura. Se utilizó un entorno integrado de gestión de bases de datos y data warehouse (DB2 versión 9.5), obtenido de la empresa IBM mediante los Acuerdos firmados entre dicha empresa y la UNNE; dicho entorno permite la extracción de conocimiento en bases de datos y DW mediante técnicas de DM como ser clustering (o agrupamiento de datos) que consiste en la partición de un conjunto de individuos en subconjuntos lo más homogéneos posible, el objetivo es maximizar la similitud de los individuos del cluster y maximizar la diferencia entre clusters. El cluster demográfico es un algoritmo desarrollado por IBM e implementado en el IM, componente del DWE, entorno antes mencionado, que resuelve automáticamente los problemas de definición de métricas

de distancia / similitud, proporcionando criterios para definir una segmentación óptima [16], [17], [18], [19].

Los pasos realizados durante el presente trabajo han sido los siguientes:

- Recolección de los datos.
- Tratamiento y depuración de los datos.
- Preparación de la base de datos y del DW correspondiente sobre la plataforma de trabajo seleccionada.
- Selección de la técnica de minería de datos para la realización del estudio (predominantemente clustering).
- Generación de diferentes gráficos para el estudio de los resultados.
- Estudio de los resultados obtenidos.
- Obtención de las conclusiones.

En esta etapa se trabajó con una porción (Data Mart: DMa) del DW, cuya estructura se muestra en la Fig. 1.

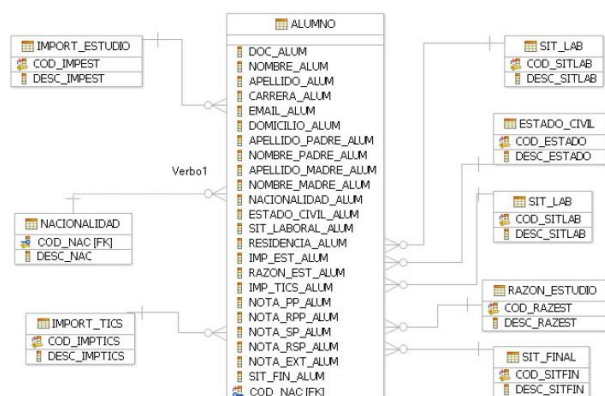


Fig. 1. Estructura del DMa utilizado, parte del DW.

En la minería de clustering se agrupó el total de la población (173 casos) en cuatro clases: 1) alumnos que no responden; 2) alumnos que estudian para aprender integralmente; 3) alumnos que estudian para aprender a aprender; 4) alumnos que estudian para aprobar; 5) alumnos que tienen otra motivación para estudiar. Se analizó la composición de cada una de estas clases en aspectos tales como: a) importancia otorgada al estudio; b) estado civil; c) opinión acerca de las TICs y d) resultados obtenidos en la asignatura Sistemas Operativos.

6 Resultados Obtenidos

Se obtuvieron diferentes clasificaciones mediante la utilización (preferentemente) de técnicas de clustering, según diferentes criterios de agrupación de los datos.

Se utilizó la Tabla 1 de nombres y significados de variables.

Tabla 1. Siglas y nombres de variables.

Nombre de Variable	Significado
SIT_LABORAL_ALUM	Situación laboral del alumno
IMP_EST_ALUM	Importancia dada al estudio por el alumno
RAZÓN_EST_ALUM	Razón para estudiar según el alumno
IMP_TICS_ALUM	Importancia dada a las TICs por el alumno
NOTA_PP_ALUM	Nota primer parcial
NOTA_RPP_ALUM	Nota recuperatorio primer parcial
NOTA_SP_ALUM	Nota segundo parcial
NOTA_RSP_ALUM	Nota recuperatorio segundo parcial
NOTA_EXT_ALUM	Nota recuperatorio extraordinario
SIT_FIN_ALUM	Situación final del alumno luego del cursado

De un primer agrupamiento surge que un 49,71% de los alumnos no respondieron a la pregunta respecto de las razones para estudiar, un 28,32% dice estudiar para aprender integralmente, un 15,03% de los alumnos dice que estudia para aprender a aprender, un 5,20% de los alumnos afirman que estudian para aprobar y un 1,74% dice tener otra motivación para estudiar (Fig. 2 y Fig. 3).

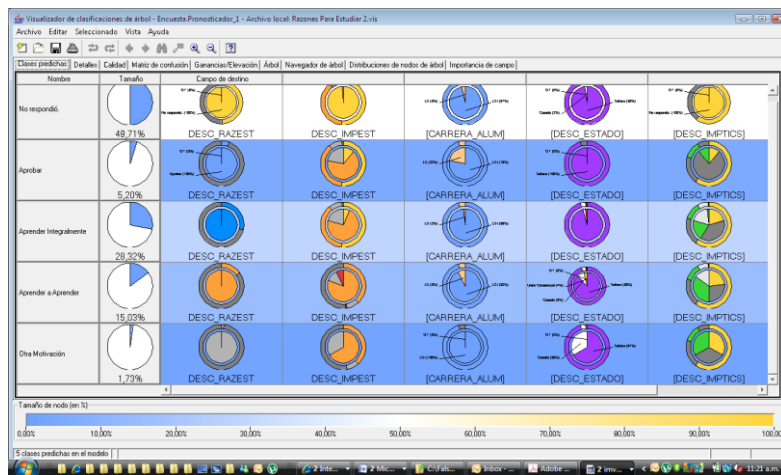


Fig. 2. Razones para estudiar - 1.

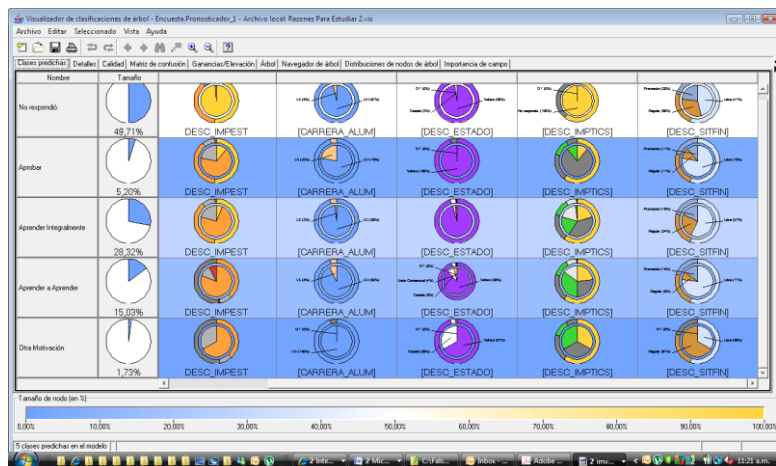


Fig. 3. Razones para estudiar - 2.

Los valores de calidad del modelo de minería de datos utilizado son los siguientes (Fig. 4):

- Calidad del modelo: 0,764 (mide la calidad general del modelo).
- Precisión: 0,988 (mide la probabilidad de que sea correcta una predicción de la clase predicha).
- Clasificación: 0,942 (mide la capacidad del modelo para ordenar correctamente registros en base a propiedades predichas).

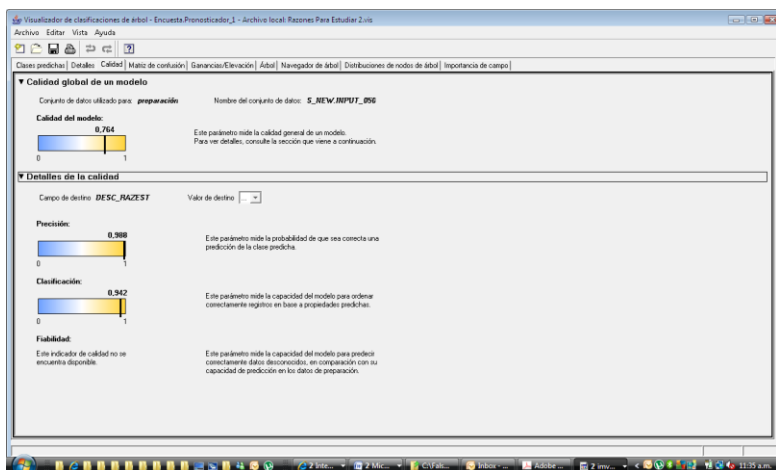


Fig. 4. Calidad del modelo de minería de datos.

De los alumnos que *no respondieron* la pregunta relativa a las razones para estudiar (49,71% de la población):

La distribución de la importancia otorgada al estudio es la siguiente: el 99% de este grupo no contestó la pregunta; el 1% manifestó que otorga al estudio más importancia que a la diversión.

El 98% de este grupo son solteros y el 2% casados.

Respecto de la importancia concedida a las TICs, el 100% no respondió.

La situación final del alumno lograda por quienes no respondieron la pregunta relativa a las razones para estudiar es como sigue: se observa que el 23% de este grupo promocionó la asignatura, el 30% quedó en condición de alumno regular y 47% quedó en condición de alumno libre.

De los alumnos que dijeron *estudiar para aprender integralmente* (un 28,32% de la población), se observa:

Respecto de la importancia otorgada al estudio, en contraste con el grupo anterior, se observa que sólo el 6% no respondió, el 21% dice otorgar más importancia al estudio que al trabajo y el 73% más que a la diversión.

Al analizar el estado civil en este grupo, los totales se distribuyen de la siguiente manera: solteros el 96%, casados el 2% y divorciados legalmente un 2%.

La actitud de estos alumnos frente a las TICs presenta los siguientes resultados: el 39% considera que facilitan los estudios, el 20% que será imprescindible su dominio, el 18% que son una realidad, el 3% que están de moda; el 20% no respondió.

La situación final del alumno lograda por quienes dicen estudiar para aprender integralmente es la siguiente: el 19% de los alumnos de esta clase promovió la asignatura, el 24% de esta clase regularizó y el 57% quedó en condición de alumno libre.

De los alumnos que dijeron *estudiar para aprender a aprender* (un 15,03% de la población):

La distribución de la importancia otorgada al estudio es: el 81% dice otorgar más importancia al estudio que a la diversión, el 12% más que al trabajo y el 7% más que a la familia. Es de destacar que en esta clase el porcentual de abstención en la respuesta es nulo.

El estado civil de este grupo es como sigue: el 88% soltero, el 8% casado y el 4% en unión consensual.

La actitud de estos alumnos frente a las TICs muestra los siguientes resultados: el 27% considera que facilitan los estudios, el 35% que será imprescindible su dominio, el 15% que son una realidad y el 23% no respondió.

La situación final de los alumnos que dicen estudiar para aprender a aprender es la siguiente: el 15% Promovió, el 8% Regularizó y el 77% de los alumnos quedó Libre.

De los alumnos que dijeron *estudiar para aprobar* (un 5,20% de la población):

La distribución de la importancia otorgada al estudio es: el 67% dice otorgar más importancia al estudio que a la diversión, el 22% más que al trabajo; el 11% no respondió.

El estado civil de este grupo es 100% soltero.

La actitud de estos alumnos frente a las TICs muestra los siguientes resultados: el 78% considera que facilitan los estudios, el 11% que será imprescindible su dominio, el 11% no respondió.

La situación final de los alumnos que dicen estudiar para aprobar es la siguiente: el 11% Promovió, el 11% Regularizó y el 78% de los alumnos quedó Libre.

De los alumnos que dijeron *estudiar por otra motivación* (un 1,73% de la población):

La distribución de la importancia otorgada al estudio es: el 67% dice otorgar más importancia al estudio que a la diversión y el 33% más que al trabajo.

El estado civil de este grupo es 67% soltero y 33% casado.

La actitud de estos alumnos frente a las TICs indica los siguientes resultados: el 33% considera que facilitan los estudios, el 33% que será imprescindible su dominio, el 34% no respondió.

La situación final de los alumnos que dicen estudiar por otra motivación es como sigue: el 67% Regularizó y el 33% de los alumnos quedó Libre.

7 Conclusiones y Líneas Futuras

Se ha podido comprobar las grandes ventajas de la utilización de tecnologías y software de última generación que soportan sistemas multiplataforma.

Se hace notar que los resultados mostrados corresponden a una parte de la etapa preliminar de diversos estudios que se están realizando.

Se destaca que los mayores porcentajes de “éxito académico” (regularizados + promocionados) corresponden a las siguientes categorías *respecto de las razones para estudiar*: aprender integralmente (42% de éxito académico), aprender a aprender (23% de éxito académico), aprobar (22% de éxito académico), otras motivaciones (67% de éxito académico (todos regularizados, ninguno promocionado)), no respondieron (53% de éxito académico).

Se tiene previsto desarrollar las siguientes líneas futuras de acción:

- Avanzar en la investigación con la utilización de minería de datos como ser redes neuronales, redes bayesianas, arboles de decisión, etc., aplicadas al almacén de datos utilizado hasta ahora generalmente con las técnicas de clustering.
- Aplicar las técnicas de minería de datos utilizadas, pero sobre otras bases de datos de alumnos de otras asignaturas y carreras para comparar los resultados obtenidos.

Reconocimientos. El presente trabajo se encuadra en el Proyecto de Investigación “El Desigual Aprovechamiento de las TICs en el Proceso de Enseñanza – Aprendizaje de los Sistemas Operativos en la FACENA de la UNNE”, acreditado por la Secretaría de Ciencia y Técnica de la UNNE como PI-120-07 (Res. N° 369/08 CS).

Referencias

1. Kubski, M.: Aplicación Orientada al Descubrimiento de Conocimiento en Bases de Datos. Trabajo Final de Aplicación de la Licenciatura en Sistemas de Información dirigido por el Prof. David Luis la Red Martínez. Corrientes. Argentina. (2005)

2. Cutro, A.: Minería de Datos Aplicada a la Encuesta Permanente de Hogares. Trabajo Final de Aplicación de la Licenciatura en Sistemas de Información dirigido por el Prof. David Luis la Red Martínez. Corrientes. Argentina. (2008)
3. Inmon, W. H.: Data Warehouse Performance. John Wiley & Sons. USA. (1992)
4. Inmon, W. H.: Building the Data Warehouse. John Wiley & Sons. USA. (1996)
5. Simon, A.: Data Warehouse, Data Mining and OLAP. John Wiley & Sons. USA. (1997)
6. Trujillo, J. C., Palomar M. & Gómez, J.: Applying Object-Oriented Conceptual Modeling Techniques To The Design of Multidimensional Databases and OLAP Applications. First International Conference On Web-Age Information Management (WAIM'00). Lecture Notes in Computer Science 1846:83-94. (2000)
7. Fayyad, U.M., Grinstein, G. & Wierse, A.: Information Visualization in Data Mining and Knowledge Discovery. Morgan Kaufmann. Harcourt Intl. (2001)
8. Fayyad, U.M., Piatetskiy-Shapiro, G., Smith, P., Ramasamy, U.: Advances in Knowledge Discovery and Data Mining. AAAI Press / MIT Press. USA. (1996)
9. Han, J. & Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann. (2001)
10. Hand, D.J., Mannila, H. & Smyth, P.: Principles of Data Mining. The MIT Press. USA. (2000)
11. Gutiérrez, J. M.: Data Mining, Extracción de Conocimiento en Grandes Bases de Datos. España. (2001)
12. IBM Software Group.: Enterprise Data Warehousing whit DB2: The 10 Terabyte TPC-H Benchmark. IBM Press. USA. (2003)
13. Berson, A. & Smith, S. J.: Data Warehouse, Data Mining & OLAP. Mc Graw Hill. USA. (1997)
14. Frawley, W. J., Piatetsky-Shapiro, G & Matheus, Ch. J.: Knowledge Discovery in Database An Overview. AI Magazine. (1992)
15. White, C. J.: IBM Enterprise Analytics for the Intelligent e-Business. IBM Press. USA. (2001)
16. Grabmeier, J. & Rudolph, A.: Techniques of Cluster Algorithms in Data Mining version 2.0. IBM Deutschland Informationssysteme GmbH. GBIS (Global Business Intelligence Solutions). Germany. (1998)
17. Baragoin, C., Chan, R., Gottschalk, H., Meyer, G., Pereira, P. & Verhees, J.: IBM International Technical Support Organization Enhance Your Business Applications. Simple Integration of Advanced Data Mining Functions. IBM Press. (2002)
18. Ballard, Ch., Rollins, J., Ramos, J., Perkins, A., Hale, R., Dorneich, A., Cas Milner, E. & Chodagam, J.: Dynamic Warehousing: Data Mining Made Easy. IBM International Technical Support Organization. IBM Press. USA. (2007)
19. Ballard, Ch., Beaton, A., Chiou, D., Chodagam, J., Lowry, M., Perkins, A., Phillips, R. & Rollins, J.: Leveraging DB2 Data Warehouse Edition for Business Intelligence. IBM International Technical Support Organization. IBM Press. USA. (2006)