

Tipo de artículo: Artículo original
Temática: Ingeniería y Gestión de software
Recibido: 7/06/2013 | Aceptado: 4/09/2013

Descubrimiento de conocimiento en lecciones aprendidas documentadas en los procesos de cierre de proyectos informáticos

Knowledge discovery in lessons learned documented of closure processes in IT projects

Eliana Bárbara Ril Valentin^{1*}, Rafael Rodríguez Puente², Pedro Y. Piñero Pérez³ Hugo A. Martínez Noriegas²

¹ Facultad 2. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños km, 2 ½. Torrens, Boyeros, La Habana, Cuba. CP.:19370

² Facultad 3. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños km, 2 ½. Torrens, Boyeros, La Habana, Cuba. CP.:19370

³ Facultad 5. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños km, 2 ½. Torrens, Boyeros, La Habana, Cuba, CP.:19370

*Autor para la correspondencia: ebril@uci.cu

Resumen

La documentación de lecciones aprendidas permite visualizar los errores del pasado y mejorar el trabajo futuro. Aprender de la experiencia de proyectos anteriores contribuye a disminuir los riesgos, evitar problemas ya identificados y reutilizar las buenas prácticas para reducir el número de proyectos no exitosos. El objetivo de esta investigación es descubrir conocimiento a partir de la aplicación de árboles de decisión y tablas de contingencia en lecciones aprendidas documentadas en los procesos de cierre de proyectos informáticos, para aumentar el conocimiento respecto a la identificación de buenas prácticas y posibles problemas, así como la asociación entre buenas prácticas y problemas en la ejecución de los proyectos. Finalmente, se exponen los resultados derivados de la aplicación de árboles de decisión y tablas de contingencia a las lecciones aprendidas documentadas en 48 proyectos informáticos en proceso de cierre.

Palabras clave: Descubrimiento de conocimiento, lecciones aprendidas, procesos de cierre, proyectos informáticos.

Abstract

The documentation of learned lessons allows us to see past mistakes and improve future work. Learning from experience of previous projects help us to reduce risks, to avoid problems already identified and to reuse best practices to reduce the number of unsuccessful projects. The goal of this research is to discover knowledge from the application of decision trees and contingency tables on documented lessons learned in closure processes of informatics projects, to increase knowledge regarding the identification of good practice, potential problems and the relationship between best practices and problems in development projects. Finally, the results of applying decision trees and contingency tables on documented lessons learned from 48 IT projects in the closure process are shown.

Keywords: knowledge discovery, lessons learned, closing process, IT project.

Introducción

Los conocimientos y experiencias de los equipos de proyectos constituyen un valioso activo para las organizaciones que buscan mejorar sus prácticas. Un estudio realizado por Alan Murphy y Ann Ledwith a través de un cuestionario enviado a 200 pequeñas y medianas empresas de alta tecnología en Irlanda, muestra que 4.18 (siendo 5 el mayor valor) apoya la afirmación que experiencias previas son un factor clave para dirigir un proyecto (Murphy and Ledwith, 2007). Sin embargo, los enfoques existentes para gestionar dicha experiencia se basan esencialmente en la creación y mantenimiento de repositorios de experiencias; pero no prescriben la manera ni el momento en que los diferentes procesos de gestión del conocimiento deben llevarse a cabo (Matturro, 2010).

Por otra parte, el manejo de grandes volúmenes de información que permita conocer el entorno y predecir su evolución, supone un enorme reto para las organizaciones. Tomando en consideración la importancia de extraer los conocimientos "perdidos" en los datos que almacenan las organizaciones, ha surgido desde hace un tiempo lo que se conoce como minería de datos; la cual se ha utilizado dentro de las organizaciones con el fin de realizar exploración y análisis de datos enfocados en el descubrimiento de conocimiento (Reyes, 2009). En este escenario, con frecuencia se cuenta con un gran número de datos y se quiere aprender y descubrir la relación entre los mismos a partir de la aplicación de técnicas de minería de datos (Ojeda, 2010).

Dado que la minería de datos es la etapa más distintiva dentro del proceso de descubrimiento de conocimiento, muchas veces se utiliza esta etapa para nombrar todo el proceso. En este sentido, algunos autores (Kantardzic, 2003; Larose, 2005) tratan como sinónimos ambos términos, sin embargo conceptualmente en este trabajo se evidencia que existen claras diferencias, ya que la minería de datos es solo una etapa dentro del proceso de descubrimiento de conocimiento (Valcarcel, 2004).

Entre la literatura dedicada al tema de descubrimiento de conocimiento, se pueden encontrar varias definiciones para este término. Según (López, 2007; Molina, 2002) el descubrimiento de conocimiento es el proceso de extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de un conjunto de datos. Incluye no solo el análisis inteligente de los datos con técnicas de minería de datos, sino también los pasos previos, como el filtrado y preprocesamiento de los datos (Gervilla *et al.*, 2009) y los posteriores, como la interpretación, validación y postprocesamiento del conocimiento extraído (Durán and Costaguta, 2007; Gómez, 1998).

Como parte de la identificación de la esencia del problema de esta investigación, se diseñó un instrumento en forma de encuesta. El mismo fue aplicado a miembros de equipos de dirección de proyectos informáticos (18 jefes de proyectos, 13 especialistas de planificación de proyectos, 10 especialistas económicos de proyectos y 11 líderes de subproyectos) con 3,5 años de experiencia en la gestión de proyectos como promedio y 1,5 años como promedio en dicho rol, obteniendo los siguientes resultados:

- El 60 % no conocía antes de ocupar su rol, problemas y buenas prácticas más frecuentes durante la gestión de proyectos anteriores.
- El 69 % refiere que las lecciones aprendidas del equipo de dirección durante la gestión de un proyecto solo se transfieren entre los integrantes del mismo proyecto.
- El 98% reconoce que no se documentan las lecciones aprendidas como parte de las actividades que se realizan en el cierre de un proyecto.

Por otra parte, mundialmente se han creado entidades encargadas de analizar y recopilar las mejores prácticas en proyectos, permitiendo la creación de estándares a nivel internacional, para minimizar los errores más frecuentes en la gestión de los mismos. Entre estas entidades sobresalen debido a la cantidad de referencias que poseían en la literatura estudiada, así como el alcance que tienen a nivel mundial a partir de las certificaciones que promueven las siguientes:

- *Association for Project Management* (Asociación para la Gestión de Proyectos, APM por sus siglas en inglés). Esta institución promueve la metodología APMBOK.
- U.K.'s *Office of Government Commerce* (Oficina de Comercio del Gobierno del Reino Unido, OGC por sus siglas en inglés). Esta entidad promueve Prince2, una metodología genérica enfocada en los productos a entregar.
- *Project Management Institute* (Instituto de Gestión de Proyectos, PMI por sus siglas en inglés). Esta organización difunde el modelo más reconocido en el mundo sobre Gestión de Proyectos, PMBOK (Gomez, 2010).

Luego del análisis realizado de metodologías tan generalizadas como APMBOK, Prince2 y PMBOK, se pudo concluir que no aparecen estudios relacionados con el descubrimiento de conocimiento en la Gestión de Proyectos, ni referencias en el uso de técnicas de minería de datos en lecciones aprendidas de proyectos informáticos. Es por ello que teniendo en cuenta la problemática descrita anteriormente, se puede afirmar que las insuficiencias en el descubrimiento de conocimiento durante los procesos de cierre de los proyectos informáticos, está afectando la gestión del conocimiento en los mismos.

Materiales y métodos

En la Figura 1 se muestran las etapas esenciales que componen el proceso de descubrimiento de conocimiento en base de datos que servirá de base a la propuesta realizada en el presente trabajo.

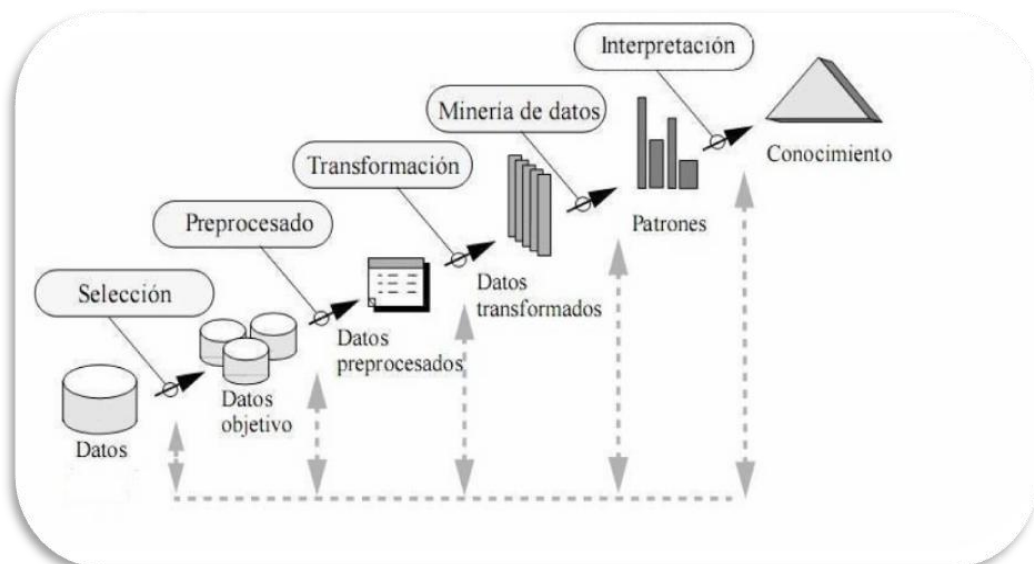


Figura 1. Etapas para el proceso de extracción de conocimiento (Fayyad *et al.*, 1996; Gómez, 1998)

Selección de los datos

En el marco de esta investigación los datos de los que se parte para realizar el proceso de descubrimiento de conocimiento, serán las lecciones aprendidas. Las mismas son conocimientos derivados de la experiencia que pueden

ser positivas o negativas; por ejemplo: problemas, buenas prácticas y oportunidades de mejora en el desarrollo de software.

Para realizar la selección de los datos, teniendo como base los resultados de la encuesta realizada, se utilizó el procedimiento de Análisis de experiencias presentado en (Ril, 2012), con el objetivo de obtener las lecciones aprendidas por cada miembro del proyecto, discutir las y documentarlas. Este procedimiento fue aplicado a 48 proyectos informáticos en proceso de cierre. Identificándose un total de 120 lecciones aprendidas que han servido para la toma de decisiones proactivas en los proyectos que se encuentran actualmente en ejecución. Dichas lecciones además se utilizan como referencia en la impartición del curso básico de Gestión de proyectos en la Maestría del mismo nombre que promueve la Universidad de Ciencias Informáticas y conforman la entrada para el preprocesado de los datos.

Preprocesado de los datos

Las técnicas de minería de datos son sensibles a la calidad de la información sobre la que se pretende extraer conocimiento. Cuanto mayor sea esta calidad, mayor será la calidad de los modelos de toma de decisiones generados a partir de dicha información. En este sentido, la obtención de información útil para ser posteriormente procesada es un factor clave. Aparece por tanto en el proceso de descubrimiento una etapa de preprocesamiento de datos previa a la Minería (Cano, 2004; Kim *et al.*, 2003).

El preprocesado consiste en estudiar los datos seleccionados y aplicar técnicas de limpieza a los mismos para entender el significado de los atributos, detectar errores de integración, estandarizar datos, hacer agrupaciones, etc. (González, 2006). Los datos se filtran de forma que se eliminan valores incorrectos, no válidos o desconocidos y se obtienen muestras de los mismos en busca de una mayor velocidad de respuesta del proceso.

Para el preprocesado de los datos se propone lo siguiente:

- Seleccionar las variables más influyentes en el problema.
- Seleccionar los atributos relevantes.
- Eliminar o corregir los datos incorrectos.
- Seleccionar una estrategia con los datos incompletos.

Transformación de los datos

Luego del preprocesado de los datos, se transforman las lecciones aprendidas a un formato común, que pudiera ser un libro de cálculo o haciendo uso de un programa estadístico. Se recomienda el uso del SPSS versión 13.0 para Windows.

Posteriormente se deben cuantificar los atributos de las variables para facilitar el uso de técnicas que requieren tipos de datos específicos. Es por ello, que se hace necesario reemplazar por enteros los valores de la variable proyecto y se definen para el caso de las variables dicotómicas problemas y buenas prácticas, 1 cuando “se presenta” el problema o la buena práctica y 0 cuando “no se presenta” el problema o la buena práctica.

Minería de datos

El minado de datos es la etapa principal del proceso de descubrimiento de conocimiento a partir de los datos. Durante esta fase, se aplican las diferentes técnicas de extracción de conocimiento, a los datos que ya han sido preprocesados y transformados en etapas anteriores.

En el presente trabajo se han seleccionado las siguientes técnicas:

- Árboles de decisión. Los modelos de árboles de decisión son comúnmente usados en la minería de datos para examinar los datos e inducir las reglas para realizar predicciones (Porcel *et al.*, 2010; Riquelme *et al.*, 2009; Sarango, 2012). Constituyen una de las principales técnicas de minería de datos (Solarte Martínez and Soto Mejía, 2011). Son fáciles de usar, tolerantes al ruido, a atributos no significativos y a valores faltantes. Permiten además reducir la cantidad de variables para realizar el análisis. Representan una gran ventaja respecto a las demás técnicas de clasificación, ya que permiten representar el conocimiento extraído en un conjunto de reglas. Se encuentran implementados en multitud de sistemas, muchos de ellos gratuitos, como: PSPP, Weka o Rapidminer, entre otros.
- Tablas de contingencia. Cuando se trabaja con variables categóricas, los datos suelen organizarse en tablas de doble entrada en las que cada entrada representa un criterio de clasificación (una variable categórica). Como resultado de esta clasificación, las frecuencias (el número o porcentaje de casos) aparecen organizadas en casillas que contienen información sobre la relación existente entre ambos criterios. A estas tablas de frecuencia se les llama tablas de contingencia (Navarro *et al.*, 2010).

La selección de estas técnicas ha estado motivada porque se pretende encontrar si existe asociación entre las variables. Se ha comprobado que estos estadísticos son los más adecuados por la naturaleza de los datos de esta investigación. De esta forma, se pudiera concluir cuáles problemas están asociados con determinadas buenas prácticas.

Interpretación de los datos

En esta etapa se interpretan el conjunto de reglas derivadas de la aplicación de árboles de decisión y las tablas de contingencia obtenidas. Se considera como interpretación de las reglas la reescritura de las mismas en lenguaje natural para su mejor análisis.

Conocimiento

Por último, se hace necesario una base de datos para el almacenamiento y recuperación de la información básica de los proyectos terminados y el conocimiento extraído en este trabajo. Dicha base de datos normalizada, almacena datos primarios sobre los proyectos: nombre abreviado, año en que inició el proyecto y cliente. Además se guardan los subproyectos por cada uno de los proyectos, así como los problemas presentados, las buenas prácticas y oportunidades de mejoras. Para el diseño de dicha base de datos se propone utilizar el diagrama de clases persistentes.

Resultados y discusión

En este epígrafe se muestran los resultados obtenidos a partir de la aplicación del método seleccionado para descubrir conocimiento en las lecciones aprendidas documentadas en los procesos de cierre de proyectos informáticos.

Preprocesado de los datos

Como parte del preprocesado, se seleccionaron las variables más influyentes en el problema: proyecto, subproyectos, año, problemas y buenas prácticas; así como sus atributos más relevantes. Por último, se corrigieron el 100 % de los datos incompletos.

Luego de realizar el preprocesamiento de los datos, se redujo las lecciones aprendidas. Quedando finalmente 32 problemas y 29 buenas prácticas.

Transformación de los datos

Las lecciones aprendidas documentadas como parte de la selección de los datos, se encontraban en los dictámenes finales de análisis de experiencia realizados a los 48 proyectos informáticos en proceso de cierre. En esta etapa se transformaron dichas lecciones a un mismo formato (haciendo uso del SPSS versión 13.0 para Windows) y se reemplazaron por enteros los valores de las variables proyecto, problemas y buenas prácticas.

Minería de datos

A partir de las lecciones aprendidas (problemas y buenas prácticas) preprocesadas y los años en que iniciaron los proyectos (2006, 2007 y 2008) se estudiaron 48 casos y se definieron 62 variables categóricas, como se muestra en la Tabla 1.

Tabla 1. Variables consideradas en el análisis.

Cantidad de variables	Medida
Años	ordinal
32 problemas	nominal
29 buenas prácticas	nominal

Utilizando el SPSS versión 13.0, se construyeron 4 árboles de decisión con el método de crecimiento de Detección automática de interacciones mediante Chi cuadrado (CHAID, por sus siglas en inglés). Exactamente, se utilizó CHAID Exhaustivo, una modificación de CHAID que examina todas las posibles asociaciones para cada predictor (Vallejo and Tenelanda, 2012) con el objetivo de reducir la cantidad de variables a analizar.

En este sentido, uno de los árboles construidos se realizó con el objetivo de encontrar los problemas y buenas prácticas más asociados a los proyectos que iniciaron en un determinado año; tomando como variable dependiente año y como posibles variables predictoras 32 problemas y 29 buenas prácticas, a través de la validación cruzada. Se personalizaron los criterios de crecimiento del árbol (máximo número de niveles del árbol igual a 10, el número de casos mínimo de nodos parentales igual a 5 y el número de casos mínimo de nodo filial igual a 4). Obteniéndose un árbol como se muestra en la Figura 2, con 3 niveles de profundidad y 9 nodos, de ellos 5 hojas o nodos terminales que resultaron ser puros.

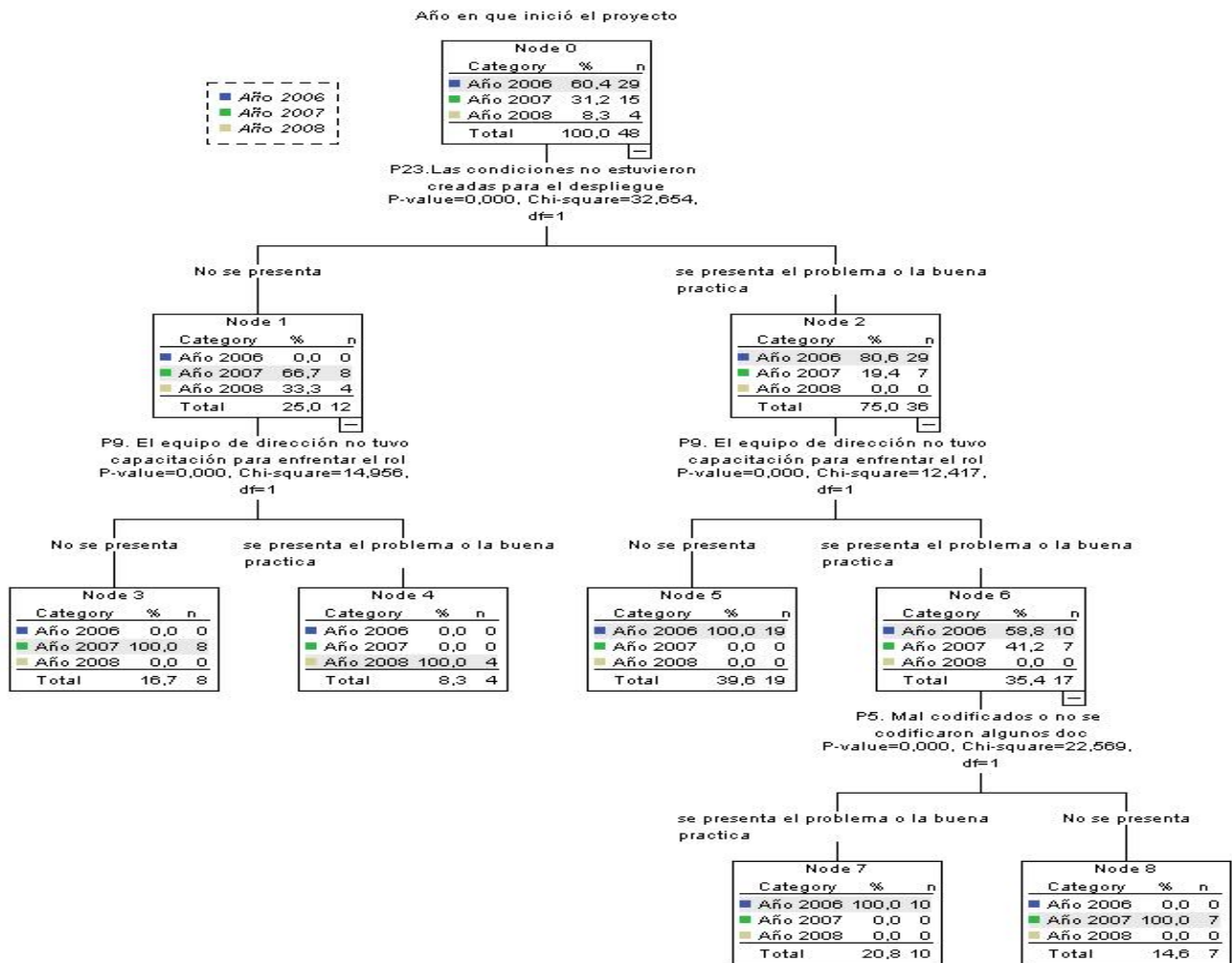


Figura 2. Árbol que asocia el año de inicio de los proyectos con problemas y buenas prácticas.

A partir del árbol de decisión obtenido, además de segmentar la población, se generaron reglas de clasificación. Las diferentes trayectorias de dicho árbol conducen a 5 nodos terminales. Asociados con cada uno de dichos nodos, se define una regla que predice el año en que iniciaron los proyectos:

/* Node 3 */

IF (P23.Las condiciones no estuvieron creadas para el despliegue = "No se presenta") AND (P9. El equipo de dirección no tuvo capacitación para enfrentar el rol != "se presenta el problema o la buena práctica")

THEN Node = 3

Prediction = 7

Probability = 1.000000

/* Node 4 */

IF (P23.Las condiciones no estuvieron creadas para el despliegue = "No se presenta") AND (P9. El equipo de dirección no tuvo capacitación para enfrentar el rol = "se presenta el problema o la buena práctica")

THEN Node = 4

Prediction = 8

Probability = 1.000000

/* Node 5 */

IF (P23.Las condiciones no estuvieron creadas para el despliegue != "No se presenta") AND (P9. El equipo de dirección no tuvo capacitación para enfrentar el rol != "se presenta el problema o la buena práctica")

THEN Node = 5

Prediction = 6

Probability = 1.000000

/* Node 7 */

IF (P23.Las condiciones no estuvieron creadas para el despliegue != "No se presenta") AND (P9. El equipo de dirección no tuvo capacitación para enfrentar el rol = "se presenta el problema o la buena práctica") AND (P5. Mal codificados o no se codificaron algunos doc != "No se presenta")

THEN Node = 7

Prediction = 6

Probability = 1.000000

/* Node 8 */

IF (P23.Las condiciones no estuvieron creadas para el despliegue != "No se presenta") AND (P9. El equipo de dirección no tuvo capacitación para enfrentar el rol = "se presenta el problema o la buena práctica") AND (P5. Mal codificados o no se codificaron algunos doc = "No se presenta")

THEN Node = 8

Prediction = 7

Probability = 1.000000

Además se construyeron otros 3 árboles de decisión, tomando como variables dependientes las variables con mayor asociación en el árbol expuesto anteriormente, es decir, los problemas:

- P23. Las condiciones no estuvieron creadas para el despliegue.
- P9. El equipo de dirección no tuvo capacitación para enfrentar el rol.
- P5. Mal codificados o no se codificaron algunos documentos.

Por otra parte, respecto a las tablas de contingencia, el grado de relación existente entre dos variables no puede ser establecido simplemente observando las frecuencias (aunque pudiera ser útil como un primer paso en el estudio de la relación entre las dos variables), ya que estos porcentajes no permiten cuantificar o probar esa relación. Es por ello, que para determinar si dos variables están relacionadas se debe utilizar alguna medida de asociación, preferiblemente acompañada de su correspondiente prueba de significación.

Entre la amplia variedad de procedimientos estadísticos diseñados para evaluar el grado de asociación existente entre dos variables se ha valorado el uso de otros estadígrafos que complementarían y corroborarían los resultados obtenidos, como son: Chi cuadrado, el Test exacto de Fisher y la V de Cramer.

En ese sentido, se construyeron 10 tablas de contingencia con 48 casos válidos (sin casos perdidos), generadas para cada par de variables que resultaron asociadas a partir de la aplicación de la técnica árboles de decisión. A continuación se muestran en la Tabla 2, los niveles de significación exacta de los estadígrafos Chi cuadrado y el Test exacto de Fisher, así como el valor de la V de Cramer para cada par de variables, problemas (P) y buenas prácticas (BP).

Tabla 2. Resultados de los estadígrafos.

<u>Tablas</u>	<u>Variables</u>	<u>Chi cuadrado</u> <i>Exact Sig.</i>	<u>Test exacto de Fisher</u> <i>Exact Sig.</i>	<u>V de Cramer</u> <i>Value</i>
1	P23-BP20	0,000	0,000	,591
2	P23-BP10	,510	,510	,121
3	P14-BP20	1,000	1,000	,030
4	P14-BP10	1,000	,715	,056
5	P9-BP20	,369	,369	,163
6	P9-BP10	0,000	0,000	1,000
7	P6-BP20	,637	,637	,123
8	P6-BP10	0,000	0,000	,531
9	P5-BP20	,637	,637	,123
10	P5-BP10	0,000	0,000	,531

Interpretación de los datos

En esta etapa se puede concluir, que los problemas que caracterizaron a los proyectos que iniciaron en un determinado año, según el árbol obtenido fueron:

Año 2006:

- En 29 proyectos (representando el 60,4% de la muestra) se presentó el problema P23 (Las condiciones no estuvieron creadas para el despliegue).
- En 29 proyectos (representando el 80,6 % de la muestra) se presentó el problema P9 (El equipo de dirección no tuvo capacitación para enfrentar el rol).
- En 10 proyectos (representando el 58,8 % de la muestra) se presentó el problema P5 (Algunos documentos fueron mal codificados o no se codificaron).

Año 2007:

- No asocia problemas.

Año 2008:

- En el 100 % de los proyectos se presentó el problema P9 (El equipo de dirección no tuvo capacitación para enfrentar el rol).

Nótese que se repite el problema P9 (El equipo de dirección no tuvo capacitación para enfrentar el rol) en los años 2006 y 2008.

Por otra parte, a partir de las tablas de contingencia construidas y teniendo en cuenta los casos en que el valor de la significación exacta del Chi cuadrado es 0,000 (coincide con la significación del Test exacto de Fisher), se decide rechazar la hipótesis de independencia con nivel de significación del 1% y concluir que las variables que se relacionan a continuación están asociadas:

- El problema P23 (Las condiciones no estuvieron creadas para el despliegue) y la buena práctica BP20 (Se concilió el cronograma de liberaciones con Calidad UCI y se realizaron estimaciones considerando riesgos en este sentido). Esta asociación sugiere que al realizar una adecuada gestión de los riesgos y si se tienen en cuenta todos los factores que intervienen en el acabado de un producto (ejemplo revisión de la calidad) entonces se pueden realizar mejores planificaciones de los despliegues.
- El problema P6 (No existe correspondencia entre el expediente del proyecto físico y el digital) y la buena

práctica BP10 (Se gestionaron con rigurosidad los riesgos, reevaluando constantemente el entorno y ajustando las acciones para mitigarlos).

- El problema P5 (Algunos documentos fueron mal codificados o no se codificaron) y la buena práctica BP10 (Se gestionaron con rigurosidad los riesgos, reevaluando constantemente el entorno y ajustando las acciones para mitigarlos).
- El problema P9 (El equipo de dirección no tuvo capacitación para enfrentar el rol) y la buena práctica BP10 (Se gestionaron con rigurosidad los riesgos, reevaluando constantemente el entorno y ajustando las acciones para mitigarlos).

Ya que la V de Cramer es igual a 1,0 se puede concluir que existe entre este último par de variables (P9 - BP10), la mayor fortaleza en la asociación.

Teniendo en cuenta los casos en que la significación exacta del Chi cuadrado es 1,0 y el Test exacto de Fisher es mayor que 0,7 se decide aceptar la hipótesis de independencia con nivel de significación del 1% y concluir que:

- No se descubrió asociación entre el problema P14 (Alcances ejecutados sin respaldo legal) con la buena práctica BP20 (Se concilió el cronograma de liberaciones con Calidad UCI y se realizaron estimaciones considerando riesgos en este sentido). Este elemento indica que en los procesos de liberación por calidad de los productos y/o servicios generados en el proyecto no se tuvo en cuenta el alcance de los contratos, provocando la afectación de la calidad total en la ejecución del proyecto.
- Se identificó además que no existe asociación entre el problema P14 (Alcances ejecutados sin respaldo legal) con la buena práctica BP10 (Se gestionaron con rigurosidad los riesgos, reevaluando constantemente el entorno y ajustando las acciones para mitigarlos). Lo que sugiere que en los proyectos analizados hay deficiencias en la gestión de los riesgos.

Por otra parte, el hecho de que no se determinó una asociación clara entre el problema P23 (Las condiciones no estuvieron creadas para el despliegue) y la buena práctica BP10 (Se gestionaron con rigurosidad los riesgos, reevaluando constantemente el entorno y ajustando las acciones para mitigarlos) reafirma que en los proyectos analizados hay deficiencias en la gestión de los riesgos.

La asociación débil entre el problema P9 (El equipo de dirección no tuvo capacitación para enfrentar el rol) y la buena práctica BP20 (Se concilió el cronograma de liberaciones con Calidad UCI y se realizaron estimaciones considerando riesgos en este sentido), indica que con independencia de la aplicación o no de actividades de capacitación del equipo de dirección, se gestionaron en la mayoría de los casos el cronograma de liberación con Calidad UCI y este elemento pudo estar motivado por la propia organización del proceso de producción y de las directivas asociadas con la calidad en la organización.

El problema P6 (No existe correspondencia entre el expediente del proyecto físico y el digital) y la buena práctica BP20 (Se concilió el cronograma de liberaciones con Calidad UCI y se realizaron estimaciones considerando riesgos en este sentido), aparentemente no están asociados. Este elemento confirma que durante los procesos de revisión de la calidad de los proyectos en la fase de liberación de sus productos o servicios no se chequea la correspondencia entre los expedientes físico y digital.

Conocimiento

En la Figura 2 se muestra el diseño de la base de datos obtenida utilizando el diagrama de clases persistentes. Esta base de datos contiene la información básica de los proyectos terminados y el conocimiento extraído en este trabajo. La misma facilita el acceso a los datos desde diferentes aplicaciones que pudieran requerir su consulta posterior.

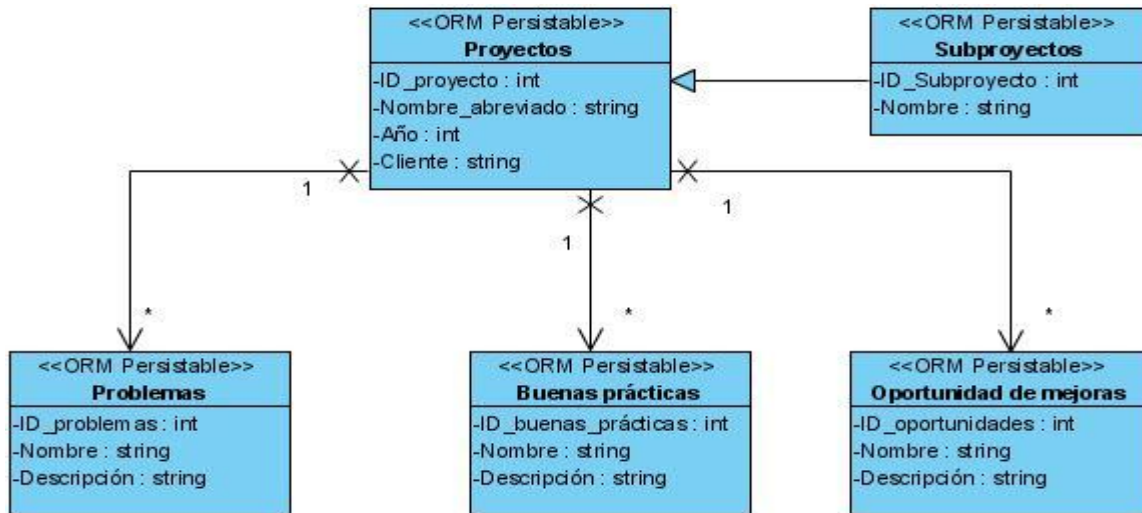


Figura 2. Diagrama de clases persistentes.

Conclusiones

A partir del trabajo realizado en esta investigación y en base a los resultados obtenidos se arribó a las siguientes conclusiones:

- La principal limitante identificada a partir del análisis crítico de metodologías generalizadas como APMBOK, Prince2 y PMBOK, se concreta en la ausencia de estudios relacionados con el descubrimiento de conocimientos en la Gestión de Proyectos.
- A partir de la etapa de preprocesamiento se pudo organizar la información disponible, dándole forma y coherencia a los datos originales, así como también, se extrajo el componente de variables que se usó para realizar el estudio.
- Siguiendo el orden propuesto en el presente trabajo para la aplicación de las técnicas de minería de datos, se disminuye a 10 la cantidad de tablas de contingencia y a 30 estadígrafos. De lo contrario, se debían interpretar 61 tablas de contingencia y 183 estadígrafos para encontrar la asociación entre las variables problemas y buenas prácticas con la variable que representa el año en que iniciaron los proyectos. De esta forma se reduce la dimensionalidad de los datos a analizar sin pérdida de información relevante.
- Se identificaron dificultades en el conocimiento de los jefes de proyectos en las áreas de conocimiento de gestión de riesgos y gestión de la calidad, fundamentalmente desde la perspectiva de los elementos legales y la importancia de la gestión documental de los expedientes de proyectos.

Referencias

- CANO DE AMO, J. R. Reducción de Datos basada en Selección Evolutiva de Instancias para Minería de Datos. In *Departamento de Ciencias de la Computación e Inteligencia Artificial*. Granada: Universidad de Granada, 2004.

- DURÁN, E.; COSTAGUTA, R. Minería de datos para descubrir estilos de aprendizaje. Revista Iberoamericana de Educación, 10 de marzo 2007, vol. 42, no. 2, p. 10.
- FAYYAD, U.; Piatetsky-Shapiro, G., *et al.* From Data Mining to Knowledge Discovery in Databases. Artificial Intelligence Magazine, 1996, p. 37-54.
- GERVILLA GARCÍA, E.; JIMÉNEZ LÓPEZ, R., *et al.* La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes Adicciones: Revista de sociodrogalcohol, 2009, vol. 21, no. 1, p. 65-80.
- GÓMEZ, C.H. COMPILACION BIBLIOGRAFICA PMBOK, OPM3 ®. 2010. Disponible en: [<http://auditoriauc20102miju02.wikispaces.com/file/view/PMBOK201021700421228.pdf>].
- GÓMEZ FLECHOSO, J.A. Inducción de conocimiento con incertidumbre en bases de datos relacionales borrosas. In *Escuela Técnica Superior de Ingenieros de Telecomunicación*. Madrid: Universidad Politécnica de Madrid, 1998.
- GONZÁLEZ, C.G. *Tratamiento de datos*. Edtion ed.: Díaz de Santos, 2006. ISBN 9788479787363.
- KANTARDZIC, M. Data Mining: Concepts, Models, methods, and Algorithms [online]. Segunda. [New York]: John Wiley & Sons, Inc., 2003. Disponible en: [<http://www.certified-easy.com/aa.php?isbn=ISBN:1118029135&name=Data Mining, Concepts, Models, Methods, and Algorithms>].
- KIM, W.; CHOI, B., *et al.* A taxonomy of dirty data. Data Mining and Knowledge Discovery. Data Mining and Knowledge Discovery, 2003, vol. 7, p. 81-89.
- LAROSE, D.T. *Discovering Knowledge in Data: An Introduction to Data Mining*. edited by J. WILEY. Edtion ed. New York: John Wiley & Sons, Inc., 2005. ISBN 0-471-66657-2.
- LÓPEZ SALLABERRY, J.M.V. La metodología data mining como subsistema de información y su funcionalidad en contexto kdd. Una aproximación empírica In. España: Universidad de Deusto, 2007.
- MATTURRO MAZONI, G. Modelo para la gestión del conocimiento y la experiencia integrada a las prácticas y procesos de desarrollo software. In *Facultad de Informática*. Madrid: Universidad Politécnica de Madrid, 2010, p. 395.
- Molina Félix, L.C. Data mining: torturando a los datos hasta que confiesen. 2002. Disponible en: [<http://www.uoc.edu/molina1102/esp/art/molina1102/molina1102.html>].
- MURPHY, A. ; LEDWITH, A. Project management tools and techniques in hightechnology SMEs. 2007, p. 153-166.
- NAVARRO CÉSPEDES, J.M.; Casas Cardoso, G.M., *et al.* Análisis de componentes principales y análisis de regresión para datos categóricos. Aplicación en la hipertensión arterial. Revista de Matemática: Teoría y Aplicaciones, 2010, vol. 17, no. 2, p. 199-233.
- OJEDA MAGAÑA, B. Aportación a la extracción de conocimiento aplicada a datos mediante agrupamientos y sistemas difusos. In *Señales, Sistemas y Radiocomunicaciones*. E.T.S.I. Telecomunicación (UPM), 2010.
- PORCEL, E.A.; DAPOZO, G.N., *et al.* Predicción del rendimiento académico de alumnos de primer año de la FACENA (UNNE) en función de su caracterización socioeducativa. Revista electrónica de investigación educativa, Noviembre 2010, vol. 12, no. 2, p. 1-21.
- REYES DUEÑAS, M.X. Minería de datos espaciales en búsqueda de la verdadera información. Ingeniería y Universidad, 2009, vol. 13, no. N°. 1.
- RIL VALENTIN, E.B. Documentación de lecciones aprendidas en el cierre de proyectos de ALBET. In *UCIENCIA 2012*. 2012.

- RIQUELME, J.C.; RUIZ, R., *et al.* Finding Defective Software Modules by Means of Data Mining Techniques. IEEE LATIN AMERICA TRANSACTIONS, Julio 2009, vol. 7, no. 3.
- SARANGO SEDAMANOS, M.Y. Aplicación de técnicas de minería de datos para identificar patrones de comportamientos relacionados con las acciones del estudiante con el EVA de la UTPL. In *ESCUELA DE CIENCIAS DE LA COMPUTACIÓN*. Loja-Ecuador: UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA, 2012, p. 213.
- SOLARTE MARTÍNEZ, G.R. ; SOTO MEJÍA, J.A. Árboles de decisiones en el diagnóstico de enfermedades cardiovasculares. Scientia et Technica, 2011, vol. 3, no. 49, p. 104-109.
- VALCARCEL ASENCIOS, V. DATA MINING Y EL DESCUBRIMIENTO DEL CONOCIMIENTO. Industrial Data, 2004, vol. 7, p. 83-86.
- VALLEJO PÉREZ, D. ; TENELANDA VEGA, G. Minería de datos aplicada en detección de intrusos. USBMed, 2012, vol. X, p. 14.