

UNIDAD DE COMPATIBILIZACIÓN, INTEGRACIÓN Y DESARROLLO
DE PRODUCTOS INFORMÁTICOS PARA LA DEFENSA

TESIS DE MAESTRÍA

MODELO DE UN SISTEMA DE RAZONAMIENTO BASADO EN CASOS
PARA EL ANÁLISIS EN LA GESTIÓN DE RIESGOS

AUTOR: Ing. SERGIO MICHEL RIVERA RODRÍGUEZ
TUTOR: Dr. PEDRO YOBANIS PIÑERO PÉREZ

CIUDAD HABANA
SEPTIEMBRE 2010

Declaración de autoría

Declaro que soy el único autor de este trabajo y autorizo a la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Ing. Sergio Michel Rivera Rodríguez

Autor

DrC. Pedro Yobanis Piñero Pérez

Tutor

Dedicatoria.

A mi hija, mis padres, y mis hermanos.

A mima, pipo y nena, mis abuelos.

A Elito y Carlos, mis tíos y maestros.

*“Los obstáculos son esas cosas que las personas ven
cuando dejan de mirar sus metas.”*

E. Joseph Crossman.

“Caminante, no hay camino, se hace camino al andar.”

Antonio Machado.

Agradecimientos.

A mi familia, por apoyarme en todos los momentos importantes de mi vida, en especial a mi madre.

A mi novia Meylin, por su ayuda, consejos y paciencia infinita.

A mis compañeros de trabajo, por su preocupación y ánimo.

A los todos los profesores de la maestría, en especial a mi tutor.

A todos, muchas gracias.

Resumen.

Aunque los diversos enfoques de gestión del riesgo aparecieron hace más de una década, sigue siendo evidente la poca utilización de sus técnicas en los proyectos de desarrollo de software actuales. El desarrollo de software por deformación profesional induce a "buscar e ignorar" el riesgo. Los desastres del software pueden ser evitados si nos preocupamos por identificar y tratar tempranamente los elementos de alto riesgo.

La investigación presentada examina cuestiones referentes al papel del proceso de análisis en la gestión de riesgos en proyectos de software, identificando conceptos fundamentales, centros de pensamiento, modelos principales, herramientas y técnicas empleadas en dicha etapa.

Se define un modelo de razonamiento basado en casos para asistir la búsqueda y reutilización de riesgos con características similares, definiendo el alcance y un conjunto de principios y premisas para aplicarlo en entornos de desarrollo de software. Se propone el empleo de una base de conocimiento de riesgos y a partir de esta se describen los procesos de inferencia y aprendizaje, que garantizan la recuperación y empleo de riesgos similares en el proceso de análisis.

Se implementó una herramienta que automatiza los procesos descritos en el modelo propuesto y se validó en bases de datos internacionales para determinar la tasa de acierto del clasificador propuesto.

Palabras claves:

Gestión de riesgos, proceso de análisis, razonamiento basado en casos, riesgo.

Abstract.

Although different risk management approaches appeared more than a decade ago, is still apparent low utilization of their skills in software development projects today. Software development by professional deformation induces a "search and ignore" the risk. Software disasters can be avoided if we care for early identification and treatment of high risk items.

The research presented examines issues concerning the role of the analysis process in risks management in software projects, identifying key concepts, thinking centers, main models, tools and techniques used at that stage.

It defines a case based reasoning model to assist the search and reuse of risks with similar characteristics, defining the scope and a set of principles and assumptions to be applied in software development environments. It proposes the use of a knowledge base of risks and from this we describe the inference and learning processes, guaranteeing the recovery and use of similar risks in the process of analysis.

We implemented a tool that automates the processes described in the model proposed and validated in international databases to determine the rate success of the proposed classifier.

Keywords:

Risk management, analysis process, case based reasoning, risk.

Índice.

Declaración de autoría.....	I
Dedicatoria.....	II
Agradecimientos.....	III
Resumen.....	IV
Abstract.....	V
Índice.....	VI
Índice de figuras.....	VIII
Índice de tablas.....	IX
Introducción.....	1
Capítulo 1. Fundamentación teórica.....	8
1.1 Gestión de riesgos.....	8
1.2 Marcos de referencia para la gestión de riesgos.....	9
1.3 Centros de pensamiento.....	10
1.4 Principales modelos de gestión de riesgos.....	12
1.4.1 Modelo de Boehm.....	12
1.4.2 Continuous Risk Management.....	13
1.4.3 Software Risk Evaluation.....	14
1.4.4 PMBOK.....	16
1.4.5 Microsoft Solution Framework.....	18
1.4.6 Magerit.....	20
1.4.7 MOGERI.....	21
1.5 Técnicas y herramientas empleadas en la etapa de análisis en la gestión de riesgos.....	22
1.6 Sistemas de razonamiento basados en casos y su vinculación con la gestión de riesgo.....	29
1.7 Conclusiones.....	34
Capítulo 2. Modelo de RBC para el análisis en la Gestión de Riesgos.....	36
2.1 Introducción.....	36
2.2 Alcance, principios y premisas.....	37
2.3 Características del modelo.....	39
2.3.1 Construcción de la base de conocimiento.....	42
2.3.1.1 Construcción de la base de conocimiento de riesgos.....	42
2.3.1.2 Selección de los atributos.....	45
2.3.2 Máquina de inferencia.....	46
2.3.2.1 Recuperación de riesgos semejantes.....	47
2.3.2.1.1 Funciones de semejanza entre atributos.....	47
2.3.2.1.2 Función de semejanza entre casos propuesta.....	53
2.3.2.1.3 Algoritmo propuesto para la recuperación de riesgos similares.....	54
2.3.2.2 Adaptación de los casos.....	55
2.3.2.3 Explicación de la inferencia.....	56
2.3.3 Módulo de aprendizaje.....	56
2.3.3.1 Validación y refinamiento de la solución.....	56
2.3.3.2 Actualización de la base de conocimiento.....	57
2.4 Conclusiones.....	57
Capítulo 3. Validación de la propuesta.....	60
3.1 Introducción.....	60
3.2 Herramienta elaborada.....	60
3.2.1 Selección de los atributos.....	60
3.2.2 Recuperación de riesgos semejantes.....	61
3.2.3 Explicación y adaptación.....	62

3.3 Caso de estudio 1: Validación del clasificador propuesto.....	62
3.3.1 Bases de datos empleadas para probar el clasificador.....	63
3.3.2 Análisis de los resultados del clasificador.....	64
3.4 Caso de estudio 2: Puesta a prueba en la UCID.....	66
3.4.1 Implementación del método Delphi.....	67
3.4.2 Análisis de los resultados.....	69
3.5 Caso de estudio 3: Puesta a prueba en la Dirección Técnica de la Infraestructura Productiva de la UCI.....	73
3.8 Conclusiones.....	73
Conclusiones.....	75
Recomendaciones.....	76
Bibliografía.....	77
Anexos.....	83
Anexo 1. Encuesta realizada a los especialistas en torno a la gestión de riesgos.....	83
Anexo 2. Encuesta de validación de la propuesta.....	84
Anexo 3. Cuestionario sobre la validación del modelo propuesto.....	85

Índice de figuras.

Figura 1: Modelo propuesto por Boehm de la Gestión de Riesgo. Adaptado de (Boehm, 1991).....	13
Figura 2: Modelo propuesto por el CRM para la Gestión de Riesgos. Adaptado de (Rosenberg et al., 1999; Rosenberg, Al Gallo, Ted Hammer, & Parolek, 2000).....	14
Figura 3: Etapas de la Gestión de Riesgos según (Williams et al., 1999).....	15
Figura 4: Procesos definidos por Microsoft para la Gestión de Riesgos. Adaptado de (Microsoft, 2004).....	19
Figura 5: Etapas propuestas por Magerit para la Gestión de Riesgos. Tomado de (MAP, 2006a).....	21
Figura 6: Modelo de procesos de gestión de riesgos. Tomado de (Véliz, 2007).....	21
Figura 7: Ejemplo de distribuciones de probabilidad comúnmente usadas. Tomado de (PMI, 2004).....	24
Figura 8: Matriz de probabilidad e impacto. Tomado de (Inteco, 2008).....	27
Figura 9: Representación de cómo RCB da solución a un problema. Adaptado de (Leake, 1996).....	30
Figura 10: El ciclo de razonamiento basado en casos (Aamodt & Plaza, 1994).....	33
Figura 11: Primer proceso: Construcción de la base de conocimientos.....	40
Figura 12: Segundo proceso: Máquina de inferencia.....	41
Figura 13: Tercer proceso: Módulo de aprendizaje.....	41
Figura 14: Representación vectorial del riesgo. Tomado de (García-Hernández, 2008).....	43
Figura 15: Modelo conceptual para el análisis de los riesgos.....	45
Figura 16: Vista de la selección de los rasgos predictivos y objetivos del un determinado vector de riesgo.....	61
Figura 17: Vista de la selección de riesgos a los que se les hallará riesgos similares.....	61
Figura 18: Vista del proceso de revisión de riesgos semejantes.....	62
Figura 19: Personal involucrado en la encuesta realizada.....	66
Figura 20: Técnicas de análisis de gestión de riesgos conocidas.....	67

Índice de tablas.

Tabla 1: Resultado de la aplicación del método H por remuestreo al clasificador.....	64
Tabla 2: Comparación con otros métodos propuestos para la base de datos Iris.....	65
Tabla 3: Comparación de métodos de clasificación empleando la base de datos Car Evaluation.....	65
Tabla 4: Comparación de métodos de clasificación empleando la base de datos Wine Quality.....	66
Tabla 5: Patrón de influencia.....	68
Tabla 6: Resultados del índice de experticidad.....	68
Tabla 7: Resultados de la encuesta.....	69
Tabla 8: Frecuencia absoluta respecto a los resultados obtenidos.....	70
Tabla 9: Frecuencia absoluta acumulada respecto a los resultados obtenidos.....	70
Tabla 10: Frecuencias relativas mediante la Distribución Normal Estándar Invertida.....	71
Tabla 11: Concordancia de los resultados.....	72

Introducción.

Los riesgos están presentes siempre en todo tipo de actividad que ha hecho el hombre, pues cada una tiene presente elementos que atentan en alguna medida contra su desarrollo satisfactorio o están propensas a que algunos eventos sucedan y causen efectos no deseados.

De esta forma en todos los aspectos de la vida diaria tenemos riesgos presentes. Como proceso del pensamiento lógico, nos cuestionamos estos elementos y eventos y tomamos medidas para que las consecuencias de su ocurrencia sean lo menos negativas posible.

O sea que haciendo uso de nuestra capacidad previsoras nos preparamos para que cuando aquellos elementos o eventos que no podemos asegurar del todo ocurran, minimizar los efectos negativos que nos puedan causar.

A este proceso donde se identifican los riesgos, se toman medidas para minimizar los efectos y otras para prepararse en caso que ocurran y se monitorean se denomina precisamente “Gestión de Riesgos”.

La Gestión de Riesgos ha evolucionado pasando por un proceso inherente, a un proceso consciente, a un proceso lógico, sistemático y explícito como tenemos hoy en día (Lovejoy, 2005), apoyado por las matemáticas y metodologías desarrolladas en este sentido.

Se gestionan los riesgos para la salud, la seguridad nacional, los cambios climáticos, el medio ambiente, la construcción, el transporte.

Los riesgos son por naturaleza inciertos (Litten, 2009; Perminova, Gustafsson, & Wikström, 2008), es decir, una vez que identificamos un riesgo este tiene determinada probabilidad de ocurrir, y los esfuerzos siempre se concentran en encontrar la probabilidad con la mayor certidumbre posible para poder prepararse mejor para su ocurrencia.

En el mundo de los negocios se comienza a formalizar la Gestión de Riesgos como un mecanismo de protección y control para los proyectos, para minimizar los eventos negativos así como sus efectos, para conocer el grado de certidumbre o cuán probable es que fracase el proyecto, o tener la seguridad que se está preparado para situaciones emergentes. También dada la necesidad de una medición constante de los indicadores claves para el éxito del proyecto, de un análisis de los resultados y control periódico, para tener un panorama en tiempo real (o lo más real posible) del estado del proyecto y así tomar las decisiones correctas en el momento adecuado.

En general todas las empresas necesitan controlar los riesgos de una forma u otra, y nuevos riesgos surgen en la medida que evolucionan las industrias y aparecen nuevas necesidades y desafíos que enfrentar (García-Hernández, 2008).

El software es complejo por naturaleza (Pressman, 2005): la situación de la vida real que un software modela o intenta resolver, tiene una complejidad que generalmente es alta, modelarla para que sea soluble por un sistema informático es complejo, además de los requerimientos del software en sí, se unen muchos otros que aumentan la complejidad: flexibilidad, encapsulamiento, alta cohesión, bajo acoplamiento, reusabilidad; la comunicación entre desarrolladores y clientes adiciona complejidad. Unido a esto, durante el desarrollo del proyecto existen muchos riesgos de diferentes categorías que atentan contra sus objetivos (Carol, 2010).

Las estadísticas indican que una parte significativa de los proyectos de software que fracasan, no tuvieron, o tuvieron una pobre Gestión de Riesgos, y los problemas que enfrentaron hubieran sido eliminados, o su impacto reducido, si se hubiera tenido una Gestión de Riesgos formal. No quiere decir que solamente aplicando esta disciplina en la gestión de proyectos se llegue al éxito, pero sí es una práctica importante (García-Hernández, 2008).

Hasta el momento se han propuesto y utilizado diferentes enfoques de gestión de riesgo desde que Barry Boehm (Boehm, 1989) atrajo a la comunidad de ingeniería de software hacia la gestión de riesgo. Sin embargo, es evidente que pocas organizaciones utilizan todavía de una forma explícita y sistemática métodos específicos para gestionar los riesgos en sus proyectos de software (Esteves, Pastor, Rodriguez, & Roy, 2005). En un estudio realizado en el 2001 con 268 organizaciones, mostró que el 3% de las mismas no utilizaba ningún marco de gestión de riesgos, el 37% de los participantes habían utilizado algún marco informal, el 28% utilizaban procedimientos repetitivos y sólo un 14% usaba un enfoque formal para identificar riesgos (Kulik & Weber, 2001). Según este estudio, las razones más comunes para utilizar un marco informal son la falta de procedimientos, necesidades del proyecto no adecuadas, organización inmadura y compromiso del equipo. Según Hoffman, citado por (Esteves et al., 2005), existen organizaciones que demuestran una pobre gestión de riesgos en el ámbito general de los sistemas de información. En el mismo artículo y citando a Kontio y Basili, se exponen tres razones para la baja tasa de divulgación de tecnologías de gestión de riesgo: la falta de conocimiento sobre posibles métodos y herramientas, limitaciones prácticas y teóricas sobre los marcos de gestión de riesgo que entorpecen la facilidad de uso

de estos métodos, y que todavía hay pocos informes con evaluaciones sistemáticas o científicas que proporcionen retroalimentación empírica sobre su viabilidad y beneficios.

Según (McManus & Wood-Harper, 2008) en un estudio realizado por los autores en empresas del Reino Unido, en un periodo de tiempo que cubre desde 1998 hasta 2005, fueron realizados 214 proyectos de software en varios sectores de la sociedad. De ellos, 51 proyectos fueron cancelados (23,8%). Entre las razones que se argumentan se apunta a una insuficiente gestión de riesgos dado por la poca cultura que muestran las organizaciones para llevar a cabo este proceso de manera formal.

Hoy día aumenta el uso del software en actividades críticas para los negocios, los gobiernos; varían las tecnologías y es necesaria una respuesta rápida a los clientes para aprovechar la oportunidad de negocio; las técnicas tradicionales de planeación necesitan de disciplinas de soportes más fuertes para que la gestión de proyectos responda a las necesidades actuales, y en este sentido la Gestión de Riesgos se alza e incluso hay tendencias a dirigir el desarrollo del proyecto guiado por los riesgos.

La Gestión de Riesgos.

Según Tom DeMarco la gestión de riesgos equivale a la gestión de proyectos para adultos (DeMarco & Lister, 2003). En esta se identifican varias etapas o procesos fundamentales que varían según la tendencia o el autor pero en general se consideran: planeación, identificación, análisis, mitigación, control y seguimiento.

En la etapa de planificación, es donde se decide qué estrategia se utilizará para la Gestión de Riesgos. Una vez que el equipo de proyecto establece las disposiciones sobre cómo se debe manejar la Gestión de Riesgos, y durante el proyecto y en varias iteraciones se hará una identificación de los riesgos. Este es un proceso donde se descubren los riesgos involucrados en el proyecto, y se documentan. Con todos los riesgos “más visibles” identificados y documentados se realiza un análisis de los mismos (el nivel de profundidad del análisis y los elementos a analizar así como la estrategia para hacerlo fue establecido en la etapa de planificación). Durante esta etapa de mitigación, se elaboran planes de contingencias para los riesgos que se decidan afrontar. En todo el ciclo de vida del proyecto se monitorea el estado de los riesgos, y se identifican nuevos comenzando el proceso otra vez, o se desechan algunos que ya no constituyen riesgos, se mantienen actualizados los documentos de la Gestión de Riesgos y se reevalúan.

La Unidad de Compatibilización, Integración y Desarrollo de soluciones informáticas para la

Defensa (UCID) se erige hoy en día como el centro por excelencia del desarrollo de herramientas para la toma de decisiones de los altos jefes del Ministerio de las Fuerzas Armadas.

La misma cuenta con una metodología de desarrollo y proceso de gestión de proyectos implantado por los cuales se viene desarrollando desde hace algún tiempo, habiendo definido entre sus procesos de gestión de proyectos la gestión de riesgo como uno de los procesos básicos a implementar en el desarrollo de los proyectos.

En entrevistas y encuestas realizadas a líderes de proyectos se detectó que aunque trabajan siguiendo lo establecido como pauta para la gestión de riesgo, se desaprovecha la información sobre riesgos antiguos, útil para la resolución de nuevos riesgos en proyectos futuros. Esto sucede, fundamentalmente, porque las herramientas de análisis de riesgos que se emplean no consideran esta vital información.

Ahora bien, ante la problemática de la utilización de riesgos guardados en forma de experiencia previa prevalece la utilización de técnicas de inteligencia artificial para realizar un tratamiento especial con esta información.

Dentro de la inteligencia artificial hay dos ramas que se destacan por sus aportes en el aprendizaje y clasificación: los algoritmos de aprendizaje basados en árboles de decisión y los sistemas basados en el conocimiento.

En estos últimos destacan los sistemas basados en reglas (SBR), los basados en frames (SBF), los basados en casos (SBC), los basados en probabilidades (SBP), las redes expertas y los sistemas borrosos (Piñero-Pérez, 2005).

Los dos pilares del enfoque de los SBC son dos suposiciones fundamentales acerca de la naturaleza del mundo:

- Su regularidad. De situaciones similares se extraen conclusiones y se aprenden lecciones similares. En consecuencia, las conclusiones y lecciones que acompañan experiencias previas pueden ser la base de las que correspondan a situaciones nuevas.
- La recurrencia de las experiencias. Es altamente probable que las situaciones futuras sean variantes de las actuales.

Apoyando la suposición básica de que “problemas parecidos tienen soluciones parecidas” y que “los nuevos problemas son similares a problemas previamente resueltos”, la resolución

de problemas basada en casos saca partido de las relaciones entre dos tipos de similitud. Estos tipos de similitud se aplican al espacio de la descripción de los problemas y de su solución. El SBC identifica un problema previo cuya descripción sea la más similar a la del nuevo en función de las características más relevantes que serán identificadas como índices para guiar el proceso de recuperación. La solución del problema considerado más similar se utiliza como punto de partida para generar la solución al nuevo problema, solución que no deberá ser muy diferente a la inicial (Agudo, 2002).

Por lo que el problema queda planteado de la siguiente forma: La ausencia de un modelo basado en la experiencia para el análisis de riesgos está afectando la correcta estimación de los valores que caracterizan al riesgo y la ayuda a la toma de decisiones en la gestión de riesgos en la UCID.

El objeto de estudio de esta investigación lo constituye la gestión de riesgos en proyectos de software y el campo de acción está relacionado con la etapa de análisis que se emplea en dicho proceso.

La hipótesis que se maneja es la siguiente: el desarrollo de un modelo de razonamiento basado en casos orientado al proceso de análisis en la gestión de riesgos en los proyectos de software de la UCID posibilitará correcta estimación de los valores que caracterizan al riesgo y la ayuda a la toma de decisiones en la gestión de riesgos.

El objetivo general que se persigue es desarrollar un modelo que, haciendo uso de técnicas de razonamiento basado en casos, permita un mejor proceso de análisis de la gestión de riesgos de software en los proyectos desarrollados en la Unidad de Compatibilización, Integración y Desarrollo de Productos Informáticos para la Defensa (UCID), y los objetivos específicos son:

- elaborar el marco teórico de la investigación identificando las principales tendencias, limitaciones y ventajas de diferentes modelos para la gestión de riesgos y en particular el análisis de los mismos,
- desarrollar un modelo que implemente técnicas de razonamiento basado en casos orientado al análisis en la gestión de riesgos, y
- validar el modelo desarrollado a partir de una prueba piloto y del criterio de expertos.

Marco conceptual.

Sistemas Basados en Casos (SBC): Los SBC están inspirados, en gran medida, en el papel

que juega el recuerdo en el razonamiento humano. Así pues, el SBC es un mecanismo de razonamiento que se basa en recordar situaciones o experiencias (casos) similares acontecidos en el pasado y almacenados en una base de experiencias, y adaptar la lección extraída de ellas a la situación actual.

Gestión de riesgos: Conjunto de pasos que ayudan al equipo del proyecto a detectar y gestionar los riesgos que aparecen durante el desarrollo de un proyecto. Los objetivos de la gestión de riesgos son aumentar la probabilidad y el impacto de los eventos positivos, y disminuir la probabilidad e impacto de los eventos adversos para el proyecto.

Análisis de riesgo: Proceso de la gestión de riesgos que sigue a la identificación de los riesgos. En este proceso se analiza el efecto de los riesgos y les asigna la calificación numérica o cualitativa.

Como parte del diseño metodológico fue seguida una estrategia explicativa pues los conocimientos precedentes acerca del problema han sido suficientes para plantear una hipótesis explicativa y la representación del problema es clara en lo referente a la caracterización del fenómeno en sus aspectos externos.

Se utilizaron diferentes métodos de investigación que se detallan en el desarrollo del documento y de los que se pueden citar, entre los métodos teóricos: el histórico-lógico, el analítico-sintético y el hipotético-deductivo; entre los métodos empíricos: la observación, la entrevista y la encuesta, y también se emplearon métodos estadísticos en la validación del clasificador propuesto y en los criterios de expertos.

Este trabajo presenta un modelo basado en técnicas de inteligencia artificial para apoyar el proceso de análisis en la gestión de riesgo, el cual se basa en la experiencia acumulada de los riesgos en los proyectos de software contribuyendo a una mejor gestión de eventos que pueden afectar los objetivos principales de los proyectos.

El documento que a continuación se muestra está estructurado en tres capítulos. En el primero se hace un estudio de los principales marcos de gestión de riesgos resaltando de ellos los principales centros de pensamiento que llevan a cabo estudios sobre la gestión de riesgo, las etapas de análisis de riesgo, algunas de las herramientas que existen que tratan estos problemas y las técnicas que proponen, además se hace un estudio introductorio sobre los sistemas de razonamiento basados en casos.

En el segundo capítulo se detalla el modelo de razonamiento basado en casos propuesto para darle solución al problema presentado, definiendo alcance, principios y premisas para

su aplicación, se propone una base de conocimiento de riesgos y se describe en detalle cada uno de los procesos que lo componen en modelo.

En el tercer capítulo se presenta una herramienta elaborada que cumple con el modelo propuesto, se valida la propuesta mediante el método Holdout o H para el caso del clasificador y método Delphi para la validación de los expertos.

Capítulo 1. Fundamentación teórica.

1.1 Gestión de riesgos.

La definición de Gestión de Riesgos se puede aceptar de acuerdo a las posiciones de quienes, en el mundo, estudian y desarrollan estos temas, como:

- Identificar, tratar y eliminar riesgos de software antes de que se tornen en amenazas que afecten objetivos esenciales para el proyecto (Rosenberg, Theodore Hammer, & Albert Gallo, 1999).
- El proceso formal en el que los factores de riesgos son sistemáticamente identificados, evaluados y mitigados (Williams, Pandelios, & Behrens, 1999).
- La identificación y tratamiento de riesgos en etapas tempranas de desarrollo de los proyectos reducen costos a largo plazo y ayudan a prevenir desastres (Boehm, 1991, 1989).
- Incluye los procesos relacionados con la planificación de la gestión de riesgos, la identificación y el análisis de riesgos, las respuestas a los riesgos, y el seguimiento y control de riesgos de un proyecto. Los objetivos son aumentar la probabilidad y el impacto de los eventos positivos, y disminuir la probabilidad y el impacto de los eventos adversos para el proyecto (PMI, 2004).
- Según Roger Pressman citando a Robert Charette (Pressman, 2005): el análisis y la gestión de riesgos son una serie de pasos que ayudan al equipo del software a comprender a gestionar la incertidumbre.
- La práctica compuesta de procesos, métodos y herramientas y que provee de un entorno disciplinado para la toma de decisiones proactiva en base a determinar constantemente que puede ir mal (riesgos), se emplea para identificar cuáles son los riesgos más importantes en los cuales enfocarse e implementar estrategias para gestionarlos (Murphy et al., 2008).

En sentido general, la investigación en la gestión de riesgos en el ámbito del software procura formalizar conocimiento orientado a minimizar y/o evitar los riesgos, mediante la generación de principios y buenas prácticas de aplicación realista (Zulueta-Véliz, 2007). Hasta el momento se han propuesto y utilizado diferentes perspectivas de gestión de riesgo desde que Barry W. Boehm atrajo a la comunidad de Ingeniería del Software hacia esta

rama. Sin embargo, es evidente que pocas organizaciones utilizan todavía de una forma explícita y sistemática métodos específicos para gestionar los riesgos en sus proyectos de software.

En esta investigación se asume la definición de gestión de riesgos dada por el Project Management Institute, por la cual el riesgo de un proyecto es un evento o condición inciertos que, si se produce, tiene un efecto positivo o negativo sobre al menos un objetivo del proyecto, como tiempo, coste, alcance o calidad (PMI, 2004).

1.2 Marcos de referencia para la gestión de riesgos.

Varios autores han trabajado en temas como la evolución de las teorías sobre Riesgos y la comparación de modelos, métodos y metodologías. Por ejemplo, se ha dividido la GR en generaciones, para exponer mejor las características y progreso de cada una. La adscripción a una generación puede diferir para los autores debido a que la evolución es continua y no a saltos y también porque los modelos estudiados pueden variar en el tiempo, es decir, puede cambiar la forma en que es abordada la GR en un modelo.

En “Estudio exploratorio sobre los métodos de gestión de proyectos de alto riesgo” (Cocho, Adam, & Torralba, 2003), se plantea la evolución de los modelos de gestión de los riesgos en forma de sucesivas generaciones.

La primera generación -G1- “casuística o tradicional” de identificación de riesgos en los proyectos con técnicas basadas en cuestionarios, listas de incidencias y de las medidas para contrarrestarlas. Se identificaban casos de riesgo y se extrapolaban a otros proyectos. Se apunta que en la década del sesenta, el análisis cuantitativo de los riesgo se realizaba fundamentalmente por procesos markovianos para describir el comportamiento de sistemas complejos con fallos ensayables y aleatorios.

La vigente segunda generación -G2- “taxonómica” de análisis de riesgos en los proyectos (Cocho et al., 2003), traduce a ese sector el análisis de riesgos en los sistemas, desarrollado en los ochenta junto a intentos poco articulados de análisis de riesgo en los negocios. En esta se ha introducido los riesgos pero se ha limitado a analizarlos al inicio del proyecto y a planificar medidas para cambiarlo. Esta visión preventiva suele considerarse “teorizante” por los gestores de proyectos avezados a incorporar medidas curativas, improvisadas en mayor o menor medida, durante su avance, para paliar los riesgos según se presentan. Los principales modelos contenidos en esta generación son los propuestos por: Boehm, Hall, el SEI, SPR, PMI y McFarlan.

La tercera generación -G3- “causal o emergente” nació a mediados de los noventa (Cocho et al., 2003) y se refiere en particular a sistemas de información. Surge de forma simultánea en Europa y Estados Unidos partiendo de la preocupación por proyectos de tanto riesgo como la adquisición o el desarrollo de software . Articula una causalidad más explicativa y por tanto más predictiva entre los elementos del modelo, sobre todo entre los factores de riesgo y sus medidas reductoras o salvaguardas. Esta causalidad prepara el paso a la gestión de proyectos por los riesgos. Se apoya en modelos sistémicos, relacionales y proactivos en el aseguramiento de los proyectos. Algunos de los modelos incluidos en esta generación son: el Eurométodo, MAGERIT e ISPL.

1.3 Centros de pensamiento.

Los centros de pensamiento más importantes, en materia de gestión de riesgos, identificados son:

- Society for Risk Analysis (SRA): La Sociedad para el Análisis de Riesgos, fundada en 1981, ofrece un área de colaboración para profesionales de varias disciplinas vinculados al tema, es la organización líder en la investigación académica en el campo de los riesgos. Es una sociedad multidisciplinaria, erudita, internacional que mantiene un foro abierto a todos aquellos que estén interesados en el análisis de riesgo. Entre sus objetivos se encuentra el intercambio de información, ideas y metodologías para el análisis de riesgos. Su publicación más importante es: “Risk analysis, an International Journal”.
- International Project Management Association (IPMA): Es uno de los principales repositorios y fuente bibliográfica sobre análisis de riesgos. Creado en 1965, la Asociación Internacional de Dirección de Proyectos, fue en su momento la primera asociación de dirección de proyecto del mundo. Desde entonces IPMA ha crecido a nivel mundial y se ha vuelto una red internacional para las sociedades de dirección de proyectos. Representa a varias asociaciones de dirección de proyecto nacionales en el nivel internacional. Certifican a gerentes del proyecto, premian los equipos de proyectos exitosos e investigadores, y proporcionan varias publicaciones de dirección de proyectos. IPMA es una de las principales referencias de la gestión formal de proyectos, y uno de sus segmentos principales es la gestión de riesgos.
- Project Management Institute (PMI): Este Instituto fue fundado en 1969. Durante este tiempo el PMI, a través del comité de estándares y colaboradores (entre ellos

empresas, universidades, asociaciones de profesionales, especialistas y consultores en proyectos) ha realizado el estudio, evaluación y revisión de los estándares generalmente aceptados a nivel internacional, dando como resultado los estándares que representan el cuerpo de conocimientos de la Dirección de Proyectos, cuyo título original es “Project Management Body of Knowledge” (PMBOK). En 1987 se publicó su primera edición. Una de sus áreas es la descripción de los procesos relacionados con el desarrollo de la gestión de riesgos de un proyecto, y dentro de esta se encuentra como uno de los procesos principales el análisis cualitativo y cuantitativo de los riesgos.

- Software Engineering Institute (SEI): El SEI, fundado en 1984, es una unidad de la Universidad Carnegie Mellon, dedicado al desarrollo de la ingeniería de software. Encargado de la investigación y creación de modelos para la mejora del desarrollo de software, comienza a principios de los años 90 a desplegar trabajos en la gestión de riesgos. Uno de los primeros es el Continuous Risk Management (CRM), un método de gestión de riesgos que incluye los principios básicos que concuerdan en gran medida con el modelo de Boehm. Otros trabajos son: Software Risk Evaluation (SRE), que define un proceso para la identificación, análisis y desarrollo de estrategias de mitigación; Taxonomy-based Risk Identification, un método para facilitar el proceso de identificación de los riesgos; y Team Risk Management (TRM), que define la estructura y operaciones a realizar para gestionar los riesgos en una organización por un equipo.
- Microsoft Corporation: Fundada en 1975, corporación multinacional productora de software líder en el mundo. En el 2002 promueve Microsoft Solutions Framework, un conjunto de guías, prácticas, procedimientos y herramientas para el desarrollo de aplicaciones, que incluye una disciplina para la Gestión de Riesgos: Microsoft Solutions Framework – Risk Management Discipline.
- Institute of Electrical and Electronics Engineers (IEEE): La IEEE es una asociación técnico-profesional mundial dedicada a la estandarización, entre otras tareas. Está formada por profesionales de las nuevas tecnologías, como ingenieros eléctricos, ingenieros en electrónica, científicos de la computación, ingenieros en informática e ingenieros en telecomunicación. Aunque su creación se remonta a 1884, en 1963 fue que adoptó su nombre. Esta organización ha formalizado la gestión de riesgos con el modelo propuesto por Boehm y continúan desarrollando investigaciones y publicaciones en este campo. La IEEE definió además un estándar para la gestión de

riesgos: 1540-2001, que puede combinarse con el estándar 12207, también llamado: de procesos de ciclo de vida de software, o puede ser usado independientemente.

- International Standard Organization (ISO). Es la Organización Internacional para la Estandarización. Creado en 1947, es el organismo encargado de promover el desarrollo de normas internacionales de fabricación, comercio y comunicación para todas las ramas industriales a excepción de la eléctrica y la electrónica. Su función principal es la de buscar la estandarización de normas de productos y seguridad para las empresas u organizaciones a nivel internacional. ISO desarrolló una norma para la gestión de riesgos: ISO/IEC 16085:2006, que define un proceso para la administración de riesgos y puede ser aplicada en el ciclo de vida de un proyecto de software.
- Risk World: Es uno de los primordiales repositorios de materiales de riesgos. Reúne una gran variedad de referencias a noticias, publicaciones internacionales, libros, bases de datos de riesgos de diferentes sectores, herramientas para ayudar en la gestión de riesgos; asociaciones e institutos en el mundo enfocados a la gestión de riesgos, grupos de discusión, artículos y otros materiales. Durante algunos años fue el publicador del "Risk Analysis, an International Journal" de la Society for Risk Analysis (SRA).

1.4 Principales modelos de gestión de riesgos.

En la gestión de riesgo existen varios modelos que sintetizan cada uno de los pasos que se deben tener presentes a la hora de llevar a cabo la Gestión para realizar un proceso más organizado y eficaz.

A continuación se presentan unas series de modelos, donde cada uno en particular, esquematiza la gestión de riesgos y la etapa de análisis de una forma diferente.

1.4.1 Modelo de Boehm.

La IEEE fue una de las primeras instituciones en formalizar la Gestión de Riesgos con el modelo en espiral propuesto por Boehm en 1989 (Boehm, 1991), que ha evolucionado incorporando técnicas y elementos para lograr un modelo más efectivo pero en esencia se mantienen las bases iniciales (Figura 1).

El modelo en espiral es un modelo evolutivo que conjuga la naturaleza iterativa del Modelo de Prototipos con los aspectos controlados y sistemáticos del Modelo Cascada. Proporciona potencial para desarrollo rápido de versiones incrementales. En el modelo Espiral el software

se construye en una serie de versiones incrementales. En las primeras iteraciones la versión incremental podría ser un modelo en papel o bien un prototipo. En las últimas iteraciones se producen versiones cada vez más completas del sistema diseñado.

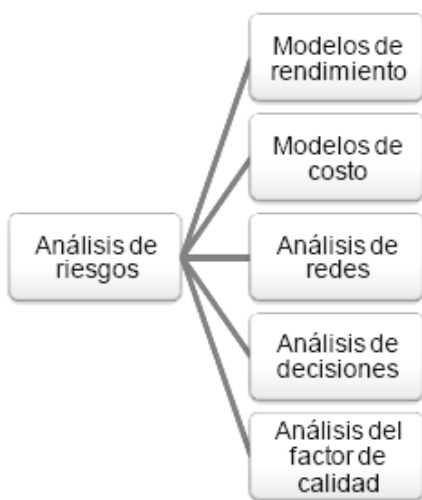


Figura 1: Modelo propuesto por Boehm de la Gestión de Riesgo. Adaptado de (Boehm, 1991)

El modelo de gestión de riesgos de Boehm propone procesos y técnicas divididas en dos partes principales: medición de los riesgos y control de los riesgos. Dentro de la medición incluye procesos como Identificación, Análisis y Priorización. Estos procesos están bien ubicados en esta categoría de acuerdo a su intención, sin embargo la priorización de los riesgos debe considerarse como un proceso dentro del análisis, como una función del análisis que utilizando los datos generados organiza los riesgos y no como un proceso al mismo nivel del resto (Boehm, 1989).

Para el proceso de análisis propone modelos de rendimiento, modelos de costo, análisis de redes, análisis del factor de calidad. Este proceso se considera importante y con el objetivo de extraer o descubrir la mayor cantidad de datos posibles de los riesgos y el uso de estas técnicas contribuye en este sentido, sin embargo, la descripción precisa y objetiva del uso de estas técnicas se describen abstractamente y carecen de claridad para ser usadas.

Este modelo de gestión de riesgos requiere mucha experiencia y habilidad para la evaluación de los riesgos, lo cual es requisito para el éxito del proyecto.

1.4.2 Continuous Risk Management.

The Software Assurance Technology Center de la NASA desarrolló en conjunto con el SEI el Continuous Risk Management a principios de los 90, un método de gestión de riesgos extendido a la comunidad de la NASA. El método reconoce la importancia de la gestión de riesgos como una práctica necesaria y define seis principios o funciones para ello (Figura 2).

El propósito del análisis es convertir los datos en la información base de la toma de decisiones, es el proceso de examinar cada riesgo en detalle para determinar su extensión, cómo se relaciona con otros, y cuáles son más importantes. El análisis de los riesgos tiene tres procesos básicos: la evaluación de los atributos de los riesgos (impacto, probabilidad y margen de tiempo), la clasificación y la priorización o ranking de los riesgos (Rosenberg et al., 1999).



Figura 2: Modelo propuesto por el CRM para la Gestión de Riesgos. Adaptado de (Rosenberg et al., 1999; Rosenberg, Al Gallo, Ted Hammer, & Parolek, 2000).

En la evaluación se provee de un mejor entendimiento del riesgo cualificando el impacto esperado, la probabilidad y el margen de tiempo. Se entiende por impacto al efecto negativo que podría tener el proyecto de ocurrir un riesgo, la probabilidad es la posibilidad que el riesgo ocurra y al margen de tiempo se le llama al período de tiempo en que se toma una acción para mitigar el riesgo. En la clasificación no se definen categorías. El principal propósito de este proceso es entender la naturaleza o fuente de los riesgos, eliminar los riesgos duplicados o combinar los equivalentes (Rosenberg, Al Gallo, Ted Hammer, & Parolek, 2000).

En la priorización no se define un criterio o algún método, solo se destaca el uso del sentido de los pocos vitales (vital few) para decidir el orden de los riesgos, que es usado posteriormente para asignar recursos al proyecto para mitigarlos pues las condiciones y las prioridades cambian constantemente en el proyecto. Esta técnica de priorización carece de argumentos objetivos pues no puede juzgar cuáles riesgos serán más importantes simplemente por un criterio arbitrario. (Rosenberg et al., 1999; Murphy et al., 2008).

1.4.3 Software Risk Evaluation.

Es un método para identificar, analizar, evaluar y desarrollar estrategias de mitigación para los riesgos; es una herramienta de diagnóstico para crear una línea base de riesgos para un proyecto. El Software Engineering Institute (SEI) propone como método a utilizar en conjunto con CRM y el Team Risk Management (TRM)

Está concebido como un servicio a prestar a otras empresas, es implementado en cinco fases: contratación (actividades necesarias para identificar los objetivos del proyecto, acordar elementos sobre el método y coordinar los recursos), identificación y análisis de los riesgos (los riesgos son licitados de acuerdo a un procedimiento bien definido y documentado, son analizados y priorizados según lo define el método, y son entregados al gerente del proyecto), reporte interino (las áreas de riesgo se analizan y se preparan recomendaciones para las que se analizarán en las fases siguientes), planeación de estrategia de mitigación (un plan de mitigación de alcance global es realizado) y reporte final (contiene todos los datos relevantes recogidos durante el proceso y son el resultado del método).

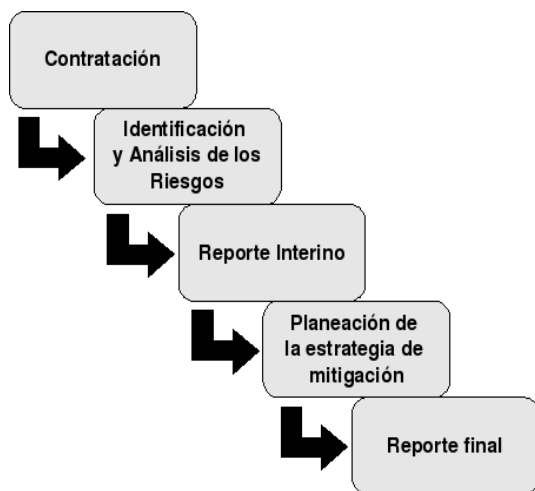


Figura 3: Etapas de la Gestión de Riesgos según Williams et al., 1999

El método describe cómo tiene que estar estructurado el equipo de SRE para aplicarlo. Describe cada fase definiendo su objetivo, los elementos que maneja (entradas y salidas), las personas involucradas, las herramientas a utilizar, las tareas a realizar y el tiempo promedio necesario para realizarlas. (Williams et al., 1999)

En la fase de identificación y análisis de los riesgos se solicitan e identifican los riesgos mediante entrevistas a los miembros del proyecto (Figura 3), luego se analizan y categorizan en áreas. La fase define 8 actividades:

Conducir Resumen del Proyecto (donde el equipo conoce datos del proyecto), Conducir el Resumen de Inicio (donde los miembros del equipo entienden los objetivos del método y las actividades que se estarán realizando), Preparar el Equipo (donde se imparte un pequeño entrenamiento a los miembros del equipo no experimentados en la aplicación de SRE), Conducir Entrevistas (donde se identifican los riesgos), Evaluación de los Participantes, Analizar la Sesión, Consolidación, Resumen de Confirmación de Datos.

El método define un grupo de sesiones de 3 horas de duración donde se identifican definiciones de riesgos documentados de la forma condición-consecuencia.

Para las entrevistas se define y describe en detalle un protocolo y el método es estricto en su cumplimiento para lograr mayor efectividad. Se definen roles durante la entrevista y sus actividades (Entrevistador, Entrevistado, Documentador de Riesgos, Documentador de Sesión, Compilador de Datos) así como las estrategias que deben seguir durante la entrevista basados en la cantidad de riesgos encontrados, en las áreas cubiertas y en la calidad de las definiciones.

Luego de que el riesgo sea definido por el entrevistado, este debe evaluarlo y revisarlo hasta que el equipo observe que el entrevistado sienta el riesgo como que él lo creó. Este aspecto humanístico es importante para crear una cultura hacia la Gestión de Riesgos y en especial hacia este método.

Los riesgos son priorizados de acuerdo a su exposición y al criterio de los entrevistados, en este sentido el método no define técnicas “fuertes” para la priorización (Higuera & Haines, 1996). Como resultado de esta priorización se obtiene un documento con la lista de riesgos

que es entregado al gerente de proyecto.

En la consolidación se reorganizan los riesgos en áreas, se revisan los contextos y se reconcilia la exposición. Este proceso consiste en determinar un promedio entre los criterios de impacto y probabilidad definidos para el riesgo según cada miembro. Esta técnica puede ser efectiva aunque se pudieran utilizar técnicas más precisas por ejemplo una suma ponderada donde se exprese la importancia de un objetivo sobre otro.

En el resumen de confirmación de datos se presentan los datos recogidos a los participantes y los resultados del análisis hecho. El método describe los elementos que debe incluir dicha presentación.

1.4.4 PMBOK.

El PMBOK®, un cuerpo de conocimiento para la gestión de proyectos y estándar nacional para los Estados Unidos desde 2001 (ANSI/PMI 99-001-2000) propone una guía genérica básica para la gestión de los riesgos que analizaremos a continuación.

Propone la siguiente serie de procesos involucrados en la gestión de riesgos: Planeación de la Gestión de Riesgos, Identificación de los riesgos, Análisis cualitativo, Análisis cuantitativo, Planeación de la respuesta a los riesgos y Monitoreo y control.

Esta edición incluye algunos elementos nuevos respecto a la edición anterior que la hacen más integral:

- Estrategias de respuesta a las oportunidades de los riesgos (explotar, mejorar, aceptar, compartir). Menciona que los riesgos no son siempre negativos, los riesgos son eventos que pueden o no ocurrir, y de hacerlo tienen un impacto positivo o negativo. Es importante identificarlos todos para poder tomar los riesgos que pueden traer beneficios si se está claro también del impacto negativo que puede provocar y por supuesto, tomar las medidas pertinentes para los negativos. Existen varios trabajos sobre la “oportunidad” de los riesgos positivos como “Software Development Risk: Opportunity, Not Problem”.
- Algunas de las relaciones de los procesos de la gestión de riesgos con otros procesos de la gestión de proyectos (como el Control Integrado de Cambios), otros elementos (como los factores ambientales de la empresa, bases de datos comerciales, políticas y procedimientos, base de conocimientos), y otras tareas que detalla un poco más.
- Categorización de los riesgos.

Análisis Cualitativo

El Análisis Cualitativo de Riesgos evalúa la prioridad de los riesgos identificados usando la probabilidad de ocurrencia, el impacto correspondiente sobre los objetivos del proyecto si los riesgos efectivamente ocurren, así como otros factores como el plazo y la tolerancia al riesgo de las restricciones del proyecto como coste, cronograma, alcance y calidad (PMI, 2004). Es normalmente una forma rápida y rentable de establecer prioridades para la Planificación de la Respuesta a los Riesgos, y sienta las bases para el Análisis Cuantitativo de Riesgos, si fuera necesario. El Análisis Cualitativo de Riesgos deberá ser revisado continuamente durante el ciclo de vida del proyecto para que esté actualizado con los cambios en los riesgos del proyecto.

En esta etapa PMBOK propone el empleo de las siguientes herramientas: evaluación de probabilidades e impacto de los riesgos, matriz de probabilidad e impacto, evaluación de la calidad de los datos sobre los riesgos y la evaluación de la urgencia del riesgo.

Análisis Cuantitativo

En este modelo, el Análisis Cuantitativo de los Riesgos se realiza respecto a los riesgos priorizados en el proceso de Análisis Cuantitativo de Riesgos por tener un posible impacto significativo sobre las demandas concurrentes del proyecto. Este proceso analiza el efecto de esos riesgos y les asigna una calificación numérica. Tiene como objetivos: cuantificar los posibles resultados del proyecto y sus probabilidades, evaluar la probabilidad de lograr los objetivos específicos del proyecto, identificar los riesgos que requieren una mayor atención mediante la cuantificación de su contribución relativa al riesgo general del proyecto, identificar objetivos de coste, cronogramas o alcance realistas y viables, dados los riesgos del proyecto, y determinar la mejor decisión de dirección de proyectos cuando algunas condiciones o resultados son inciertos.

El Análisis Cuantitativo de Riesgos generalmente sigue al proceso Análisis Cualitativo de Riesgos, si bien algunos directores de riesgos experimentados a veces lo realizan directamente después de la Identificación de Riesgos. En algunos casos, es posible que no sea necesario el mismo.

En (PMI, 2004) se recomienda que este proceso debe repetirse después de la Planificación de la Respuesta a los Riesgos, también como parte del Seguimiento y Control de Riesgos, para determinar si el riesgo general del proyecto ha sido reducido satisfactoriamente.

Entre las herramientas que propone figuran como técnicas de recopilación y representación

de datos: entrevistas, distribuciones de probabilidad, y juicio de expertos; y como técnicas de análisis cuantitativo de riesgos y de modelado: análisis de sensibilidad, análisis mediante árbol de decisiones, y modelado y simulación.

En sentido general, de las técnicas descritas por PMBOK, solo se describen con un nivel aceptable de detalle la referente a la matriz de probabilidad e impacto en el análisis cualitativo y el árbol de decisiones en el análisis cuantitativo. En esta última el objeto de análisis está enfocado en los costes, sin ejemplificar siquiera el tiempo, el alcance o los recursos humanos.

1.4.5 Microsoft Solution Framework.

Microsoft Solutions Framework (MSF) es un marco de trabajo propuesto por Microsoft que cubre el ciclo de vida del proyecto y propone métodos y herramientas para el desarrollo del mismo. Dentro de sus disciplinas proponen “Risk management discipline”, como un mecanismo para enfrentar la incertidumbre en los proyectos, proponen un enfoque proactivo donde continuamente se evalúen los riesgos.

Se basa en los principios fundamentales de MSF en general y además la Gestión de Riesgos específicamente, hace énfasis en el principio de “mantenerse ágil, esperar el cambio”: este principio no contradice el mantenerse proactivo, se refiere a que cuando se espera que las cosas pueden cambiar, se está más abierto a identificar elementos que pueden hacer que las cosas cambien, y actuar antes de que ocurran.

El principio de las comunicaciones abiertas, es fundamental y al igual que otros modelos, MSF enfatiza la necesidad de que participen tanto el equipo de desarrollo como los interesados claves en un ambiente sincero de comunicación.

El principio de “aprender de todas las experiencias” ofrece un punto de retroalimentación a su disciplina; consiste en aumentar el conocimiento a través de un análisis para en futuros proyectos disminuir la incertidumbre sobre algunas situaciones y contribuir a la toma de mejores decisiones en general.

El último principio es la “responsabilidad compartida”, a diferencia de otros modelos o tendencias MSF propone que nadie es dueño de un riesgo. Todos en el equipo son responsables por activamente participar en la Gestión de Riesgos (no quiere esto decir que alguien modere y controle este proceso). (García-Hernández, 2008)

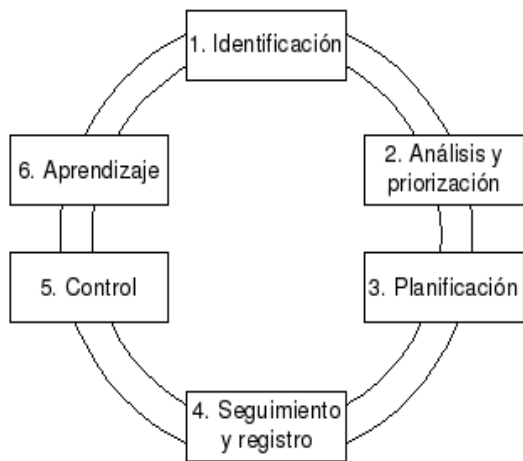


Figura 4: Procesos definidos por Microsoft para la Gestión de Riesgos. Adaptado de (Microsoft, 2004)

MSF define seis procesos para la gestión de riesgos (Figura 4):

El objetivo del proceso “Análisis y priorización” es convertir los datos del riesgo en un formulario utilizable en la toma de decisiones, y priorizar los riesgos obteniendo un ranking para centrarse posteriormente en los más altos riesgos.

Durante esta etapa se analiza la probabilidad del riesgo, para ello MSF define una escala con valores probabilísticos, verbales y numéricos para utilizar.

La asignación de valores numéricos es un aspecto importante pues calculan la exposición del riesgo a partir de estos valores como se verá posteriormente. Respecto a otros modelos, la asignación de valores numéricos es un aspecto positivo en MSF.

También en este proceso se determina el impacto del riesgo, es decir la magnitud de la pérdida o la ganancia para determinados objetivos. MSF también propone una escala de valores para identificar la magnitud en cuanto a costo y tiempo. A cada valor de la escala se le asignan valores numéricos lineales, no utilizan ninguna otra “función” lo que indicaría el grado de aceptación de los valores y puede resultar interesante en el análisis de los riesgos.

MSF incluye un elemento interesante: la exposición del riesgo. La exposición de un riesgo es una medida de su amenaza, y se define una función para calcular la exposición dependiendo del grado de aceptación de la entidad. MSF propone una forma de calcularla y destaca que la utilización de una matriz (probabilidad x impacto) ayuda a los “decision-makers” pues se puede hacer una escala de colores para ilustrar las exposiciones altas o bajas.

Los riesgos se deben ordenar por la exposición, y adicionalmente plantean una fórmula para que determinadas propiedades de los riesgos influyan en su posición del Rank en mayor o menor medida.

Como resultado de este proceso se obtiene la “master risk list” que contienen una lista de los riesgos con los atributos primarios identificados y evaluados. Adicionalmente definen otros métodos y otra representación de los riesgos en esta etapa.

Esta representación de los riesgos es una de las más completas encontradas ya que recoge

varios atributos de los riesgos en una estructura única y entendible. (Microsoft, 2005)

Adicionalmente proponen que se mantenga una lista pequeña con los riesgos más altos: “top-risks”, la intención de esta lista no es otra que mantener un registro de las altas amenazas.

Aunque no lo describen en su representación del riesgo, MSF define un estado que pueden tener (desactivado) aquellos riesgos que se decida que su gestión no merita la pena, no se define ningún criterio o métrica para esta decisión. (García-Hernández, 2008)

Este modelo incluye como un proceso de la gestión de riesgos que realiza una etapa de aprendizaje, en la cual sugiere se cuente con una base de datos de riesgos donde al concluir cada proyecto se actualice con los nuevos riesgos identificados.

1.4.6 Magerit.

El Consejo Superior de Administración Electrónica ha elaborado Magerit y promueve su utilización como respuesta a la percepción de que la Administración (y en general toda la sociedad) depende de forma creciente de las tecnologías de la información para el cumplimiento de su misión. La razón de ser de Magerit está directamente relacionada con la generalización del uso de los medios electrónicos, informáticos y telemáticos, que supone unos beneficios evidentes para los ciudadanos; pero también da lugar a ciertos riesgos que deben minimizarse con medidas de seguridad que generen confianza.

En el periodo transcurrido desde la publicación de la primera versión de Magerit (1997) hasta la fecha, el análisis de riesgos se ha venido consolidando como paso necesario para la gestión de la seguridad. Conocer el riesgo al que están sometidos los elementos de trabajo es, simplemente, imprescindible para poder gestionarlos.

Magerit persigue los objetivos de concienciar a los responsables de los sistemas de información de la existencia de riesgos y de la necesidad de atajarlos a tiempo, ofrecer un método sistemático para analizar tales riesgos, ayudar a descubrir y planificar las medidas oportunas para mantener los riesgos bajo control y apoyar la preparación a la Organización para procesos de evaluación, auditoría, certificación o acreditación, según corresponda en cada caso.

Así mismo, se ha cuidado la uniformidad de los informes que recogen los hallazgos y las



Figura 5: Etapas propuestas por Magerit para la Gestión de Riesgos. Tomado de (MAP, 2006a)

conclusiones de un proyecto de análisis y gestión de riesgos: modelo de valor, mapa de riesgos, evaluación de salvaguardas, estado de riesgo, informe de insuficiencias, y plan de seguridad. (MAP, 2006a)

Las tareas de análisis y gestión de riesgos no son un fin en sí mismas sino que se encajan en la actividad continua de gestión de la seguridad.

El análisis de riesgos proporciona un modelo del sistema en términos de activos, amenazas y

salvaguardas, y es la piedra angular para controlar todas las actividades con fundamento. La gestión de riesgos es la estructuración de las acciones de seguridad para satisfacer las necesidades detectadas por el análisis.

En esta fase se proponen las siguientes técnicas: (MAP, 2006b)

- Técnicas específicas: análisis mediante tablas, análisis algorítmico, árboles de ataque.
- Técnicas generales: análisis de coste – beneficio, diagrama de flujo de datos, técnicas gráficas(diagramas GANTT, gráficos de radar), planificación de proyectos (diagramas PERT) y valoración Delphi.

1.4.7 MOGERI.

El Modelo de gestión de Riesgos para proyectos de Software en la UCI (MOGERI) consta de seis procesos (Figura 6) (Zulueta-Véliz & Despaigne-Herrera , 2007): Planificación de la gestión de los riesgos, Identificación de los riesgos, Análisis de los riesgos, Planificación de las respuestas a los riesgos, Seguimiento y Control de los riesgos, y Comunicación de la información de los riesgos.

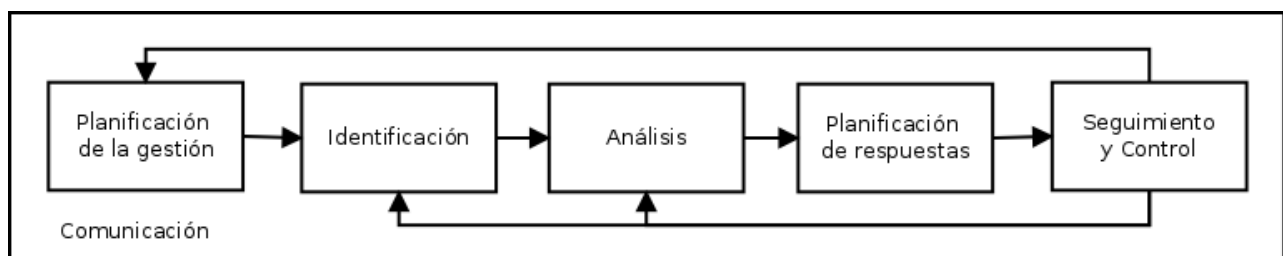


Figura 6: Modelo de procesos de gestión de riesgos. Tomado de (Véliz, 2007)

El funcionamiento del modelo se basa en la realización de un conjunto de actividades por cada proceso, y a su vez el cumplimiento de cada actividad está representado por el desarrollo de diferentes tareas. En este, el proceso de análisis se divide en cuatro actividades (Zulueta-Véliz, 2007):

- análisis cualitativo: se estima la probabilidad y se priorizan los riesgos,
- análisis cuantitativo: se cuantifica la probabilidad de ocurrencia e impacto de los riesgos,
- análisis de los atributos de los riesgos: se verifica la exactitud de los datos, estimaciones y cálculos realizados,
- comunicar resultados: se comunican los resultados al equipo del proyecto y se documentan experiencias.

La aplicación de un análisis cualitativo como primer paso, puede servir para realizar un trabajo de priorización previo y enfocar el análisis cuantitativo posterior a los riesgos de mayor trascendencia según los objetivos del proyecto. En este caso las herramientas propuesta son la matriz de probabilidad e impacto y las reuniones de análisis.

Los riesgos luego de un análisis cualitativo pueden ser listados por prioridades separadamente por el costo, tiempo o alcance y el gestor del proyecto puede decidir cuáles gestionar primero o realizar el análisis cuantitativo para tener datos más precisos sobre las pérdidas que pueden darse.

Si bien el análisis cuantitativo de los riesgos puede ser más costoso por el tiempo que requiere y la preparación del personal acerca de las técnicas a emplear, es muy necesario en ocasiones, para esclarecer los atributos del riesgo y minimizar la incertidumbre que entraña el tema “riesgo”. Aunque es importante acotar que las técnicas matemáticas pueden causar una falsa impresión si no son correctamente aplicadas (Zulueta-Véliz, 2007). Las herramientas propuestas para estos casos son: método de estado natural, análisis de sensibilidad, árboles de decisión, valor monetario esperado, sumas estadísticas, simulación, opinión experta y reuniones de análisis.

1.5 Técnicas y herramientas empleadas en la etapa de análisis en la gestión de riesgos

Árboles de decisión.

Los árboles de decisión son una técnica de recuperación y organización de la información que permite, a partir de una serie de ejemplos, clasificarlos de acuerdo a los valores de sus atributos, para determinar el valor de otro de sus atributos. Esta clasificación se realiza de modo descendente de acuerdo a los valores de los diferentes atributos. El árbol creado se corresponde con una clasificación de acuerdo al valor de los atributos de las instancias con las que se ha realizado el aprendizaje, por lo que al aparecer nuevas instancias se espera que el árbol tenga un alto acierto de clasificación, aunque no asegura que vaya a clasificar todas las instancias correctamente. No obstante, la recuperación con árboles de decisión es una técnica que da buenos resultados dentro de la recuperación y organización de la información (Litten, 2009).

Sus aplicaciones enfocadas al tratamiento de riesgos son muy variadas, entre ellas se puede encontrar en el riesgo crediticio, en la cual se presenta la utilización de los árboles de decisión como herramienta para el cálculo de probabilidades de incumplimiento en crédito de bancos (Cardona, 2004), con el propósito de realizar una evaluación del riesgo agroambiental en los suelos (Morin, 2007), para abordar el problema de la detección precoz de la insolvencia empresarial (Díaz-Martínez, Fernández-Menéndez, & Segovia-Vargas, 2004). Además se encuentra como método a emplear en el modelo de gestión de riesgos propuesto por Boehm (Boehm, 1991), por el Project Management Institute (PMI, 2004), se propone como técnica específica en MAGERIT (MAP, 2006a) y se tiene en cuenta en una propuesta sobre la gestión de riesgos en proyectos de tecnología de la información realizada en (Gallegos, 2006).

Simulación Montecarlo

El método de Montecarlo es una herramienta de investigación y planeamiento; básicamente es una técnica de muestreo artificial, empleada para operar numéricamente sistemas complejos que tengan componentes aleatorios (Laurent, 1966).

Gracias a la constante evolución de las microcomputadoras, en lo que se refiere a su capacidad de procesamiento de la información, el método de Montecarlo es cada vez más frecuentemente utilizado.

Para ello son realizadas diversas simulaciones donde, en cada una de ellas, son generados valores aleatorios para el conjunto de variables de entrada y parámetros del modelo que están sujetos a incertidumbre. Tales valores aleatorios generados siguen distribuciones de probabilidades específicas que deben ser identificadas o estimadas previamente (Fiorito,

2006).

Este método es sugerido en el modelo de referencia propuestos por el PMI, y se recomienda su empleo para determinar coste de elementos del proyecto o duración de actividades (PMI, 2004).

Análisis de sensibilidad

El análisis de sensibilidad ayuda a determinar qué riesgos tienen el mayor impacto posible sobre el proyecto. Este método examina la medida en que la incertidumbre de cada elemento del proyecto afecta al objetivo que está siendo examinado, cuando todos los demás elementos inciertos se mantienen en sus valores de línea base. Una representación típica del análisis de sensibilidad es el diagrama con forma de tornado, que es útil para comparar la importancia relativa de las variables que tienen un alto grado de incertidumbre con aquellas que son más estables (PMI, 2004).

En (Houston, Mackulak, & Collofello, 2001), los autores proponen esta técnica como apoyo a los métodos de análisis de regresión, sistemas expertos y modelación estocástica en los procesos de identificación y seguimiento de los riesgos.

Boehm, por su parte (Boehm, 1991), vincula el análisis de sensibilidad a los árboles de decisiones y señala que este método es muy útil en situaciones donde las probabilidades de ganancia o pérdida no pueden ser estimadas lo suficiente como para garantizar un análisis preciso.

Distribuciones de probabilidad

Según (PMI, 2004), las distribuciones continuas de probabilidad representan la incertidumbre de los valores, como las duraciones de las actividades del cronograma y los costes de los componentes del proyecto, mientras que las distribuciones discretas pueden usarse para

representar eventos inciertos, como el resultado de una prueba o un posible escenario en un árbol de decisiones.

La Figura 7 muestra dos ejemplos de distribuciones continuas ampliamente usadas. Estas distribuciones asimétricas representan formas que son compatibles con los datos generalmente desarrollados

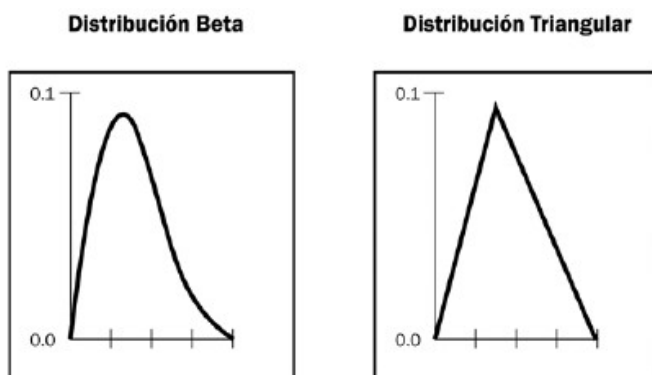


Figura 7: Ejemplo de distribuciones de probabilidad comúnmente usadas. Tomado de (PMI, 2004).

durante el análisis de los riesgos del proyecto. Las distribuciones uniformes pueden usarse si no hay ningún valor obvio que sea más probable que cualquier otro entre límites altos y bajos especificados, como en la etapa inicial de concepto de diseño. Las distribuciones beta y triangular se usan frecuentemente en el análisis cuantitativo de riesgos. Los datos que se muestran aquí son un ejemplo de una familia de dichas distribuciones determinadas por dos parámetros de forma. Otras distribuciones comúnmente usadas incluyen la uniforme, la normal y la lognormal. En estos diagramas los ejes horizontales representan los valores posibles de tiempo y coste, y los ejes verticales la probabilidad relativa.

Juicio de expertos

El juicio de expertos se establece recopilando opiniones emitidas por informantes calificados acerca de los niveles de validez de una técnica, entendiéndose por validez la coherencia entre lo que la técnica observa y lo que con ella se pretende observar. En otras palabras, lo que se busca es constatar si es coherente la relación entre las preguntas que incluye el formato de la técnica, los indicadores, temas o preguntas orientadoras, con los resultados o dimensiones de análisis.

Para llevar a cabo este proceso es muy frecuente el uso de formatos que permiten a los expertos convocados visualizar la coherencia entre los elementos mencionados.

En términos muy puntuales, luego de convocar la participación de quienes actuarán como expertos y de informarles sobre cómo deben desarrollar el proceso de valoración, el proceso requiere los siguientes pasos:

- Valorar la pertinencia de los indicadores o preguntas orientadoras respecto de los resultados o dimensiones de análisis.
- Valorar la pertinencia de las preguntas de la técnica respecto de los indicadores o preguntas orientadoras.
- Emitir una opinión sobre cada pregunta, utilizando la escala y los espacios de opinión literal (comentarios).

El procedimiento de "juicio de expertos" es utilizado para la validación de técnicas cuantitativas (como las encuestas) y cualitativas (como las entrevistas o los grupos focales).

Ampliamente utilizado por PMI, forma parte del conjunto de técnicas usadas en todas sus áreas de conocimientos, en la documentación del SRE (Williams et al., 1999) se incluye también el uso de la experticia como apoyo fundamental a la toma de decisiones.

Delphi

La técnica o método Delphi, original de la Rand Corporation (Research and Development), comenzó a aplicarse desde 1948 en un proyecto avanzado de las Fuerzas Aéreas de los Estados Unidos y la Compañía Douglas de Aviación, orientándose desde entonces a los estudios prospectivos de investigación espacial. De forma paulatina la técnica diseñada por la Rand Corporation ha ido ampliando sus campos de aplicación: así esta "reflexión intuitiva de expertos", puede ser utilizada con éxito en multitud de campos y sectores (Estévez & Gallastegui, 2002).

- Es una técnica netamente cualitativa que relativamente permite tratar con alta precisión problemas técnicamente complejos.
- Está planteada como una reflexión organizada de expertos sobre un tema concreto, reflexión que permite recoger las ideas y opiniones más cualificadas en el ámbito de la seguridad (valoración de activos e identificación de amenazas e impactos).
- Se desarrolla a partir de un cierto 'escenario inicial' de modo que permita una adecuada recapitulación e identificación de los problemas que ya existen actualmente.
- Desarrolla una prospectiva mucho más rica que la mera identificación de la opinión mayoritaria, por medio de un proceso de convergencia de opiniones que se consigue mediante rondas sucesivas de entrevistas (Ricardo Mauricio Aflaro-López, Víctor Danilo Castro-Rivera, & Balmero Osmar Romero-Lainez, 2005).
- Garantiza satisfactoriamente la 'limpieza' de la investigación, impidiendo el predominio de unos expertos sobre otros por razones ajenas a la calidad de sus opiniones.

La técnica Delphi es un instrumento de uso múltiple que se utiliza con muy variados objetivos (Trujillo-Cabezas, 2004): identificar problemas, desarrollar estrategias para la solución de problemas, fijando un rango de alternativas posibles, identificar factores de resistencia en el proceso de cambio, establecer previsiones de futuro sobre la evolución de las tendencias que se observan en un determinado campo o sector, y/o contrastar opiniones en un tema abarcando un amplio campo de disciplinas o sectores.

Esta técnica se propone en las "Guías de Técnicas" de MAGERIT (MAP, 2006a) donde se expone con una amplia documentación al respecto. PMI (PMI, 2004), por su parte, indica su uso en la etapa de Identificación de los riesgos.

Matriz de probabilidad e impacto.

Aunque es importante identificar el mayor número posible de riesgos del proyecto, en muchos casos el número de riesgos identificados puede ser abrumador, y lógicamente el equipo de trabajo no podrá realizar un seguimiento ni una gestión efectiva de todos ellos. Una solución sería agrupar los riesgos en función de sus prioridades de tal forma que el equipo pueda centrarse en los más críticos.

La evaluación de la importancia de cada riesgo y, por consiguiente, de su prioridad, generalmente se realiza usando una matriz de probabilidad e impacto (matriz P-I). Esta matriz asignará categorías a los riesgos basándose en la combinación de dichos factores (probabilidad e impacto) que llevan a la calificación de los riesgos como de prioridad baja, moderada o alta (Figura 8). Pueden usarse términos descriptivos o valores numéricos, dependiendo de la preferencia de la organización. Las reglas para calificar los riesgos pueden adaptarse al proyecto específico en el proceso planificación de la gestión de riesgos (PMI, 2004).

		Probabilidad de riesgo			
		Muy alto	Alto	Medio	Bajo
Impacto del riesgo	Muy alto	Muy alto	Alto	Alto	Alto
	Alto	Alto	Alto	Medio	Medio
	Medio	Alto	Medio	Medio	Bajo
	Bajo	Bajo	Bajo	Bajo	Bajo

Una vez asociado a cada riesgo un valor estimado cualitativo de probabilidad e impacto, se propone un ejemplo de una Matriz P-I con tres distintas clasificaciones de riesgos,

Figura 8: Matriz de probabilidad e impacto. Tomado de (Inteco, 2008).

basados en niveles de probabilidad e impacto de riesgo, para definir el valor cualitativo de cada riesgo.

De esta forma podemos seleccionar qué riesgos merecen un mayor estudio, esfuerzo y respuesta. Algunos riesgos bajos o incluso medios, son posibles candidatos a ser tratados como suposiciones. Es recomendable realizar un análisis cualitativo de las suposiciones que haya en el proyecto ya que pueden convertirse en riesgos (Inteco, 2008).

En caso de que haya un número alto de riesgos dentro de la categoría 'Alto', es recomendable priorizar dichos riesgos identificando los "n" riesgos más altos dentro de esta categoría, siendo este número determinado por la organización en función de su situación. Entre los indicadores de prioridad pueden incluirse: las categorías determinadas para el riesgo, el impacto de los riesgos identificados, el número de riesgos, el tipo de riesgos o el tiempo para dar una respuesta a los riesgos

El resto de riesgos no seleccionados se deberán vigilar en otras fases ya que pueden cambiar su estado (aumente su probabilidad o cambie su impacto potencial).

Además de las técnicas presentadas, existen una gran cantidad de herramientas software de gestión de riesgos disponibles en el mercado, las cuales siguen diferentes metodologías. Las más significativas son las siguientes:

EAR/PILAR

Software que pone en práctica y amplía la Metodología MAGERIT (ver 1.4.6 Magerit. página 20) y es diseñado para apoyar el proceso de gestión de riesgos a lo largo de períodos largos, proporcionando el análisis incremental. Proporciona cálculos rápidos, y genera una gran cantidad de texto y gráfico. Plantean que los activos están expuestos a amenazas que, cuando se materializan, degradan el activo, produciendo un impacto. Si se estima la frecuencia con que se materializan las amenazas, se puede deducir el riesgo al que está expuesto el sistema.

Proveedor: Centro Criptológico Nacional (España). Idioma: español, inglés, francés, portugués, italiano y otros. Sistemas Operativos: Windows, Unix. Licenciamiento: Comercial. Última versión: 4.3 (2009).

Technical Risk Identification and Mitigation System (TRIMS).

Herramienta integrada de Gestión de Riesgos que emplea ingeniería de conocimientos y se enfoca en la identificación, medición y seguimiento de riesgos técnicos de proyectos.

Proveedor: Best Manufacturing Practices. Plataforma: Win32. Sistemas Operativos: Windows 98, ME, NT, 2000, y XP. Idioma: inglés. Licenciamiento: Libre. Posee gerenciamiento integrado de riesgos y generación de reportes. Orientada a categorías de riesgos predefinidas para sectores específicos.

@Risk.

Herramienta de Inteligencia de Negocios, que muestra todos los resultados posibles de una situación de negocios e indica la probabilidad de que ocurran.

Proveedor: Palisade. Sistemas Operativos: Windows 98, 2000 y XP. Idioma: Español, inglés, alemán, francés y japonés. Licenciamiento: Comercial.

Chinchón.

Proveedor: Free. Plataforma: Java. Licenciamiento: Libre. Es una herramienta para analizar

cuantitativamente el riesgo de un sistema de información. La herramienta sigue el modelo MAGERIT 1.0. Su versión 1.3 fue elaborada por el Dr. José Antonio Mañas, profesor de la Escuela Técnica Superior de Ingenieros de Telecomunicaciones de la Universidad Politécnica de Madrid, España. La entrada se escribe en XML, y realiza un análisis de la posición de riesgo, sirviendo de apoyo a su gestión. Los derechos de propiedad intelectual pertenecen al autor, quien ha puesto la herramienta en el dominio público.

SCRAM99.

SCRAM99 es un software que permite evaluar el nivel del calendario y los riesgos de costo de un proyecto, identificando las áreas del proyecto que se beneficiarían con una mayor atención de gestión, identificando las tendencias de riesgos con el propósito de prevenir los problemas antes de que se produzcan, establece y mantiene el plan de contingencia, y analiza sus impactos potenciales. Es un software propietario. Es un software para ejecutar simulación de Monte Carlo en MS Project con MS Excel. Realiza análisis probabilístico de riesgos usando estimación fuzzy para predecir los resultados de un proyecto. Simula la ejecución de proyectos, estima la duración y costo con los criterios: Optimista, Más probable, y Pesimista. Define los puntos de decisión y uso de lógica de decisión para modelos de proyectos exactos con pruebas, y planes de contingencia.

Valoración sobre las herramientas actuales.

Las herramientas actuales para el análisis cuantitativo de riesgos están caracterizadas por un efectivo y rápido análisis de riesgos. Presentan los resultados con gráficos que facilitan la comprensión y generan detallados reportes, los cuales le permiten al cliente saber qué puede o esté marchando mal en su organización. La mayoría de estas herramientas poseen la desventaja de que no se comercializan libremente y el enfoque está definido más desde el punto de vista cualitativo que cuantitativo. No se observó que alguna implementará técnicas de razonamiento basado en casos o empleo de la experiencia acumulada para determinar soluciones a los nuevos riesgos.

1.6 Sistemas de razonamiento basados en casos y su vinculación con la gestión de riesgo.

Los sistema de razonamiento basados en casos tienen un grupo de características que pueden ser aprovechadas en el proceso de análisis de riesgos en proyectos de software. Entre ellas se encuentran:

- razonan desde episodios específicos: la identificación de riesgos similares no pueden ser explicadas mediante reglas generalizables y es mucho más fácil describir las características de los riesgos que ya se han tratado,
- el mantenimiento del conocimiento no requiere la intervención del experto: la experiencia de riesgos tratados puede ser evaluada y si es correcta se puede incorporar a la base de casos,
- reutilizan soluciones previas: muchos riesgos que se presentan durante el desarrollo de los proyectos tienen soluciones similares y se puede reutilizar su solución.

El Razonamiento Basado en Casos (RBC) representa un nuevo método para resolver problemas no estructurados, en el cual el razonamiento se realiza a partir de una memoria asociativa que usa un algoritmo para determinar una medida de semejanza entre dos objetos. En este paradigma, la base del comportamiento inteligente radica en recordar situaciones similares existentes en el pasado (Kolodner & Leake, 1996). Debe destacarse que es una técnica, en la cual la memoria se sitúa como fundamento de la inteligencia artificial y más concretamente de los sistemas basados en el conocimiento.

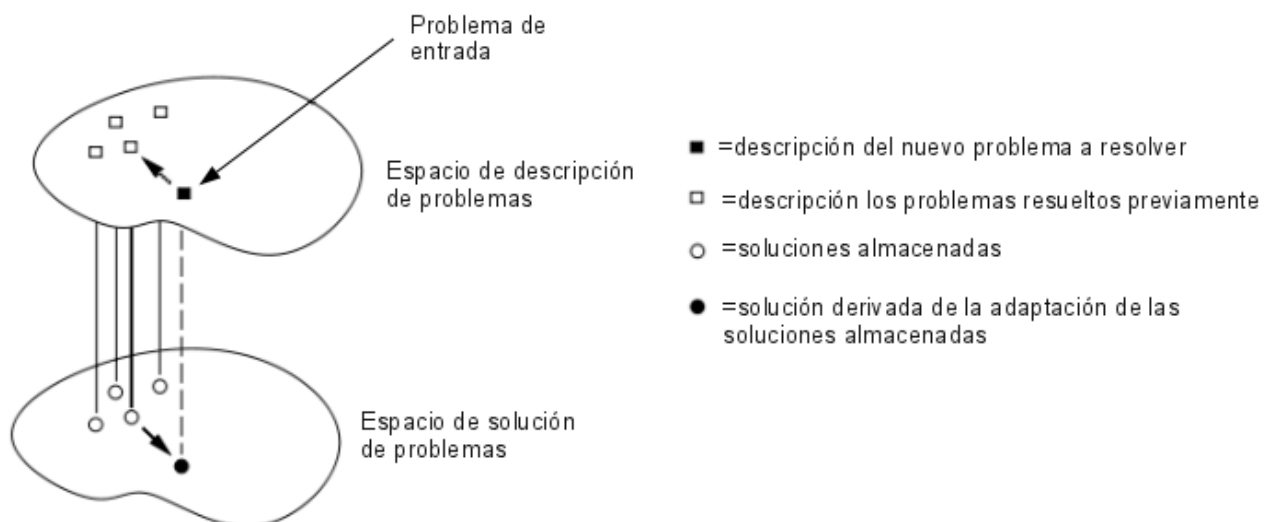


Figura 9: Representación de cómo RCB da solución a un problema. Adaptado de (Leake, 1996)

El RBC no es más que otro paradigma de resolución de problemas, pero son precisamente las diferencias con el resto de los acercamientos de la inteligencia artificial las que lo hacen tan especial. En lugar de confiar únicamente en el conocimiento general del dominio del problema, o realizar asociaciones a lo largo de relaciones entre descripciones del problema y conclusiones, este paradigma es capaz de utilizar conocimiento específico de experiencias previas, es decir, situaciones de un problema concreto (casos). Ante el planteamiento de un problema no abordado con anterioridad, se intenta localizar un caso pasado similar y adaptar

su solución a la situación del problema nuevo (Figura 9). De esta adaptación podemos obtener una nueva experiencia a la hora de resolver problemas con ciertas similitudes, lo que nos lleva a una segunda diferenciación del RBC con respecto al resto de tendencias, el aprendizaje incremental, ya que las nuevas adaptaciones se almacenan como nuevos casos, relacionados, y disponibles para comparaciones futuras (Leake, 1996).

Con las pocas nociones vistas hasta ahora ya se puede formular una primera definición para el Razonamiento Basado en Casos: Resolver un problema nuevo recordando una situación similar previa y reutilizando su información y conocimiento.

El RBC sugiere un modelo de razonamiento que incorpora los aspectos ya mencionados de resolución de problemas, entendimiento y aprendizaje e integra todo ello en meros procesos de memoria.

Según (Agudo, 2002) la calidad de un razonador basado en casos depende de:

1. la experiencia que tiene,
2. la habilidad para entender situaciones nuevas en términos de experiencias pasadas,
3. su capacidad de adaptación,
4. su capacidad de evaluación y preparación, y
5. su habilidad para integrar nuevas experiencias en su memoria adecuadamente.

El principal activo de un sistema RBC es la base de casos. Según Janet Kolodner y David Leake (Kolodner & Leake, 1996):

“Un caso es un fragmento contextualizado de conocimiento que representa una experiencia que enseña una lección importante para conseguir los objetivos del razonador”

En esta breve definición aparecen varias ideas importantes. Un caso contiene información útil en un contexto concreto; el problema es identificar los atributos que caracterizan al contexto y detectar cuándo dos contextos son similares. Un caso es una experiencia que enseña algo, ya que pueden haber experiencias que no aporten nueva información al sistema, lo cual plantea el problema de identificar cuándo dos casos superficialmente distintos, contienen información redundante entre sí. Y, por último, lo que el caso enseña es relativo a determinados objetivos, y por tanto, un caso puede ser útil o inútil dependiendo del objetivo, o resultar útil para distintos fines.

En su formulación más general un caso se compone de:

- la descripción de un problema,
- la descripción de una solución asociada al caso, y
- el resultado de aplicar la solución

En algunas propuestas se incluye una componente adicional: la justificación de un caso, proceso mediante el cual se justifican las decisiones adoptadas.

Las formas de representar la solución, la justificación y el resultado varían según el dominio de aplicación del sistema RBC y la aproximación utilizada. En (Althoff, Aamodt, Magaldi, & Milne, 1995) distinguen dos grandes tipos de representaciones para los casos:

Representaciones planas: En este tipo de representación se define una serie de atributos, cada uno con un conjunto de posibles valores de tipos simples -cadenas de caracteres, números, símbolos-. En general no se define relación alguna entre los atributos ni entre sus valores, esto es, si utilizamos n atributos en la representación, una caso vendrá descrito por la *n-tupla* de valores de los atributos del caso en cuestión (Arean-Rodríguez, 2003).

Representaciones estructuradas: También aquí se analizan listas de atributos y valores asociados, pero a diferencia de las representaciones planas, los valores pueden ser objetos que a su vez tienen atributos. De esta forma se consiguen expresiones más expresivas en las que se definen relaciones entre los atributos y/o entre los valores. Para implementar este tipo de descripciones se suelen utilizar cálculos de predicados, redes semánticas, lenguajes basados en marcos, lenguajes orientados a objetos y lógicas descriptivas, entre otros (Bergmann & Wilke, 1998; Schank, 1999; Zdrahal & Motta, 2001).

Además de los dos tipos anteriores existen otros, por ejemplo, la aproximación textual que utiliza el lenguaje natural como mecanismo de representación de casos, variando en la utilización de distintas estructuras de texto, o aproximaciones multimedia que incluyen la representación de gráficos, videos y/o sonidos (Bergmann & Stahl, 1998; Bregón et al., 2005; Dapena, 2006; Piñero-Pérez, 2005). Por supuesto, el tipo y la estructura de representación de casos influyen en los procesos encargados de manejar dichos casos (Olvera-López, Martínez-Trinidad, & Carrasco-Ochoa, 2005; Stanfill & Waltz, 1986; Wilson & Martinez, 2000).

Los sistemas RBC se descomponen en varias fases que se ejecutan ante la aparición de una nueva situación (Figura 10):

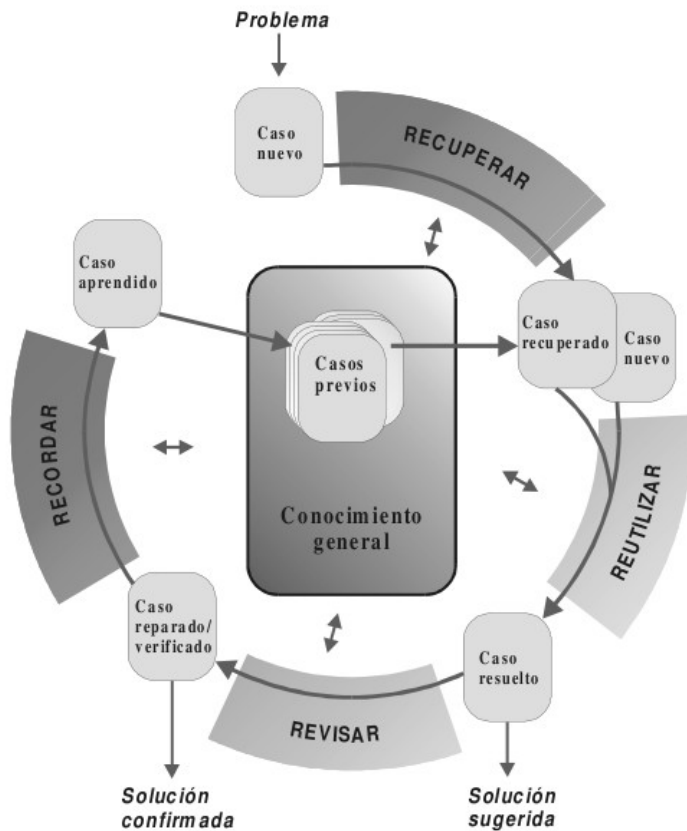


Figura 10: El ciclo de razonamiento basado en casos (Aamodt & Plaza, 1994).

El objetivo de la *recuperación* de casos es encontrar un caso o un pequeño conjunto de casos en la librería de caso que contengan una solución útil para el problema o la situación actual.

Una vez que un caso adecuado es recuperado de la base de casos, la solución sugerida por este caso es objeto de una tentativa de *reutilización* para la solución del problema actual. Durante este paso, se da una reutilización del conocimiento (la descripción de la solución) del caso anterior, conocido, para el caso actual, sin embargo este todavía no ha sido solucionado (Althoff et al., 1995; Leake & Plaza, 1997).

Una vez que el caso recuperado de la base de casos puede no satisfacer completamente los requisitos dados por la nueva situación, se puede tornar necesario adaptar la solución descrita en el caso recuperado antes de la aplicación al caso actual. En la mayoría de las circunstancias y dominios de aplicación de RBC, no entonces, generalmente será suficiente que se copie la solución del caso encontrado para el caso actual y se aplique esta solución, o en tanto que se adapte el caso manualmente (Pal & Shiu, 2004).

Cuando una solución para un caso generado en la fase de reutilización no es correcta, surge una oportunidad para el aprendizaje a partir de esta falla. Esto ocurre en la fase del RBC llamada *revisión*.

1. **Recuperar** el caso o casos pasados más similares a la nueva situación presentada.
2. **Reutilizar** la información y conocimiento de este caso o casos recuperados para resolver el nuevo problema.
3. **Revisar** la solución propuesta.
4. **Recordar** o almacenar la nueva solución una vez ha sido confirmada o validada de una manera que pueda ser útil para resolver problemas futuros.

El objetivo de la *recuperación* de casos es encontrar un caso o un pequeño conjunto de casos en la librería de caso que

La revisión consiste de dos tareas (Aamodt & Plaza, 1994):

- Avalar con un criterio cuidadoso la solución generada por la reutilización. Si fuera considerada como correcta, aprenda con el éxito y continúe con la retención del nuevo caso en la base de casos.
- En caso contrario, reparar la solución para el caso, utilizando conocimiento específico sobre el dominio de aplicación o información suministrada por el usuario.

Recordar, almacenar o retener los casos es el proceso de incorporación, al conocimiento ya existente, de aquello que es útil de un nuevo episodio de solución de un problema. El objetivo de que se retenga continuamente el conocimiento toda vez que un nuevo problema es resuelto, es el de constantemente actualizar y extender la base de casos. Esto permite a un sistema de RBC continuamente incrementar su conocimiento y convertirse en un solucionador de problemas más poderoso (Bergmann & Stahl, 1998; Leake & Plaza, 1997; Schank, 1999).

1.7 Conclusiones.

En cuanto a los modelos de gestión de riesgos estudiados:

- En el modelo de Boehm las técnicas y herramientas se describen abstractamente y carecen de claridad para ser usadas, además que este modelo requiere mucha experiencia y habilidad para la evaluación de los riesgos.
- En Continuous Risk Management la técnica de priorización propuesta carece de argumentos objetivos pues no puede juzgar cuáles riesgos serán más importantes simplemente por un criterio arbitrario.
- En Software Risk Evaluation los riesgos son priorizados de acuerdo a su exposición y al criterio de los entrevistados, en este sentido no se definen técnicas bien justificadas para la priorización. En el proceso de consolidación se pudieran utilizar técnicas más precisas por ejemplo una suma ponderada donde se exprese la importancia de un objetivo sobre otro.
- Las técnicas descritas por PMBOK, solo se describen con un nivel aceptable de detalle la referente a la matriz de probabilidad e impacto en el análisis cualitativo y el árbol de decisiones en el análisis cuantitativo. En esta última el objeto de análisis está enfocado en los costes, sin ejemplificar siquiera el tiempo, el alcance o los recursos humanos.

- El modelo propuesto por Microsoft Solution Framework incluye como un proceso de la gestión de riesgos que realiza una etapa de aprendizaje, en la cual sugiere se cuente con una base de datos de riesgos donde al concluir cada proyecto se actualice con los nuevos riesgos identificados. Aunque no lo describen en su representación del riesgo, MSF define un estado que pueden tener aquellos riesgos que se decida que su gestión no merita la pena.
- La documentación del modelo Magerit describe en gran medida un conjunto de técnicas y herramientas a utilizar en el análisis de los riesgos, enfocadas al análisis estadístico.
- Por su parte el MOGERI propone en el proceso de análisis la documentación de las experiencias, pero no describe como llevarlo a cabo.

En los modelos estudiados, los riesgos se caracterizan por un conjunto de atributos comunes, como son probabilidad, impacto, tiempo de mitigación, categorización, estrategias de respuestas, prioridad y exposición. En este sentido, MOGERI define un vector de representación de los riesgos con un conjunto de atributos, pero incluye dos campos (Indicadores y Modos de evaluación) en los cuales no se define el tipo de información a tratar.

De aquí se puede concluir que en el proceso de gestión de riesgos, la etapa de análisis es la que, en nuestro entorno, está menos desarrollada, en gran medida por lo complejo que resulta aplicar los métodos propuestos en la literatura, y en este sentido es necesario proponer nuevas técnicas y herramientas mucho más eficaces.

De lo anterior se puede afirmar que en ningún modelo o herramienta/técnica estudiada existen o se proponen métodos que tribute al proceso de análisis en la gestión de riesgo que posibiliten el empleo de la experiencia almacenada para brindar formas de mitigación a los riesgos en los proyectos de software. Por este motivo parece aconsejable utilizar el razonamiento basado en casos para acumular la experiencia en la identificación y tratamiento de los riesgos de software y emplearla ante la aparición de nuevos riesgos.

Capítulo 2. Modelo de RBC para el análisis en la Gestión de Riesgos.

2.1 Introducción.

La propuesta de este trabajo es justamente la definición de un modelo para el análisis en la GR que incluye mecanismos para la acumulación de experiencia utilizando la técnica de RBC. Las razones por las que se ha decidido emplear esta técnica de inteligencia artificial son las siguientes:

- A los roles del proyecto involucrados en la GR les resulta muy difícil establecer cadenas de reglas generalizables que permitan inferir las características de los riesgos en un nuevo proyecto, y les es más fácil expresarlo en términos de casos ya ocurridos.
- La modelación del riesgo está matizada por características del medio que pueden generar diferentes soluciones con solo variar uno de estos atributos, como por ejemplo la prioridad y la probabilidad asignadas por el equipo de proyecto. Estos atributos son transferidos a los riesgos que pueden generarse durante el proceso de desarrollo, pues lo que constituye un riesgo priorizado en un proyecto pudiera no serlo en otro con solo cambiar algunos de sus atributos.
- Existe una alta frecuencia de repetición de riesgos similares en proyectos cuyas características son similares. Los riesgos que se generan durante el proceso de desarrollo en general pueden ser empleados para prevenir su ocurrencia en proyectos con características similares a aquellos en los que fueron generados.

En este capítulo se propone un modelo de razonamiento basado en casos para emplearlo en la etapa de análisis en la GR, de forma que se pueda aprovechar la experiencia acumulada en el análisis de riesgos en el desarrollo de los proyectos de software.

El presente apartado está compuesto por una sección de alcance, principios y premisas donde se exponen las características del modelo propuesto, los principios por los que deben guiarse los proyectos y equipos de desarrollo en torno a los roles involucrados, enfoques a tomar, y las premisas que se deben llevar a cabo para lograr una mejor aplicación del modelo. Posteriormente se caracteriza el modelo propuesto, describiendo detalladamente los procesos que lo componen.

2.2 Alcance, principios y premisas.

Este modelo es aplicable a proyectos de producción de software que deseen implementar una gestión de riesgos periódica con el objetivo de realizar un análisis de los riesgos ya identificados previamente, para determinar los riesgos más similares y con ello sus respuestas o planes de mitigación, que permitirá disminuir las amenazas y aumentar las oportunidades.

El modelo ofrece descripción y guía objetiva para la realización de los procesos involucrados en el análisis de los riesgos, especificando sus entradas, salidas, participantes, descripción del flujo de trabajo y las actividades que se realizan durante su ejecución. Define métodos a aplicar para dicha realización, así como la implementación de los algoritmos que se propone usar con este propósito.

Los principios aquí expuestos tributan a un adecuado manejo de la GR en proyectos de software y por ende del proceso de análisis de los mismos. Estos son:

- personal involucrado: para que el proceso sea efectivo es necesario que el personal involucrado en este proceso se sienta comprometido con la validez de los resultados; se propone que los roles involucrados sean:
 - el líder del proyecto: es el que mejor debe conocer cómo se llevará a cabo la gestión de riesgos en el proyecto, pues quien debió elaborar el plan de gestión de riesgos; además es el responsable de cumplimentar los cronogramas y compromisos productivos, y de supervisar y guiar el proceso de desarrollo de software; le corresponde organizar y controlar el trabajo de los miembros, y participar en la toma de decisiones estratégicas; debe mantener el control y seguimiento periódico de los indicadores de su proyecto, y determinar medidas correctivas cuando sea necesario; además debe ser el que gestione los recursos materiales necesarios para el proyecto y para el equipo de desarrollo;
 - el analista de sistemas: es el que debe dominar el negocio y es el velador del cumplimiento de los requisitos de los clientes, a pesar de los riesgos que aparezcan en el proyecto; debe participar en la definición del proyecto, y ser el responsable de traducir la comunicación entre clientes y desarrolladores;
 - el arquitecto: es el que mejor debe conocer el desarrollo y qué consecuencias en él podría traer determinado riesgo; debe ser el líder técnico y el que posea la

autoridad para tomar decisiones de este tipo; constantemente comunica el valor de lo que se está haciendo y entiende el proceso de desarrollo; le corresponde ser el responsable de que las políticas y estándares, definidos en la arquitectura, se cumplan; ayuda al líder del proyecto a estimar los costos, y ayuda a ubicar al personal según sus posibilidades;

- el planificador: debe participar en el proceso de análisis porque es el responsable de planificar y controlar las tareas que se asignan, los planes de iteraciones, y de actualizar el cronograma, plan de proyecto y fechas de entregas, es el que debe llevar el control de la ejecución del proyecto, medir la eficiencia del desarrollo, y controlar imprevistos y contratiempos;
 - el arquitecto de datos: es el responsable del diseño del almacén de datos persistentes utilizado por el sistema, realiza las pruebas de concepto de la base de datos, que garanticen la integridad de los datos, el rendimiento y otros principios.
 - gestor de configuración: rol encargado de administrar la configuración del proceso de desarrollo garantizando el acceso de todos sus miembros a los artefactos del proyecto.
- enfoque de mejora continua: el principal elemento que contribuye a la mejora es el aprendizaje; el modelo asume un aprendizaje sistemático, pues cada cierto tiempo establecido por el equipo, se realiza una retroalimentación de los posibles riesgos a los que se ve expuesto, actualizando de esta forma los atributos de los riesgos detectados, permitiendo un mejor proceso de análisis; y
 - centrado en los clientes: el modelo se centra en obtener un beneficio visible para los clientes logrando así una mayor involucración de los mismos en este proceso.

Las siguientes premisas son fundamentales para un futura aplicación del modelo:

- personal capacitado en el proceso de análisis en gestión de riesgos: es necesario que se asigne un equipo capacitado en el tema de la gestión de riesgos en sentido general, independientemente de que cubran alguna otra responsabilidad en el proyecto, dicho equipo es el encargado de todo el flujo, son los que identifican, y priorizan los riesgos, entradas fundamentales para el proceso de análisis, además de que tendrían una visión general de los mismos;
- plan de gestión de riesgos: denota la involucración activa de la dirección del proyecto

con la gestión de riesgos, así como las tareas generadas durante el proceso de análisis en los planes y mecanismos de planificación y control del proyecto, y asegura una vinculación con el resto de los procesos organizativos, de control y de desarrollo del proyecto.

2.3 Características del modelo.

El modelo propuesto posee una serie de actividades que se deben realizar para obtener los riesgos similares y posteriormente utilizar su información para adaptar sus características a los nuevos riesgos.

Hacer cualquier inferencia basada en casos incluye los pasos siguientes:

1. Presentar como entrada al sistema una descripción del problema a resolver o problema actual.
2. Encontrar aquellos casos en memoria que resuelven problemas similares al actual.
3. Determinar cuales partes del viejo caso se deben focalizar para resolver el subproblema actual.
4. Derivar una solución adaptando la solución previa a las restricciones del nuevo problema.
5. Chequear la consistencia del valor derivado con la descripción del problema a resolver, y aceptar o rechazar este.
6. Si se considera apropiada la solución validada se añade a la base de casos para ser usada en el futuro si se considera conveniente.

Basado en la estructura básica de los sistemas basados en conocimiento (SBC) en la cual se cumple que estos están compuestos por la base de conocimientos (BC), la máquina de inferencia (MI), la interfaz de usuario o interfaz de explicación y el módulo de aprendizaje (MA) ($SBC=BC+MI+IE+MA$), y de los procesos del RBC en particular, las actividades y artefactos propuestos son los siguientes. A continuación se describen los procesos fundamentales del modelo:

- Construcción de la base de conocimientos (Figura 11): consta de dos subprocesos, la construcción de la base de conocimiento de riesgos y la selección de atributos, a continuación de este comienza el proceso de inferencia:

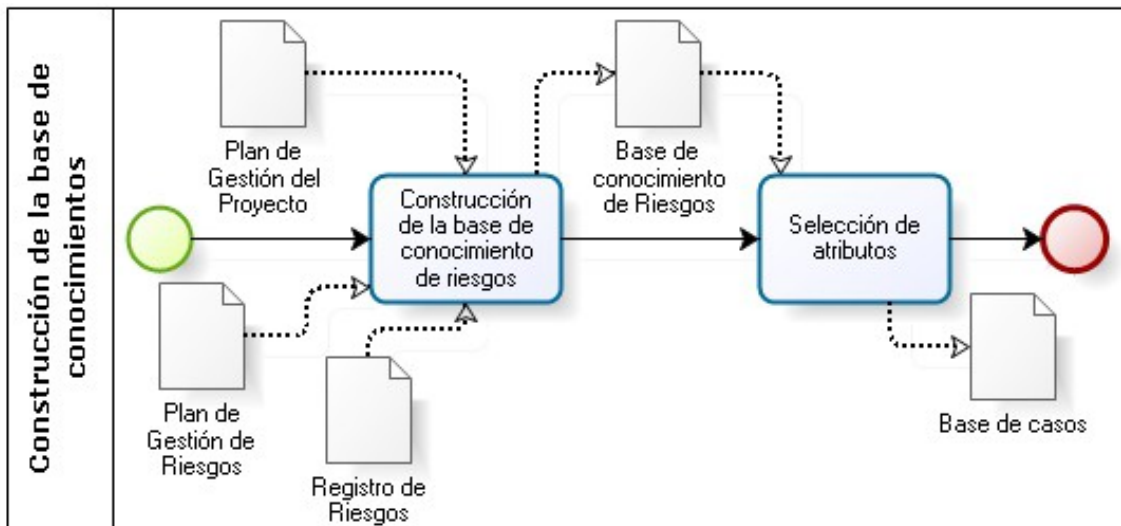


Figura 11: Primer proceso: Construcción de la base de conocimientos.

- Construcción de la base de conocimiento de riesgos: para la correcta aplicación del modelo, basado en la técnica de inteligencia artificial prevista, es necesario contar con una base de conocimiento de riesgos (BCR), obtenidos principalmente, resultado de la experiencia de proyectos anteriores, los cuales a su vez contarán con un grupo de indicadores a través de los cuales serán evaluados.
 - Selección de atributos: se lleva a cabo la selección de los atributos tanto de entrada como de salida que participarán en el cálculo de la similitud y la definición de sus pesos, representando los niveles de importancia de cada uno ante la similitud calculada.
- Máquina de inferencia (Figura 12): compuesto de tres procesos, la recuperación de riesgos semejantes, la explicación de la inferencia y la adaptación de los riesgos, para posteriormente darle paso al proceso de aprendizaje:
 - Recuperación de riesgos semejantes: en este proceso fundamental se definen los riesgos similares (RS) sobre los cuales se procederá según los tipos de atributos que componen la descripción de los casos.
 - Adaptación de los riesgos: una vez determinados los RS se procede reutilizar sus atributos, en este casos las respuestas u otro atributo de interés.
 - Explicación de la inferencia: sirve para explicar al usuario el conocimiento aplicado en la resolución del problema.

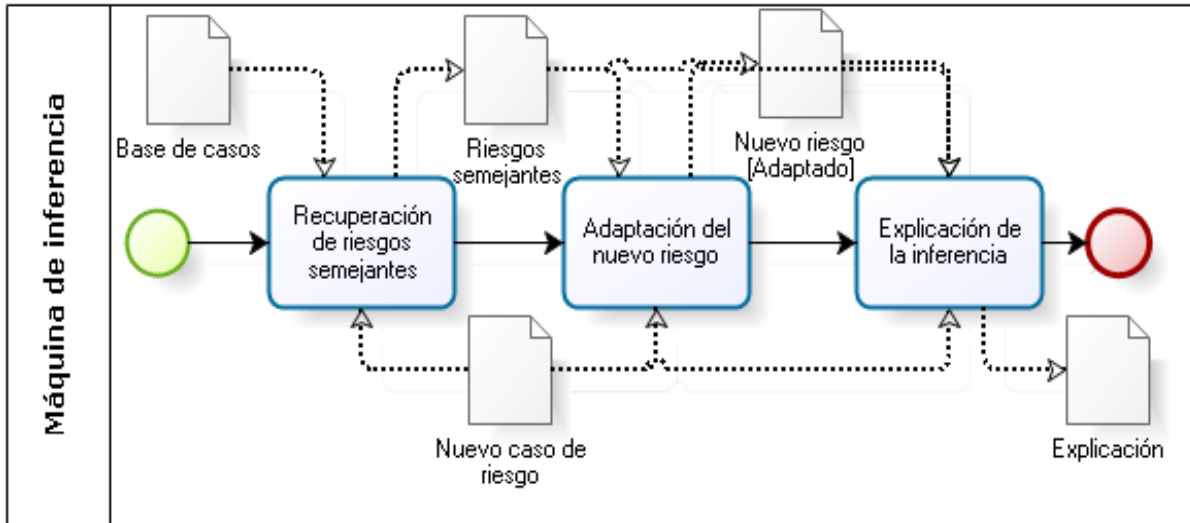


Figura 12: Segundo proceso: Máquina de inferencia.

- Módulo de aprendizaje (Figura 13): último proceso propuesto, consta de dos subprocesos, el refinamiento y ajuste de la solución y la actualización de la base de conocimiento:

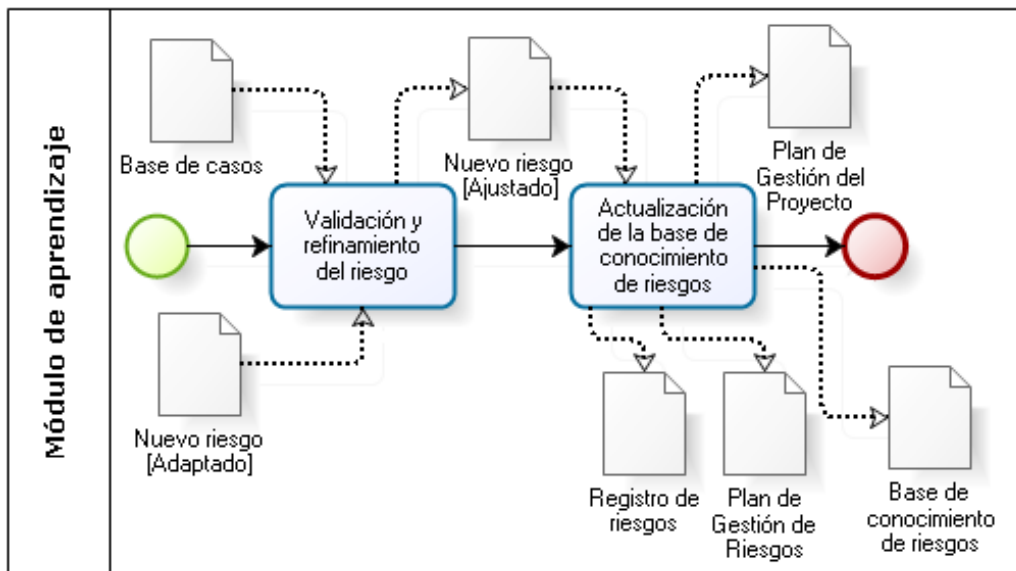


Figura 13: Tercer proceso: Módulo de aprendizaje.

- Validación y refinamiento del riesgo: en la validación se somete el riesgo adaptado a una evaluación del equipo del proyecto, el refinamiento se puede ver como una reimplementación del paso anterior, se lleva a cabo cuando los atributos de los RS son aplicados a los nuevos riesgos, pero estos no tienen resultados satisfactorios y se hace necesario reparar la solución, modificando y evaluando nuevas características.
- Actualización de la base de conocimiento: se registran los nuevos riesgos con sus

nuevos atributos adaptados y modificados, además de las soluciones que no resultaron satisfactorias puesto que de estas se puede aprender también.

A continuación se procede a explicar en detalle cada uno de los procesos propuestos.

2.3.1 Construcción de la base de conocimiento.

Una base de conocimiento es un tipo especial de base de datos para la gestión del conocimiento. provee los medios para la recolección, organización y recuperación computarizada de conocimiento.

2.3.1.1 Construcción de la base de conocimiento de riesgos.

La primera de las actividades del modelo es la construcción de la BCR, para ello se necesitan consultar una serie de artefactos de entrada. En un entorno de desarrollo de aplicaciones existen otras fuentes de información para realizar esta actividad, como son registros de proyectos anteriores, métricas y mediciones realizadas a los proyectos en desarrollo o en cualquiera de sus etapas. Los involucrados en el análisis utilizan, para realizar las actividades, los artefactos mínimos siguientes:

- el plan de gestión del proyecto,
- el plan de gestión de riesgos, y
- el registro de riesgos.

El plan de gestión del proyecto (PGP) facilita información acerca del plan de gestión del cronograma, plan de gestión de costes, entre otros. El plan de gestión de riesgos (PGR) permite conocer los roles, responsables, planificación de las actividades a realizar, presupuesto, técnicas a usar, entre otros datos, que permiten conocer y ubicar a los involucrados en el contexto del proyecto, y poder llenar algunos valores de la base de riesgos. El registro de riesgos (RR) es el artefacto que mayor información provee para la base de riesgos, pues contiene, de cada riesgo un conjunto de atributos que lo caracterizan: identificador, etapa, estado, prioridad, categoría, entrevistado, evento, consecuencia, causa, posible probabilidad, posible impacto, marco de tiempo, estrategia de respuesta, responsables, riesgos relacionados, observaciones, entre otros.

Posteriormente, con el desarrollo de las actividades de gestión de riesgos, algunos de estos artefactos pueden ser actualizados, tal es el caso del RR cuya información puede resultar modificada, en especial los valores de algunos atributos del riesgo.

La base de riesgos (BR) elaborada debe contener información relativa a los riesgos gestionados que permita contar con la experiencia acumulada en los procesos definidos en el PGR proceso para que pueda ser utilizada en el desarrollo y revisión de nuevos riesgos.

Para la representación de los riesgos se usó una propuesta entendible, usable, práctica y que contribuye a conocer más datos sobre el riesgo (Brito-González & Bon-Pérez, 2009; García-Hernández, 2008). Esta es una representación vectorial del riesgo la cual es fácil de representar computacionalmente, apoya al análisis y a las búsquedas. Sus parámetros pueden sufrir variaciones de acuerdo al proyecto en cuestión.

Identificador	Etapas	Estado	Prioridad	Categoría	Entrevistado	□ □ □
---------------	--------	--------	-----------	-----------	--------------	-------

Figura 14: Representación vectorial del riesgo. Tomado de (García-Hernández, 2008).

Las propiedades del vector se describen a continuación, proponiendo dividir las en dos grupos como datos de entrada y datos de salida para su posterior procesamiento:

- Datos de entrada:
 - **Identificador:** identificador único del riesgo, debe ser único. Puede ser numérico consecutivo o bien puede ser una combinación o concatenación de otros identificadores, como por ejemplo de línea de producción, factoría, proyecto, entre otros.
 - **Etapas:** etapa del proyecto donde se manifiesta el riesgo, refiere un tipo de dato ordinal (Inicio, Elaboración, Construcción, Cierre, Soporte).
 - **Estado:** uno de estos valores ordinales: identificado, analizado, activo, cerrado, reabierto.
 - **Proyecto:** proyecto en el cual se detectó. Este atributo bien puede ser meramente textual, aunque de existir una nomenclatura de proyectos pasa a ser nominal.
 - **Entrevistado:** atributo textual que refiere la persona que identificó el riesgo.
 - **Evento:** descripción textual del suceso incierto que causa preocupación. No debe ser ambiguo, debe ser corto pero no debe limitarse su descripción por acortarlo.
 - **Causa:** descripción textual de la causa que origina el riesgo. Es importante que sea precisa, descriptiva y objetiva.
 - **Consecuencia:** descripción textual de las consecuencias que trae la ocurrencia del evento. Debe expresar claramente el impacto sobre los objetivos del proyecto, no

tiene por qué ser sobre todos los objetivos. (Nota: estas dos propiedades: evento y consecuencia, al leerlas debe tener sentido, el riesgo será enunciado utilizando estas palabras)

- *Palabras claves*: una lista de las palabras claves que componen la descripción del evento.
- *Personas involucradas*: personas relacionadas al desarrollo de los riesgos.
- *Riesgos padres*: una lista de los riesgos de los cuales se puede derivar el riesgo actual.
- *Riesgos Relacionados*: una lista de otros riesgos relacionados cuya causa o evento esté muy influenciada por este riesgo.
- *Certidumbre de la Información*: valor que indica cuán seguro se está de la información obtenida sobre el riesgo.
- *Tiempo de exposición*: valor numérico que indica que tiempo se espera que demore ocurriendo.
- *Observaciones*: notas importantes sobre el riesgo, su historial.
- Datos de salida:
 - *Prioridad*: orden de prioridad para atender este riesgo. Es un valor numérico diferente para cada riesgo indicando su prioridad, 0 muy priorizado, > 0 menos priorizado. También se puede considerar un conjunto de datos ordinal, donde los rangos de prioridad se den en particiones separables.
 - *Categoría*: identificador nominal de la categoría del riesgo (una de las definidas por el proyecto o línea de desarrollo).
 - *Probabilidad*: valor numérico indicando la probabilidad de ocurrencia.
 - *Impacto*: valor de impacto para cada objetivo de proyecto.
 - *Frecuencia*: el factor de aparición del riesgo en el tiempo de proyecto. Expresa un valor porcentual respecto al tiempo total del proyecto indicando con qué frecuencia aparece este riesgo.
 - *Marco de tiempo*: valor numérico que indica el tiempo mínimo en el que debe ejecutarse los planes de contingencia o explotación para este riesgo, debe expresarse en la unidad de medida acordada en el PGR.

- *Tiempo de ocurrencia*: valor numérico que indica el tiempo que resta para que ocurra el riesgo.
- *Complejidad de resolución*: valor numérico u ordinal que indica cuán complejo resulta resolver el riesgo.
- *Respuesta*: descripción textual de la respuesta utilizada para mitigar el riesgo.
- *Éxito*: valor binario que denota si la respuesta aplicada tuvo éxito o no.

Este tipo de representación de la información de riesgos coincide a plenitud con la propuesta de representación plana de la información en los casos (Althoff, Auriol, Barletta, & Manago, 1995; Arian-Rodríguez, 2003), por lo que, para su posterior tratamiento, se emplearán medidas de similitud que garanticen una correcta recuperación de los casos más semejantes, descritas más adelante.

A continuación se expone una propuesta de diseño de la BR que contiene la experiencia de los riesgos y que permite gestionar la información que se deriva de ellos.

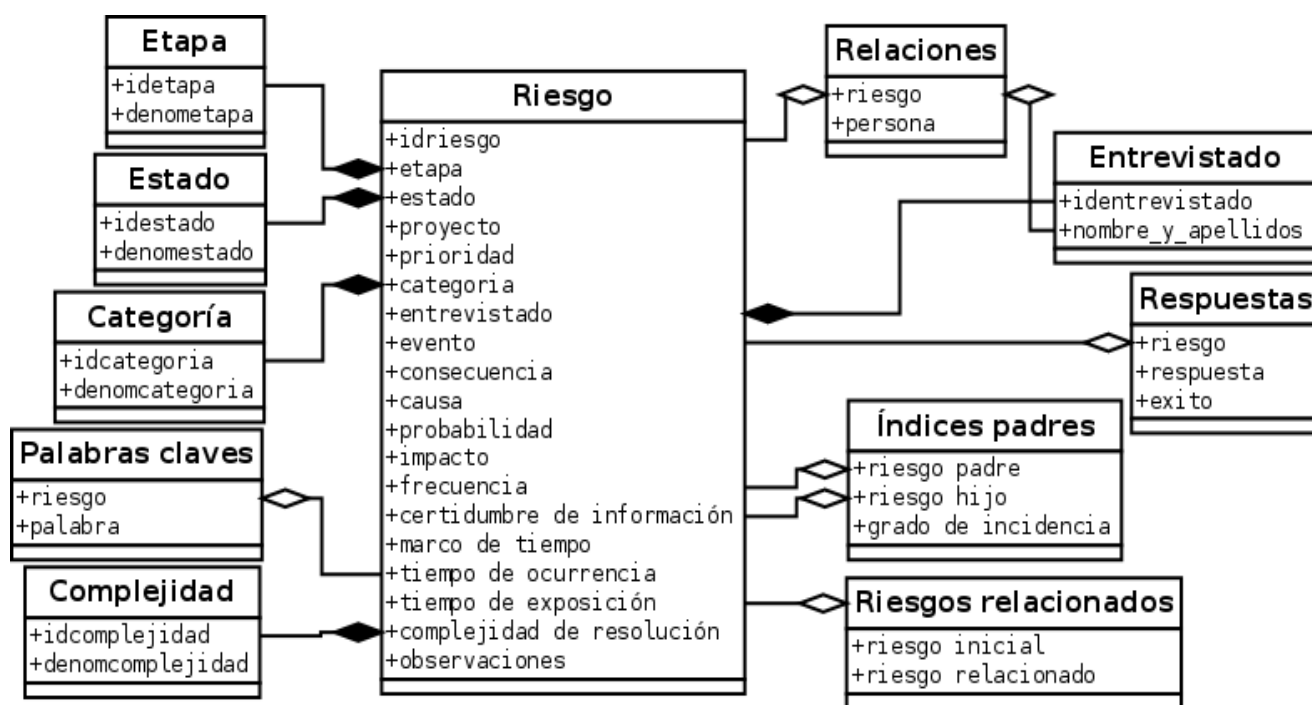


Figura 15: Modelo conceptual para el análisis de los riesgos.

A partir de esta representación se puede obtener los valores señalados en el vector de representación del riesgo.

2.3.1.2 Selección de los atributos.

Identificar un problema puede involucrar simplemente darse cuenta de sus atributos de

entrada, pero a menudo, y particularmente en métodos de gestión de conocimiento, se toma un enfoque más elaborado, en el que se hace un intento de comprender el problema en su contexto. Entender el problema involucra filtrar o desechar los atributos que pueden causar ruido al sistema, para poder inferir otras características importantes del problema y para chequear cuando los valores de los atributos toman sentido dentro del contexto del problema, entre otras. Otros descriptores que no sean los que se ofrecen de entrada, puede deducirse mediante un modelo de conocimiento general, o mediante la recuperación de un problema de descripción similar de la base de casos.

En este sentido, seleccionar los atributos adecuados para efectuar la recuperación de los riesgos más semejantes, puede depender de varios factores como son el tipo de proyecto, la prioridad, los objetivos del mismo, la situación geográfica, los recursos materiales y humanos a su disposición, entre otros.

Otra actividad a tener en cuenta es definir los pesos de significación o importancia de los atributos de entrada. Esto incide directa y sensiblemente en la función de similitud, por lo que se debe prestar mucha atención al valor asignado a cada uno de ellos.

En el enfoque propuesto, el usuario jugará un importante papel en este sentido, ya que a partir del vector de riesgos propuesto (Figura 14) se deberán escoger las características que servirán como base para la posterior etapa de recuperación de casos de riesgos semejantes. En este sentido es importante que los roles implicados en este proceso (ver apartado 2.2 Alcance, principios y premisas.) participen activamente en la definición de dichas características, principalmente el gestor de configuración y el arquitecto de datos. Al mismo tiempo se definirán los pesos de cada atributo del vector de representación del riesgo y se determinarán los atributos que posteriormente serán presentados en forma de explicación de la similitud calculada. De este proceso queda elaborada la base de casos (BC) con los atributos que servirán posteriormente para el cálculo de la similitud.

2.3.2 Máquina de inferencia.

Es un módulo que se encarga de las operaciones de búsqueda y selección de la información a utilizar en el proceso de razonamiento.

En esta propuesta involucra tres procesos fundamentales: recuperación de riesgos semejantes, la explicación de la inferencia realizada y la adaptación de los casos de riesgo.

2.3.2.1 Recuperación de riesgos semejantes.

Cualquier método de recuperación combina algún procedimiento para valorar el grado de similitud entre dos casos y algún procedimiento para buscar el caso más similar a la consulta. Respecto al procedimiento de búsqueda la investigación se refiere a limitar o acotar la búsqueda para poder encontrar el mejor caso de forma más eficiente sin reducir la calidad del resultado. Respecto a la valoración de la similitud entre los casos existen dos aproximaciones básicas y numerosas variantes suyas:

- la aproximación representacional: donde los casos residen en estructuras de datos complejas, y la similitud se mide sobre la proximidad a dicha estructura, y
- la aproximación computacional: la cual computa los valores de similitud durante la ejecución del proceso de recuperación.

Esta última aproximación, simple y muy utilizada, consiste en llevar a cabo una búsqueda exhaustiva en la base de datos comparando cada caso con la consulta, pudiendo resultar de una eficiencia intolerable sobre una recuperación lineal con un gran número de casos. Por este motivo, la comunidad RBC propone el uso de arquitecturas paralelas que se apoyan en un aumento de recursos de hardware.

Tomando como referencia el vector de riesgo propuesto en el presente trabajo en forma de memoria plana, se tiene entonces que la similitud respecto a otros riesgos de la base de casos se computará en forma de una aproximación computacional, calculando la similitud de cada atributo para llegar a una medida de similitud global, también llamado método del vecino más cercano (Liu, Moore, Gray, & K. Yang, 1998; Lozano & Fernández, 2004; Zhang & Srihari, 2002), k-vecinos cercanos o knn (del inglés *K nearest neighbors*). En este sentido es necesario emplear funciones de similitud cuyo resultado se exprese en el mismo rango para todos los tipos de atributos que componen el vector de riesgo; en este caso se propone que dicho rango se encuentre entre cero y uno.

2.3.2.1.1 Funciones de semejanza entre atributos.

La selección de un caso semejante a un nuevo problema entre un grupo de casos almacenados, exige el uso de alguna herramienta que indique cuan semejantes son los casos analizados. Una medida que determine el grado de proximidad entre los casos, se conoce como "Medida de semejanza". En términos matemáticos esto significa definir un espacio métrico entre los casos x e y . Se plantea que debe cumplirse que exista una función $d(x, y)$ real no negativa y un conjunto U , tal que se cumplan las siguientes

condiciones (Moreno, 1998; Michavila, 1986):

Siendo $d(x, y), U \in \mathbb{R}^+$, $\forall x, y \in U$ se cumple:

$$d(x, y) = 0 \Leftrightarrow x = y \quad (\text{Axioma de la separación})$$

$$d(x, y) = d(y, x) \quad (\text{Axioma de la simetría})$$

$$d(x, y) \geq 0 \Leftrightarrow x \neq y$$

$$d(x, y) \leq d(x, z) + d(z, y) \quad \text{para } z \in U \quad (\text{Desigualdad triangular})$$

Si $d(x, y)$ es una función de distancia, cuantifica el valor de semejanza de forma inversa, es decir, mientras menor sea la distancia, mayor será la semejanza entre los casos (Kolodner & Leake, 1996).

a) Función de semejanza para atributos numéricos.

Distancia de Manhattan

La distancia absoluta, (llamada de Manhattan o de City Block) que se representa como la diferencia absoluta sobre todas las dimensiones. En la mayoría de las aplicaciones, esta genera resultados similares a la distancia euclidiana, pero con más bajo nivel computacional. Está dada por la siguiente expresión:

$$d(x, y) = |x - y| \quad [2.3.2.1.1-1]$$

Esta función retorna valores en dependencia del valor de sus parámetros, por lo que para lograr que su resultado se ajuste al rango propuesto se debe ajustar de la siguiente forma:

$$d(x, y) = \left| \frac{x - y}{r_{\text{máx}} + r_{\text{mín}}} \right| \quad [2.3.2.1.1-2]$$

Donde $r_{\text{máx}}$ y $r_{\text{mín}}$ son los valores de rango máximo y mínimo respectivamente del conjunto de valores existentes.

b) Funciones de semejanza para atributos nominales y textuales.

Cuando los atributos son simbólicos o nominales, la forma tradicional de adaptar el método del vecino más cercano es recurrir a una métrica general, la cual cuenta el número de atributos que tienen valores diferentes para los casos. Una de estas distancias es la distancia de Hamming.

Para el cálculo de la similitud de los atributos en forma de descripciones textuales, se realiza una normalización del texto, agregando al modelo de datos una tabla en la cual se almacenen un conjunto de palabras que no reporten interés a la descripción (stop words),

sobre las cuales no se procede a realizar cálculo alguno. Según algunos estudios sobre el tema (Beall, 2008; Feldman & Sanger, 2006; Köhler, Philippi, Specht, & Rüegg, 2006; Kuhn, Ducasse, & Gîrba, 2007; Lefever, Fayruzov, Hoste, & De Cock, 2010; Y. Yang & Pedersen, 1997), se propone que estas sean fundamentalmente artículos gramaticales (determinados: el, la, los, las; indeterminados: un, una, unos, unas; neutro: lo; y contracto: al, del), proposiciones, signos de puntuación, entre otros. Se procede a realizar un filtro en el cual quedan eliminados del atributo textual los elementos (tokens) anteriores, quedando una descripción textual más depurada para el cálculo de similitud. Con esta representación se elaboran vectores de representación del texto y se lleva cabo el cálculo de la medida de distancia entre vectores (Han & Kamber, 2006).

Otros estudios relacionados a este tema (Gelbukh, 2008; Hand, Keim, & Ng, 2002; Han & Kamber, 2006) señalan el empleo de una técnica llamada *stemming*, la cual es un método para reducir una palabra a su raíz basándose en remover sufijos o prefijos de una palabra. El algoritmo más común para stemming es el algoritmo de Porter (Kowalski & Maybury, 2000). Existen además métodos basados en análisis lexicográfico y otros algoritmos similares (KSTEM, stemming con cuerpo, métodos lingüísticos, entre otros) (Antonín, Montraveta, & García, 2003).

Distancia de Hamming

Para obtener la semejanza entre los casos, se cuenta el número de atributos diferentes. Por ejemplo sean 2 vectores de 4 componentes:

$$c1 = (a, b, a, c) \text{ y } c2 = (a, d, a, c)$$

La distancia entre $c1$ y $c2$ es 1.

Esta distancia también puede ser utilizada cuando todos los atributos son booleanos. La cantidad de atributos que tienen en común los ejemplos, indican los valores de semejanza. Los casos con mayor cantidad de atributos iguales, serán los más semejantes. Con esta idea esta distancia es utilizada en otras aplicaciones como puede ser el análisis de los códigos de errores. Para medir

Medida del coseno

La similitud del coseno es una medida de similitud basada en vectores. La cadena de entrada es transformada en un espacio vectorial en la cual la regla del coseno euclídeo se puede utilizar para determinar la similitud. La similitud del coseno a menudo se emplea con otros

enfoques para limitar la dimensionalidad del problema. Por ejemplo, con simples cadenas de texto, la lista de palabras vacías o *stop words* se utilizan para excluir esta dimensionalidad de la comparación. En teoría, este problema tiene tantas dimensiones como términos de existir (Bilenko & Mooney, 2003; Cohen, Ravikumar, & Fienberg, 2003).

$$d(x, y) = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}} \quad [2.3.2.1.1-3]$$

Donde i representa cada dimensión del vector de la palabra elaborado y n la cantidad de términos o dimensiones en el vector.

Value Difference Metric (VDM)

En esta métrica (Wilson & Martínez, 1997) la similitud entre dos valores $x_i(O_1), x_i(O_2)$ del atributo i con respecto a los objetos O_1, O_2 viene dada por la siguiente expresión:

$$vdm_i(x_i(O_1), x_i(O_2)) = \sum_{c=1}^M \left(\frac{N_{i,x_i(O_1),c}}{N_{i,x_i(O_1)}} - \frac{N_{i,x_i(O_2),c}}{N_{i,x_i(O_2)}} \right) \quad [2.3.2.1.1-4]$$

Donde $N_{i,x_i(O_1)}$ es el número de veces (en el conjunto de entrenamiento) que el atributo i tiene el valor $x_i(O_1)$, $N_{i,x_i(O_1),c}$ es el número de veces que el atributo i tiene el valor $x_i(O_1)$ en la clase c , y M es el número total de clases.

Distancia de Jaccard.

Esta es otra medida de similitud empleada en espacios vectoriales, similar a la distancia coseno. La similitud Jaccard compara dos conjunto de tokens (palabras), en cuanto a la cantidad de tokens que está presente en cada uno por sobre la cantidad total de ellos en ambos conjuntos.

$$D(X, Y) = \frac{(|X \cap Y|)}{(|X \cup Y|)} \quad [2.3.2.1.1-5]$$

Donde X e Y representan los conjuntos de palabras antes mencionados.

c) Funciones de semejanza para dominios de atributos heterogéneos e información incompleta.

Se considera como información incompleta, los valores desconocidos en un momento dado para un atributo.

Heterogeneous Euclidean Overlap Metric (HEOM)

Una de las funciones desarrolladas para analizar los problemas con atributos numéricos y simbólicos con o sin información incompleta es la Métrica Heterogénea Euclidiana Generalizada (conocida por sus siglas en inglés HEOM)(Wilson & Martínez, 1997).

En este caso la distancia entre “x” e “y” para un valor de atributo “i” se define como:

$$d_i(x_i, y_i) = \left\{ \begin{array}{l} 1 \text{ si } x_i \text{ o } y_i \text{ son desconocidos} \\ \text{overlap}(x_i, y_i) \text{ si } i \text{ es un atributo nominal} \\ \text{valor del rango normalizado } (rn_{dif}(x_i, y_i)) \text{ en otro caso} \end{array} \right\} \quad [2.3.2.1.1-6]$$

Donde los valores desconocidos del atributo, se manejan devolviendo el valor de distancia 1 (máximo valor de distancia), la función “overlap” analiza los atributos nominales (simbólicos) y se define como:

$$\text{overlap}(x_i, y_i) = \left\{ \begin{array}{l} 0 \text{ si } x_i = y_i \\ 1 \text{ en otro caso} \end{array} \right\} \quad [2.3.2.1.1-7]$$

Para los valores numéricos los valores se calcula el valor del rango normalizado, el cual está dado por la siguiente expresión:

$$rn_{dif}(x_i, y_i) = \frac{|x_i - y_i|}{rango_i} \quad [2.3.2.1.1-8]$$

donde $rango_i = max_i - min_i$ siendo max_i y min_i los valores máximos y mínimos observados en el conjunto de entrenamiento para en atributo i .

Para la comparación entre todos los atributos que conforman el caso se utiliza la siguiente expresión:

$$HEOM(x, y) = \sqrt{\sum_{i=1}^n d_i(x_i, y_i)^2} \quad [2.3.2.1.1-9]$$

Como se puede apreciar esta función utiliza diferentes funciones de distancia para diferentes tipos de atributos, si todos los atributos son simbólicos la simplificación de la función conduce a la distancia de Hamming, pues la sumatoria general incluirá solo el valor del “overlap”. Por el contrario, si todos los valores son numéricos se calcula el valor del rango normalizado.

Función de Gower

Para el caso de características no numéricas o información incompleta mezclada con características numéricas (McCane & Albert, 2008; Piccarreta, 2010), en 1971, Gower

demuestra que la siguiente expresión constituye una distancia:

$$D(x, y) = \sqrt{2(1 - S_{xy})} \quad [2.3.2.1.1-10]$$

En este caso S_{xy} , es el coeficiente de semejanza de Gower, entre los casos "x" e "y". Precisamente la importancia de este coeficiente radica en que posibilita el cálculo de la semejanza en problemas caracterizados por variables heterogéneas.

El coeficiente de semejanza está dado por la siguiente expresión:

$$S_{xy} = \frac{\sum_{i=1}^n S_{xy i}}{\sum_{i=1}^n W_{xy i}} \quad [2.3.2.1.1-11]$$

En dependencia del tipo de característica que se analice se asignan los valores de $S_{xy i}$ y $W_{xy i}$.

Para el caso de variables binarias:

$$S_{xy i} = \begin{cases} 1 & \text{si } x_i \text{ e } y_i \text{ están presentes} \\ 0 & \text{si } x_i \text{ e } y_i \text{ están ausentes o alguno es desconocido} \end{cases} \quad [2.3.2.1.1-12]$$

$$W_{xy i} = \begin{cases} 1 & \text{si } x_i \text{ e } y_i \text{ están presentes} \\ 0 & \text{si } x_i \text{ e } y_i \text{ están ausentes o alguno es desconocido} \end{cases} \quad [2.3.2.1.1-13]$$

Para el caso de variables cualitativas los valores de $S_{xy i}$ y $W_{xy i}$, están dados por:

$$S_{xy i} = \begin{cases} 1 & \text{si } x_i = y_i \\ 0 & \text{en otro caso} \end{cases} \quad [2.3.2.1.1-14]$$

$$W_{xy i} = \begin{cases} 1 & \text{si } x_i \text{ e } y_i \text{ están presentes} \\ 0 & \text{si } x_i \text{ e } y_i \text{ están ausentes o alguno es desconocido} \end{cases} \quad [2.3.2.1.1-15]$$

Para el caso de variables cuantitativas los valores de $S_{xy i}$ y $W_{xy i}$, están dados por:

$$S_{xy i} = \begin{cases} 1 - \frac{|x_i - y_i|}{\text{rango}_i} & \text{si la información está completa} \\ 0 & \text{si } x_i \text{ o } y_i \text{ son desconocidos} \end{cases} \quad [2.3.2.1.1-16]$$

$$W_{xy i} = \begin{cases} 1 & \text{si no falta algún valor} \\ 0 & \text{si } x_i \text{ o } y_i \text{ son desconocidos} \end{cases} \quad [2.3.2.1.1-17]$$

Donde $\text{rango}_i = \max_i - \min_i$

Esta función ha sido utilizada para el reconocimiento de patrones supervisado.

Función de Argelio

Para este caso la función propuesta define la distancia como (Arean-Rodríguez, 2003):

$$D(x, y) = \sum_{i=1}^n \left(1 - \frac{1}{1 + d(x_i, y_i)}\right) \quad [2.3.2.1.1-18]$$

Donde en dependencia del tipo de variables se tiene para variables cualitativas:

$$d_i(x_i, y_i) = \begin{cases} 0 & \text{si } x_i = y_i \\ 1 & \text{si } x_i \neq y_i \end{cases} \quad [2.3.2.1.1-19]$$

Y para variables numéricas:

$$d_i(x_i, y_i) = \begin{cases} \frac{|x_i - y_i|}{\text{rango}_i} & \text{si son conocidos los valores de } x_i \text{ e } y_i \\ 1 & \text{si la información es desconocida} \end{cases} \quad [2.3.2.1.1-10]$$

2.3.2.1.2 Función de semejanza entre casos propuesta.

La propuesta sobre el tipo de funciones de similitud a emplear para cada tipo de atributo queda entonces de la siguiente forma:

- atributos numéricos: Manhattan ajustada,
- atributos nominales: Hamming.
- atributos textuales: Medida del coseno, Euclidiana, Jaccard.
- atributos ausentes: en estos casos se procede de la siguiente manera: cuando faltan los dos atributos no se lleva a cabo el cálculo de la distancia, pero cuando falta un atributo la distancia se hace 1, también llamada distancia trivial (Moreno, 1998).

De esta forma queda definida la distancia entre casos de la siguiente manera:

$$f(W, D(X, Y)) = \frac{\sum_{i=1}^n (w_i \cdot D(X(x_i), Y(y_i)))}{\sum_{i=1}^n w_i} \quad [2.3.2.1.2-1]$$

Siendo w_i el nivel de importancia, significación o peso de cada atributo $X(x_i)$ en el vector de representación del riesgo X , o de los atributos previamente seleccionados, debe ser numérico.

Para el caso particular en que cada $w_i \in [0, 1]$ y se cumpla que $\sum_{i=1}^n w_i = 1$, entonces la ecuación [2.3.2.1.2-1] queda definida como:

$$f(W, D(X, Y)) = \sum_{i=1}^n (w_i \cdot D(X(x_i), Y(y_i))) \quad [2.3.2.1.2-2]$$

La distancia $D(X(x_i), Y(y_i))$ se calcula de manera heterogénea según el tipo de atributo, pudiendo simultanear con las distintas medidas de similitud propuestas.

2.3.2.1.3 Algoritmo propuesto para la recuperación de riesgos similares.

Para computar los riesgos similares se procede a efectuar el siguiente algoritmo propuesto.

Algoritmo 1. Selección de riesgos similares.

Entradas:

- $X = (x_1, x_2, \dots, x_n)$ vector que contiene los n atributos del nuevo riesgo,
- $\bar{Y} = (Y_1, Y_2, \dots, Y_m)$ lista de los riesgos de la base de casos, siendo $Y_1 = (y_{1,1}, y_{1,2}, \dots, y_{1,n})$, m la cantidad de riesgos en la base de casos, e $y_{m,n}$ el atributo n del m riesgo seleccionado,
- k cantidad de riesgos semejantes,
- $W = (w_1, w_2, \dots, w_n)$ vector con los pesos relativos de cada atributo.

Salidas:

- $\bar{S} = (S_1, S_2, \dots, S_k)$ lista de los k riesgos más similares, con $S_k = (s_{k,1}, s_{k,2}, \dots, s_{k,n})$,

Pasos:

Para cada atributo $X(x_i)$ de X y en dependencia del tipo de dato que lo caracterice,

1. Para cada atributo $Y(y_i)$ del riesgo Y_j de la base de casos,
2. Calcular la distancia $D(X(x_i), Y(y_i))$, en dependencia del tipo de atributo.
3. Con el peso w_i del atributo $X(x_i)$, calcular $w_i \cdot D(X(x_i), Y(y_i))$.
4. Repetir los pasos del 1 al 3 hasta completar los n atributos del riesgo.
5. Guardar en una lista \bar{S} el valor de la distancia entre X y cada Y_j calculado y la caracterización de Y_j .

6. Ordenar la lista de forma creciente según la distancia calculada y obtener solo los primeros k elementos que la componen.

De esta forma se obtiene una caracterización de los riesgos similares y sus distancias respecto al vector de riesgo de entrada.

2.3.2.2 Adaptación de los casos.

Después de realizar la búsqueda y selección de los riesgos similares, se debe verificar el grado de similitud que tengan los k riesgos encontrados respecto al riesgo inicial para poder reutilizar los atributos de S_i en función de X .

Una técnica de reutilización propuesta para este caso es el ajuste de parámetros, el cual se emplea para interpolar los valores de una solución previa a una nueva solución. Es decir, dada una solución previa y un caso nuevo que difiere en algún grado del caso previo, la solución previa se modifica en la medida en que difieren los dos casos. Para realizar los cambios en los parámetros de la solución previa, se debe tener en cuenta en qué grado de diferencia hay con el nuevo caso y en qué consisten estas diferencias. Otra técnica que pudiera emplearse es la definición, en el caso de atributos numéricos, de una media aritmética entre el nuevo riesgo y los riesgos semejantes analizados, en pos de lograr una mejor aproximación a un valor que permita elaborar una estrategia para lograr mitigarlo.

Para esto se procedería como indica el Algoritmo 2.

Algoritmo 2. Ajuste de parámetros.

Entradas:

- $X = (x_1, x_2, \dots, x_n)$: vector que contiene los n atributos del riesgo.
- $\bar{S} = (S_1, S_2, \dots, S_k)$: lista de los k riesgos más similares.

Salidas:

- $X' = (x'_1, x'_2, \dots, x'_n)$: vector de riesgo resultante de aplicar la reutilización a X .

Pasos:

Para cada elemento S_i :

1. Mostrar los valores de los atributos de $X(x_i)$ y de S_i con las distancias entre los atributos.

2. Ajustar manualmente los valores del atributo $X(x_i)$ respecto a los de S_i resultando $X'(x'_i)$.

Cuando se utilizan atributos de un riesgo similar encontrado en la base de conocimiento es común que, en estos casos, se puedan reutilizar respuestas similares, como un atributo priorizado, en los nuevos riesgos detectados.

2.3.2.3 Explicación de la inferencia.

A continuación se procede a elaborar la explicación en torno a la inferencia realizada de los riesgos semejantes respecto al nuevo riesgo adaptado en la base de conocimiento.

Para esto se cuenta básicamente con los atributos escogidos en el proceso “Selección de los atributos” y los pesos asociados a los atributos de entrada al cálculo de la similitud. La explicación de la inferencia, entonces, será mostrada como una lista de las características de entrada similares al nuevo riesgo refiriendo el peso y cálculo de las distancias entre atributos de cada vector de caso. De esta forma se puede apreciar la influencia de los pesos de significación sobre cada atributo en la determinación de la función de similitud global del riesgo.

2.3.3 Módulo de aprendizaje.

2.3.3.1 Validación y refinamiento de la solución.

Validación de la solución.

La tarea de evaluación toma el resultado de aplicar los cambios a los casos en un entorno real o supervisado. Este paso usualmente queda fuera del ciclo RBC, ya que involucra la aplicación de la solución propuesta o modificada al problema real. En determinados sistemas la aplicación de la solución puede tomar algún tiempo en arrojar resultados.

Como el modelo propuesto se basa fundamentalmente en un enfoque proactivo de gestión de riesgo, evidentemente no da tiempo a que sean visibles los resultados de la aplicación de la respuesta elaborada a un riesgo, por lo que se propone que la evaluación de la solución y el paso anterior, la modificación de la solución, se lleven a cabo en forma de reunión de grupo de trabajo con los roles implicados en este modelo o con el equipo de desarrollo. O sea, una vez riesgos similares sean determinados, se lleva a cabo la elaboración y/o modificación de las respuestas a los riesgos y en este preciso momento se avala por un equipo de trabajo que los atributos modificados, son los más propicios para un riesgo determinado.

Refinamiento.

El refinamiento o ajuste (también llamado “reparación”) de casos involucra detectar errores en las soluciones actuales para poder efectuar una modificación sobre ellas. En el caso del modelo propuesto, estas actividades están asociadas a un constante seguimiento por parte del líder del proyecto a las respuestas dadas a los riesgos del proyecto, o a otras características que puedan influir de forma positiva sobre los objetivos del proyecto.

De este seguimiento se puede inferir que la reparación de casos de riesgos estará relacionada a la efectividad de la aplicación de las respuestas y/o contingencias a los riesgos que se deberán estar chequeando constantemente, por lo que se puede afirmar que dicho ajuste se asemeja a una reimplementación de la fase de adaptación a una solución dada de un riesgo determinado.

2.3.3.2 Actualización de la base de conocimiento.

Este es el proceso en que se incorpora lo que se considera útil retener del nuevo riesgo y su solución a la base de casos existente como conocimiento persistente. Se debe tener en cuenta los éxitos y fallos de las soluciones propuestas. A partir de ambos se puede incrementar el conocimiento referente al tratamiento de los nuevos riesgos.

En el modelo propuesto, los riesgos se mantienen actualizados todo el tiempo en la base de casos y mediante el atributo "estado" se indica el estado en que se encuentra la gestión del nuevo riesgo. En este sentido la etapa de retención, recuerdo o aprendizaje se mantiene activa todo el tiempo, desde que el nuevo riesgo ingresa a la base de casos hasta que se repara la solución propuesta. De esta forma el resto de los proyectos pueden contar con todas las respuestas disponibles en todo momento en la base de casos.

Una vez se determina que las respuestas aplicadas a un nuevo riesgo tienen resultados satisfactorios se debe cambiar el estado del riesgo a uno que denote que la solución es válida.

Las respuestas que resultaron un fracaso para el riesgo también se deben retener, ya que a partir de estas se pueden generar otras variantes de respuestas. Para esto se cambia el valor del atributo "éxito" en la clase Respuestas de la base de conocimiento propuesta.

2.4 Conclusiones.

Con el desarrollo del modelo se lograron establecer los pasos necesarios para realizar el análisis de los riesgos, haciendo uso de técnicas de inteligencia artificial, específicamente de

razonamiento basado en casos. Particularmente se especificó el alcance del mismo, los principios y premisas para su aplicación y los elementos que garantizan el funcionamiento del modelo, así como una representación del mismo y de sus actividades:

- En la construcción de la base de casos se realizó una propuesta de modelo conceptual de acuerdo con modelos estudiados y sugerencias propias.
- En el cálculo de similitud se evaluaron medidas de distancias usadas para tipos de atributos numéricos, nominales, textuales y listas, los cuales conforman la representación del vector de casos que se utiliza en el modelo. Para esto, se realizó el ajuste de una función para atributos numéricos para que su resultado fuera calculado en un rango de valores determinado.
- El resto de los pasos propuestos en el modelo tienen una muy dependiente relación con los usuarios. En el caso de la selección de los atributos, los usuarios deben seleccionar cuáles atributos se emplearán para medir la similitud con los riesgos en la base de casos, sus pesos de significación y los atributos de salida; la adaptación y evaluación de la solución se lleva a cabo al determinar los riesgos similares, de la misma forma con el refinamiento y ajuste de la solución.
- Finalmente la actualización de la base de conocimiento de riesgos se lleva a cabo de forma automática, permitiendo contar con las respuestas aplicadas a los riesgos en todo momento, incluyendo las que no resultaron satisfactorias.

Este modelo tiene como limitante principal que sobre el tipo de vector de información propuesto, el cual se adecua a un tipo de representación de la información plana, se lleva a cabo el cálculo de la similitud mediante medidas basadas fundamentalmente en métodos de vecinos cercanos, pudiendo atentar contra el rendimiento de la recuperación a medida que esta crece. Una solución viable a este problema radica en la construcción de un árbol de decisión previamente preindexado, el cual se construye en torno a los criterios que conforman el agrupamiento de los índices. En nuestro caso los índices se pudieran conformar de acuerdo al peso de significación de cada atributo sobre el resto, y de esta forma construir el resto de los niveles del árbol de decisión. Esta técnica tiene la desventaja de que si los atributos de un caso son desconocidos se hace complejo hallar la similitud en este tipo de estructura. En este sentido las técnicas basadas en vecinos cercanos son menos sensibles a información desconocida.

Se recomienda incorporar al modelo funciones de similitud que tengan en cuenta conjuntos

borrosos ya que representan matemáticamente la imprecisión intrínseca de ciertas decisiones a tomar.

Igualmente, se propone la definición de algoritmos que consigan “olvidar” la información que no se utiliza y de esta forma aligerar la cantidad de información a procesar por el razonador.

Capítulo 3. Validación de la propuesta.

3.1 Introducción.

El presente capítulo está dividido en cinco secciones. En la primera de ellas se presenta una herramienta elaborada a partir del modelo propuesto como parte de un plugin adaptado al software Redmine, se muestran sus pantallas principales y se caracterizan dos entornos en los que está a prueba. En la siguiente sección se muestra la validación de la solución propuesta haciendo un análisis de los resultados que arrojó el sistema después de haber sido probado con un conjunto de bases de conocimientos mediante un análisis piloto a partir del clasificador propuesto. Posteriormente se muestra la evaluación realizada a un conjunto de expertos a través del método Delphi y el análisis de sus resultados. Finalmente las conclusiones del presenta capítulo.

3.2 Herramienta elaborada.

A partir del modelo propuesto se elaboró una herramienta en forma de plugin del software Redmine, considerado entre las mejores herramientas de gestión y seguimiento de proyectos actualmente en la Web, empleada además en la UCID, contando está con un plugin para la identificación de riesgos. Está desarrollada mediante el framework Ruby on Rails. Es multiplataforma y soporta múltiples motores de bases de datos relacionales. Es software libre bajo los términos de la licencia GNU General Public License v2 (GPL).

3.2.1 Selección de los atributos.

Una vez que el usuario se encuentra dentro del sistema Redmine y selecciona el proyecto donde trabajará, se procede a seleccionar los atributos que desea sean analizados por la herramienta implementada. Para esto se muestra una lista de los atributos que componen el vector de riesgo. Se debe señalar que esto se puede hacer de forma independiente para cada proyecto.

Al mismo tiempo se asignan los pesos a los atributos y se definen los atributos que se mostrarán a la salida del cálculo de similitud.

Se selecciona la función de similitud por la cual se procesará la similitud del atributo de acuerdo al tipo de dato que represente.

Las funciones de similitud a seleccionar se habilitan de acuerdo a si están implementadas en una librería de clasificación con que cuenta la herramienta elaborada, que está compuesta

por funciones de clasificación para atributos numéricos, nominales, tipo y textuales (Figura 16).

Vistazo Actividad Peticiones Nueva petición Noticias Documentos Wiki

Selección de atributos

BASE TABLE - attachments

Tabla	Atributo	Peso	Relevante	Función de semejanza	Salida
project_risks	id	1.0	<input type="checkbox"/>	[VACIO]	<input type="checkbox"/>
project_risks	name	10.0	<input checked="" type="checkbox"/>	Coseno	<input checked="" type="checkbox"/>
project_risks	description	10.0	<input checked="" type="checkbox"/>	Jaccard	<input checked="" type="checkbox"/>
project_risks	impact	1.0	<input checked="" type="checkbox"/>	Manhattan	<input checked="" type="checkbox"/>
project_risks	probability	1.0	<input checked="" type="checkbox"/>	Manhattan	<input checked="" type="checkbox"/>
project_risks	mitigation	1.0	<input type="checkbox"/>	[VACIO]	<input checked="" type="checkbox"/>
project_risks	mitigation_status	1.0	<input type="checkbox"/>	[VACIO]	<input type="checkbox"/>
project_risks	contingency	1.0	<input type="checkbox"/>	[VACIO]	<input type="checkbox"/>
project_risks	contingency_responsable	1.0	<input type="checkbox"/>	[VACIO]	<input checked="" type="checkbox"/>
project_risks	project_id	1.0	<input type="checkbox"/>	[VACIO]	<input type="checkbox"/>

Figura 16: Vista de la selección de los rasgos predictivos y objetivos de un determinado vector de riesgo.

3.2.2 Recuperación de riesgos semejantes.

Una vez que se tengan todos los datos de la representación del vector, se procede a buscar los casos semejantes al nuevo caso. Para esto primeramente se muestra una lista de los riesgos del proyecto con el riesgo y su descripción como indicadores básicos, a continuación se sugiere una cantidad determinada de riesgos semejantes a seleccionar.

Vistazo Actividad Peticiones Nueva petición Noticias Documentos Wiki Archivos Riesgos Configuración

Riesgos a seleccionar:

Riesgo	Descripción	Cantidad similar	Nuevos similares	Borrar previos	
Mala comunicación con el cliente	Se hace compleja la comunicación con el cliente. Viene poco, falta a reuniones de despacho del proyecto, no avisa contratiempos.	8 <input type="text"/>	<input type="button" value="Aceptar"/>	<input type="button" value="Borrar"/>	<input type="button" value="Revisar"/>
Fecha de entrega poco realista	La fecha pactada con el usuario se concibió de forma muy optimista, no se tuvieron en cuenta factores personales del equipo de trabajo	8 <input type="text"/>	<input type="button" value="Aceptar"/>	<input type="button" value="Borrar"/>	<input type="button" value="Revisar"/>

Figura 17: Vista de la selección de riesgos a los que se les hallará riesgos similares.

Una vez seleccionados se presiona el botón Aceptar y se calculan los riesgos semejantes mediante los atributos seleccionados previamente y se habilitan las opciones Borrar y Revisar.

3.2.3 Explicación y adaptación.

El resultado de este proceso es la lista de riesgos similares al nuevo riesgo seleccionado. La siguiente imagen (Figura 18) muestra una vista panorámica del resultado de este proceso.

Vistazo Actividad Peticiones Nueva petición Noticias Documentos Wiki Archivos Riesgos Configuración				
Revisión:				
Nuevo riesgo	Característica	Anterior	Riesgo similar	Próximo
2	IDENTIFICADOR	9		
	DISTANCIA	0.90909090909091		
Mala comunicación con el cliente	NAME	Problemas para contactar con el cliente		
Se hace compleja la comunicación con el cliente. Viene poco, falta a reuniones de despacho del proyecto, no avisa contratiempos.	DESCRIPTION	Existen serias dificultades para contactar con el cliente principal del proyecto, puesto que este responde por varias responsabilidades en su centro laboral		
4	IMPACT	4		
3	PROBABILITY	3		
	MITIGATION	Concertar reunión con el director de la empresa para proponer la liberación parcial del cliente, debido a la alta prioridad del proyecto.		
Jorge L Cuellar	CONTINGENCY_RESPONSABLE	Alberto Perez Suárez		

Figura 18: Vista del proceso de revisión de riesgos semejantes.

En este paso se habilita una navegación hacia los riesgos próximos y anteriores, ordenados por la distancia hacia el riesgo seleccionado. Una vez revisados los riesgos semejantes se procede a actualizar la información referente a los nuevos riesgos a partir de la adaptación de los atributos de los riesgos hallados en la base de conocimiento de riesgos.

Como se comentó anteriormente, esta herramienta se ha desarrollado como plugin de Redmine, por lo que para ponerla a prueba se decidió integrarla al sistema de seguimiento de proyectos de la UCID y de la Dirección Técnica de la Infraestructura Productiva de la Universidad de las Ciencias Informáticas (UCI).

3.3 Caso de estudio 1: Validación del clasificador propuesto.

El objetivo de la clasificación supervisada es la inducción de modelos clasificatorios que tengan una buena capacidad generalizadora, es decir, modelos clasificatorios que ante un nuevo caso del que se conocen los valores de las variables predictoras sean capaces de clasificarlo correctamente con una alta probabilidad.

Para llevar a cabo la validación del modelo de clasificación se empleó el método H (Holdout) también conocido como método de entrenamiento-testeo, el cual se basa en particionar el conjunto de entrenamiento de tamaño N en dos subconjuntos de tamaños respectivos N_1 y N_2 . El primero de dichos subconjuntos (N_1) se denomina de entrenamiento, ya que a partir

del mismo se induce el modelo clasificatorio. La estimación de probabilidad de éxito de dicho modelo frente a casos nuevos se obtiene por medio del porcentaje de casos bien clasificados obtenidos en el segundo subconjunto (N_2). A N_2 se le denomina subconjunto de testeo, ya que es en el que se testa la bondad del modelo clasificatorio.

Suele ser habitual utilizar porciones $2/3$ y $1/3$ para los subconjuntos de entrenamiento y testeo respectivamente, o sea $N_1=2N/3$ y $N_2=N/3$.

Posteriormente el método H se aplica B veces sobre particiones independientes del conjunto de casos. A esto se le denomina remuestreo aleatorio (Kohavi, 1995). En nuestro casos se determinó $B=10$.

En cada una de las B veces se procede de forma análoga. Se induce aleatoriamente el subconjunto N_1 y se aplica el clasificador. El porcentaje de casos bien clasificados, p_1, p_2, \dots, p_B , obtenido en cada uno de los B subconjuntos de testeo se promedian, obteniéndose p_M , cuyo valor constituye una estimación de la probabilidad de llevar a cabo una clasificación correcta inducida por el algoritmo a partir del conjunto de casos.

3.3.1 Bases de datos empleadas para probar el clasificador.

Para evaluar el modelo clasificador, se emplearon cinco bases de datos publicadas en el Machine Learning Repository de la Universidad de California (Center for Machine Learning and Intelligent Systems):

- Iris, suministrada por Michael Marshall, 1988. Base de datos que contiene cuatro atributos numéricos (anchura y longitud del sépalo y pétalo) y uno nominal (tipo de iris) que determina el tipo de planta iris. El conjunto de datos contiene tres clases, cada una de cincuenta instancias las cuales refieren características de uno de los tipos de planta iris.
- Car Evaluation (Evaluación de automóviles), suministrada por Marko Bohanec, 1997. Base de datos derivada de un simple modelo de decisión jerárquico, compuesto por seis atributos nominales (precio de venta, tipo de mantenimiento, puertas, personas, zapata de arranque y seguridad) y uno nominal que determina el tipo de característica del auto.
- Wine Quality (Calidad del vino blanco), suministrada por Paulo Cortez, Universidad de Minho, Guimarães, Portugal, 2009. Base de datos compuesta por once atributos numéricos (acidez fija, acidez volátil, ácido cítrico, azúcar residual, cloruros, dióxido de

azufre libre, dióxido de azufre total, densidad, pH, sulfatos y alcohol) y otro numérico a inducir la calidad del vino en una escala del uno al diez.

- Abalone, suministrada por Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn y Wes B Ford, 1995. Base de datos sobre una especie de caracol marino llamado Haliotis, presenta siete atributos numéricos (longitud, diámetro, ancho, peso total, peso extraído, peso de viseras y peso shell), uno entero (cantidad de anillos) como rasgos descriptores y uno nominal a inducir que identifica el sexo.
- Plantas, suministrada por el departamento de Agricultura de Estados Unidos, 2008. Contiene todas las plantas (especies y géneros) presentes en los estados de Estados Unidos y Canadá. Las columnas que presenta están conformadas por la especie, el género de la planta y a continuación los estados (en abreviatura) en los que se encuentra, por lo que se puede asumir que presenta tres columnas, siendo la tercera la lista de estados. Este conjunto tiene datos ausentes.

3.3.2 Análisis de los resultados del clasificador.

A forma de resumen se muestra en la Tabla 1 los resultados obtenidos en la evaluación del clasificador propuesto. En este caso se presenta el clasificador, empleando para datos de tipo de datos numéricos la función de Manhattan ajustada, para datos de tipo nominales la función de Hamming y para datos de tipo textuales la función Coseno.

Base de datos	Instancias	Atributos	Tipos de datos	Tasa de acierto (%)	Tasa de error (%)
Iris	150	5	Reales	95,1	4,9
Car Evaluation	1728	6	Nominales	90,4	9,6
Wine Quality	4898	12	Reales	93,2	6,8
Abalone	4177	9	Enteros y reales	92,8	7,2
Plantas	22632	70	Texto	78,6	21,4

Tabla 1: Resultado de la aplicación del método H por remuestreo al clasificador.

Como resultado de aplicarle el método H o método de Holdout con remuestreo aleatorio al sistema se concluye lo siguiente: se prueba que estadísticamente el clasificador es correcto, ya en cada uno de los casos la tasa acierto supera el 90% para datos de tipo numérico, reales y nominales, y 78% para datos de tipo texto con ausencia de campos.

Adicionalmente se presentan resultados de comparaciones realizadas entre el clasificador

propuesto y otros encontrados en la literatura científica revisada.

Para el caso de la base de datos Iris, se revisaron las referencias * (Chen & Fang, 2005) y ** (Lai, 2002), en las cuales se aplican fundamentalmente técnicas fuzzy y algoritmos evolutivos. Los resultados son los siguientes ordenados por la tasa de acierto:

Algoritmos	Instancias de entrenamiento	Instancias de prueba	Remuestreo	Tasa de acierto (%)	Tasa de error (%)
Método de Cheng Fang *	150	150	1	97,33	2,67
Método de Cheng Fang *	120	30	200	96,72	3,28
Método de Hong y Cheng *	150	150	1	96,67	3,33
Método de Castro *	120	30	200	96,6	3,4
Método de Cheng Fang *	75	75	200	96,28	3,72
Método de Wu y Cheng *	75	75	200	96,21	3,79
Método de Chang y Cheng *	75	75	200	96,07	3,93
Método de Lai **				95,73	4,27
Método de Hong y Lee **	75	75	200	95,57	4,43
<i>Método propuesto</i>	<i>100</i>	<i>50</i>	<i>10</i>	<i>95,1</i>	<i>4,9</i>
Método de Dasarathy *	150	150	1	94,67	5,33
Método de Ishibuchy **				94,67	5,33

Tabla 2: Comparación con otros métodos propuestos para la base de datos Iris.

En el caso de la base de datos Car Evaluation las referencias revisadas fueron * (Hoffmann, Kwok, & Compton, 2001) y ** (Sameh & Magdy, 2010), estos últimos aplicando algoritmos de colonias de hormigas. Los resultados son los siguientes:

Algoritmos	Instancias de entrenamiento	Instancias de prueba	Remuestreo	Tasa de acierto (%)	Tasa de error (%)
Método de Sameh y Magdy **				93	7
Método de Hoffman et. al. *				91,94	8,06
<i>Método propuesto</i>	<i>1152</i>	<i>576</i>	<i>10</i>	<i>90,4</i>	<i>9,6</i>

Tabla 3: Comparación de métodos de clasificación empleando la base de datos Car Evaluation.

En el caso de la base de datos Wine Quality se revisó el artículo en conferencia de (Senge &

Hüllermeier , 2010),

Algoritmos	Instancias de entrenamiento	Instancias de prueba	Remuestreo	Tasa de acierto (%)	Tasa de error (%)
Método de Senge y Hüllermeier				94,12	5,88
<i>Método propuesto</i>	3266	1632	10	93,2	7,8

Tabla 4: Comparación de métodos de clasificación empleando la base de datos Wine Quality.

3.4 Caso de estudio 2: Puesta a prueba en la UCID.

La UCID no cuenta con suficiente personal como para asumir la gestión de riesgos con rigurosidad, pues apenas alcanzan para desarrollar los proyectos que se les asignan. En esta situación es necesario encontrar una solución al problema utilizar el recuerdo en la detección de respuestas a los riesgos que permita a los equipos de desarrollo asumir los proceso de análisis y de mitigación de gestión de riesgos de sus propios proyectos, teniendo en cuenta los riesgos tratados con anterioridad y las soluciones encontradas en ellos.

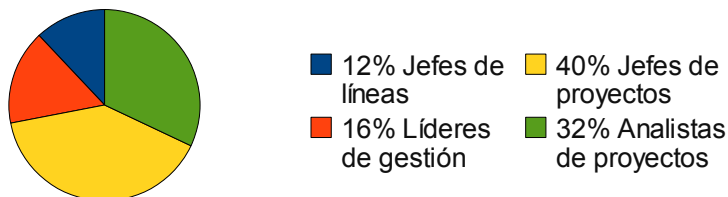


Figura 19: Personal involucrado en la encuesta realizada.

En un diagnóstico realizado por el autor en octubre del 2009 a 25 especialistas del centro los cuales cumplen distintos roles en el proceso de desarrollo de software como indica la Figura 19 (ver Anexo 1. Encuesta realizada a los especialistas en torno a la gestión de riesgos.), se detectó lo siguiente:

- Todos dominan los términos riesgo y gestión de riesgos.
- Todos afirman que en los proyectos que trabajan se gestionan riesgo.
- El 60% lleva a cabo la gestión de riesgos de acuerdo a una metodología formal y el resto lo hace adecuando la que está implantada.
- Aunque se conocen varias técnicas de análisis de gestión de riesgo (Figura 20), la mayoría (92%) coincide que solo se trabaja con la matriz de probabilidad e impacto por ser la más fácil de entender y aplicar.
- Todos afirman documentar las experiencias recogidas cuando los proyectos sufren el

impacto de los riesgos.

- Todos consideran oportuno aprender de la gestión de riesgos realizadas a proyectos anteriores, a fin de evitar problemas en los proyectos actuales.
- Se establece una buena comunicación con el equipo en torno al conocimiento de los riesgos que afectan al proyecto y su estudio.

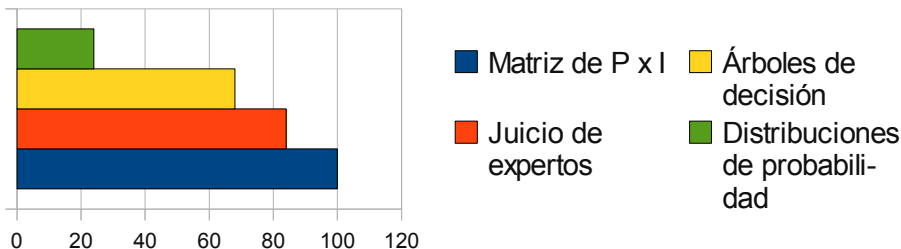


Figura 20: Técnicas de análisis de gestión de riesgos conocidas.

A raíz de la implementación de la herramienta descrita en el apartado 3.2 (ver página 60), se decidió someterla a una revisión por parte del grupo de calidad y ponerla a prueba por el grupo de soporte, ambos pertenecientes al centro.

3.4.1 Implementación del método Delphi.

Para llevar a cabo la validación del modelo propuesto se utilizó el método de expertos Delphi. Para ello se contó con la participación de nueve especialistas a los cuales se les evaluó sus conocimientos y habilidades relacionado con los temas Gestión de Proyectos, Gestión de Riesgos y Herramientas y/o técnicas de análisis de riesgos. De ellos cinco son doctores en ciencias técnicas, tres máster en informática aplicada, siete son profesores titulares, seis son investigadores titulares y tres investigadores auxiliares.

Para esto se llevaron a cabo dos rondas de preguntas, estableciendo el coeficiente de competencia o índice de experticidad (K) para cada experto (ver Anexo 2. Encuesta de validación de la propuesta.) a partir de la fórmula:

$$K = \frac{(K_c + K_a)}{2}$$

Donde

K_c es el coeficiente de conocimiento o información que tiene el experto acerca del problema, calculado sobre la valoración del propio experto en una escala del uno al diez de forma creciente y multiplicado por 0,1.

K_a es el coeficiente de argumentación o fundamentación de los criterios del experto calculado con el promedio de los grados de influencia asignados en una tabla patrón mostrada a continuación:

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus criterios		
	Alto	Medio	Bajo
Análisis teóricos realizados por usted (AT)	0,3	0,2	0,1
Su experiencia obtenida (EO)	0,5	0,4	0,2
Trabajo de autores nacionales (AN)	0,05	0,05	0,05
Trabajo de autores extranjeros (AE)	0,05	0,05	0,05
Su propio conocimiento del estado del problema en el extranjero (PC)	0,05	0,05	0,05
Su intuición (I)	0,05	0,05	0,05

Tabla 5: Patrón de influencia.

Los resultados mostrados del cálculo del índice de experticidad se muestran a continuación:

Expertos	Kc	AT	EO	AN	AE	PC	I	Ka	K	Competencia
1	0,8	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,7	Media
2	0,9	0,2	0,5	0,05	0,05	0,05	0,05	0,9	0,9	Alta
3	0,8	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,7	Media
4	0,9	0,2	0,5	0,05	0,05	0,05	0,05	0,9	0,9	Alta
5	0,8	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,7	Media
6	0,8	0,2	0,4	0,05	0,05	0,05	0,05	0,8	0,8	Alta
7	0,7	0,2	0,4	0,05	0,05	0,05	0,05	0,8	0,75	Media
8	0,7	0,2	0,2	0,05	0,05	0,05	0,05	0,6	0,65	Media
9	0,9	0,3	0,4	0,05	0,05	0,05	0,05	0,9	0,9	Alta

Tabla 6: Resultados del índice de experticidad.

Los rangos establecidos para determinar la competencia se establecen de la siguiente forma:

- Alto: si K es mayor o igual que 0,8,
- Media: si K es mayor o igual que 0,5 y menor que 0.8, y
- Baja: si K es menor que 0,5.

Como los índices de experticidad se encuentran por encima de 0,25 (considerado el rango mínimo) se decidió contar con todos los los expertos seleccionados en la primera ronda de preguntas.

El cuestionario elaborado consta de seis preguntas encaminadas a evaluar el modelo de razonamiento basado en casos propuesto como técnica o herramienta para el proceso de análisis en la gestión de riesgos.

Las respuestas a estas preguntas se categorizaron en:

- Muy adecuado (C1),
- Bastante adecuado (C2),
- Adecuado (C3),
- Poco adecuado (C4), e
- Inadecuado (C5).

El cuestionario se entrego en copia dura a los expertos, y se les anexo un documento con la descripción del trabajo, características y métodos propuestos.

3.4.2 Análisis de los resultados.

Para llevar a cabo el análisis de los resultados se utilizaron métodos estadísticos tal y como indica el método Delphi. La siguiente tabla muestra los resultados obtenidos en la encuesta realizada (Anexo 3. Cuestionario sobre la validación del modelo propuesto.):

No	Aspectos	C1	C2	C3	C4	C5	Total
1	Propuesta de una herramienta RBC para apoyar el proceso de análisis en la GR	8	1				9
2	Estructura de la base de conocimientos de riesgos	7	1	1			9
3	Fases propuestas en el modelo	8		1			9
4	Funciones de semejanza	6	1	1	1		9
5	Algoritmo de recuperación	7	1	1			9
6	Algoritmo de reutilización	7	1	1			9

Tabla 7: Resultados de la encuesta.

A partir de la frecuencia absoluta se puede extraer la frecuencia absoluta acumulada la que se obtiene sumando los valores de cada fila con el valor anterior, quedando de la siguiente manera:

No	Aspectos	C1	C2	C3	C4	C5
1	Propuesta de una herramienta RBC para apoyar el proceso de análisis en la GR	8	9	9	9	9
2	Estructura de la base de conocimientos de riesgos	7	8	9	9	9
3	Fases propuestas en el modelo	8	8	9	9	9
4	Funciones de semejanza	6	7	8	9	9
5	Algoritmo de recuperación	7	8	9	9	9
6	Algoritmo de reutilización	7	8	9	9	9

Tabla 8: Frecuencia absoluta respecto a los resultados obtenidos.

La frecuencia absoluta acumulada se divide entre la cantidad de expertos y se obtiene la frecuencia relativa acumulada. En la siguiente representación desaparece la columna C5.

No	Aspectos	C1	C2	C3	C4
1	Propuesta de una herramienta RBC para apoyar el proceso de análisis en la GR	0.8888	0.9999	0.9999	0.9999
2	Estructura de la base de conocimientos de riesgos	0.7777	0.8888	0.9999	0.9999
3	Fases propuestas en el modelo	0.8888	0.8888	0.9999	0.9999
4	Funciones de semejanza	0.6666	0.7777	0.8888	0.9999
5	Algoritmo de recuperación	0.7777	0.8888	0.9999	0.9999
6	Algoritmo de reutilización	0.7777	0.8888	0.9999	0.9999

Tabla 9: Frecuencia absoluta acumulada respecto a los resultados obtenidos.

Por último se buscan las imágenes de las frecuencias relativas acumuladas por medio de la Distribución Normal Estándar Invertida y se adicionan las siguientes columnas:

- Sumatoria (Suma): sumatoria de cada fila y columna.
- Promedio (P): promedio de los criterios de los expertos
- N: es el resultado de dividir la sumatoria de las sumas entre el producto del número de categorías por el número de pasos
- N-P: es entonces el valor promedio que le otorgan los expertos consultados a cada paso de la metodología
- Punto de corte: es el promedio de cada columna.

La siguiente tabla resume estos resultados:

No	Aspectos	C1	C2	C3	C4	Suma	N		Competencia
							P	N-P	
1	Propuesta de una herramienta RBC para apoyar el proceso de análisis en la GR	1,22	3,72	4,26	4,26	13,47	3,37	-0,76	Muy Adecuado
2	Estructura de la base de conocimientos de riesgos	0,76	1,22	4,26	4,26	10,51	2,63	-0,02	Muy Adecuado
3	Fases propuestas en el modelo	1,22	1,22	4,26	4,26	10,97	2,74	-0,13	Muy Adecuado
4	Funciones de semejanza	0,43	0,76	1,22	4,26	6,68	1,67	0,94	Bastante Adecuado
5	Algoritmo de recuperación	0,76	1,22	4,26	4,26	10,51	2,63	0,02	Muy Adecuado
6	Algoritmo de reutilización	0,76	1,22	4,26	4,26	10,51	2,63	0,02	Muy Adecuado
Suma		5,16	9,36	22,55	25,59	62,66			
Puntos de corte		0,86	1,56	3,76	4,26				

Tabla 10: Frecuencias relativas mediante la Distribución Normal Estándar Invertida.

Los puntos de corte sirven para determinar la categoría o grado de adecuación de cada criterio según la opinión de los expertos consultados. La determinación del grado de adecuación de cada aspecto a validar se realiza de la siguiente forma:

- Si N-P es menor o igual que 0,86 entonces es Muy Adecuado,
- Si N-P es mayor que 0,86 y menor o igual que 1,56 es Bastante Adecuado,
- Si N-P es mayor que 1,56 y menor o igual que 3,76 es Adecuado, y
- Si N-P es mayor que 4,26 es Poco Adecuado.

A partir de este análisis todos los aspectos del modelo propuesto fueron catalogados de muy adecuados a excepción de las funciones de semejanza propuestas cuyos señalamientos principales van encaminados al uso otras funciones para tipos de datos nominales y dirigidas al análisis de texto.

La evaluación de la concordancia se realizó mediante el cálculo del coeficiente de Kendall (W), y el estadígrafo Chi Cuadrado. Los análisis realizados se muestran a continuación:

	E1	E2	E3	E4	E5	E6	E7	E8	E9	Ep	$\sum E$	ΔC	ΔC^2
P1	5	5	5	5	5	5	5	5	4	4,89	44	2	4
P2	5	4	5	3	5	5	5	5	5	4,67	42	0	0
P3	5	3	5	5	5	5	5	5	5	4,78	43	1	1
P4	5	4	5	5	5	3	5	5	2	4,33	39	-3	9
P5	5	5	5	4	3	5	5	5	5	4,67	42	0	0
P6	5	5	5	4	3	5	5	5	5	4,67	42	0	0
Total	30	26	30	26	26	28	30	30	26		252		14

Tabla 11: Concordancia de los resultados.

Donde:

C: es la cantidad de criterios presentes,

Ep: es el promedio de los criterios dados (filas),

$\sum E$: es la sumatoria de los criterios dados (filas)

ΔC : es la diferencia entre el total de $\sum E$ entre C y $\sum E$,

Δ : es el cuadrado de ΔC .

Con estos datos se calcula S según:

$$S = \sum \left(\sum E - \sum \frac{\sum E}{C} \right)^2$$

Y a continuación el coeficiente de Kendall (W) y Chi Cuadrado (X^2) según:

$$W = \frac{S}{\frac{E^2}{12}(C^3 - C)}$$

$$X^2 = E(C - 1)W$$

Arrojando como resultado un coeficiente de Kendall (W) de 0.0098, y para 6 criterios manejados un valor hallado de Chi Cuadrado de 0.44 donde, comparado con las tablas estadísticas, para los mismos datos se obtiene un Chi Cuadrado de 0,554 determinando concordancia entre los criterios de los expertos.

Los comentarios aportados por los expertos para mejorar el modelo estuvieron enfocados a conjugar la propuesta presentada con otras herramientas para así darle una mayor

connotación a los resultados que se podrían obtener en el proceso de análisis. Principalmente expresaron que el uso de árboles de decisiones en la etapa propuesta de revisión permitiría hacer una corrida del comportamiento de los riesgos sobre los objetivos del proyecto.

3.5 Caso de estudio 3: Puesta a prueba en la Dirección Técnica de la Infraestructura Productiva de la UCI.

La Dirección Técnica de la Infraestructura Productiva de la Universidad emplea la herramienta GESPRO, la cual es un Paquete para la Gestión de Proyectos desarrollado por la UCI y comercializable desde las empresas comercializadoras asociadas a la Universidad, además es un producto registrado en el Centro Nacional de Derecho de Autor (CENDA).

El paquete GESPRO v1.0 ha sido desplegado y es utilizado con éxito por toda la red de centros de la Universidad de Ciencias Informáticas (Red Centros UCI) para la gestión de los proyectos, que incluye más de 17 centros de desarrollo de software, con contratos nacionales e internacionales que superan los 150 proyectos. En este entorno gestiona la actividad de más de 6300 usuarios; el 75% de los centros ubicados en el campus universitario y el 25 % en regiones distantes del país; es la plataforma para la gestión de las actividades de los proyectos así como el control y seguimiento de los mismos por la dirección de la Universidad; cada centro dispone de un paquete con total independencia de los demás, que permite agilidad e independencia para las decisiones tácticas y operativas; la información de todos los centros se concentra en un Almacén de Datos Operacional (ODS – Operational Data Store) para su procesamiento y toda de decisiones estratégicas a través de un DATAMART por la máxima dirección de la Universidad y sus Empresas Asociadas, y más de 20 tipos de reportes personalizados para la toma de decisiones, y la generación dinámica de nuevos reportes a partir de solicitudes de clientes.

Esta herramienta presenta beneficios agregados sobre los siguientes procesos: gestión de integración, alcance, duración de tiempo, recursos humanos, comunicaciones, riesgos, costos, calidad y adquisiciones de proyectos. En este caso la herramienta propuesta se puede integrar a esta solución para contar con una propuesta más de herramienta para el análisis de los riesgos de software que se registren en la misma.

3.8 Conclusiones.

Se elaboró una herramienta que aun continúa en desarrollo, la cual implementa las

funcionalidades básicas del modelo propuesto, la cual está integrada a los sistemas de gestión de proyectos de la UCID y de la Dirección Técnica de IP de la UCI.

El modelo propuesto de razonamiento basado en casos se presentó a un conjunto de expertos que valoraron un conjunto de características que se proponen en el presente trabajo.

A partir un análisis estadístico se validó de muy adecuado el modelo propuesto en cuanto a:

- la propuesta de contar con una herramienta de razonamiento basado en casos que apoye el proceso de análisis en la gestión de riesgos de software,
- la estructura propuesta de base de conocimiento de riesgos,
- la fases con que cuenta el modelo propuesto,
- el algoritmo de recuperación de riesgos, y
- el algoritmo de reutilización.

De este análisis se derivó también que las funciones de semejanza propuestas se caracterizan como bastante adecuadas, puesto que los expertos argumentaron que también se deben proponer otras para el análisis de texto y tipos de datos nominales.

Conclusiones.

Una vez concluida esta investigación se concluye lo siguiente:

- Investigaciones realizadas demuestran que aunque se establece la recopilación de la información respecto a la gestión de riesgos, esta no se reutiliza posteriormente para aplicarla a nuevos riesgos, solo un modelo de gestión de riesgos concibe una etapa de aprendizaje pero no define con claridad los artefactos y procesos para poner esta información en uso en favor del análisis en la gestión de riesgos.
- En la UCID aumenta el número de proyectos a realizar y con ello aumentan las actividades implicadas en la gestión de proyectos, trayendo como consecuencia que la gestión de riesgos se torne compleja, al manejar demasiada información.
- El modelo de razonamiento basado en casos propuesto utiliza la información “pasiva” de los riesgos anteriores en función del proceso de análisis de la gestión de riesgos.
- Se propone una representación de la base de conocimiento de riesgos, sobre la cual debe trabajar el razonador basado en casos.
- El modelo propone un conjunto de algoritmos para recuperar, adaptar y aprender de la información de los riesgos con el objetivo de determinar posibles soluciones o caracterizar riesgos actuales.
- Se realizó una validación del clasificador propuesto en una prueba piloto con bases de datos publicadas en el Machine Learning Repository de la Universidad de California, obteniendo en cada uno de los casos una tasa de acierto superior al 90 por ciento para atributos de tipo numérico y superior al 78 por ciento para atributos textuales.
- Los expertos concuerdan en la efectividad del modelo para darle solución a la problemática planteada.

Recomendaciones.

Con el objetivo de mejorar la solución propuesta se recomienda lo siguiente:

- Implementar el modelo propuesto en una herramienta automatizada, de la cual ya se tiene un gran avance.
- Continuar investigando en las medidas de similitud y modelos de memoria para lograr un mejor desempeño en la clasificación de los riesgos.
- Realizar nuevas propuestas sobre el modelo conceptual de riesgos, con el fin de hacerlo más genérico.
- Incorporar al modelo funciones de similitud que tengan en cuenta conjuntos borrosos ya que representan matemáticamente la imprecisión intrínseca de ciertas decisiones a tomar.
- Definir algoritmos para “olvidar” la información de los riesgos que no son útiles.

Bibliografía.

- Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches. *AI Communications*. IOS Press, 7(1), 39-59.
- Agudo, M. B. D. (2002). *Una aproximación ontológica al desarrollo de sistemas de razonamiento basado en casos*. Tesis Doctoral, Universidad Complutense de Madrid. España. Recuperado Mayo 16, 2008, a partir de <http://www.ucm.es/eprints/4674/>.
- Althoff, K., Aamodt, A., Magaldi, R., & Milne, R. (1995). Evaluating case-based reasoning systems. En *Case-based reasoning: A new force in advanced systems development*. Unicom Seminars & AI Intellingence, London, United Kingdom.
- Althoff, K., Auriol, E., Barletta, R., & Manago, M. (1995). *A Review of industrial case-based reasoning tools*. AI perspectives report. (Oxford). Recuperado Marzo 23, 2010, a partir de http://openlibrary.org/b/OL15216465M/Review_of_industrial_case-based_reasoning_tools.
- Antonín, M. A. M., Montraveta, A. F., & García, G. V. (2003). *Lexicografía computacional y semántica*. Edicions Universitat Barcelona, España.
- Arean-Rodríguez, Y. (2003). *Una herramienta de búsqueda inteligente en bases de datos utilizando técnicas de razonamiento basado en casos*. Tesis de Maestría, CEIS, Instituto Superior Politécnico José Antonio Echevarría, Cuba.
- Beall, J. (2008). The Weaknesses of Full-Text Searching. *The Journal of Academic Librarianship*, 34(5), 438-444. doi: 10.1016/j.acalib.2008.06.007.
- Bergmann, R., & Stahl, A. (1998). Similarity measures for object-oriented case representations. En *Advances in Case-Based Reasoning*, Lecture Notes in Computer Science (Vol. 1488, págs. 25-36). London, UK: Springer Berlin / Heidelberg. Recuperado Marzo 4, 2010, a partir de <http://dx.doi.org/10.1007/BFb0056319>.
- Bergmann, R., & Wilke, W. (1998). Towards a New Formal Model of Transformational Adaptation in Case-Based Reasoning. *UNIVERSITY OF ROSTOCK*, 53--57.
- Bilenko, M., & Mooney, R. J. (2003). Adaptive Duplicate Detection Using Learnable String Similarity Measures. *IN PROCEEDINGS OF THE NINTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (KDD-2003)*, 39--48.
- Boehm, B. (1989). Software Risk Management. En *ESEC '89*, Lecture Notes in Computer Science (Vol. 387, págs. 1-19). IEEE Computer Society Press.
- Boehm, B. (1991). Software Risk Management: Principles and Practices. *IEEE Software*, 32-41. doi: <http://doi.ieeeecomputersociety.org/10.1109/52.62930>.
- Bregón, A., Simón, A., Alonso, C., Pulido, B., Moro, I., & Juan José Rodríguez. (2005). Un sistema de razonamiento basado en casos para la clasificación de fallos en sistemas dinámicos. En *Actas del III Taller Nacional de Minería de Datos y Aprendizaje* (págs. 203-211). TAMIDA.
- Brito-González, D., & Bon-Pérez, M. (2009). *Propuesta de Modelo para el Análisis Cuantitativo de Riesgos basado en técnicas de Softcomputing*. Tesis de Grado, Universidad de las Ciencias Informáticas, Cuba.
- Cardona, P. A. (2004). Aplicación de árboles de decisión en modelos de riesgo crediticio. *Revista Colombiana de Estadística, Colombia*. Recuperado Noviembre 3, 2008, a

partir de <http://ideas.repec.org/a/col/000163/004075.html>.

- Carol, G. (2010, Enero). Project Risk Management for Absolute Beginners. Recuperado Junio 4, 2010, a partir de <http://www.pmhut.com/project-risk-management-for-absolute-beginners>.
- Chen, S., & Fang, Y. (2005). A new approach for handling the iris data classification problem. *International Journal of Applied Science and Engineering*, 3(1), 37-49.
- Cocho, J. M., Adam, M. R., & Torralba, J. M. (2003). Estudio exploratorio sobre los métodos de gestión de proyectos de alto riesgo (págs. 343-359). Presented at the Primer congreso Soporte del Conocimiento con la Tecnología, Universidad Politécnica de Valencia, España.
- Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. *Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03)*, 73--78.
- Dapena, M. D. D. (2006). *Un modelo para la gestión de revisiones en proyectos de software utilizando razonamiento basado en casos*. Tesis Doctoral, Instituto Superior Politécnico José Antonio Echeverría, Cuba.
- DeMarco, T., & Lister, T. (2003). Risk Management during Requirements. *IEEE Software*, 20(5), 99-101. doi: 10.1109/MS.2003.1231163.
- Díaz-Martínez, Z., Fernández-Menéndez, J., & Segovia-Vargas, M. J. (2004). *Sistemas de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradoras*. Universidad Complutense de Madrid, Facultad de Ciencias Económicas y Empresariales. España. Recuperado Noviembre 3, 2008, a partir de <http://ideas.repec.org/p/ucm/doctra/04-09.html>.
- Esteves, J., Pastor, J., Rodríguez, N., & Roy, R. (2005, Marzo). Implementación y mejora del método de gestión de riesgos del SEI en un proyecto universitario de desarrollo de software. *IEEE Latin America Transactions*, 3(1).
- Estévez, M. D. L. B., & Gallastegui, J. J. A. (2002). El método Delphi. Su implementación. Una estrategia didáctica para la enseñanza de las demostraciones geométricas. *Revista Iberoamericana de Educación*, 11.
- Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook, Advanced Approaches in Analyzing Unstructured Data*. New York. United States of America: Cambridge University Press.
- Fiorito, F. (2006). La simulación como una herramienta para el manejo de la incertidumbre. Universidad del Cema.
- Gallegos, J. D. C. (2006). Análisis del riesgo en la administración de proyectos de tecnología de la información. *Ind. Data, Perú*, 9(1), 104-107.
- García-Hernández, E. (2008). *Propuesta de modelo para la gestión de riesgos en los proyectos de producción de software*. Tesis de Grado, Universidad de las Ciencias Informáticas, Cuba.
- Gelbukh, A. (2008). *Computational linguistics and intelligent text processing*. Springer Publishing Company, Incorporated, Haifa, Israel.
- Hand, D., Keim, D., & Ng, R. (2002). *Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Canada: ACM.

- Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*. Data Management Systems (2^o ed., pág. 770). ACM New York, NY, USA .
- Higuera, R. P., & Haimés, Y. Y. (1996). *Software risk management*. Technical Report, CMU/SEI-96-TR-012, Software Engineering Institute, Carnegie Mellon University, USA.
- Hoffmann, A. G., Kwok, R. B. H., & Compton, P. (2001). Using Subclasses to Improve Classification Learning. En *Proceedings of the 12th European Conference on Machine Learning* (págs. 203-213). Springer-Verlag, London UK.
- Houston, D. X., Mackulak, G. T., & Collofello, J. S. (2001). Stochastic simulation of risk factor potential effects for software development risk management. *Journal of Systems and Software*, 59(3), 247-257. doi: 10.1016/S0164-1212(01)00066-8.
- Inteco. (2008). *Guía avanzada de gestión de riesgos*. Instituto Nacional de Tecnologías de Comunicación, España.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. En *Proceedings of the 14th international joint conference on Artificial intelligence* (Vol. 2, págs. 1137-1143). Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc. Recuperado Junio 23, 2010, a partir de <http://portal.acm.org/citation.cfm?id=1643047>.
- Köhler, J., Philippi, S., Specht, M., & Rüegg, A. (2006). Ontology based text indexing and querying for the semantic web. *Knowledge-Based Systems*, Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, 19(8), 744-754. doi: 10.1016/j.knosys.2006.04.015.
- Kolodner, J., & Leake, D. (1996). A Tutorial Introduction to Case-Based Reasoning. En D. Leake (Ed.), *Case-Based Reasoning: Experiences, Lessons, and Future Directions* (págs. 31-65). AAAI Press / The MIT Press, United States of America.
- Kowalski, G., & Maybury, M. T. (2000). *Information storage and retrieval systems*. Springer.
- Kuhn, A., Ducasse, S., & Gîrba, T. (2007). Semantic clustering: Identifying topics in source code. *Information and Software Technology*, Butterworth-Heinemann Newton, MA, USA, 49(3), 230-243. doi: 10.1016/j.infsof.2006.10.017.
- Kulik, P., & Weber, C. (2001). *Software Risk Management Practices - 2001*. KLCI Research Group, United States of America.
- Lai, F. (2002). *GA-Based fuzzy clustering applied to irregular shape*. Tesis de Maestría, Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan.
- Laurent, P. J. (1966). *Les methodes de Monte-Carlo*. Faculte des Sciences de Grenoble et Institut Polytechnique. Francia.
- Leake, D. B. (Ed.). (1996). CBR in Context: The Present and Future. En *Case-Based Reasoning: Experiences, Lessons and Future Directions*. Menlo Park: AAAI Press/MIT Press, Indiana University, United States of America.
- Leake, D. B., & Plaza, E. (Eds.). (1997). *Lecture Notes in Artificial Intelligence: Case-Based Reasoning Research and Development* (Vol. 1266). Providence, United States of America.
- Lefever, E., Fayruzov, T., Hoste, V., & De Cock, M. (2010). Clustering web people search results using fuzzy ants. *Information Sciences*, Elsevier Science Inc. New York, NY, United State of America, 180(17), 3192-3209. doi: 10.1016/j.ins.2010.05.018.

- Litten, D. (2009, Octubre). Project Risk and Risk Management. Recuperado Junio 4, 2010, a partir de <http://www.pmhut.com/project-risk-and-risk-management>.
- Liu, T., Moore, A. W., Gray, A., & Yang, K. (1998). An investigation of practical approximate nearest neighbor algorithms. School of Computer Science, Carnegie-Mellon University, Pittsburg, United States of America.
- Lovejoy, K. G. (2005). The Evolution of Operational Risk Management. *Sarbanes-Oxley Compliance Journal*, New Jersey, Estados Unidos de América. Recuperado Junio 4, 2010, a partir de http://www.s-ox.com/dsp_getFeaturesDetails.cfm?CID=568.
- Lozano, L., & Fernández, J. (2004). Razonamiento basado en casos: una visión general. Universidad de Valladolid, España.
- MAP. (2006a). Magerit- versión 2. Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información. III - Guías de Técnicas. Ministerio de Administraciones Públicas, España.
- MAP. (2006b). Magerit- versión 2. Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información. I - Método. Ministerio de Administraciones Públicas, España.
- McCane, B., & Albert, M. (2008). Distance functions for categorical and mixed variables. *Pattern Recognition Letters*, Elsevier Science Inc. New York, NY, USA, 29(7), 986-993. doi: 10.1016/j.patrec.2008.01.021.
- McManus, J., & Wood-Harper, T. (2008, Junio). A study in project failure. *The Chartered Institute for IT*. Recuperado a partir de <http://www.bcs.org/server.php?show=ConWebDoc.19584>.
- Michavila, F. (1986). *Fundamentos del calculo numerico I: Topología numérica*. Editorial Reverte, SA, Barcelona, España.
- Microsoft. (2005). Microsoft Operations Framework - MOF: The Risk Management Discipline for IT Operations White Paper. Recuperado Octubre 2, 2008, a partir de <http://www.microsoft.com/technet/solutionaccelerators/cits/mo/mof/mofrisk.msp>.
- Moreno, J. M. D. (1998). *Introducción a la topología de los espacios métricos*. Universidad de Cadiz, España: Editorial Servicio Publicaciones.
- Morin, D. R. L. (2007). Evaluación del riesgo agroambiental de los suelos de las comunidades indígenas del estado Anzoátegui, Venezuela. *Ecosistemas*, 16(1), 69-79.
- Murphy, R. L., Alberts, C. J., Williams, R. C., Higuera, R. P., Dorofee, A. J., & Walker, J. A. (2008). Continuous Risk Management Guidebook. Software Engineering Institute, Carnegie Mellon University, USA. Recuperado Septiembre 27, 2008, a partir de <http://www.sei.cmu.edu/publications/books/other-books/crm.guidebk.html>.
- Olvera-López, J. A., Martínez-Trinidad, J. F., & Carrasco-Ochoa, J. A. (2005). Edition Schemes Based on BSE. *Progress in Pattern Recognition, Image Analysis and Applications*, Lecture Notes in Computer Science, 3773, 360-367. doi: 10.1007/11578079_38.
- Pal, S. K., & Shiu, S. C. K. (2004). *Foundations of Soft Case-Based Reasoning*. Jonh Wiley & Sons, Inc.
- Perminova, O., Gustafsson, M., & Wikström, K. (2008). Defining uncertainty in projects - a new perspective. *International Journal of Project Management*, 26(1), 73-79. doi: 10.1016/j.ijproman.2007.08.005.

- Piccarreta, R. (2010). Binary trees for dissimilarity data. *Computational Statistics & Data Analysis*, Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, 54(6), 1516-1524. doi: 10.1016/j.csda.2009.12.011.
- Piñero-Pérez, P. Y. (2005). *Un modelo para el aprendizaje y la clasificación automática basado en técnicas de softcomputing*. Tesis Doctoral, Universidad Central de Las Villas, Cuba.
- PMI. (2004). *Guía de los Fundamentos de la Dirección de Proyectos (Guía del PMBOK)* (Tercera Edición.). Project Management Institute, Inc. Newtown Square, Pennsylvania 19073-3299 USA.
- Pressman, R. S. (2005). *Ingeniería de Software. Un enfoque práctico* (Quinta Edición.). Editorial Félix Valera, La Habana, Cuba.
- Ricardo Mauricio Aflaro-López, Víctor Danilo Castro-Rivera, & Balmero Osmar Romero-Lainez. (2005). *Pronóstico Delphi*. Universidad de El Salvador. El Salvador.
- Rosenberg, L., Gallo, A., Hammer, T., & Parolek, F. (2000, Febrero). Continuing Risk Management at NASA. *CrossTalk, The Journal of Defense Software Engineering. USA*, 13(2). Recuperado Mayo 13, 2010, a partir de <http://www.stsc.hill.af.mil/Crosstalk/2000/02/rosenberg.html>.
- Rosenberg, L. H., Hammer, T., & Gallo, A. (1999). Continuous Risk Management at NASA. NASA Software Assurance Technology Center, USA. Recuperado Septiembre 27, 2008, a partir de http://satc.gsfc.nasa.gov/support/ASM_FEB99/crm_at_nasa.html.
- Sameh, A., & Magdy, K. (2010). Data Mining Ant Colony for Classifiers. *International Journal of Basic & Applied Sciences*, 10(3), 28-39.
- Schank, R. C. (1999). *Dynamic memory revisited*. New York. United States of America: Cambridge University Press.
- Senge, R., & Hüllermeier, E. (2010). Pattern Trees for Regression and Fuzzy Systems Modeling. En *Draft of a paper published in Proc. WCCI-2010*. Presented at the World Congress on Computational Intelligence, Barcelona.
- Stanfill, C., & Waltz, D. (1986). Toward memory-based reasoning. *Commun. ACM, New York, NY, USA*, 29(12), 1213-1228. doi: 10.1145/7902.7906.
- Trujillo-Cabezas, R. (2004). *Aplicaciones del Método Delphi*. Universidad Externado de Colombia.
- Williams, R. C., Pandelios, G. J., & Behrens, S. G. (1999). *Software risk evaluation method description (Version 2.0)*. Software Engineering Institute, Carnegie Mellon University. United States of America.
- Wilson, R., & Martínez, T. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, AI Access Foundation, USA, 6(1), 1-34.
- Wilson, R., & Martinez, T. R. (2000). Reduction Techniques for Instance-Based Learning Algorithms. *MACHINE LEARNING*, Kluwer Academic Publishers Hingham, MA, USA, 38(3), 257-286.
- Yang, Y., & Pedersen, J. (1997). A Comparative Study on Feature Selection in Text Categorization. En *ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning* (págs. 412-420). Morgan Kaufmann Publishers Inc. San Francisco, California, United States of America.

- Zdrahal, Z., & Motta, E. (2001). Improving Competence by Integrating Case-Based Reasoning and Heuristic Search. Knowledge Media Institute, The Open University Walton Hall, United Kingdom. Recuperado Agosto 8, 2008, a partir de <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/zdrahal/cbr.html#cbd>.
- Zhang, B., & Srihari, S. N. (2002). A fast algorithm for finding k-nearest neighbors with non-metric dissimilarity (págs. 13 - 18). Presented at the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02), CEDAR, Department of Computer Science and Engineering, State University of New York at Buffalo, United States of America. doi: 10.1109/IWFHR.2002.1030877 .
- Zulueta-Véliz, Y. (2007). *Modelo de gestión de riesgos en proyectos de desarrollo de software*. Tesis de Maestría, Universidad de las Ciencias Informáticas, Cuba.
- Zulueta-Véliz, Y., & Despaigne-Herrera , E. (2007). Modelo de gestión de riesgos en proyectos de desarrollo de Software. En *UCIENCIA 2007*. Presented at the III Taller de Ingeniería, Arquitectura y Gestión de Software, Universidad de las Ciencias Informáticas, Cuba.

Anexos.

Anexo 1. Encuesta realizada a los especialistas en torno a la gestión de riesgos.

1. ¿Domina los términos riesgo y gestión de riesgos?
 - Respuestas: Sí ___ No ___ Otro criterio _____
2. ¿En los proyectos que ha trabajado o trabaja se han gestionado los riesgos?
 - Respuestas: Sí ___ No ___ Otro criterio _____
3. ¿Se lleva a cabo la gestión de riesgos de acuerdo a algún marco formal?
 - Respuestas: Sí ___ No ___ Otro criterio _____
4. ¿Realizan el proceso de análisis de riesgo?
 - Respuestas: Sí ___ No ___ Otro criterio _____
5. ¿Qué herramientas o técnicas conocen para analizar los riesgos?
 - Respuestas: _____
6. ¿Qué herramientas o técnicas utiliza para analizar los riesgos?
 - Respuestas: _____
7. ¿Se aplican todas las técnicas o herramientas que se conocen para analizar los riesgos?
 - Respuestas: Sí ___ No ___ ¿Por qué? _____
8. ¿Cuándo el proyecto ha identificado algún riesgo, se han documentado las experiencias sobre su solución?
 - Respuestas: Sí ___ No ___ Otro criterio _____
9. ¿Considera oportuno aprender de las soluciones dadas a riesgos anteriores para evitar percances en los proyectos actuales?
 - Respuestas: Sí ___ No ___ Otro criterio _____
10. ¿Qué importancia le concede a los siguientes elementos para el éxito del proyecto?
 - N: Ninguna, P: Poca, M: Media, A: Alta

Acciones	Criterio
----------	----------

El proyecto conoce los eventos que pueden dañarlo	
El proyecto conoce los eventos que pueden dañarlo y conoce la forma de enfrentarlos	
El proyecto investiga sobre problemas que han afectado a otros proyectos y planifica acciones para evitarlos.	
Los miembros del equipo se reúnen a debatir sobre los riesgos.	

Anexo 2. Encuesta de validación de la propuesta.

Compañera(o):

Teniendo en cuenta su experiencia profesional y sus características personales se necesita que colabore en la validación de una investigación, de la cual se anexa una síntesis de la misma.

1. Años de experiencia profesional.
2. Años de experiencia en la actividad objeto de estudio.
3. Categoría científica.
4. Categoría docente.
5. Marque con una cruz en una escala creciente del uno al diez, el valor del grado de conocimiento e información que usted tiene sobre el tema objeto de investigación:

1	2	3	4	5	6	7	8	9	10

6. Realice una autovaloración, según la tabla que a continuación se ofrece, de sus niveles de argumentación o fundamentación sobre el tema que se investiga. Marque con una cruz:

Fuentes de argumentación	Alto	Medio	Bajo
Análisis teóricos realizados por Usted (AT)			
Experiencia adquirida (EA)			
Conocimiento de trabajo de autores nacionales (AN)			
Conocimiento de trabajo de autores extranjeros (AE)			
Su propio conocimiento del estado del problema en el extranjero (PE)			
Su intuición (I)			

Muchas gracias por su ayuda.

Anexo 3. Cuestionario sobre la validación del modelo propuesto.

1. ¿Considera usted que el empleo de las herramientas actuales contribuyen a obtener un mejor proceso de análisis en la gestión de riesgos de software?

- Completamente ____
- En gran medida ____
- Poco ____
- Muy poco ____
- Nada ____
- ¿Por qué?

2. ¿Cual es su opinión sobre la propuesta de una herramienta de razonamiento basado en casos como apoyo al proceso de análisis en la gestión de riesgos?

- Muy adecuado ____
- Bastante adecuado ____
- Adecuado ____
- Poco adecuado ____
- Inadecuado ____
- ¿Por qué?

3. ¿Cual es su opinión sobre la estructura de base de conocimiento de riesgo de software propuesto?

- Muy adecuado ____
- Bastante adecuado ____
- Adecuado ____
- Poco adecuado ____
- Inadecuado ____

- ¿Por qué?

4. ¿Cual es su opinión sobre las fases propuestas en el modelo?

- Muy adecuado ____
- Bastante adecuado ____
- Adecuado ____
- Poco adecuado ____
- Inadecuado ____
- ¿Por qué?

5. ¿Cómo considera las funciones de semejanza empleadas?

- Muy adecuado ____
- Bastante adecuado ____
- Adecuado ____
- Poco adecuado ____
- Inadecuado ____
- ¿Por qué?

6. ¿Cómo considera el algoritmo de recuperación propuesto?

- Muy adecuado ____
- Bastante adecuado ____
- Adecuado ____
- Poco adecuado ____
- Inadecuado ____
- ¿Por qué?

7. ¿Cómo considera el algoritmo de reutilización propuesto?

- Muy adecuado ____
- Bastante adecuado ____
- Adecuado ____
- Poco adecuado ____
- Inadecuado ____
- ¿Por qué?

8. Elabore un comentario general sobre el modelo propuesto que aporte a la mejora del mismo.
