

Universidad de las Ciencias Informáticas
Facultad 5
Centro de Consultoría y Desarrollo de Arquitecturas Empresariales
Laboratorio de Gestión de Proyectos



ALGORITMO BASADO EN CASOS PARA EVALUAR COMPETENCIAS PROFESIONALES

Trabajo final presentado en opción al título de
Máster en Gestión de Proyectos Informáticos

Autor:

Ing. Aymé Perdomo Alonso

Tutores:

MSc. Surayne Torres López

Dr.C Pedro Yobanis Piñero Pérez

La Habana, diciembre del 2013

AGRADECIMIENTOS

A mi familia, en especial a mis padres y hermana por su apoyo incondicional a lo largo de mi carrera.

A Ernesto, por su paciencia y por ser el motor que me impulsa a ser cada día mejor.

A mis tutores por su apoyo y guía a lo largo de este trabajo.

Al Dr. Maikel Yelandi y su esposa Karina por las sabias sugerencias realizadas.

A mis compañeros del CDAE.

A todos los que de una forma u otra han contribuido a mi crecimiento profesional.

DECLARACIÓN JURADA DE AUTORÍA

Declaro por este medio que yo Aymé Perdomo Alonso, con carné de identidad 85101219377, soy la autora principal del trabajo final de maestría ALGORITMO BASADO EN CASOS PARA EVALUAR COMPETENCIAS PROFESIONALES, desarrollada como parte de la Maestría en Gestión de Proyectos Informáticos y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en La Habana a los días ____ del mes de diciembre del año 2013.

Ing. Aymé Perdomo Alonso
(Autora)

MSc. Surayne Torres López
(Tutora)

Dr. Pedro Yobanis Piñero Pérez
(Tutor)

RESUMEN

El razonamiento basado en caso constituye una técnica que ha recibido una creciente atención debido a sus potencialidades para la resolución de problemas basado en ejemplos almacenados en una memoria de casos. Sin embargo persisten un conjunto de dificultades que limitan su utilización en la toma de decisiones, relacionadas fundamentalmente con la organización de la memoria de los casos. En el presente trabajo se expone un algoritmo que permite el aprovechamiento de las experiencias previas en la organización para evaluar competencias profesionales mediante el uso del modelo lingüístico basado en 2-tuplas combinada con elementos básicos de la teoría de los operadores de agregación. Su implementación posibilita aumentar la aplicabilidad y fiabilidad de esta técnica en la toma de decisiones. Para corroborar la pertinencia y demostrar su efectividad se aplica la solución propuesta en el Paquete de Herramientas para la Gestión de Proyectos GESPRO 13.05, se desarrolla un cuasiexperimento, se aplican test estadísticos y se obtiene retroalimentación de los usuarios evidenciándose la correspondencia entre el objetivo y los resultados obtenidos.

Palabras Clave: algoritmo, competencias profesionales, modelo lingüístico basado en 2-tuplas, razonamiento basado en casos, operadores de agregación.

ABSTRACT

Case-based Reasoning is a technique that has received increasing attention due their potential to solve problems based on samples stored in a memory of cases. However persist a set of difficulties that limit their use in decision making, related to the organization of memory cases. In this work is exposed an algorithm that allows the use of previous experiences in the organization to assess competencies professional by means of a linguistic model based on 2-tuple combined with basic elements of the theory of aggregation operators. Its implementation allows to increase the applicability and reliability of this technique in decision making. To corroborate the relevance and demonstrate its effectiveness in the proposal is applied solution in ToolsPack for Project Management GESPRO 13.05, is develops a quasi-experiment, apply statistical tests and user feedback is obtained evidencing correspondence between the objective and results obtained.

Keywords: aggregation operators, algorithm, case-based reasoning, competencies professional, linguistic model based on 2-tuple.

ÍNDICE GENERAL

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN.....	9
1.1 Análisis bibliométrico.....	9
1.2 La evaluación de competencias profesionales	10
1.2.1 Análisis de las tendencias del proceso de evaluación de competencias profesionales	11
1.2.2 Diagnóstico del proceso de evaluación de competencias en la UCI.....	13
1.3 Sistemas Basados en Casos.....	15
1.3.1 Componentes de un Sistema Basado en Casos	16
1.3.2 Problemas actuales de los Sistemas Basados en Casos.....	19
1.4 La toma de decisiones.....	21
1.5 La incertidumbre y su tratamiento en la toma de decisiones	22
1.5.1 Agregación de información	25
1.6 Análisis de las herramientas de gestión de proyectos existentes.....	26
Conclusiones del capítulo	27
CAPÍTULO 2: ALGORITMO BASADO EN CASOS PARA EVALUAR COMPETENCIAS	28
1.1 Estructura general de la propuesta	28
1.2 Estructura general del algoritmo.....	29
1.2.1 Elementos principales del algoritmo.....	30
1.2.2 Objetivo y función principal del algoritmo.....	30
1.2.3 Precondiciones.....	30
1.2.4 Entradas y salidas del algoritmo	30
1.3 Descripción formal del algoritmo.....	31
1.3.1 Paso 1. Construir la base de casos	31
1.3.2 Paso 2. Recuperar los casos más semejantes.....	37
1.3.3 Paso 3. Recomendar la solución	42
1.3.4 Paso 4. Verificar y validar la solución	43
1.3.5 Paso 5. Retener la experiencia en la base de caso.....	44
Conclusiones del capítulo	45
CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA	46
3.1 Solución propuesta.....	46
3.2 Validación del algoritmo	48
3.3 Cuasiexperimento para validar el algoritmo	49
3.3.1 Descripción del cuasiexperimento	49
3.3.2 Ejecución del cuasiexperimento	50

3.3.3 Análisis de los resultados	51
3.4 Resultados experimentales en la recuperación de los casos	51
3.5 Validación estática.....	55
3.6 Validación de la variable dependiente eficiencia en el proceso de evaluación	58
3.7.1 Comparación de funcionalidades entre versiones de GESPRO	59
3.7 Análisis económico.....	60
3.8 Análisis social	62
Conclusiones del capítulo	63
CONCLUSIONES.....	64
RECOMENDACIONES	65
BIBLIOGRAFÍA.....	66
ANEXOS.....	72

INTRODUCCIÓN

Antecedentes y situación problemática

La industria del software y los servicios informáticos (ISWSI) basa sus producciones en el empleo de los conocimientos de los recursos humanos que la componen sobre las tecnologías existentes. El perfeccionamiento vertiginoso de esta industria ha alcanzado un auge significativo en las últimas décadas. Cada día los proyectos están expuestos a mayores restricciones en cuanto a tiempo, costo, calidad y riesgos, elementos que han permitido una evolución acelerada de las prácticas de gestión de proyectos debido a la creciente complejidad de los sistemas de información y el contexto económico actual.

El estudio *The CHAOS Report* más referenciado sobre el éxito y fracaso de los proyectos es el que realiza la consultora Standish Group (Group 2009), del cual se tiene evidencias desde el año 2004 con discreta disminución, muestra cifras que oscilan entre el 44% y el 68% del total de proyectos son cancelados y/o fracasados. En este sentido varios autores (André 2009; Charette 2005; Piorun 2003; Ryan 2007) concluyen que entre las principales causas que conllevan a estos resultados se encuentran los problemas asociados con factores humanos que constituyen uno de los de mayor incidencia en el resultado exitoso de los proyectos. En (Pressman 2004) se establece como elementos clave para lograr una eficaz gestión de proyectos de software a las personas, el producto, el proceso y el proyecto, siendo este orden intencional, demostrando la importancia de los recursos humanos en el desarrollo de software. En los últimos años se ha empezado a considerar a los recursos humanos como un recurso estratégico, por lo que una buena gestión persigue crear valor para la empresa, más que reducir costes (Alles 2000; Mondy and Noe 2005).

Todo lo relacionado con la disciplina *peopleware*¹ influye fuertemente en la productividad (DeMarco and Lister 1999). El conocimiento y la experiencia ya no son, por sí solos, elementos lo suficientemente diferenciadores para crear ventaja competitiva y añadir valor a la empresa, sino que también se debe tener en cuenta la motivación, el compromiso, la conducta de las personas (Bohlander et al. 2003). Es por ello que se plantea el reto de obtener una perfecta adecuación entre el trabajador y el puesto de trabajo de manera que se logre un desempeño excelente y no meramente satisfactorio de las tareas.

La Gestión de Recursos Humanos (GRH) contiene un grupo de procesos que van desde su selección, formación, hasta la evaluación y remuneración (Chiavenato 2002; Curtis et al. 2009; PMI 2008). Ha sido una tendencia en los últimos tiempos la gestión de los recursos humanos por competencias, ya que facilita el uso de un lenguaje común, permite conocer comportamientos futuros de las personas, facilita la formación y evaluación (Cuesta 2001; Cuesta 2010a; Cuesta 2010b; Hayes et al. 2000).

No son pocas ya las organizaciones que están adoptando prácticas y sistemas de recursos humanos basados en competencias. Esta orientación viene provocada por la necesidad de encontrar soluciones a los problemas de gestión y motivación de su personal en los nuevos entornos laborales. El desarrollo de software es un proceso que involucra a varias personas y en el cual es significativo la conformación del equipo, el

¹ Disciplina que se refiere al papel de las personas en el desarrollo o uso de software y sistemas hardware

establecimiento de roles dentro de los equipos de trabajo, la conceptualización de estos roles, la definición de las competencias para el desempeño de cada uno de ellos, y su evaluación (André 2009; Verdecia 2011). La evaluación es la herramienta con que cuentan los procesos de GRH para determinar el nivel de las competencias de las personas que integran los equipos de proyectos y planificar futuras acciones para mejorar su desempeño. A medida que se implementan esfuerzos de desarrollo se necesita conocer si las actividades y estrategias implementadas han sido efectivas. Es aquí donde la evaluación se convierte en una actividad esencial para los procesos de desarrollo y gestión de los equipos de proyectos informáticos (PMI 2008).

El proceso de evaluación puede ser llevado a cabo usando diferentes métodos. Varios autores (Espinilla et al. 2010; Herrera et al. 2004; Jiménez et al. 2003) plantean que el uso de enfoques de decisión puede ser satisfactoriamente aplicado para resolver problemas de evaluación.

La evaluación de desempeño es una actividad habitual en las organizaciones, siendo las competencias profesionales un área de información evaluable. Actualmente los procesos de evaluación del desempeño se realizan mediante métodos no formales, lo que provoca falta de precisión y objetividad en los resultados. Asimismo, en los sistemas tradicionales los evaluadores se ven forzados a emitir valoraciones numéricas sobre los atributos a evaluar, aunque dichos atributos sean difícilmente cuantificables debido a su naturaleza cualitativa (De Andrés 2009).

En el proceso de evaluación de competencias profesionales, la incertidumbre está caracterizada por la presencia definitiva e importante del juicio humano en la decisión. La teoría clásica de la decisión proporciona modelos probabilísticos para manejar la incertidumbre, sin embargo es fácil observar que en todos los problemas este modelo es adecuado (Martínez 2010).

Para la solución de problemas de evaluación de competencias profesionales en la actualidad se atiende el empleo de técnicas novedosas en su solución como la minería de datos, la inteligencia artificial (IA), la lógica difusa, la computación con palabras (del inglés, *Computing with Words: CWW*) y en general las técnicas de *soft computing*, específicamente en los procesos de evaluación del personal se pueden encontrar trabajos (Andrés et al. 2010; Fazhu et al. 2010; Jing 2009; Li et al. 2012; Xiaofan and Fengbin 2010) que reafirman esta tendencia, aplicando la modelación lingüística.

En estos procesos la información dada por los expertos se modela con palabras, pues es el lenguaje común que ellos utilizan. El uso de CWW (Martínez 2010; Zadeh 1998) surge de la necesidad de operar con términos lingüísticos. El empleo de técnicas que permitan esta operatoria facilita el tratamiento de la información cualitativa y de la incertidumbre, manteniendo en la medida de lo posible, la naturaleza de la información que se maneja.

Existen varios modelos para implementar la CWW, y en los últimos años se ha convertido en un tópico donde se investiga activamente, así se demuestra en (Jin-Hsien and Jongyun 2006; Martínez 2012; Martínez 2010; Mendel 2010; Safarzadegan 2011). La relación de la CWW con la toma de decisión lingüística, a partir de que

proporciona herramientas más cercanas a los procesos del razonamiento humano, es una de las áreas más tratadas en la actualidad (Martínez 2010).

Con la intención de superar los problemas típicos de la construcción de sistemas expertos, (Kolodner 1993) propuso un modelo revolucionario: un sistema de razonamiento basado en casos (SRBC) que utiliza como base una forma de razonamiento humano. Desde entonces, los SRBC se han aplicado con éxito a un amplio espectro de problemas. La aplicabilidad de esta técnica como sistema para la toma de decisiones ha sido analizada por diferentes autores (Bridge et al. 2005; García 1997; Gutiérrez 2003; Van Setten et al. 2004).

La idea básica de un SRBC es recuperar, adaptar y validar las soluciones encontradas en experiencias previas en un intento de relacionarlas con un problema actual. Las experiencias previas están representadas en una base de caso (BC) que reside en memoria, que consiste en cierto número de problemas junto a sus correspondientes soluciones. Cuando se enfrenta con un nuevo problema, el sistema recupera un caso similar, y la solución del caso se adapta al nuevo problema en un intento para resolverlo.

Diversos autores (Alles 2000; Ayala 2004; Chiavenato 2007; Cuesta 2005a; Morales 2009; Pérez 2010; PMI 2008; Rodríguez 2012; Torres 2011) han definido modelos teóricos para la GRH que plantean qué hacer, pero no cuentan con un respaldo tecnológico que permita aplicar el modelo utilizando las técnicas de *soft computing* existentes de manera que puedan nutrirse de las experiencias previas de los procesos de evaluación y evalúen las competencias de los recursos humanos durante el desarrollo de los proyectos de software sin centrar la evaluación a través de expertos para no interferir en la eficiencia del proceso.

Actualmente existen un gran número de herramientas que facilitan y automatizan los procesos de gestión de proyectos de software (Stang and Handler 2013). En su mayoría tratan temas de gestión de las tareas y el personal, pero no soportan los aspectos referidos a la evaluación de las competencias profesionales. Cubren en el mejor de los casos solo los procesos de planificación y adquisición, centrándose en la asignación del personal a las tareas o roles, la conformación de equipos y a la evaluación de las tareas.

Nuestro país demanda la urgencia de contar con mecanismo eficiente de control en el empleo del personal, los recursos materiales y financieros que permitan contribuir a la toma de decisiones, un salto cualitativo en este sentido se evidencia en los Lineamientos de la Política Económica y Social del Partido Comunista de Cuba (PCC 2011).

El uso de herramientas para la gestión de proyectos en la Universidad de las Ciencias Informáticas (UCI), lugar donde se enmarca la presente investigación, se ha ido incrementando con el objetivo de facilitar el trabajo de los especialistas en los proyectos de software. En este centro de altos estudios convergen en la ejecución de los proyectos; especialistas, estudiantes y profesores en una población de más de 5000 usuarios que contribuyen al desarrollo de los mismos; cifra que en diciembre 2012, según el Diagnóstico aplicado por la entidad Calisoft a la actividad productiva alcanzó un total de 145 proyectos, de ellos 32 de servicios y 113 de desarrollo (Calisoft 2012).

Con este fin existe una red de 14 centros de producción dónde se organizan los recursos humanos, que colaboran y desarrollan soluciones informáticas bajo una estrategia única de gestión de proyectos. A partir del año 2010 comienza a utilizarse la herramienta para la Gestión de Proyectos (GESPRO); actualmente registrada con la versión 12.05. Su desarrollo se sustenta por el estándar propuesto por el Project Management Institute (PMI 2008) y las buenas prácticas del programa de formación del Máster en Gestión de Proyectos de la UCI (Piñero et al. 2013).

La herramienta GESPRO 12.05 posee un módulo para la GRH que incluye procesos para realizar un uso más eficiente de las personas involucradas con el proyecto, el cual tiene como fin la planificación organizacional, la adquisición del personal, la evaluación y el desarrollo del equipo. Como parte de las funcionalidades para la toma de decisiones se incluyen más de 60 reportes, permite el trabajo con indicadores y la toma de decisiones a tres niveles: proyecto, entidad ejecutora y corporación (Piñero, et al. 2013).

Teniendo en cuenta que el proceso de evaluación es continuo y periódico se tienen almacenadas un gran cúmulo de experiencias previas dentro de la organización. Existen casos de personas con sus competencias ya evaluadas pero dicha información no se usa y explota adecuadamente para la evaluación de los nuevos casos. Esta situación provoca que ante la pérdida de expertos por envejecimiento o migración de puestos de trabajo se pierda el conocimiento. Además provoca deficiencias a la hora de definir una estrategia óptima para elevar el nivel profesional. A su vez dificulta la toma de decisiones respecto a la ejecución de actividades de formación debido a su rentabilidad y aplicabilidad.

En general el proceso de evaluación de competencias actual se torna:

- Engorroso por su complejidad de aplicación y la cantidad de personas que intervienen.
- Se manifiesta un alto nivel de subjetividad por la intervención de expertos humanos, resultando difíciles de evaluar de forma precisa mediante escalas numéricas.
- Falta de aprovechamiento de las experiencias en los procesos de evaluación anteriores.
- Se necesitan modelos de evaluación que sean capaces de ofrecer resultados entendibles e interpretables.

Problema científico

Las insuficiencias en los métodos y técnicas para la evaluación de competencias profesionales respecto al aprovechamiento de las experiencias previas está afectando la eficiencia del proceso.

Objeto de investigación

Evaluación de competencias profesionales.

Objetivo general

Desarrollar un algoritmo basado en casos para la evaluación de competencias profesionales que permita mejorar la eficiencia del proceso de evaluación.

Objetivos específicos

- Caracterizar el proceso de evaluación de competencias profesionales teniendo en cuenta la utilización de las experiencias previas.
- Desarrollar un algoritmo utilizando técnicas de razonamiento basado en caso para la evaluación de competencias profesionales.
- Validar a partir de los métodos científicos definidos en la investigación el algoritmo propuesto.

Campo de acción

Algoritmo basado en casos para la evaluación de competencias profesionales.

Hipótesis

Si se desarrolla un algoritmo basado en casos para la evaluación de competencias profesionales que incluya una estructura de indexación no secuencial de la base de casos y el uso del método lingüístico 2-tuplas para agregar información cualitativa, entonces se contribuirá a la mejora de la eficiencia del proceso de evaluación.

Operacionalización de las variables dependientes e independientes

Variable independiente: algoritmo basado en casos para la evaluación de competencias profesionales.

Variable dependiente: eficiencia del proceso de evaluación.

Variable independiente	Dimensión	Indicadores	Unidades de medida
Algoritmo basado en casos para la evaluación de competencias profesionales	Fiabilidad	Capacidad para recomendar valores cercanos a la evaluación del decisor	Alta, Media, Baja
	Sistema basado en casos	Capacidad de localizar soluciones relevantes al problema actual	Alta, Media, Baja
Variable dependiente	Dimensión	Indicadores	Unidades de medida
Eficiencia del proceso de evaluación	Método de evaluación de competencias	Mecanismo de recolección de datos	Automático, Semiautomático, Manual
		Utiliza indicadores	Sí, No
		Tipo de procesamiento	Automático, Semiautomático, Manual
	Tiempo para evaluar las competencias profesionales	Horas, Minutos, Segundos	
	Herramientas de evaluación	Capacidad de automatizar el trabajo de los evaluadores	Sí, No
		Cantidad de funcionalidades implementadas para el desarrollo de los recursos humanos	Porcentaje

Tipo de investigación

Correlacional: Pretenden saber cómo se puede comportar un concepto o variable conociendo el comportamiento de otros u otras. Miden el grado de relación que existe entre dos o más conceptos o variables.

Métodos de investigación

Métodos teóricos

- Análisis - Síntesis: Permite llegar a conclusiones a partir del estudio realizado sobre los procesos de evaluación de competencias profesionales.
- Inductivo - Deductivo: Para realizar el tránsito de lo general a lo particular y viceversa al establecer los nexos entre las técnicas para el modelado lingüístico, los sistemas basados en casos y las evidencias de las competencias evaluadas. Este método fue una importante vía científica para arribar de los análisis particulares a las generalizaciones.
- Histórico - Lógico: Para conocer el comportamiento y evolución de las diferentes posiciones respecto al estudio de los procesos de evaluación de competencias profesionales de los recursos humanos, los sistemas basados en caso, así como de la integración de estos dos elementos en las herramientas de gestión de proyectos existentes.
- Modelación: Contribuye al desarrollo del algoritmo y su implementación.

Métodos empíricos

- Medición: Permite medir los resultados de la comparación del algoritmo propuesto con otros métodos existentes para analizar la eficiencia del proceso de evaluación.
- La técnica de **ladov** para validar y obtener retroalimentación de los usuarios sobre el nivel de satisfacción lograda por la propuesta.
- Experimental: Se utiliza con datos provenientes de casos reales. Se aplican pruebas estadísticas debidamente fundamentadas para analizar los resultados. Se establecen indicadores adecuados que permiten realizar correctas mediciones de los resultados.

Muestreo

Población

Las evidencias de los recursos humanos evaluados en la herramienta GESPRO 12.05 (164 casos personas con sus competencias evaluadas de un centro de desarrollo)

Muestra

El 100% de las evidencias de los recursos humanos evaluados, correspondiente al centro de desarrollo (método no probabilístico intencional).

Diseño de experimentos

Para demostrar la efectividad del algoritmo se realizará un diseño *cuasiexperimental* con post prueba y grupo de control.

Se escoge este tipo de experimento fundamentado en que no es posible realizar un experimento puro, los cuasiexperimentos se diferencian de los experimentos en que la asignación de participantes a los grupos no se hace en forma aleatoria, ni por emparejamiento. Ocurre cuando los grupos están previamente confeccionados (grupos intactos) (Grau 2004).

G1	X	O1
G2	---	O2a

Descripción de las variables

G1: Grupo de experimentación compuesto por las evidencias de las competencias evaluadas de los recursos humanos de un centro de desarrollo.

X: Aplicación de la propuesta al grupo de experimentación a través de la aplicación del algoritmo.

O1: Observación de la efectividad de aplicar el algoritmo propuesto.

G2: Grupo de control compuesto por las evidencias de las competencias de los recursos humanos de un centro de desarrollo.

O2a: Observación después de evaluar las competencias de los recursos humanos por grupos de expertos decisores.

Análisis estadístico a realizar

En los experimentos para la comparación de las muestras se utilizará Kruskal-Wallis Test, aplicando una simulación de Monte Carlo con un intervalo de confianza del 99%. (SPSS-Inc 2004).

Instrumentos

Herramienta pgAdminIII versión 1.14.0 (PGDG 2013) y R versión 2.15.3 (RDCT 2013), matriz de comparación de funcionalidades para herramientas de gestión de proyectos y encuestas para la medición de la fiabilidad de la propuesta.

Novedad

Algoritmo para la evaluación de competencias profesionales utilizando técnicas de *soft computing*.

Aporte práctico de la investigación.

- Algoritmo desarrollado en lenguaje PL/R para el cálculo de la función de semejanza.
- Algoritmo desarrollado en lenguaje PL/R para la evaluación de las competencias profesionales de los recursos humanos.
- Diseño de un sistema basado en casos para la evaluación de competencias profesionales utilizando una estructura de árbol binario.

Listado de publicaciones, eventos y avales de la investigación

1. PERDOMO, A. y MENDOZA, M. Registro de Partos y Nacimientos. En *IV Taller hacia la informatización de los Servicios de Salud. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2008. La Habana. 2008.*
2. PERDOMO, A. y JORDÁN, E. Selección de una herramienta de Gestión de Procesos de Negocios, haciendo uso de la técnica multicriterio AHP. En *Novena Semana Tecnológica de Fordes. La Habana. 2009.*
3. PERDOMO, A.; JORDÁN, E., *et al.* Selection of Business Process Management tools based on Analytic Hierarchy Process approach. En *International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE 09). 2009.*
4. PERDOMO, A. y MENDOZA, M. Registro de Partos y Nacimientos para el Sistema Nacional de Salud. En *VII Congreso Internacional de Informática en la salud. Informática Habana 2009. 2009.*

5. PERDOMO, A. Métodos de evaluación de arquitecturas de software. Dificultades y soluciones. En *V Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2010. La Habana. 2010.*
6. PERDOMO, A. y CASTELL, J. Experiencias de una iniciativa orientada a procesos para el Centro de Inmunología Molecular. En *Décima Semana Tecnológica de Fordes. La Habana. 2010.*
7. PERDOMO, A. y SÓÑORA, W. ¿Es la Arquitectura de Negocio necesaria para SOA? En *V Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2010. La Habana. 2010.*
8. SÓÑORA, W.; PERDOMO, A., *et al.* La Arquitectura Empresarial: sinergia entre el negocio y la tecnología. En *V Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2010. La Habana. 2010.*
9. PERDOMO, A.; CASTELL, J., *et al.* Desarrollo de un Proyecto Piloto basado en SOA y BPM para las operaciones industriales biotecnológicas. En *V Congreso Latinoamericano de Ingeniería Biomédica. La Habana. 2011.*
10. PERDOMO, A. y PÉREZ, Y. Caso de Estudio: Aplicación práctica de la técnica de descomposición de procesos propuesta por IBM. En *VI Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2012. La Habana. 2012.*
11. PERDOMO, A. Métodos de evaluación de arquitecturas de software. Dificultades y soluciones. In *V Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2010. La Habana, 2010.*
12. PERDOMO, A. y PÉREZ, Y. Procedimiento para la captura de requerimientos del negocio para iniciativas orientadas a servicios. En *VI Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2012. La Habana. 2012.*
13. PERDOMO, A. y PÉREZ, Y. ¿Cómo identificar servicios candidatos? En *VI Taller de Calidad, Ingeniería, Arquitectura y Gestión de Software. Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA 2012. La Habana. 2012.*

Estructura de la tesis

La tesis consta de introducción, tres capítulos, conclusiones, recomendaciones, bibliografía y un cuerpo de anexos, que permite abundar en los temas tratados en la misma.

En el capítulo uno se presentan los resultados del estudio para establecer el marco teórico referencial de la investigación, en él se incluyen: proceso de evaluación de competencias, sistemas basados en caso para la toma de decisiones, modelado lingüístico y se realiza un análisis comparativo de las herramientas de gestión de proyectos existentes.

En el capítulo dos se presenta el algoritmo propuesto, sus características y estructura general. Además se esbozan los pasos lógicos del mismo: construcción de la BC, recuperación de los casos más semejantes, recomendación de la evaluación, validar la solución y retener la experiencia. Los algoritmos se describen utilizando un lenguaje natural.

En el capítulo tres se exponen los resultados alcanzados de la validación del algoritmo en GESPRO 13.05. Se valida la eficiencia y efectividad de algoritmo, a partir de la aplicación de un diseño *cuasiexperimental*, la aplicación de test estadísticos, la comparación con otros modelos y métodos de evaluación de competencias y la retroalimentación de los usuarios apoyado por la técnica de ladov.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN

En el presente capítulo se realiza un estudio de las tendencias en el proceso de evaluación de competencias profesionales. Se aborda qué se entiende por sistema basado en casos, cuáles son los aspectos que hay que tener en cuenta para desarrollarlos y las etapas por las que hay que transitar para su construcción. Se analizan los problemas actuales existentes en estos sistemas relacionados con la representación del conocimiento y métodos de recuperación de los casos más semejantes. Se aborda el proceso de evaluación como un problema de la toma de decisiones y se plantea como solución la modelación lingüística o la computación con palabras para el manejo de la incertidumbre. Se concluye con un análisis comparativo de las herramientas de gestión de proyectos existentes.

1.1 Análisis bibliométrico

En la Tabla 1.1 se esboza el análisis bibliométrico relacionado con el tema de investigación. Para la búsqueda de información se consultaron las bases de datos de la IEEE² y SCOPUS³, se utilizaron las herramientas Scirus⁴ y Google Scholar.

Tabla 1.1. Resumen cuantitativo de la bibliografía consultada.

Clasificación	Últimos 5 años	Años anteriores
Libros y monografías	9	21
Tesis de doctorados	5	2
Tesis de maestrías	5	-
Artículos en Revistas referenciadas en Web of Science, SCOPUS	20	31
Memorias de eventos	5	-
Artículos publicados en la web	14	32
Reportes técnicos y conferencias	5	2
Entrevistas personales	1	-

De la bibliografía consultada se puede concluir que el tema ha sido abordado en investigaciones de maestrías y doctorado dentro del período de los últimos cinco años, demostrando la pertinencia del mismo. Los criterios de búsqueda utilizados fueron “evaluación de competencias profesionales”, “razonamiento basado en casos”, “herramientas de gestión de proyectos”, “lenguaje R”, computación con palabras”. La mayoría de la bibliografía referenciada y consultada fueron artículos en revistas referenciadas representando el 34% del total. Nótese que más del 40% se encuentra en dentro de los últimos 5 años.

² Es una base de datos en línea producida y distribuida por el Institute of Electrical and Electronics Engineers, IEEE.

³ Es una base de datos bibliográfica de resúmenes y citas de artículos de revistas científicas. Cubre aproximadamente 18.000 títulos de más de 5.000 editores internacionales, incluyendo la cobertura de 16.500 revistas revisadas por pares de las áreas de ciencias, tecnología, medicina y ciencias sociales, incluyendo artes y humanidades. Está editada por Elsevier.

⁴ Es un motor de búsqueda específico de contenido científico, fue el primer buscador de su tipo, creado por la editorial Holandesa Elsevier.

1.2 La evaluación de competencias profesionales

La GRH en una organización es el área que define los procesos y actividades relacionados con los recursos humanos (Chiavenato 2002; Curtis, et al. 2009; PMI 2008). Las actividades que se definen como parte de este proceso, aunque con diferentes nombres en la literatura (Biles et. al. 1986; Chiavenato 2002; Cuesta 2005b; Mondy and Mondy 2009; Pino et al 2008; Puchol 2007; Rul et al 1996; Wayne et al 2009; Williamson et al. 2008), son de manera general: reclutamiento y selección, formación, evaluación y remuneración.

El desarrollo de los recursos humanos es uno de los procesos que se plantea realizar dentro de la GRH, intrínsecamente se evalúa el desempeño del personal (Alles 2005; Chiavenato 2007; Cuesta 2005a; Cuesta 2012; PMI 2008). Es el área que retroalimenta a los demás los procesos y actividades relacionados con la gestión del personal en una organización.

Generalmente, en los procesos de evaluación del desempeño se consideran tres grandes categorías de información evaluables: información basada en rasgos, información basada en el comportamiento que incluye (los comportamientos, las competencias y el potencial de mejora) y la información basada en los resultados obtenidos (De Andrés 2009).

Diferentes autores (Alles 2005; André 2009; Cuesta 2010b; Tuning 2007) han aludido el tema de Gestión por Competencias (GPC), confiriéndole un gran valor para el logro de las aspiraciones empresariales. La (GPC) se perfila como una herramienta crucial en la medida en que supone una gestión del personal más ajustada a las características claves que posee el capital humano y a aquellas que hacen, que un determinado puesto de trabajo se desarrolle más eficazmente. Las competencias son evaluadas a lo largo de la vida laboral del sujeto, como medio para valorar su desempeño en el puesto de trabajo o sus posibilidades de desarrollo. La norma cubana (NC) define como competencia al conjunto sinérgico de conocimientos, habilidades, experiencias, sentimientos, actitudes, motivaciones, características personales y valores, basado en la idoneidad demostrada, asociada a un desempeño superior del trabajador y de la organización en correspondencia con las exigencias técnicas, productivas y de servicios. Es requerimiento esencial que estas competencias sean observables, medibles y que contribuyan al logro de los objetivos de la organización (OCN 2007).

Se asumen en la presente investigación el concepto planteado por la NC, ya que está en correspondencia con los procedimientos de gestión organizacional propios del país. Teniendo en cuenta que las competencias podrán ser evaluadas independientemente de su clasificación sean genéricas o técnicas mientras cumplan con los requerimientos de la norma.

Los sistemas de evaluación del desempeño basados en competencias además de evaluar en función de los estándares de actuación en el trabajo y de los resultados más tradicionales, agregan las conductas de trabajo necesarias para realizar tareas del puesto de trabajo. Las personas serán seleccionadas, capacitadas, desarrolladas y evaluadas en función de dichos comportamientos que estarán asociados a un desempeño exitoso en el trabajo y permitirán a la organización alcanzar sus objetivos.

Las competencias son la expresión del accionar de una persona, el desempeño permite ofrecer una valoración de la competencia del recurso humano, es por ello que en su propia definición se evidencia la incertidumbre que lo rodea. Las valoraciones de las competencias expresadas por los expertos y evaluadores de las personas se hacen mayormente con palabras, no expresando números o cifras exactas. Estas afirmaciones se evidencian en investigaciones que asocian la GRH, la competencia y la incertidumbre (Canós et al 2008; Canós et. al. 2007; Safarzaghan 2011).

1.2.1 Análisis de las tendencias del proceso de evaluación de competencias profesionales

En (Chiavenato 2007) se plantea que la evaluación es el elemento integrador de las prácticas de la GRH ya que sirve de enlace a los demás. Afirma además que una de las principales tendencias en la evaluación del desempeño son los indicadores para establecer una visión global necesaria.

Según los autores (Ayala 2004; Chiavenato 2007; De Andrés 2009; Fisher et al. 2006) los procesos de evaluación tiene dentro de principales funciones: la planificación de los recursos humanos, la selección de personal, la capacitación y desarrollo, la planificación y desarrollo de la carrera profesional, los programas de compensación, las relaciones internas con los empleados y la evaluación del potencial de los empleados. La evaluación del desempeño mediante competencias consiste en estimar aproximadamente el grado de eficacia y eficiencia con el que los trabajadores realizan sus actividades, cumplen los objetivos y son responsables de su puesto de trabajo, es decir, estiman el rendimiento global del empleado. Estos aspectos quedan recogidos en (Armstrong 2003; Chiavenato 2007; Cuesta 2012; Murphy and Cleveland 1991)

Los procesos de evaluación del desempeño se pueden realizar siguiendo diferentes metodologías, aunque podemos clasificarlos en dos grandes categorías en función de la información que utilicen: métodos de evaluación basados en información objetiva o métodos de evaluación basados en información subjetiva (Chiavenato 2007; Cuesta 2005a; De Andrés 2009; Fisher, et al. 2006).

Los métodos de evaluación del desempeño existentes, presentan una serie de problemas comunes, que son la base de la mayoría de las críticas realizadas a los procesos de evaluación del desempeño (Cuesta 2005a; De Andrés 2009; Fisher, et al. 2006; Mondy and Noe 2005), entre otros:

- La subjetividad aparece en los métodos de evaluación que utilizan información no objetiva en el proceso.
- Uno de los problemas más típicos en los procesos de evaluación son los prejuicios, más conocidos como errores halo⁵.
- La indulgencia en la valoración de los empleados de manera positiva y alta, sin ser merecida por el trabajador.
- Los evaluadores clasifican el desempeño de todos los trabajadores como una tendencia central.
- Existencia de prejuicio de comportamiento reciente.
- Los resultados obtenidos en el proceso de evaluación puede ser fácilmente manipulado por los evaluadores.

⁵ Se refiere a un sesgo o prejuicio cognitivo por el cual la percepción de un rasgo particular es influenciada por la percepción de rasgos anteriores en una secuencia de interpretaciones.

Dentro de los métodos más comunes utilizados según datos de encuestas realizadas en Estados Unidos demuestran que más del 90% de las empresas que aparecen en la lista Fortune 1000 han desarrollado sistemas de evaluación integral o de 360 grados (De Andrés 2009), pero no es menos cierto que se requiere de capacitación para que el sistema funcione con eficacia y además de la participación de múltiples grupos de expertos.

Como tendencia en las últimas décadas se plantean soluciones al proceso de evaluación empleando técnicas de AI. Con el objetivo de cuantificar la imprecisión e incertidumbre se han integrado a técnicas de *soft computing*, aumentando la capacidad de procesar la información cualitativa y de diversa naturaleza que está involucrada en el proceso.

En (Chen and Tai 2005) se propone un modelo de evaluación para medir el nivel del capital intelectual de una organización basado en el modelo de representación lingüística difusa 2-tuplas. Presenta como punto débil el tiempo que requiere y la intervención de expertos durante todo el proceso para determinar la importancia de cada uno de los criterios y de cada elemento con respecto a cada criterio. Y luego, para determinar la calificación del desempeño de cada elemento de evaluación con respecto a cada criterio.

En (De Andrés 2009) se plantea un modelo lingüístico de evaluación del desempeño de 360 grados fundamentado en la teoría de la toma de decisiones, donde las opiniones de los evaluadores son expresadas mediante diversas escalas cualitativas, definidas a través de términos lingüísticos, los cuales permiten capturar la imprecisión y la incertidumbre con la que los evaluadores se enfrentan a la hora de emitir sus juicios basados en percepciones.

En (Wang and Jiang 2010) se propone un modelo de red neuronal de tipo propagación de errores hacia atrás para la evaluación de los recursos humanos, mediante 14 factores previamente definidos y evaluados por expertos humanos. Este método tiene su principal limitante en la técnica de AI empleada que funciona como una caja negra, por lo que no ofrece resultados que puedan ser interpretables para su discusión.

En (Ligen and Zhenlin 2010) se propone un marco para la evaluación de competencias básicas empresariales propuestas por (Prahalad and Hamel 1990) desde la perspectiva del capital intelectual mediante el uso de un modelo de evaluación basado en una medida de incertidumbre. Se apoya en el método multicriterio AHP (del inglés, *Analytic Hierarchy Process*) para la construcción de la matriz de comparación mediante el cálculo del valor de peso.

En (Mkaouer et al. 2010) se describe un marco de trabajo para la selección del mejor candidato para realizar una tarea determinada teniendo en cuenta la evaluación de sus competencias utilizando el modelo lingüístico difuso 2-tuplas. Para la selección del mejor candidato se hace uso del método multicriterio TOPSIS (del inglés, *Technique for Order Preference by Similarity to Ideal Solution*). Las evaluaciones lingüísticas son generadas por un grupo de supervisores teniendo en cuenta la competencia necesaria y los recursos humanos.

La minería de datos ha tenido un crecimiento acelerado en los últimos tiempos en el ámbito empresarial. En (Xiaofan and Fengbin 2010) se introducen soluciones basadas en árboles de decisión mediante la

identificación de los factores que afectan el rendimiento de los empleados. Como resultado del análisis se obtienen un conjunto de reglas de rendimiento que permite predecir el comportamiento del personal. En (Li, et al. 2012) se hace un análisis y se plantea el problema de la evaluación del personal a partir de la definición de atributos, utilizando para su representación en un árbol de decisión utilizado el algoritmo ID3 difuso. En (Espinilla et al. 2012) se plantea un modelo para la evaluación del desempeño aplicando el método de 360 grados. Permite manejar información heterogénea y la diversidad de criterios que intervienen, la evaluación es expresada en términos lingüísticos utilizando el método 2-tuplas.

La lógica difusa y el modelado lingüístico son los métodos más generalizados en los trabajos revisados, aunque mayormente se emplea en combinación con otras técnicas de la AI para lograr una solución más cercana a la realidad del problema. El empleo de métodos matemáticos se puede diferenciar de acuerdo al tratamiento de la incertidumbre que pueda existir, es por ello que entre los trabajos estudiados se encuentran algunos, aunque muy pocos, que emplean métodos determinísticos.

Los métodos analizados ofrecen mecanismos y técnicas a partir de las cuales se puede evaluar el personal atendiendo a criterios previamente definidos por expertos. La principal deficiencia es que se limitan solamente a utilizar la información reciente que se tiene del comportamiento de los empleados, y no así la histórica. Además la información evaluable no se basa en el comportamiento a través de la definición de perfiles de competencias o puestos de trabajo sino en factores que son definidos y evaluados por expertos. No cuentan con un soporte tecnológico de aplicación en las herramientas de gestión de proyectos existentes.

1.2.2 Diagnóstico del proceso de evaluación de competencias en la UCI

Para caracterizar el proceso se toma como referencia dos encuestas aplicadas por (Cuza 2013) y (Pérez 2010). El análisis de sus resultados muestra que:

- De los encuestados el 100% no conoce o aplica algún proceso de evaluación de competencias del personal desempeñándose en un determinado rol en su área.
- El 100% plantea que no se aplican de manera sistemática los procesos de evaluación de competencias, señalando como los principales factores que influyen en ello: no estar informatizado el proceso, la necesidad de manejar gran cantidad de información, la cantidad de personas a evaluar y el poco tiempo en que debe ser ejecutado el proceso.
- La mayoría de los encuestados están de acuerdo en la necesidad de definir un proceso de evaluación de competencias, poniendo como principales ventajas un mejor aprovechamiento de los recursos humanos y la identificación de necesidades de formación para mejorar su desempeño en los proyectos de desarrollo de software.
- El 100% de los encuestados no reconocen que el proceso de evaluación de competencias profesionales no es una técnica o método para identificar las necesidades de formación.
- El 95% identifica un alto grado de importancia para el proceso productivo conocer una valoración del estado de competencias del personal vinculado a la actividad productiva.

- El 72% de los encuestados expresaron que no tienen definidas las competencias que deben tener los individuos para desarrollar con eficiencia las tareas del rol que desempeñan. Se detectó además que el 88 % no aplican el enfoque de competencias profesionales a la hora de definir un plan de formación que permita alcanzar al personal el estado óptimo en cada competencia.

Varios autores(Álvarez 2012; Pérez 2010; Rodríguez 2012; Torres 2011) han desarrollado modelos para la GRH que permiten evaluar competencias genéricas y equipos de desarrollo de software. En la Tabla 1.2 se muestra la evaluación realizada.

Tabla 1.2. Evaluación de los modelos de GRH desarrollados en la UCI.

Modelos	Descripción
(Pérez 2010)	Se fundamenta sobre la base del método de 360 grados para la evaluación de 17 competencias genéricas. Tiene la particularidad que está enfocado a la evaluación de estudiantes durante su desempeño en el proyecto.
(Torres 2011)	Permite observar el comportamiento del equipo, gestiona los conflictos y evalúa su rendimiento. Las evaluaciones son expresadas de manera cualitativa y dependen del cumplimiento y la evaluación de las tareas asignadas.
(Álvarez 2012)	El proceso de evaluación se encuentra orientado a valorar el comportamiento y avance en el tiempo de las personas involucradas en el proyecto. Se centra fundamentalmente en la evaluación de las tareas utilizando el modelo para la evaluación por competencias propuesto por (Pérez 2010) lo que permitirá posteriormente evaluar al equipo de trabajo.
(Rodríguez 2012)	Permite evaluar la composición de un equipo cuyo tamaño no exceda los siete miembros independientemente del entorno productivo donde se desarrolle. A partir de las salidas del modelo es posible conocer el estado actual de la composición de un equipo en relación al balance de competencias genéricas y el balance de los roles de equipo.

Los evaluadores valoran de acuerdo a sus experiencias cómo los resultados de los individuos influyen en el desempeño en un rol determinado, teniendo en cuenta que para la evaluación, son tomados los reportes de evaluación de las tareas, los reportes de incidencias disciplinarias del período y la evaluación del desempeño. Una brecha no cubierta por ninguna de las investigaciones analizadas es que no cuentan con un soporte tecnológico de aplicación en las herramientas de gestión de proyectos existentes. Necesitan involucrar numerosos expertos para llevar a cabo un proceso eficaz, teniendo en cuenta que se requiere de un esfuerzo considerable en horas/hombres, situación que provoca que actualmente el uso de estos modelos y/o métodos se vean limitados para su aplicación. Además al no utilizarse las evidencias previas de los procesos de evaluación no existen referencias de uso de técnicas de razonamiento basado en casos.

1.3 Sistemas Basados en Casos

Los CBS (del inglés, *Case Based Systems*: CBS), tienen su origen en los Sistemas Basados en Conocimiento (del inglés, *Knowledge Based System*: KBS) para la toma de decisiones. Estos sistemas utilizan el razonamiento basado en casos (del inglés *Case Based Reasoning*: CBR) como método de solución de problemas (MSP) para resolver nuevas situaciones. A continuación se presenta una definición de sistema de razonamiento tomada de (Gutiérrez 2003):

Un CBS para la toma de decisiones es un sistema de decisión S que se define en términos de un par $(U, X \cup Y)$ donde U es un conjunto finito no vacío de objetos o eventos llamados casos, mientras X e Y son dos conjuntos finitos, no vacíos de atributos o propiedades llamados conjuntos de rasgos predictores y objetivos respectivamente. Su representación matemática se representa en (1.1).

$$S = (U, X \cup Y) \quad (1.1)$$

Cada rasgo predictor $x_i \in X$ puede ser considerado una función que mapea elementos de U en un conjunto $M_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$, el cual se denomina conjunto de valores del rasgo predictor x_i . Su representación matemática se representa en (1.2).

$$x_i: U \rightarrow M_i \quad (1.2)$$

De manera análoga cada rasgo objetivo $y_j \in Y$ puede ser considerado una función que mapea elementos de U en un conjunto $N_j = \{y_{j1}, y_{j2}, \dots, y_{jn}\}$, el cual se denomina conjunto de valores del rasgo predictor y_j . Su representación matemática se representa en (1.3).

$$y_j: U \rightarrow N_j \quad (1.3)$$

El CBR representa un 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. El mismo denota un método en el cual la solución de un nuevo problema se realiza a partir de las soluciones conocidas para un conjunto de problemas previamente resueltos (o no resueltos) del dominio de aplicación.

El CBR tiene como una de sus justificaciones la analogía con el mundo real, donde un experto tiene más conocimiento a medida que tiene más experiencia. En este sentido en el CBR se entiende el aprendizaje como un ciclo continuo de recordar y memorizar nuevas situaciones para ir ganando experiencia. En la Figura 1.1 se representa los cuatro pasos fundamentales del proceso cíclico de un CBR.

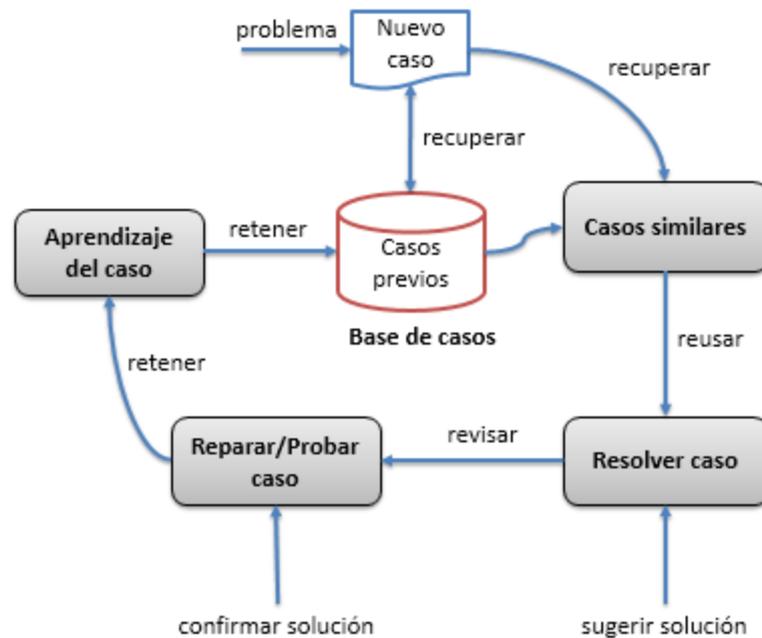


Figura 1.1. Ciclo del razonamiento basado en casos(Aamodt and Plaza 1994).

1.3.1 Componentes de un Sistema Basado en Casos

Los componentes fundamentales de un CBR son la BC, el módulo de recuperación de casos y el módulo de adaptación de las soluciones.

Base de casos

La BC contiene las experiencias, ejemplos o casos a partir de los cuales el sistema hace sus inferencias. Esta puede ser generada por entrevistas a expertos humanos o por un procedimiento automático o semiautomático que construye los casos desde datos existentes registrados, por ejemplo, en una base de datos.

No existe consenso sobre qué información exactamente se debe representar en un caso. De acuerdo con (Kolodner 1993) un caso es contextualizado como una pieza de conocimiento que representa una experiencia que enseña una lección fundamental para lograr los objetivos del razonador.

La representación de los casos en la base de conocimiento constituye el núcleo de un CBS del cual todos los demás procesos dependen. Maher identifica tres cuestiones principales (Maher and Balachandran 1994):

1. El contenido del diseño de los casos.
2. La organización en memoria de los casos.
3. La presentación de los casos diseñados al usuario.

Existe una gran variedad de formas de representación, sin embargo, los principales enfoques de representación de casos propuesto por (Bergmann et al. 2005) son: vector de rasgos, representación basada en frames, representación orientada a objetos, textual y jerárquica. En (Watson 1998) se propone catalogar a los sistemas de CBR por el tipo de tarea que desempeñan en: clasificación o síntesis.

La autora recomienda utilizar una representación del caso utilizando un vector de rasgos. Además se propone utilizar una forma de representación del conocimiento (FRC) utilizando una estructura de indexación

jerárquica de la BC mediante una representación de para la realización de tareas de clasificación de manera que facilite la recuperación eficiente y eficaz del proceso de síntesis y/o justificación. El diseño de los casos parte de una estructura multidimensional de datos, donde la cantidad de dimensiones dependerá de la cantidad de competencias que tenga asociados un individuo. Esta FRC permitirá optimizar el tiempo de la búsqueda de los casos semejantes, así como mejorar la eficiencia del proceso de recuperación.

Módulo de Recuperación

El proceso de recuperación comienza con una descripción del problema (parcial), y finaliza cuando se encuentra un caso anterior, que es el que mejor ajusta. Este se divide en varias etapas: identificación de características, emparejamiento inicial, búsqueda y selección.

Por lo general, el propio modelo de organización de la BC es quien facilita el acceso al conjunto de casos candidatos. La recuperación de los casos que mayor similitud guardan con el patrón de búsqueda, puede realizarse con la utilización de dos métodos: por semejanza parcial o por analogía. En el primer método, se emplea una función de semejanza la cual determina una medida numérica del grado de similitud de cada objeto de la memoria con respecto al nuevo problema. El segundo método busca un elemento de la memoria cuya descripción pueda hacer igual a la del nuevo problema mediante el reemplazamiento de los valores de los rasgos que tienen valores diferentes, a partir de la equivalencia entre valores definido en una red semántica.

La búsqueda en el CBR tiene dos aspectos principales: el acceso a un conjunto de elementos de memoria o casos que resulten promisorios para el problema a resolver y la recuperación de los elementos de memoria de este conjunto que sean más semejantes al problema. El método de acceso a los casos candidatos está muy vinculado al modelo de memoria empleado; por ejemplo, las técnicas de indexación permiten desarrollar un modelo de memoria y a la vez definen un método de acceso.

Para determinar qué tan similar es un caso a otro se han desarrollado varias técnicas. La más sencilla consiste en contar el número de características similares entre los dos casos. El problema de esta técnica es que la importancia de las características varía de un contexto a otro. Otra técnica consiste en utilizar un conjunto de heurísticas que permitan determinar cuáles características tienen mayor relevancia (peso) y formular una función de semejanza que involucre la similitud entre cada uno de los rasgos teniendo en cuenta el peso de los mismos (Gutiérrez et al. 2002).

Por lo general, en el proceso de recuperación de los casos más similares se emplea el enfoque conocido como "k vecino más cercano" (del inglés, *K-Nearest Neighbor: KNN*) (Cover and Hart 1967). Esta técnica necesita de una medida de similitud para determinar la proximidad entre casos, generalmente usando una función de distancia ponderada. Varios investigadores sugieren que el peso de todas las características de ser adquirida por el dominio del conocimiento de los expertos (Kolodner 1993), por técnicas de aprendizaje, tales como algoritmos genéticos (Shin and Han 1999), proceso de jerarquía analítica (del inglés, *Analytic Hierarchy*

Process: AHP) (Park and Han 2002) o por métodos estadísticos tales como el análisis discriminante múltiple y regresión.

Varios autores (Arya et al. 1998; Muja and Lowe 2009) han definido el problema de búsqueda del vecino más cercano como: dado un conjunto de puntos $P = \{p_1, \dots, p_n\}$ en un espacio vectorial X , estos puntos deben ser pre-procesados de tal manera que, dado un nuevo punto de consulta $q \in X$ la búsqueda de los puntos de P que están más cerca de q se puede realizar de manera eficiente.

El problema presentado agrega una complejidad y es la representación multidimensional de los datos, por lo que en este escenario los algoritmos KNN no son eficientes teniendo en cuenta que el resultado de la búsqueda lineal es muy costosa. Esto ha generado un interés en algoritmos que realizan una búsqueda aproximada del vecino más cercano (del inglés, Approximate Nearest Neighbors (ANN)), aunque a veces no siempre el óptimo vecino es devuelto; tal aproximación de estos algoritmos puede ser en órdenes de magnitud más rápido que la búsqueda exacta, sin dejar de ofrecer casi el óptimo en exactitud (Muja and Lowe 2009; O'Hara and Draper 2013).

La autora del presente trabajo acorde al problema presentado recomienda el uso del método de indexación ANN en el proceso de recuperación utilizando para ello una métrica de similitud basada en la distancia.

Módulo de Adaptación

La reutilización de casos es el proceso de transformar una solución recuperada en una solución apropiada para el problema actual. Se ha llegado a argumentar que la reutilización es el paso más importante del razonamiento basado en casos, ya que incorpora inteligencia a lo que, de no ser así, sería un mero proceso de reconocimiento de patrones.

Una vez seleccionados los casos similares, se efectúa el procedimiento de adaptación, que consiste en la modificación y combinación de las soluciones de los casos similares para formar una nueva solución, una interpretación, o una explicación dependiendo de la tarea que lleve a cabo el sistema. Es deseable que el sistema también genere justificaciones o explicaciones que apoyen a la solución creada para el nuevo caso. La adaptación puede realizarse por el usuario o de manera automática por el sistema. Si la adaptación la realiza el usuario, el sistema sólo realizaría la búsqueda de los casos similares en cambio si la adaptación la ejecuta el sistema, entonces debe contener algún conocimiento, tal como fórmulas o reglas.

Se considera que el proceso de adaptación es una de las partes más difíciles del ciclo de CBR. En (Wilke and Bergmann 1998) se clasifica en tres tipos principales: la adaptación nula, la adaptación de transformación y adaptación generativa.

La autora propone utilizar el método más simple de adaptación que consiste en buscar en la base de caso aquellos casos que contengan requerimientos similares al problema planteado, utilizando la técnica del ANN.

1.3.2 Problemas actuales de los Sistemas Basados en Casos

Un razonador basado en casos depende de la estructura y el contenido de la base de conocimiento. Por eso, uno de los problemas fundamentales de estos sistemas consiste en encontrar un modelo de organización e indexación de la memoria de casos que permita un almacenamiento, recuperación y reuso eficientes (Gutiérrez 2003).

Los aspectos fundamentales en la representación del conocimiento incluyen la flexibilidad en la representación y la eficiencia en la recuperación. Los conceptos de flexibilidad y eficiencia están usualmente relacionados a los conceptos de expresividad, accesibilidad, contexto y transparencia (Gutiérrez 2003).

La expresividad es relativa a cómo el contenido del conocimiento (hechos, hipótesis, heurísticas, incertidumbre etc.) puede ser representado y no a cómo pudiera ser codificado en algún lenguaje de representación. La expresividad está además relacionada con la capacidad del sistema de representar y manipular conocimiento incierto (Kolodner 1996)].

La accesibilidad se define como el problema de localizar aquellos casos en la base que son más relevantes al problema actual (Kolodner 1996). La misma está muy relacionada con el problema de la indexación de los casos de la base. Este problema ha sido reconocido como uno de los puntos más importantes del CBR ya que la selección de un buen índice resulta tan difícil, como la formulación de una buena regla en un CBR (Gutiérrez and Bello 2002).

El algoritmo de acceso a los casos debe ser rápido y eficiente. Este depende de las técnicas de indexación usadas y su diseño se vuelve un aspecto crítico cuando la BC es muy grande.

En general, hay dos tipos de estructuras de indexación que son secuencial y no secuencial. La técnica de indexación de forma secuencial es la técnica convencional que ha sido aplicada a buscar posibles casos en la BC. A través de la técnica secuencial de los casos se recuperan caso por caso siguiendo una secuencia hasta que se corresponda con el caso más similar. El procesamiento de todos los casos de forma lineal en la memoria tiene complejidad $O(n)$, donde n es el número de casos; siendo una sobrecarga si n es muy grande. Un método más eficiente consiste en dividir los casos en grupos y organizarlos jerárquicamente. En una estructura jerárquica los casos se agrupan en categorías para reducir el número de casos a buscar en una consulta. Diferentes jerarquías han sido propuestas en diferentes modelos, prevaleciendo los árboles de decisión (Gutiérrez 2003), los árboles hash (Farhan et al. 2011), los árboles de proximidad, los árboles de binarios, siendo los árboles k-d (*k-dimensional*) (Wess et al. 1994) una especificación de este último.

La representación de los casos utilizando la técnica de indexación *hash* propuesta en (Farhan, et al. 2011), tiene la ventaja de reducir la complejidad computacional considerablemente en el proceso de recuperación, la complejidad de aplicarlo radica en identificar al caso utilizando una clave teniendo en cuenta el problema de la multidimensionalidad de los atributos de los casos en el problema en cuestión.

En (Wess, et al. 1994) se propone un enfoque para reducir el tiempo de recuperación se basa en la organización de los casos en la memoria utilizando un árbol de búsqueda binaria multidimensional (del inglés,

kd-trees: KDT) para dividir la memoria de casos en grupos utilizando el diagrama de Voroni, de tal manera que cada grupo contiene los casos que son similares entre sí de acuerdo con una medida de similitud, obteniéndose un esfuerzo medio de $O(\log_2 n)$, siendo n el número de casos.

El uso de KNN en el proceso de recuperación tiene como principal desventaja la velocidad de recuperación ya que para encontrar el caso que mejor se ajusta, el caso objetivo debe compararse secuencialmente con cada caso definido en la base de conocimiento. Esto quiere decir que una comparación de similitud (distancia) debe calcularse para cada característica indexada.

Para abordar el problema de la multidimensionalidad de los datos existen numerosos métodos como: ANN (Arya, et al. 1998; Bern 1993), así como las técnicas *Hierarchical Kmeans* (HKM) (Arai and Barakbah 2007), *Locality Sensitive Hashing* (LSH) (Gionis et al. 1999), KDT (Bentley 1979) los cuales proporcionan computacionalmente procedimientos eficientes para encontrar objetos similares en grandes conjuntos de datos. Sin embargo en (O'Hara and Draper 2013) se presenta un nuevo método llamado Bosque de proximidad (del inglés, *Proximity Forest*: PF), que se basa en un bosque de métricas de árboles aleatorios donde para la recuperación y clasificación del vecino más cercano se realizan búsquedas de similitud utilizando la indexación ANN. Los resultados de la investigación evidencian comparaciones superiores respecto a los métodos HKM, KDT que utilizan el algoritmo ANN para su indexación.

En (Friedman et al. 1977) se demuestra que el tiempo de consulta con el uso de KDT puede ser $O(\log n)$ o $O(n)$ en el peor de los casos. Sin embargo, este método sufre cuando la dimensión de datos aumenta.

En (Arya, et al. 1998) se plantea que desde la perspectiva de los resultados de búsqueda en el peor de los casos, la solución ideal sería pre-procesar los puntos del espacio de búsqueda en el tiempo de $O(n \log n)$ en una estructura de datos que requiere $O(n)$, de modo que las consultas respondan en un tiempo de $O(\log n)$. Para datos con una dimensión esto es posible gracias a la clasificación de los puntos y el uso de la búsqueda binaria para responder a las consultas, también para datos con dos dimensiones, es también factible mediante el cálculo del diagrama *Voroni* para el conjunto de punto. Sin embargo, en dimensiones mayores que 2, en el peor caso la complejidad de los diagramas de *Voroni* crece en $O(n^{d/2})$.

La dificultad de obtener algoritmos que son eficientes en el peor de los casos con respecto al espacio y el tiempo de consulta para las dimensiones mayores que dos, sugiere considerar el enfoque alternativo de encontrar vecinos cercanos aproximados.

En los CBS la incertidumbre está presente en casi todos los componentes del conocimiento: en los rasgos predictores usados para describir los casos, en las relaciones entre casos y sus rasgos, en la forma de evaluar la semejanza, y en la manera que la BC se organiza y accede (Dutta and Bonissone 1990; Ruspini 1990). Su aplicabilidad en la toma de decisiones ha sido analizada por diversos autores (García 1997; Gutiérrez 2003; Gutiérrez and Bello 2003) entre otros.

1.4 La toma de decisiones

En una variedad dominios de la vida real, resolver un problema conduce a un proceso de identificación y selección de la acción adecuada para su solución. A este proceso se le denomina toma de decisiones, donde una decisión es la elección de una entre diversas alternativas.

Los problemas de toma de decisiones pueden desarrollarse en diferentes contextos o ámbitos de aplicación y pueden clasificarse atendiendo al contexto de definición en el que se formulan en (Merigó 2008):

1. Certidumbre: se produce cuando se tiene información precisa sobre todas las situaciones que pueden afectar al problema.
2. Riesgo: se puede definir como aquel en el que alguno de los factores o elementos que intervienen en el mismo están sujetos al azar, es decir, tienen asociados una probabilidad.
3. Incertidumbre: se plantea cuando la información disponible sobre las opciones elegibles es vaga o imprecisa. Este último ámbito o contexto es el que más se aproxima a la toma de decisiones características o propias de la vida real y del comportamiento humano, ya que la imprecisión es inherente a los individuos.

Otra clasificación importante en el proceso de toma de decisiones (Martínez 1999b; Sánchez 2009) es la dada a partir de:

- Cantidad de criterios: Cuando intervienen varios criterios se denomina toma de decisión multicriterio (TDMC).
- Cantidad de expertos: Cuando intervienen varios expertos o decisores se denomina toma de decisión multiexperto o en grupo.

A pesar de que la teoría de toma de decisiones tiene muchos y diferentes ámbitos de aplicación y por tanto, muchos y diferentes tipos de problemas, existe un esquema básico de resolución común a todos ellos (De Andrés 2009; Espinilla, et al. 2012; Fodor and Roubens 1994; Roubens 1997), ver Figura 1.2.

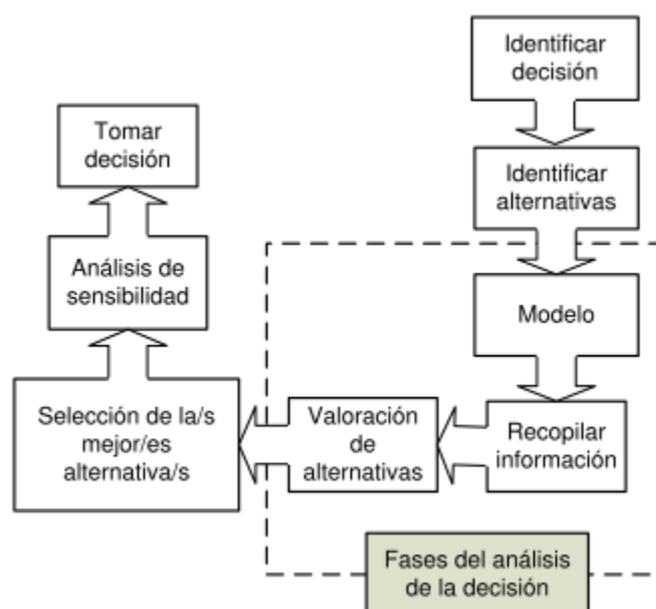


Figura 1.2. Proceso de resolución de un problema de toma de decisiones.

Los modelos de toma de decisión han evolucionado permitiendo tratar la incertidumbre, empleando mayormente la teoría de conjuntos difusos (Zadeh 1965), los conjuntos aproximados (Pawlak 1982), la modelación lingüística o la computación con palabras (CWW) (Herrera and Martínez 2000; Martínez 1999a).

1.5 La incertidumbre y su tratamiento en la toma de decisiones

La incertidumbre surge por la necesidad de modelar la imprecisión de conceptos e información involucrada en los problemas, para simular actividades propias del comportamiento y razonamiento humano, u otras inherentes a fenómenos que presentan falta de certidumbre.

La incertidumbre que rodea las preferencias, juicios y decisiones de las personas advierte de la inseguridad, influencia de la experiencia y conocimientos de las personas en la toma de decisión. Esto provoca que los juicios emitidos por los decisores no sean precisos, por lo que la elección o decisión se ve sesgada por la incertidumbre. Es más conveniente entonces, evaluar u obtener criterios en términos de intervalos o palabras que en números exactos.

Para los problemas de decisión en ambiente de incertidumbre la teoría clásica de la decisión y sus métodos no permiten dar un tratamiento adecuado. Es por esto que la aplicación de métodos de toma de decisión integrado a técnicas de *soft computing* brindan una representación real del problema, y una solución más cercana a la representación del razonamiento humano, incorporando sus valoraciones con tratamiento de la incertidumbre.

Como respuesta a la dificultad experimentada en proporcionar entradas exactas a los sistemas inteligentes, Lotfi Zadeh (Zadeh 1975a; Zadeh 1975b; Zadeh 1975c) propuso la teoría de los conjuntos difusos.

La teoría de los conjuntos difusos capta la subjetividad de comportamiento humano. La utilidad de un conjunto difuso para modelar un concepto o una etiqueta lingüística dependerá de la forma que tome su función de pertenencia. Resulta de gran importancia la determinación práctica de una función de pertenencia precisa y justificable.

Definición 1 (Zadeh 1965): Un conjunto difuso \tilde{A} en X se caracteriza por una función de pertenencia $f_{\tilde{A}}(x)$ que asocia a cada elemento de x un número real del intervalo $[0,1]$, $f_{\tilde{A}}:X \rightarrow [0,1]$, donde el valor de $f_{\tilde{A}}(x)$ en x representa el grado de pertenencia de x al conjunto \tilde{A} .

A diferencia de los conjuntos clásicos en los que un elemento pertenece o no a un conjunto, en los conjuntos difusos los elementos pueden pertenecer con determinado grado. El elemento clave es la función de pertenencia. En (Dubois 1980; Merigó 2008; Wierman 2010), se definen varios tipos de funciones de pertenencia, se utilizan dependiendo de la complejidad del problemas a tratar y de la preferencia de quien las utiliza. Entre las más empleadas están las funciones triangular y trapezoidal.

Es válido aclarar que a estas representaciones de los conjuntos difusos se les conoce como conjuntos difusos de primer orden (de tipo 1), sin embargo en los últimos años se ha venido imponiendo, la representación de los conjuntos de tipo 1 en conjuntos de 2-tuplas (Martínez 1999a; Martínez and Herrera 2012; Wang and Hao 2006) y los conjuntos difusos de tipo 2 (Chen-Tung Chen 2012; Chen-Tung Chen 2011; Ming Li 2012), que

permiten modelar mejor la incertidumbre. Lo anterior se debe a aspectos planteados en (Martínez 2010; Mendel 2010) sobre las limitaciones de los conjuntos de tipo 1 para representar las palabras, fundamentado en que utiliza funciones de pertenencia precisas y ellas son las que guían a las palabras. Uno de los elementos que se introduce con la existencia de estas teorías y que tienen su base en la teoría de conjuntos difusos es el modelado lingüístico.

Los métodos para la solución de problemas y la toma de decisión han evolucionado en función de lograr una representación lo más cercana posible a la naturaleza real de la información involucrada y al razonamiento de los decisores. Son estos los elementos que dan base a la necesidad del modelado lingüístico en los problemas de toma de decisión. Ocurre con frecuencia que la información que se maneja en la toma de decisiones no se puede valorar de forma precisa sino mediante valoraciones cualitativas o en un intervalo y no a un número exacto. En estas situaciones es adecuado el uso de las variables lingüísticas (VL).

Definición 2 (Zadeh 1975a; Zadeh 1975b; Zadeh 1975c): una VL es caracterizada por una quintupla $(N, T(N), U, G, M)$, donde N es el nombre de la variable, $T(N)$ es el conjunto de términos que constituyen los valores lingüísticos, U es el universo de discurso, G es la regla sintáctica para generar los términos en $T(N)$, y M es la regla semántica para asociar a cada término lingüístico X su significado.

Un ejemplo de variable lingüística es la “Evaluación” de una competencia, normalmente expresada en términos de intervalos que pueden ser asociados a las clasificaciones de: Muy Bajo (MB), Bajo (B), Medio (M), Alto (A) o Muy Alto (MA). En la figura 1.3. Se muestran los elementos de la variable lingüística evaluación.

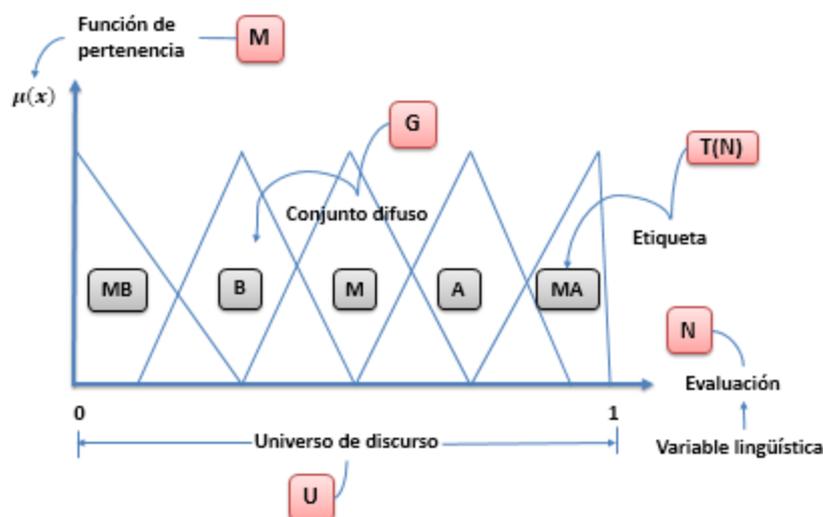


Figura 1.3. Ejemplificación de los componentes del concepto de variable lingüística (Fuente: Elaboración propia).

Es preciso buscar una expresión de la cardinalidad que permita representar la información sin afectar la precisión y discriminación de algunos valores. La cardinalidad recomendada en los modelos lingüísticos es un valor impar entre 5 y 9, sin superar los 11 o 13 términos (Bonissone et. al. 1986).

Una vez definidos los términos lingüísticos es necesario definir la semántica del conjunto de etiquetas, para ello existen varios enfoques, pero el más utilizado es el enfoque basado en funciones de pertenencia (Bonissone et. al. 1986). Este enfoque describe los conjuntos difusos mediante funciones de pertenencia (Dubois 1980; Merigó 2008; Wierman 2010). Para la representación de la información lingüística se han

desarrollado varios modelos, los más reconocidos, basados en los conjuntos difusos de tipo 1 (Zadeh 1965) y basado en los conjuntos difusos de tipo 2 (Mendel 2007b; Zadeh 1978).

El modelado lingüístico trae consigo la necesidad de operar con etiquetas lingüísticas, usando técnicas computacionales que tienen definidos operadores de agregación, comparación, negación y otros, sobre información lingüística. En la literatura se encuentran dos tipos de modelos computacionales sobre información lingüística, siendo significativo que en ambos modelos se produce pérdida de información en las operaciones (Carlsson 1999):

- Modelos basados en el Principio de Extensión.
- Modelos Simbólicos.

Basado en el concepto de traslación simbólica se desarrolló un modelo lingüístico difuso de representación para la información lingüística basado en 2-tuplas que amplía su capacidad de utilización y minimiza la pérdida de información en las operaciones (Martínez 1999a).

Para definir el modelo de representación basado en 2-tuplas se parte del concepto básico de traslación simbólica. Siendo $S = \{S_0, \dots, S_n\}$ un conjunto de términos lingüísticos, y $\beta \in [0, g]$ un valor obtenido por un método simbólico operando con información lingüística (Martínez 1999a).

Definición 3: La traslación simbólica de un término lingüístico S_i es un número valorado en el intervalo $[-0.5, 0.5)$ que expresa la “diferencia de información” entre una cantidad de información expresada por el valor $\beta \in [0, g]$ obtenido en una operación simbólica y el valor entero más próximo, $i \in \{0, \dots, g\}$ que indica el índice de la etiqueta lingüística (S_i) más cercana en S .

A partir de este concepto en (Martínez 1999a) se desarrolla un modelo de representación para la información lingüística, el cuál usa como base la representación 2-tuplas, (r_i, α_i) , $r_i \in S$ y $\alpha_i \in [-0.5, 0.5)$, donde: r_i representa la etiqueta lingüística, y α_i es un número que expresa el valor de la distancia desde el resultado original al índice de la etiqueta lingüística r_i más cercana en el conjunto de términos lingüísticos S , es decir, su traslación simbólica.

Este modelo permite las operaciones que se requiere para implementar la CWW. Un elemento significativo de este modelo es la facilidad para implementar la CWW, paradigma utilizado para dar tratamiento la incertidumbre (Herrera 2000; Jin-Hsien and Jongyun 2006; Lawry 2001; Martínez 2010; Mendel 2007a; Türksen 2007).

En (Mendel 2010) autores reconocidos en el tema de la CWW, brindan criterios sobre cuáles son sus puntos de vistas y consideraciones sobre la definición de CWW y otros elementos de interés para su implementación. De ello se puede resumir que opera con palabras o sentencias definidas en un lenguaje natural o artificial y que pueden ser instanciadas con números, simulando el proceso cognitivo humano para tratar con problemas de decisión (Martínez 2010).

La autora considera importante resaltar que la facilidad de controlar la desviación del valor real respecto al valor aproximado permite operar con los valores reales en cada momento y por tanto no hay pérdida de

información en la agregación. Independientemente del paradigma a utilizar en la operatoria y la representación de la información, es necesario, estudiar operadores que permitan agregar la información y ordenar.

1.5.1 Agregación de información

La agregación de la información supone el uso de operadores que permitan establecer una relación de preferencia colectiva. Existe variedad de operadores de agregación, algunos de ellos son tipos de operadores de agregación base, es decir, operadores sobre los cuales se definen operadores de agregación para modelos y problemas más específicos. Unido a esto, los modelos de representación lingüística estudiados de igual manera hacen una propuesta de operadores de agregación adecuados a sus características.

Definición 4 (Merigó 2008): Un operador de agregación es una función $F : [0,1]^n \rightarrow [0,1]$ que cumple las condiciones de contorno $F(0, \dots, 0) = 0$ y $F(1, \dots, 1) = 1$, debe ser continuo y monótono no decreciente. Entre los tipos de operadores de agregación más utilizados se encuentran los basados en las t-norma (conjuntiva) y t-conorma (disyuntiva), los operadores OWA y la media aritmética (Canós et. al. 2006; Canós et. al. 2007; Canós 2008; Chen-Tung Chen 2012; Doumpos 2010; Herrera 2008; Merigó 2008; Ming Li 2012). Una de las familias de operadores más empleadas son los operadores media, definidos en la ecuación 1.4:

Definición 5 (Canós et. al. 2006; Merigó 2008): Una función $M : [0,1]^n \rightarrow [0,1]$ es un operador media de dimensión n si y sólo si tiene las propiedades de ídempotencia, monotonicidad y conmutatividad.

$$f_{\alpha}(x) = h^{-1} \left[\frac{1}{n} \sum_{i=1}^n h(x_i) \right], \quad x \in I^n \quad (1.4)$$

Donde la función es $h : I \rightarrow \mathbb{R}$ y el generador f_{α} es continua y estrictamente creciente. A partir de este operador media se definen otros operadores como son la media aritmética, geométrica, armónica y otras (Canós et. al. 2006; Canós 2008). Estos operadores tienen variantes que incorporan a su definición pesos, que facilitan el tratamiento diferenciado de cada uno de los valores a agregar.

Uno de las familias de operadores más ampliamente empleados en los métodos de toma de decisión para agregar la información son los operadores OWA (del inglés, Ordered Weighted Averaging) definidos en (Yager 1988; Yager 1993). Varios son los trabajos que utilizan estos operadores o definen otros teniendo los OWA como base (Canós et. al. 2006; Canós 2008; Herrera 1997; Merigó 2008; Ming Li 2012; Pelaez ; Sadiq 2010).

Definición 6 (Yager 1988; Yager 1993): Una función $F : \mathbb{R}^n \rightarrow \mathbb{R}$ es un operador OWA de dimensión n si tiene un vector asociado W de dimensión n tal que sus componentes satisfacen: $w_j \in [0,1]$, $\sum_{j=1}^n w_j = 1$ y $F(a_1, \dots, a_n) = \sum_{j=1}^n w_j b_j$, donde b_j es el j -ésimo más grande de los a_i .

En la representación basada en 2-tuplas, se ha desarrollado toda una operatoria asociada a ella, definiéndose la comparación, negación y operadores de agregación. Algunos de los trabajos donde se definen operadores de agregación para 2-tuplas son (Herrera et al. 2008; Herrera and Martínez 2000; Martínez 1999a; Martínez and Herrera 2012).

Existe variedad de operadores de agregación, aunque en su mayoría se definen sobre la base de los

presentados en este epígrafe. La aplicación de operadores de agregación es importante en la toma de decisión, son los que permiten integrar la información que brinden los expertos y deben ser considerados para formar los criterios que permitan evaluar y ordenar las alternativas. La autora considera que para problemas que no tengan una complejidad elevada, la familia de operadores media es adecuada, además recomienda el uso de la representación utilizando 2-tuplas así como la operatoria asociada, fundamentado en la facilidad que brinda para evitar la pérdida de información.

1.6 Análisis de las herramientas de gestión de proyectos existentes

Para llevar a cabo la gestión eficaz de un proyecto se necesita contar con herramientas que permitan guiar el desarrollo exitoso del mismo, teniendo en cuenta que cada día el desarrollo de software se complejiza; se hace imprescindible la automatización de los procesos de gestión recursos humanos para medir el desempeño del personal y alcanzar el éxito del proyecto. Un problema común identificado en las investigaciones revisadas es no se encontraron indicios de aplicación de los modelos y métodos existentes en las herramientas de gestión de proyectos.

La Tabla 1.3 muestra un estudio comparativo de tres herramientas, dos de ellas líderes del mercado según el cuadrante de Gartner (Stang and Handler 2013). El estudio abarcó a un total de 138 herramientas en la Tabla 1 del Anexo I se muestra un resumen de la comparación realizada.

Tabla 1.3. Comparación y análisis de los competidores similares a GESPRO en el mercado.

Criterios analizados	Innotas	Daptiv PPM	GESPRO 12.05
Soporte sobre la web	✓	✓	✓
Tipo de licenciamiento	Propietaria	Propietaria	GPL v2.0
Entorno colaborativo	✓	✓	✓
Planificación	✓	✓	✓
Gestión del portafolio de proyectos	✓	✓	✓
Gestión de documentos	✓	✓	✓
Flujo de trabajo	✓	✓	✓
Reportes y análisis	✓	✓	✓
Gestión financiera	✓	✓	✓
Gestión de Recursos	✓	✓	✓
Manejo de incertidumbre	---	---	✓
Evaluación de competencias	---	---	---

Según (Pacelli 2004) las principales causas de una débil gestión de proyectos están referidas a la falta de conocimiento en el control y seguimiento de proyectos, así como las debilidades de las herramientas para la toma de decisiones, expresadas en carencias en las áreas de conocimiento e insuficiencias para el tratamiento de la incertidumbre de los datos y la ambigüedad en los conceptos.

Teniendo en cuenta la comparación anterior se evidencia a GESPRO 12.05 con un mayor cubrimiento de los criterios analizados. Este incorpora el manejo de la incertidumbre para comunicar de manera más objetiva el estado de los proyectos mediante el uso de indicadores de control (Lugo 2012).

Conclusiones del capítulo

Como resultado de la fundamentación teórica realizada se arrojaron las siguientes conclusiones:

- Los modelos de GRH desarrollados en la UCI, manifiestan problemas para el manejo de múltiples criterios y decisores.
- Dadas las características del proceso de evaluación de competencias se recomienda la utilización de un CBS para la utilización de las experiencias de anteriores.
- La representación de la información mediante el uso de las variables lingüísticas se considera adecuada ya que brinda grandes facilidades para tratar información cualitativa y expresar los resultados en el lenguaje de los decisores.
- Se considera adecuado el uso de métodos de distancia y la familia de operadores media para la agregación de información utilizando el modelo de representación lingüística basado en 2-tuplas.
- Se propone la herramienta GESPRO como base para el desarrollo y aplicación del algoritmo portener un mayor cubrimiento de los aspectos necesarios que aborda la investigación.

CAPÍTULO 2: ALGORITMO BASADO EN CASOS PARA EVALUAR COMPETENCIAS

En el presente capítulo se presenta el algoritmo basado en casos para evaluar competencias profesionales a partir de las evidencias de los recursos humanos en el desempeño de los roles vinculados a los proyectos desarrolladores de software. Se fundamenta y se describen los pasos y los componentes del algoritmo formado por la base de caso, el módulo de recuperación y de adaptación.

1.1 Estructura general de la propuesta

Dentro un entorno productivo de software un individuo puede desempeñarse de acuerdo a sus habilidades y competencias en uno o varios roles indistintamente. De ahí que si sus competencias son evaluadas de mal, su desempeño en el rol daría el mismo resultado.

Dado que el ámbito de estudio de este trabajo es la evaluación de competencias, es necesario señalar que los procesos de evaluación son procesos cognitivos complejos que llevan consigo diversos mecanismos que permiten identificar los elementos que van a ser evaluados, fijar el marco o el contexto en el que se va a realizar la evaluación, recopilar la información necesaria para llevar a cabo la evaluación y, finalmente, obtener una valoración de los elementos del problema. Se parte de las actividades de la fase de análisis de la decisión propuestas en la Figura 1.2.

A continuación se formaliza las actividades del análisis de la decisión que intervienen en proceso de evaluación de competencias:

1. Establecer el marco de evaluación: Se seleccionan los expertos y criterios (evidencias basadas en indicadores) y las competencias con el fin de priorizar y evaluar estas últimas. El marco de trabajo queda definido de la siguiente forma:

Se parte de un conjunto de individuos $I = \{i_1, i_2, \dots, i_m\}, m \geq 1, m \in \mathbb{N}$, de los cuales se cuenta con evidencias basadas en indicadores respecto a cada competencia teniendo en cuenta la eficacia, la planificación, los costos y el nivel de relación de las tareas con dicha competencia en el desempeño de rol, para el individuo I_j el conjunto de indicadores asociados a una competencia se define en 2.1.

$$C_j = \{c_{j1}, c_{j2}, \dots, c_{jp}\}, 1 \leq p \leq 4, p \in \mathbb{N}, 1 \leq j \leq m \quad (2.1)$$

Los atributos pertenecen al dominio $c_j \in \mathbb{R}, 0 \leq c_j \leq 1$, dada que la semántica de cada indicador está definida sobre el intervalo $[0,1]$.

Las competencias del individuo I_j se definen como: $R_j = \{r_{j1}, r_{j2}, \dots, r_{jn}\}, 1 \leq n \leq 40, n \in \mathbb{N}$

En el proceso de evaluación interviene un grupo de expertos $E = \{e_1, e_2, \dots, e_n\}, n \geq 1, n \in \mathbb{N}$, cada uno con sus propias creencias, ideales, actitudes, conocimientos y experiencias, lo que propicia diferencias al emitir las valoraciones o criterios. Los expertos no intervienen durante el proceso de recomendación de la evaluación del algoritmo, su función principal es verificar y validar la solución recomendada.

2. Recopilar información: Se obtiene información sobre las preferencias dadas por el algoritmo. Esta información representa la valoración de cada competencia con respecto a los criterios (indicadores). El vector de utilidad (Espinilla, et al. 2012) es representado de la siguiente forma: $P_j = \{p_{j1}, p_{j2}, \dots, p_{jk}\}$, donde p_{jk} , es la preferencia en relación al indicador C_k de las competencias R_j por el algoritmo.
3. Valorar alternativas: Se agrega la información y se da una valoración lingüística. Se emplean en esta etapa el modelo de representación lingüística basado en 2-tuplas con los operadores de agregación asociados.
4. Validar la información: Finalmente la información es validada por los expertos definidos en el marco de evaluación.

A partir de la definición del proceso de evaluación y los estudios realizados sobre la CWW se propone un algoritmo para la evaluación de competencias, tomando para ello las evidencias de un individuo en un determinado rol, expresada en indicadores.

1.2 Estructura general del algoritmo

La solución propuesta (ver Figura 2.1) está basada en el uso de RBC como técnica para la representación de ejemplos de problemas resueltos que constituyen evidencias basadas en indicadores evaluadas por decisores. Está compuesto por los principales algoritmos que automatizan el ciclo básico del RBC definidos en las Tablas 2.3, 2.4, 2.6, 2.7, 2.8, 2.9 y 2.10.

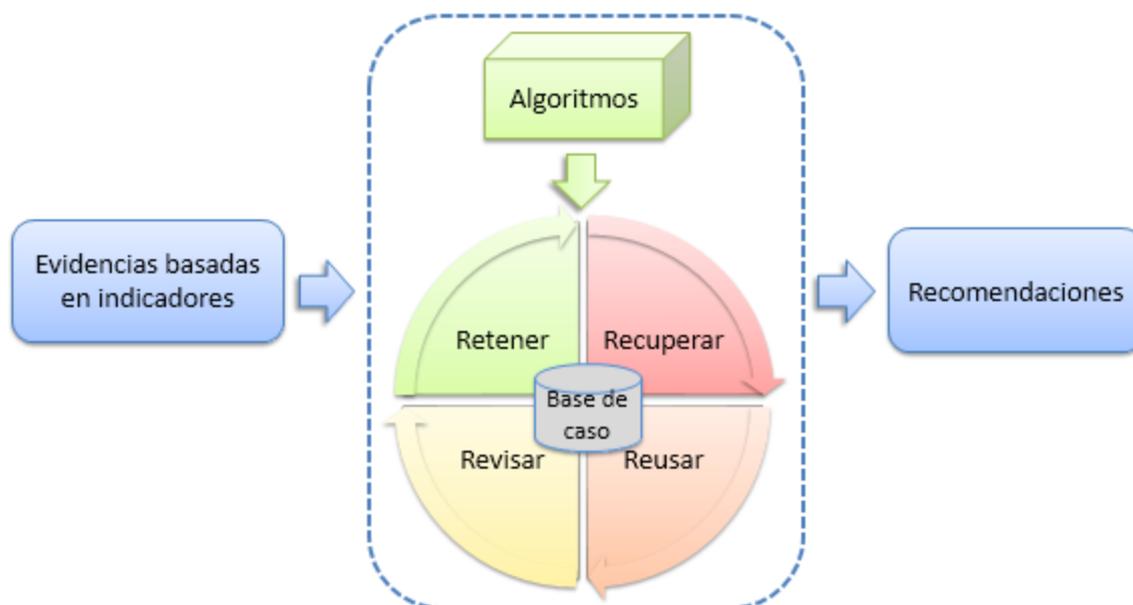


Figura 2.1. Esquema general del algoritmo (Fuente: Elaboración propia).

En la Figura 2.2 se esbozan los principales pasos lógicos del algoritmo, que serán definidos posteriormente.

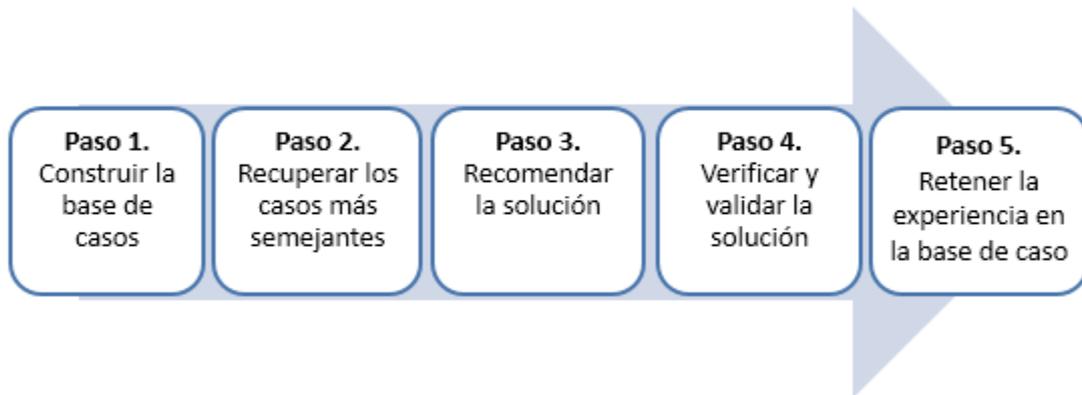


Figura 2.2. Pasos lógicos para la construcción del algoritmo (Fuente: Elaboración propia).

1.2.1 Elementos principales del algoritmo

El algoritmo que se presenta utiliza los siguientes elementos:

1. Diseño de una BC que utiliza una estructura de indexación no secuencial mediante un árbol de proximidad binario.
2. Funciones de semejanza para la recuperación los k casos más semejantes. Utiliza métricas de distancia como: Euclidean, Maximum, Manhattan, Canberra, Binary o Minkowski.
3. Modelo de representación lingüística 2-tuplas; así como los operadores de agregación asociados, fundamentado en la facilidad que brinda para evitar la pérdida de información para evaluar las distintas alternativas.

1.2.2 Objetivo y función principal del algoritmo

El objetivo principal del algoritmo es evaluar las competencias profesionales de un individuo asociada a un rol, a partir de la información existente que fue valorada por los expertos que intervienen en el proceso. La función principal está relacionada con brindar una vía rápida y eficiente que permita el análisis de la información de las personas para predecir su comportamiento en el rol e identificar sus necesidades de formación. Sirve de base para elaborar un plan personalizado a partir de las deficiencias detectadas que contribuya al desarrollo de sus competencias profesionales. Tiene como elemento principal la mejora de la actuación, elemento que se vincula estrechamente al aumento de la productividad del trabajo. Permite el seguimiento de brechas de la formación de una persona para evaluar su progreso.

1.2.3 Precondiciones

Como elementos iniciales se requiere tener definidos:

1. Los perfiles de competencias o del puesto de trabajo.
2. Conjunto de competencias evaluadas pertenecientes a una persona que constituyen casos validados por expertos.

1.2.4 Entradas y salidas del algoritmo

Las **entradas** del algoritmo son los elementos que componen la evidencia resultante del desempeño de un

individuo en un rol determinado.

Elementos que conforman una evidencia de un individuo:

- Id Usuario: Identificador del usuario.
- Competencia: Identificador de la competencia evaluada de un determinado individuo.
- Competencia Respecto a la Eficacia (CRE): Indicador que muestra la eficacia del recurso humano en una competencia.
- Competencia Respecto a la Planificación (CRP): Indicador que muestra el rendimiento del recurso humano respecto a la planificación en una competencia.
- Competencia respecto a los Costos (CRC): Indicador que muestra el rendimiento del recurso humano respecto al costo en una dimensión.
- Consistencia del conjunto (CC): Indicador que muestra el nivel de relación de las tareas con la competencia.
- Evaluación de experto: Representa la evaluación que otorga un experto a la competencia. La misma es expresada cualitativamente utilizando las etiquetas definidas para la variable lingüística evaluación como se muestra en la Figura 1.3.

En las ecuaciones (2.3), (2.6), (2.7), (2.8) y (2.11) se definen la forma de obtención de los indicadores. Un ejemplo de una de una evidencia se muestra en la Tabla 2.1.

Tabla 2.1. Representación de una evidencia.

Usuario	Competencia	CRE	CRP	CRC	CC	Evaluación del experto
Id	C_1	[0,1]	[0,1]	[0,1]	[0,1]	[MB, B, M, A, MA]
	C_2	[0,1]	[0,1]	[0,1]	[0,1]	[MB, B, M, A, MA]
	C_3	[0,1]	[0,1]	[0,1]	[0,1]	[MB, B, M, A, MA]
	C_n	[0,1]	[0,1]	[0,1]	[0,1]	[MB, B, M, A, MA]

Como **salidas** se obtiene la recomendación de la evaluación de las competencias de una persona. Se propone la variable lingüística Evaluación con las etiquetas definidas en (ver Figura 1.3).

1.3 Descripción formal del algoritmo

1.3.1 Paso 1. Construir la base de casos

La base de caso está compuesta por ocho rasgos, de ellos siete son predictores y uno es objetivo, siendo posteriormente evaluado el caso por la opinión de un experto como se muestra en la Tabla 2.2. Un caso se encuentra representado por un conjunto multidimensional de datos donde existe una correspondencia variable de competencias a evaluar por individuos.

En un diseño matricial los atributos que conforman un caso se correspondería con lo expresado en la ecuación (2.2).

$$\begin{matrix}
 & I_1 & I_2 & \dots & I_p \\
 C_1 & v_{11} & v_{12} & \dots & v_{1p} \\
 C_2 & v_{21} & v_{22} & \dots & v_{2p} \\
 \vdots & \vdots & \vdots & \dots & \vdots \\
 C_m & v_{m1} & v_{m2} & \dots & v_{mp}
 \end{matrix} \quad \text{donde } 1 \leq p \leq 4; 1 \leq m \leq 40 \quad (2.2)$$

Los elementos presentes en la matriz (2.2) son los siguientes:

- I_1, \dots, I_p : los diferentes indicadores correspondientes a una competencia respecto a la eficacia, a la planificación, a los costos y al nivel de relación de las tareas.
- C_1, \dots, C_m : las diferentes competencias de un individuo en el desempeño de un rol determinado.
- v_{11}, \dots, v_{mp} : el valor de cada indicador p correspondiente a la competencia m .

Tabla 2.2. Rasgos predictores y objetivos de las bases de casos.

Rasgos	Clasificación	Valores Posibles	Fuente de Obtención
Usuario	Predictor	ID del usuario	Exacta
Competencia	Predictor	ID de la competencia	Opinión de expertos
CRE	Predictor	[0, 1]	Exacta
CRP	Predictor	[0, 1]	Exacta
CRC	Predictor	[0, 1]	Exacta
CC	Predictor	[0, 1]	Exacta
Evaluación del sistema	Objetivo	[MB, B, M, A, MA]	Exacta
Evaluación del experto	Predictor	[MB, B, M, A, MA]	Opinión de expertos

Los atributos predictores se definen en cuatro indicadores fundamentales propuestos por (Piñero, et al. 2013) y se emplean en el proceso de evaluación de las competencias profesionales. A continuación se detalla la forma de obtención de los mismos:

Competencia Respecto a la Eficacia (CRE): Se calcula mediante la relación entre la cantidad de tareas cerradas evaluadas de Bien o Excelente relacionadas con la competencia (CTBERC) y el total de tareas cerradas relacionadas con la competencia (CTCRC) hasta la fecha de corte dd/mm/aaaa. Adquiere valores entre 0 y 1. Este índice está alineado con lo establecido por la norma cubana sobre gestión integral de capital humano. Interpretación: cercano a 1, mejor. Se calcula según la ecuación definida en (2.3).

$$CRE(dd/mm/aaaa) = \frac{CTBERC}{CTCRC} \quad (2.3)$$

Donde:

$$CTBERC = 3 * CTCPEBEC + 2 * CTCPEBEC + CTCPEBEC \quad (2.4)$$

$$CTCRD = 3 * CTCPEAC + 2 * CTCPEAC + CTCPEAC \quad (2.5)$$

A continuación se definen las variables utilizadas en las ecuaciones (2.3), (2.4) y (2.5):

- CTCABEC: Cantidad de Tareas Cerradas con Prioridad Alta evaluadas de Bien o Excelente relacionadas con la Competencia.
- CTCNBEC: Cantidad de Tareas Cerradas con Prioridad Normal evaluadas de Bien o Excelente relacionadas con la Competencia.
- CTCBBEC: Cantidad de Tareas Cerradas con Prioridad Baja evaluadas de Bien o Excelente relacionadas con la Competencia.
- CTCPAC: Cantidad de Tareas Cerradas con Prioridad Alta relacionadas con la Competencia.
- CTCNPC: Cantidad de Tareas Cerradas con Prioridad Normal relacionadas con la Competencia.
- CTCBPC: Cantidad de Tareas Cerradas con Prioridad Baja relacionadas con la Competencia.

Competencia Respecto a la Planificación (CRP): Se calcula mediante la relación entre la sumatoria de los días planificados para las tareas cerradas relacionadas con una competencia y los días dedicados a las tareas cerradas relacionadas con la competencia en la fecha de corte dd/mm/aaaa. Otra variante es tener en cuenta la misma relación pero en un período dd/mm/aaaa inicial, dd/mm/aaaa final.

Interpretación: $CRP < 1$ menor eficiencia por utilizar más tiempo, $CRP = 1$ en tiempo, $CRP > 1$ mayor eficiencia por utilizar menos tiempo del planificado. Se puede calcular utilizando una de las dos variantes que se proponen (2.6) o (2.7).

Variante 1: Calcular el acumulado utilizando una fecha de corte.

$$IRP(dd/mm/aaaa) = \frac{\sum_{i=1}^n Fecha_Fin_Tarea_RHC(i) - Fecha_Inicio_Tarea_RHC(i)}{\sum_{i=1}^n Fecha_Cierre_Tarea_RHC(i) - Fecha_Inicio_Tarea_RHC(i)} \quad (2.6)$$

Variante 2: Tener en cuenta solo un período de tiempo utilizando dd/mm/aaaa inicial, dd/mm/aaaa final.

$$IRP(dd/mm/aaaa, dd/mm/aaaa) = \frac{\sum_{i=1}^n Fecha_Cierre_Tarea_RHC(i) - Fecha_Inicio_Tarea_RHC(i)}{\sum_{i=1}^n Fecha_Fin_Tarea_RHC(i) - Fecha_Inicio_Tarea_RHC(i)} \quad (2.7)$$

Competencia respecto a los Costos (CRC): Se calcula mediante la relación entre el costo planificado (CP) para las tareas del recurso humano relacionadas con una competencia y el costo real (CR) para las tareas del recurso humano relacionadas con una competencia, en la fecha de corte dd/mm/aaaa. Otra variante es tener en cuenta la misma relación pero en un período dd/mm/aaaa inicial, dd/mm/aaaa final. Se calcula según la ecuación definida en (2.8).

Si $CRC < 1$ mayor costo, $CRC = 1$ en costo y $CRC > 1$ menor costo.

$$IRC(dd/mm/aaaa, dd/mm/aaaa) = \frac{\sum_{i=1}^n Costo_Planif_Trabajo_Planif_RHC(i)}{\sum_{i=1}^n Costo_Real_Trabajo_Real_RHC(i)} \quad (2.8)$$

Donde:

$$\text{Costo_Planif_Trabajo_Planif} = \text{Tarifa_Horaria} \times (\sum_{i=1}^n \text{Tiempo_Planif_Tarea_RHC}(i)) \quad (2.9)$$

$$\text{Costo_Real_Trabajo_Real_RH} = \text{Tarifa_Horaria} \times (\sum_{i=1}^n \text{Tiempo_Dedicado_Tarea_RHC}(i)) \quad (2.10)$$

Consistencia del conjunto (CC): Se va a constituir de acuerdo a la cercanía de la similitud de cada tarea del conjunto con el coeficiente de similitud de texto establecido. Adquiere valores entre α – corte. Este indicador ayuda al manejo de la certidumbre de la información consultada. Representa el grado de pertenencia del resto de los indicadores a la competencia.

Interpretación: cercano a 1, mejor. Se calcula utilizando un operador compensatorio basado en el NEATOWA(Liu and Lou 2006) según la ecuación (2.11):

$$CC = \sum_{\mu_i=1}^n P_i \times \mu_{D_i}(tarea_j) \quad (2.11)$$

Para calcular el peso se utiliza la ecuación (2.12):

$$P = \frac{\mu_{D_i}(tarea_j)}{\sum_{j=1}^n \mu_{D_i}(tarea_j)} \quad (2.12)$$

Donde: $\mu_{D_i}(tarea_j)$ es el coeficiente de similitud de coseno.

La relación entre los principales conceptos que se manejan en la BC se evidencia en la Figura 2.3.

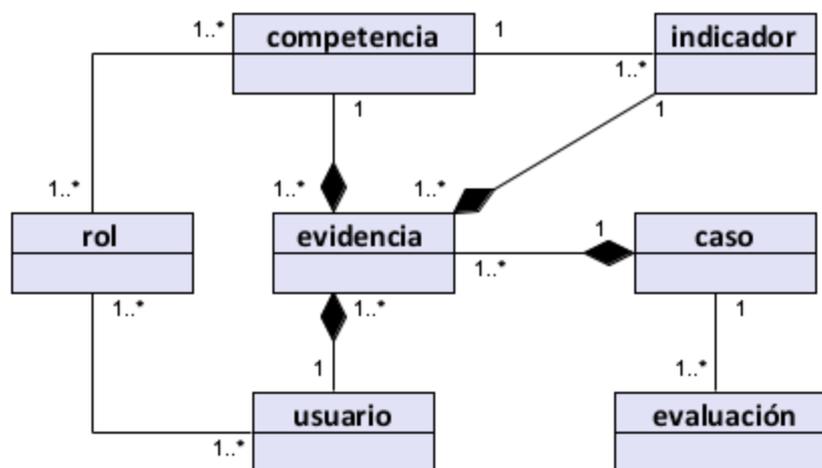


Figura 2.3. Diagrama de clases de la base de conocimiento (Fuente: Elaboración propia).

La FRC en la BC se concibe a partir de un enfoque que permite el manejo de datos jerárquicos en una base de datos relacional. A continuación se abordan los elementos fundamentales para su representación (Celko 2012).

Cada nodo del árbol representa a un caso y este consiste en una tupla que contiene un identificador de él mismo y un identificador al nodo padre. En (2.13) y (2.14) se define las dos funciones que operan en los nodos.

$$id(x, y) = x \quad (2.13)$$

$$parent(x, y) = (y, z) \quad (2.14)$$

El identificador de un nodo es único para un árbol y se especifica cómo se evidencia en la expresión (2.15).

$$\forall (x, y) \neg \exists (u, v): x = u, y \neq v \quad (2.15)$$

En (2.16) se define la raíz del árbol como un nodo que se tiene así mismo como padre:

$$root = id(x, y) = id(parent(x, y)) \quad (2.16)$$

Cada árbol es definido para tener al menos un nodo raíz. Teniendo en cuenta esto, se puede ampliar fácilmente el conjunto de definiciones de camino (*path*) y subárbol (*subtree*). Se define como camino el conjunto de padres de un nodo cualquiera hasta un nodo raíz y el subárbol de un nodo x es cualquier nodo que tiene x en su camino. En (2.17) y (2.18) se definen las expresiones matemáticas de camino y subárbol respectivamente.

$$path(x, y) = \begin{cases} \emptyset, & \text{if } (x, y) = id(parent(x, y)) \\ id(parent(x, y)) \cup path(parent(x, y)) \end{cases} \quad (2.17)$$

$$subtree(x, y) = \{(z, u) | x \in path(z, u)\} \quad (2.18)$$

La estructura jerárquica de la BC se representa mediante un árbol binario, donde cada nodo tiene como máximo dos nodos hijos, por lo general se distinguen como izquierdo y derecho. Los nodos con hijos son padres, y los nodos hijos contienen referencias a sus padres, se tiene en cuenta una referencia al nodo raíz para conocer cuáles son los antepasados de todos los nodos.

Se propone para la construcción de la estructura de la BC el uso del método de indexación ANN utilizando el algoritmo PF propuesto por (O'Hara and Draper 2013). Este algoritmo consiste en construir un conjunto de árboles de proximidad de manera aleatoria utilizando un conjunto de datos, para la obtención de los k vecinos por cada árbol del bosque. Los mejores resultados de cada árbol se comparan para devolver los vecinos más cercanos del bosque. Esta propuesta garantiza que si existe algún vecino mal localizado en un solo árbol sea mitigado mediante el uso de un bosque (O'Hara and Draper 2013).

Se valoraron dos formas de construir el bosque de árboles teniendo como entrada la cantidad de árboles a representar. La primera consiste en construir cada árbol teniendo en cuenta todos los elementos del conjunto y la segunda consiste en dividir el conjunto y con cada subconjunto construir un árbol.

La clasificación en cada nodo se produce en función de acuerdo de que si la distancia entre un elemento del conjunto y el elemento de pivote es menor que el umbral de distancia del nodo actual. En la Tabla 2.3 se define el pseudocódigo del algoritmo 1, el umbral de distancia (varía desde nodo a nodo) se basa en la distancia media entre un subconjunto de elementos para el elemento de pivote seleccionado al azar.

Tabla 2.3. Algoritmo 1. Crear el bosque de árboles.

Algoritmo 1.
<p>Entrada (s): S, Conjunto de casos para construir el árbol. k, Cantidad de árboles a representar en el bosque. $mode$, Modo de construcción del bosque de árbol. Variante 1: “split”, Variante 2: “full”. τ, Número de elementos de datos para dividir el conjunto.</p> <p>Salida (s): Booleano que indica que el bosque de árboles fue creado.</p>
<p>INICIO</p> <p>SI $mode = \text{“full”}$ ENTONCES</p> <p style="padding-left: 20px;">PARA CADA $i = 1$ hasta $i \leq k$ HACER</p> <p style="padding-left: 40px;">$P \leftarrow$ selección aleatoria de un elemento pivote de S</p> <p style="padding-left: 40px;">ProximityTree (P, S, τ) (definido en la Tabla 2.4)</p> <p style="padding-left: 20px;">FIN PARA</p> <p>SINO SI $mode = \text{“split”}$ ENTONCES</p> <p style="padding-left: 20px;">$\hat{S} \leftarrow count(S)/\tau$</p> <p style="padding-left: 20px;">PARA CADA $i = 1$ hasta $i \leq k$ HACER</p> <p style="padding-left: 40px;">$conjunto(i) =$ subconjunto de \hat{S} elementos de S</p> <p style="padding-left: 40px;">$P \leftarrow$ selección aleatoria de un elemento pivote de $conjunto(i)$</p> <p style="padding-left: 40px;">ProximityTree ($P, conjunto(i), \tau$) (definido en la Tabla 2.4)</p> <p style="padding-left: 20px;">FIN PARA</p> <p>FIN SI</p> <p>FIN</p>

La complejidad del algoritmo se define como $O(k n \log n)$, siendo k el número de árboles a construir y n la cantidad del conjunto.

Tabla 2.4. Algoritmo 2. Crear árbol del bosque.

Algoritmo 2. ProximityTree(P, S, τ)
<p>Entrada (s): S, Conjunto de casos para construir el árbol. P, pivote τ, Número de elementos de datos para dividir el conjunto.</p> <p>Salida (s): Booleano que indica que un árbol del bosque se ha creado.</p>
<p>INICIO</p> <p>δ, Función de distancia.</p> <p>SI $S < \tau$ ENTONCES /*Caso base del algoritmo recursivo */</p> <p style="padding-left: 20px;">Retornar</p> <p>SINO</p> <p style="padding-left: 20px;">$\hat{S} \leftarrow$ selección aleatoria de dos elementos de S</p> <p style="padding-left: 20px;">$D \leftarrow \{\delta(x, P)\}, \forall x \in \hat{S}$</p> <p style="padding-left: 20px;">$dt \leftarrow media(D)$</p> <p style="padding-left: 20px;">/* Partición S en subconjuntos izquierdo y derecho */</p> <p style="padding-left: 20px;">$S_{\leq} \leftarrow \{x \in S \mid \delta(x, P) \leq dt\}$</p> <p style="padding-left: 20px;">$S_{>} \leftarrow \{x \in S \mid \delta(x, P) > dt\}$</p> <p style="padding-left: 20px;">$P_i \leftarrow$ pivote de S_{\leq}</p> <p style="padding-left: 20px;">$P_d \leftarrow$ pivote de $S_{>}$</p> <p style="padding-left: 20px;">ProximityTree (P_i, S_{\leq}, τ)</p> <p style="padding-left: 20px;">ProximityTree ($P_d, S_{>}, \tau$)</p> <p>FIN SI</p> <p>FIN</p>

La estructura de representación propuesta permite reducir el número de casos a considerar durante la recuperación pues sólo se seleccionarán aquellos casos representados en el árbol que se encuentran evaluados y tienen valores similares al problema planteado.

Para calcular la complejidad de la parte del **SINO** habría que sumar dos veces el tiempo de ejecución del algoritmo para el parámetro $\frac{n}{2}$, $2T\left(\frac{n}{2}\right)$ con el tiempo del procedimiento del cálculo de la distancia. Este último, se utiliza un arreglo auxiliar, tiene una complejidad lineal.

Para la determinación de la complejidad del algoritmo recursivo se procede de la siguiente forma:

- Se asocia al procedimiento recursivo una función recurrente $T(n)$
- Se resuelve la ecuación recurrente.

En (2.19) se obtiene la siguiente ecuación recurrente, aplicando el método matemático Teorema Maestro (Cormen et al. 2001):

$$T(n) = \begin{cases} c_1; & n < \tau \text{ (caso base)} \\ 2T\left(\frac{n}{2}\right) + c_2n; & n \geq \tau \end{cases} \quad (2.19)$$

Con constantes $a \geq 1$ y $b \geq 2$. Donde $a = 2$, $b = 2$ y $f(n) = 1 \Rightarrow n^{\log_b a} = n^{\log_2 2} = n^1 = n$

El caso base posee un coste constante: $T_{\text{ProximityTree}}(0) = 1$, teniendo en cuenta que la complejidad temporal de las instrucciones de lectura, escritura, comparación, operaciones aritméticas y asignación se consideran que son $O(1)$.

Para los casos recurrentes la complejidad del algoritmo se cumple el caso 2 de teorema maestro, por lo que se concluye que la complejidad del algoritmo 2 es de orden exacto $\theta(n \log n)$.

Es importante destacar que la precisión de PF sobre otros métodos como KDT y HKM muestra mejores resultados para $\tau \geq 2$ y el rendimiento baja considerando tamaños del bosque de 3 o menor.

1.3.2 Paso 2. Recuperar los casos más semejantes

El procedimiento general para el cálculo de la distancia ha sido tomado de (Gutiérrez 2003) y que se describe a continuación en el Algoritmo 1:

Algoritmo 3: Cálculo de semejanza entre casos

1. Para cada rasgo predictor r_i :
 - i. Buscar los valores $r_i(O_i)$ y $r_i(O_j)$ (valores del rasgo r_i en los casos O_i y O_j)
 - ii. Calcular una medida de comparación $\delta_i(r_i(O_i), r_i(O_j))$ entre dichos valores.
2. Tomar en consideración el peso w_i del rasgo predictor r_i y $\delta_i(r_i(O_i), r_i(O_j))$ a través de una medida de semejanza entre un nuevo problema a resolver O_i y un caso O_j de la BC definida en (2.20):

$$\beta(O_i, O_j) = \frac{\sum_{i=1}^n w_i * \delta_i(r_i(O_i), r_i(O_j))}{\sum_{i=1}^n w_i} \quad (2.20)$$

Dónde:

- n : es el número de rasgos predictores
- w_i : Peso o relevancia del rasgo i .
- $\delta_i(r_i(O_i), r_i(O_j))$: Función de comparación entre los casos O_i y O_j atendiendo al rasgo i . Esta función puede estar definida de diferentes formas, por ejemplo:

Este algoritmo contiene dos procedimientos denotados por δ_i y f cuyo papel se describe a continuación: Para un rasgo predictor r_i , δ_i establece la semejanza o distancia entre dos valores: $r_i(O_i)$ y $r_i(O_j)$ de ese rasgo.

Finalmente f es una función de semejanza o de distancia usada para combinar los valores de las semejanzas o distancias entre los rasgos predictores ponderados por el peso. En la ecuación (2.21) se muestra un ejemplo de una función f con n rasgos predictores:

$$f(w_i, \delta_i(r_i(O_i), r_i(O_j))) = \left(\sum_{i=1}^k (w_i * \delta_i(r_i(O_i), r_i(O_j)))^\gamma \right)^{\frac{1}{\gamma}} \quad (2.21)$$

Es válido destacar que una función de semejanza o distancia es una métrica si cumple las condiciones expresada en (2.22), (2.23), (2.24) y (2.25).

$$D(O_i, O_j) > 0 \quad (2.22)$$

$$D(O_i, O_j) = D(O_j, O_i) \quad (2.23)$$

$$D(O_i, O_j) = 0 \text{ si y solo si } O_i = O_j \quad (2.24)$$

$$D(O_i, O_j) + D(O_i, O_k) \geq D(O_j, O_k) \quad (2.25)$$

En este caso las funciones de semejanzas se les llaman distancias. Si no se cumple (2.25) se le denominan pseudo-distancias.

Existe un gran número de medidas de similitud disponibles, sin embargo, existen pocos estudios comparativos de ellas y sus efectos en el agrupamiento.

La métrica Minkowski (Batchelor 1978) calcula la distancia entre dos objetos O_i y O_j , descritos por k rasgos y donde se cumple que $\gamma \geq 1$, se define en la ecuación (2.26).

$$D(O_i, O_j) = \left(\sum_{h=1}^k |O_{ih} - O_{jh}|^\gamma \right)^{\frac{1}{\gamma}} \quad (2.26)$$

La distancia de Minkowski es una métrica en el espacio euclidiano que puede considerarse como una generalización de la distancia Euclidiana y la distancia Manhattan.

Cuando $\gamma = 1$, la forma de calcular la distancia entre los objetos se llama la métrica Manhattan y se define en la ecuación (2.27).

$$D(O_i, O_j) = \sum_{h=1}^k |O_{ih} - O_{jh}| \quad (2.27)$$

Si $\gamma = 2$, nos referimos a la distancia Euclidiana, por ser una generalización natural de la idea cartesiana la cual es calculada a partir de la ecuación definida en (2.28).

$$D(O_i, O_j) = \sqrt{\sum_{h=1}^k (O_{ih} - O_{jh})^2} \quad (2.28)$$

Para los valores $\gamma \geq 2$, estamos en presencia de la métrica Supermum.

Otra forma de medir la distancia entre dos objetos O_i y O_j , descritos por k rasgos, es la métrica Canberra, calculada a partir de la ecuación definida en (2.29).

$$D(O_i, O_j) = \sum_{h=1}^k \frac{|O_{ih} - O_{jh}|}{|O_{ih} + O_{jh}|} \quad (2.29)$$

Otra forma muy utilizada de medir la distancia entre objetos es la distancia de Hamming cuando los atributos son simbólicos o nominales. Tanto en su definición binaria a partir de la cantidad de atributos en que difieren dos objetos, como en sus variantes pesadas (Duch 2000).

Para el cálculo de la semejanza entre los casos se implementó un paquete en el lenguaje estadístico matemático R, que utiliza la función `dist(x, method = "euclidean", p = 2)` definida por el propio lenguaje.

La función ***dist*** calcula y devuelve la matriz de distancia, utilizando la medida de distancia especificada. En la Tabla 2.5 se muestra una descripción de los parámetros de entradas de la función.

Tabla 2.5. Descripción de los argumentos de la función *dist* del lenguaje R.

Argumentos	Descripción
x	Matriz numérica, data.frame u objeto "dist".
$method$	Medida de distancia utilizada, por defecto aparece el valor de <i>euclidean</i> , pero puede esperar los siguientes valores: "maximum", "manhattan", "canberra", "binary" o "minkowski".
p	La potencia de la distancia de Minkowski.

Se emplea en la función de distancia el operador OWA para ello se calcula la media aritmética teniendo en cuenta el vector de distancia de los casos analizados y el vector que representa al indicador consistencia del conjunto (CC).

El peso de cada caso se representa con w y se calcula aplicando la media aritmética al vector que contiene el valor del indicador CC del nuevo caso con los restantes evaluados. Se introdujo para discriminar los casos que tenían valores semejantes, le brinda una mayor prioridad a aquellos casos que tengan un valor mayor del indicador CC. El valor de w se encuentra definido en el intervalo $[0,1]$.

En la Tabla 2.6 se detalla el pseudocódigo el algoritmo implementado en el lenguaje R que permite devolver los k casos con menor distancia respecto al nuevo caso.

Tabla 2.6. Algoritmo 4. Calcular distancia entre casos.

Algoritmo 4.
<p>Entrada(s): O_i, Nuevo caso sin evaluar. Casos O_j evaluados por los expertos. $method \leftarrow "euclidean"$ (Valor por defecto para el cálculo de la distancia). $n \leftarrow 10$, cantidad de resultados (10 es el valor por defecto). $p \leftarrow 2$ (si se define como método "euclidean"), si $p = 1$ (si se define como método "manhattan").</p> <p>Salida(s): k casos O_j ordenados con menor distancia respecto al nuevo caso.</p>
<p>INICIO</p> <p>$k \leftarrow$ Cantidad de casos evaluados. PARA $i = 1$ hasta $i \leq k$ HACER $D(O_i, O_j) \leftarrow w_{(O_i, O_j)} * dist(matrix(O_i, O_j[i]), method, n, p)$ FIN PARA</p> <p>DEVOLVER los n casos ordenados de menor a mayor respecto a la distancia calculada.</p> <p>FIN</p>

Siendo n la cantidad de casos disponibles en la BC y d la cantidad de dimensiones de un caso entonces la complejidad temporal del algoritmo 4 se define como $O(n * d)$.

El proceso de recuperación de los casos más semejantes con la estructura propuesta garantiza una mayor eficiencia en cuanto al tiempo de búsqueda de los casos más semejantes. En la Tabla 2.7 se muestra el pseudocódigo del algoritmo para recuperar los casos más semejantes.

Tabla 2.7. Algoritmo 5. Recuperar casos más semejantes.

Algoritmo 5.
<p>Entrada (s): O_i, Nuevo caso sin evaluar. k, Número de casos a recuperar.</p> <p>Salida (s): k casos recuperados ordenados por distancia descendente.</p>
<p>INICIO</p> <p>$k \leftarrow$ Cantidad de árboles del bosque.</p> <p>PARA CADA $i = 1$ hasta $i \leq k$ HACER</p> <p style="padding-left: 20px;">Encontrar los vecinos del caso entrado O_i (definido en la Tabla 2.8)</p> <p>FIN PARA</p> <p>Eliminar los vecinos repetidos</p> <p>PARA CADA vecino encontrado HACER</p> <p style="padding-left: 20px;">Calcular distancia del vecino al nuevo caso O_i</p> <p>FIN PARA</p> <p>Ordenar descendientemente por distancia</p> <p>DEVOLVER los k casos más semejantes</p> <p>FIN</p>

Siendo k la cantidad de árboles en el bosque y n la cantidad de casos evaluados la complejidad temporal del algoritmo 5 se define como $O(k \log n)$.

En la Figura 2.4 (a) se muestra un esquema de cómo sería la búsqueda de los casos más semejantes aplicando un método de indexación secuencial, mientras que en la Figura 2.4 (b) se esboza la solución propuesta planteada en el algoritmo 5.

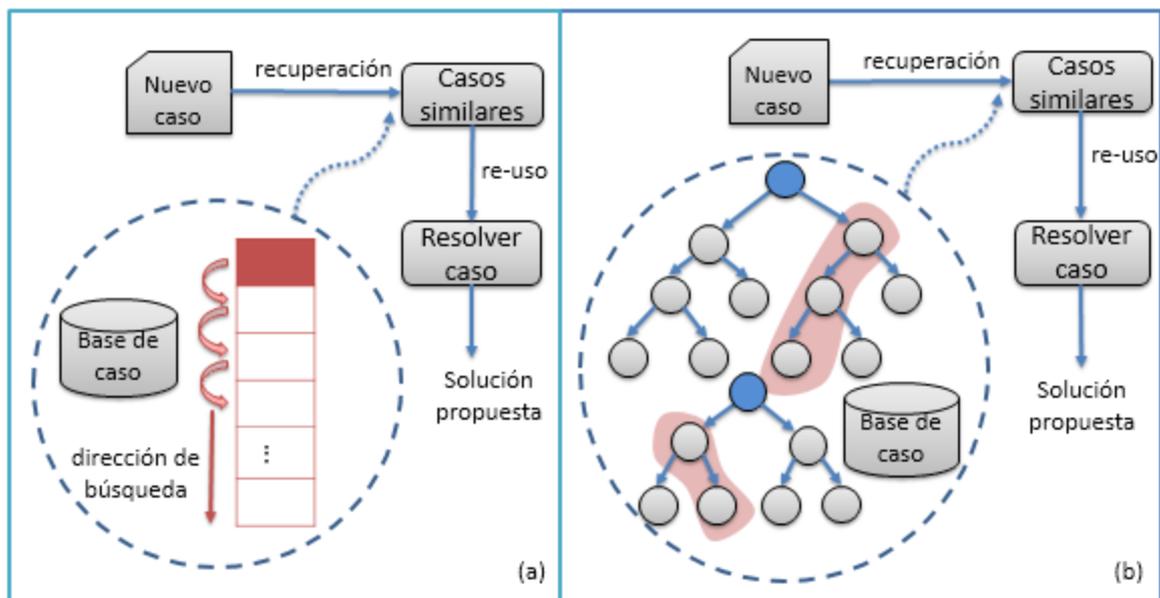


Figura 2.4 (a): Método de indexación secuencial y Figura. 3.8 (b): Método de indexación jerárquico (Fuente: Elaboración propia).

Tabla 2.8. Algoritmo 6. Encontrar vecinos de un árbol.

Algoritmo 6.
<p>Entrada (s): O_i, Nuevo caso sin evaluar. $O_i r(T_i)$, Caso O_j raíz del árbol (T_i).</p> <p>Salida(s): Lista de vecinos del árbol al caso entrado O_i.</p>
<p>INICIO $nodo_{actual} = O_i r(T_i)$ $rama$, representa lado izquierdo o derecho del árbol.</p> <p>MIENTRAS $nodo_{actual} \neq null$ HACER SI $nodo_{actual}$ es hoja ENTONCES Terminar</p> <p> SINO $rama = SeleccionarRama(nodo_{actual})$ $nodo_{actual} = SeleccionarHijo(rama)$</p> <p> FIN SI</p> <p>FIN MIENTRAS DEVOLVER $nodo_{actual}$ FIN</p>

Siendo n a cantidad de casos representados en un árbol la complejidad temporal del algoritmo 6 se define como $O(\log n)$.

1.3.3 Paso 3. Recomendar la solución

Una vez seleccionados los casos más semejantes se procede a solucionar el nuevo problema para determinar el valor del rasgo objetivo. Durante este proceso, es posible que los casos recuperados propongan soluciones diferentes por lo que se hace necesario determinar cuál de ellas debe tomarse. En la Tabla 2.9 se esboza el pseudocódigo del algoritmo 7 para evaluación de las competencias, que considera las soluciones de los casos recuperados y le aplica el método lingüístico 2-tuplas.

Tabla 2.9. Algoritmo 7. Recomendar evaluación de las competencias profesionales.

Algoritmo 7.
<p>Entrada (s): Competencias del caso O_i (sin evaluar). Distancia entre el nuevo caso y los k más semejantes evaluados. Casos O_j (evaluados). Conjunto de términos lingüísticos $S = \{MB, B, M, A, MA\}$. Tipo de método (inverse, subtract, gaussian) para darle un el peso de los casos más semejantes, por defecto es gaussian.</p> <p>Salida (s): Nuevo caso O_i con sus competencias evaluadas.</p>
<p>INICIO $k \leftarrow$ Cantidad de casos más semejantes evaluados. PARA CADA $i = 1$ hasta $i \leq k$ HACER Obtener matriz de evaluación (ME) representada en 2-tuplas asociada a la evaluación de los casos con sus competencias evaluadas $ME = [i_{pq}]_{p \times q}$ definida en (2.30).</p> <p> FIN PARA Agregar respuesta ME utilizando el operador OWA y transformar a 2-tuplas. Recomendar la evaluación de cada competencia correspondiente a una persona.</p> <p>FIN</p>

La matriz de evaluación (ME) se representa en (2.30):

$$\begin{matrix} & C_1 & C_2 & \dots & C_q \\ idCase_1 & i_{i1} & & \dots & i_{iq} \\ idCase_2 & i_{i2} & & \dots & i_{iq} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ idCase_p & i_{p1} & & \dots & i_{pq} \end{matrix} \quad \text{donde } 1 \leq q \leq 40; p \geq 1 \quad (2.30)$$

Los elementos presentes en la ME son los siguientes:

- C_1, \dots, C_q : las diferentes competencias de un individuo asociadas a un rol.
- $idCase_1, \dots, idCase_p$: los diferentes casos evaluados más semejantes respecto al caso en cuestión ordenados por peso.
- i_{11}, \dots, i_{pq} : los diferentes resultados correspondiente a la evaluación de las competencias que indican el índice de la etiqueta lingüística (s_i) más cercana en S .

Para el cálculo del peso de los vecinos más cercanos se implementan tres funciones: Inversa, Resta, Gaussiana (Segaran 2007).

La principal desventaja potencial de la función inversa es que aplica pesos muy pesados a los elementos que se encuentran cerca y cae rápidamente después de eso. Esto puede ser deseable, pero en algunos casos hará al algoritmo mucho más sensible al ruido.

La función resta supera el problema potencial de sobre-ponderar elementos cercanos, pero tiene su propia limitación. Debido a que el peso eventualmente cae a 0, es posible que haya nada lo suficientemente cerca como para ser considerado un vecino cercano, lo que significa que para algunos elementos del algoritmo hacer una predicción en absoluto.

La función final a considerar es una función gaussiana, también conocida como una curva de campana. El peso de esta función es 1 cuando la distancia es 0, y el peso disminuye a medida que aumenta la distancia. Sin embargo, a diferencia de la función de resta, el peso nunca cae hasta llegar a 0, por lo que siempre será posible hacer una predicción.

Se recomienda en el uso de la función gaussiana para el cálculo del peso porque mitiga considerablemente los problemas mencionados en la función inversa y resta.

La complejidad temporal del algoritmo 7 se define como $O(n)$, ya que el algoritmo itera n veces teniendo en cuenta la cantidad de casos semejantes recuperados.

1.3.4 Paso 4. Verificar y validar la solución

Para validar la evaluación de cada competencia perteneciente a un caso recomendada por el Algoritmo 7 de manera automática, interviene un grupo de expertos $E = \{e_1, e_2, \dots, e_n\}, n \geq 1, n \in \mathbb{N}$, que corroboran la misma y brindarán la recomendación final de la evaluación del caso.

1.3.5 Paso 5. Retener la experiencia en la base de caso

Los casos que formarán parte de la experiencia de la base de caso, serán aquellos que el experto valide. En la Tabla 2.10 se muestra el pseudocódigo del algoritmo 8 que permite retener un caso una vez que haya sido verificado y validado por el experto.

Tabla 2.10. Algoritmo 8. Retener un nuevo caso en la base de conocimiento.

Algoritmo 8.
<p>Entrada (s): O_i, Nuevo caso validado por el experto. Salida (s): Caso O_i insertado en la base de caso.</p>
<p>INICIO $\tau \leftarrow$ Representa la cantidad de divisiones de un nodo en un árbol. $k \leftarrow$ Cantidad de árboles del bosque. $modo \leftarrow$ "split" or "full". SI $modo = \text{"split"}$ ENTONCES Buscar el nodo raíz más cercano al caso O_i. Buscar el caso donde O_i puede ubicarse como hijo. Insertar el nuevo caso O_i como hijo del caso marcado como padre. SI la cantidad de hijos del caso padre $\geq \tau$ ENTONCES $S \leftarrow$ Guardar subárbol representado por el caso padre y los hijos asociados. $P \leftarrow$ hacer pivote el caso marcado como padre. Eliminar el subárbol representado por el caso padre y los hijos asociados. <i>ProximityTree</i>(P, S, τ) (definido en la Tabla 2.4) FIN SI SINO SI $modo = \text{"full"}$ ENTONCES PARA CADA $i = 1$ hasta $i \leq k$ HACER Insertar el nuevo caso O_i como hijo del caso marcado como padre del árbol i. SI la cantidad de hijos del caso padre $\geq \tau$ ENTONCES $S \leftarrow$ Guardar subárbol representado por el caso padre y los hijos asociados. $P \leftarrow$ hacer pivote el caso marcado como padre. Eliminar el subárbol representado por el caso padre y los hijos asociados. <i>ProximityTree</i>(P, S, τ) (definido en la Tabla 2.4) FIN SI FIN SI FIN SI FIN</p>

La complejidad temporal del algoritmo 8 se define como $O(k \log n)$, siendo k la cantidad de árboles. Para adaptar el nuevo caso solo se debe recorrer una rama del árbol para encontrar donde se debe insertar.

Conclusiones del capítulo

- Los algoritmos desarrollados cubren el ciclo de un CBS y se orientan a lograr un incremento de la fiabilidad en la toma de decisiones.
- Se desarrolla un algoritmo que permite construir una BC mediante una estructura de árbol de proximidad de tipo binario (Algoritmo 1) con complejidad temporal $O(k n \log n)$, que garantiza un acceso eficiente, facilita la obtención los mejores resultados de cada árbol que posteriormente se comparan para devolver los vecinos más cercanos del bosque.
- Para la recuperación de los casos más semejantes se propone un algoritmo con complejidad temporal $O(k \log n)$ que minimiza el tiempo de recuperación en la medida que los casos se incrementen en la BC (Algoritmo 3).
- Para la evaluación de las competencias se propone un algoritmo con complejidad temporal polinomial de grado uno que modela las respuesta utilizando el modelo lingüístico 2-tuplas.
- La solución propuesta permite evaluar competencias de manera automática, lo que significa una mejora con respecto a otros métodos reportados en la literatura ya que aprovecha las experiencias previas de los procesos de evaluación en las organizaciones.

CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA

En este capítulo se presenta la aplicación del algoritmo en la herramienta GESPRO 13.05. Se desarrolla en detalle un cuasiexperimento a partir de los resultados de la aplicación práctica de la propuesta. Este permite comprobar la fiabilidad de los resultados que se obtienen y demostrar que la efectividad del algoritmo es adecuado al problema que se pretende solucionar. Se concibe un experimento para validar el algoritmo en cuanto a las variables tiempo de recuperación de los casos más cercanos y la precisión.

Para la validación estática de la solución propuesta se empleó la técnica de ladov, para medir el grado de satisfacción de los usuarios. Además se valida la variable dependiente eficiencia del proceso mediante la comparación con otros métodos de evaluación para demostrar que se minimiza considerablemente el tiempo y el esfuerzo. Se realiza una comparación entre dos versiones de GESPRO para analizar el porcentaje de funcionalidades que se añaden con la implementación de la propuesta y por último se realiza una evaluación económica y social.

3.1 Solución propuesta

El algoritmo implementado se aplicó en la herramienta GESPRO 13.05. En la Figura 3.1 se puede apreciar una vista general de la BC implementada en una base de datos relacional utilizando como gestor *PostgreSQL* 9.1. La misma consta con un total de 10 funciones que cubren el ciclo del CBR. Las restantes funciones son las que permiten integrar el lenguaje PL/R y además se incluyen otras por el uso de la extensión del módulo *intarray* para la manipulación de arreglos de tipo enteros.

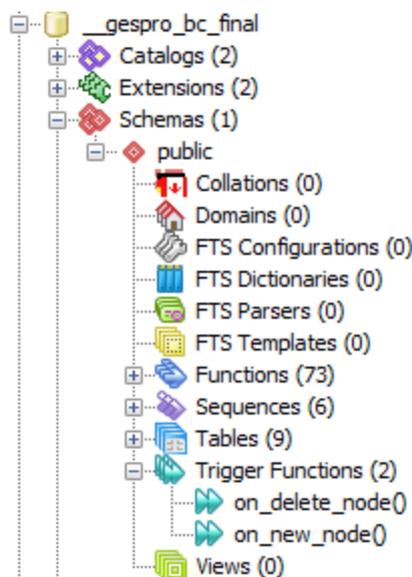


Figura 3.1. Vista general de la base de caso implementada en PostgreSQL 9.1.

En la BC se definen 4 indicadores, especificados en las ecuaciones (2.3), (2.6), (2.7), (2.8) y (2.11). El método de recolección de los datos y el tipo de procesamiento utilizado es automático. Para el cálculo de cada

indicador se implementan funciones en PL/pgSQL (del inglés, *Procedural Language/PostgreSQL Structured Query Language*) y se utilizan técnicas de la minería de datos como la medida de similaridad coseno para calcular el nivel de relación de las tareas con la competencia.

En la Figura 3.2 se muestra el modelo lógico de la base de caso que garantiza una estructura jerárquica para la representación de los casos. En el modelo se utiliza la tabla *gespro_btree_hierarchy* para recordar la trayectoria de un nodo y mantener la información sobre cada subárbol.

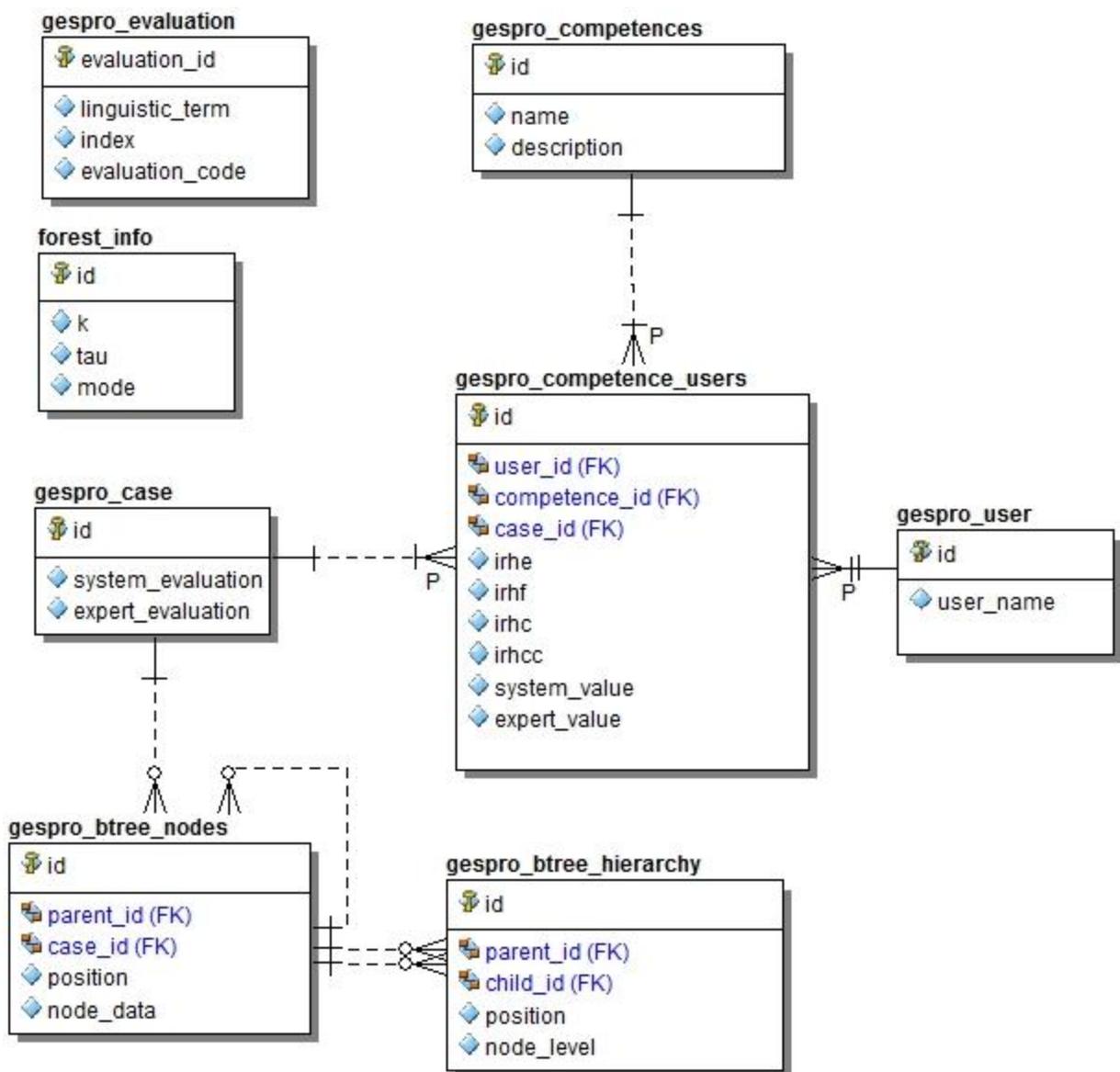


Figura 3.2. Modelo lógico relacional de la BC. (Fuente: Elaboración propia).

Se utilizó el tipo de índice *btree* en los atributos de las tablas *gespro_btree_nodes* y *gespro_btree_hierarchy* de la BC para mejorar los resultados de las búsquedas en la recuperación de los casos más semejantes.

La propuesta de solución añade dos tablas *gespro_btree_nodes* y *gespro_btree_hierarchy* y un *trigger* implementando en el lenguaje PL/pgSQL, mostrado en la Tabla 2.3.

Tabla 2.3. Trigger para adicionar nodos en la tabla gespro_btree_nodes.

```

CREATE OR REPLACE FUNCTION on_new_node()
RETURNS trigger AS
$BODY$
BEGIN
IF (@@RowCount=1) THEN
    INSERT INTO gespro_btree_hierarchy(parent_id, child_id, position, node_level)
    SELECT (CASE WHEN gespro_btree_hierarchy.parent_id IS NULL THEN
gespro_btree_hierarchy.child_id ELSE gespro_btree_hierarchy.parent_id END), (NEW.id), (CASE
WHEN gespro_btree_hierarchy.parent_id IS NULL THEN (NEW.position) ELSE
gespro_btree_hierarchy.position END), (gespro_btree_hierarchy.node_level + 1 ) FROM
gespro_btree_hierarchy
WHERE gespro_btree_hierarchy.child_id = (NEW.parent_id)
UNION ALL
SELECT NULL, NEW.id, NEW.position, 1;
ELSE
    RAISE EXCEPTION 'Multiple Insertion Not Handled in Trigger to update gespro_btree_hierarchy';
END IF;
RETURN NEW;
END;
$BODY$
LANGUAGE plpgsql VOLATILE
COST 100;
ALTER FUNCTION on_new_node()
OWNER TO postgres;

```

En la Figura 3.3 se muestra el resultado de evaluar las 3 competencias asociadas al caso con identificador 49, utilizando el algoritmo 7 definido en la Tabla 2.9. Obsérvese que la evaluación recomendada está representada por la etiqueta lingüística Muy Bajo.

	competence_id numeric	index character varying	symbolic_translation double precision	evaluation character varying	evaluation_id numeric	user_id numeric
1	23	0	0	Muy Bajo	1	196
2	25	0	0	Muy Bajo	1	196
3	26	0	0	Muy Bajo	1	196

Figura 3.3. Información de salida del algoritmo 6.

La solución implementada permite evaluar competencias profesionales a partir de las experiencias previas en las organizaciones. Brinda la posibilidad de retener el nuevo conocimiento generado en los procesos de evaluación. Ofrece resultados entendibles e interpretables y la recomendación es expresada cualitativamente mediante el uso del método lingüístico 2-tuplas.

3.2 Validación del algoritmo

En el contexto de la toma de decisiones en general la fiabilidad está relacionada con la capacidad para induir adecuadamente las preferencias del decisor, tener en cuenta de forma apropiada múltiples criterios (Singh

2011) y dar respuestas cercanas a las que daría un experto o grupo de expertos (Lamy et al. 2010), o las conocidas en la práctica (Chong and Wong 2008; Singh 2011).

En los modelos de decisión, reviste especial importancia evaluar la capacidad que brinde el modelo de ser interpretable y comprensible por los involucrados. Esto implica que, a pesar de ser un área que requiere rigor matemático, la validación pasa por juicios subjetivos (Singh 2011).

Es de destacar además que no existe un método de decisión que sea universal y aplicable a todas las situaciones. El objetivo de los métodos de decisión no debe ser mostrar una única alternativa sino permitir a los decisores una mayor comprensión del problema para posibilitar una decisión informada (Singh 2011).

3.3 Cuasiexperimento para validar el algoritmo

Con el objetivo de validar en la práctica el algoritmo propuesto se realizó un cuasiexperimento. Se muestran los resultados de la implementación de la propuesta en la herramienta GESPRO 13.05. El diseño del cuasiexperimento consiste en la aplicación del algoritmo como un clasificador y compara los resultados utilizando una medida de efectividad para la validación de los clasificadores.

3.3.1 Descripción del cuasiexperimento

Como resultado de la aplicación del algoritmo se obtuvo la recomendación de la evaluación de las competencias profesionales de los individuos en el desempeño de un rol. Los resultados obtenidos pueden ser comparados con las decisiones de evaluación tomadas por los decisores. La información con que se cuenta para la validación se especifica en la Tabla 3.1. En la Tabla 2 del Anexo II se muestra un fragmento de la información de los individuos procesados.

Tabla 3.1. Datos para la realización del cuasiexperimento.

Atributos del caso					Proceso de Evaluación	Resultado de Algoritmo
Id caso	a_1	a_2	...	a_n	Evaluación	Evaluación

Donde:

- **Id:** Identificador del caso
- **Atributos del caso:** Información de los atributos del individuo que contempla la evidencia resultante del desempeño de un individuo en un rol, representan el identificador de las competencias, los valores de los indicadores de competencia respecto a la eficacia, a la planificación, a los costos y al nivel de relación de las tareas.
- **Proceso de Evaluación: Evaluación:** La evaluación de la competencia obtenida por los decisores.
- **Resultados del Algoritmo: Evaluación:** Se toman la evaluación de la competencia que recomienda el algoritmo para cada competencia.

Considerando la información con que se cuenta se plantea para el cuasiexperimento el uso del algoritmo propuesto como un clasificador, validando su efectividad mediante una medida definida para ello en los

sistemas de clasificación.

Los clasificadores son sistemas que tienen como objetivo la clasificación de objetos o casos de acuerdo a un grupo de clases definidas a partir de información que tienen almacenada. Un clasificador es una función f que a cada N – *uplo* de atributos X asocia la clase a la cual debe pertenecer (Bonet 2008).

El algoritmo propuesto, puede funcionar como un clasificador considerando que su objetivo es recomendar la evaluación de las competencias profesionales correspondientes a un individuo en el desempeño de un rol. En este sentido se pueden homologar los componentes de los sistemas de clasificación con los elementos que definen a un clasificador:

- La evaluación constituyen las clases en las que se clasifica a la competencia.
- Las competencias constituyen los casos u objetos a clasificar.
- El método para buscar la cercanía en el algoritmo es el método por distancia, por lo que pudiera suponer que es el motor de inferencia del clasificador.

Para la evaluación y validación de un clasificador se utilizan casos cuya clase es conocida, se establece la medida de efectividad a partir de los casos clasificados correctamente:

$$\text{Efectividad} = \frac{CC}{\text{Total}} \times 100 \quad (3.1)$$

Donde: CC son los casos clasificados correctamente por el clasificador y $Total$ es el total de casos.

En función de la información que se tiene se definieron dos clasificadores, un clasificador es el decisor (CD) que evaluó previamente las competencias de un individuo y el otro el algoritmo propuesto (CA). Para el desarrollo del cuasiexperimento se emplean los resultados de información según la Tabla 3.1. Se determina la efectividad de cada clasificador y se analizan los resultados en función de validar los indicadores definidos para el modelo.

3.3.2 Ejecución del cuasiexperimento

Como primer paso se escoge la muestra de los casos para ser clasificados, escogiendo un total de 50 casos de individuos con sus competencias profesionales.

Se determina la efectividad del CA, empleando como datos los elementos de la Tabla 3 del Anexo II, que representan un fragmento de la información procesada. Se toma como clase real la evaluación dada por los decisores durante el proceso de evaluación.

Como resultado para el CA se tiene: $CC = 41$ para un $Total = 50$ con una $Efectividad = 82$

Para el determinar la efectividad de CD, la clase real se determina a partir de la evaluación de la competencia dada por los decisores. Para ello se emplea como criterio que la evaluación de Alto o Muy Alto puede ser dada a una competencia si el valor del indicador CRE que representa la eficacia se encuentran dentro del intervalo de $[0,95; 1]$ en caso contrario se le otorgará una evaluación de Medio, Bajo o Muy Bajo, siendo esta última reportada si los valores tienden a cero.

Se considera correcta la clasificación del CD si la evaluación otorgada a las competencias tiene una

correspondencia con intervalo definido, en caso contrario es incorrecta la clasificación. Un fragmento de los casos procesados se muestra en la Tabla 4 del Anexo II.

Como resultado para CD se tiene: $CC = 35$ para un $Total = 50$ y una $Efectividad = 70$.

3.3.3 Análisis de los resultados

La efectividad de CA se comporta por encima del valor de efectividad obtenido para CD. Este resultado es válido y además se considera adecuado. Es una muestra de la capacidad del algoritmo de adaptarse y la efectividad del clasificador CA, validando la capacidad que debe tener el algoritmo de brindar respuestas cercanas a las de los decisores.

Del total de 35 casos en los que el decisor evaluó a la competencia adecuadamente, CA para esos casos tuvo una efectividad de 82, el significado de este valor corresponde con la capacidad del algoritmo de brindar respuestas cercanas a los de los decisores, pero en este caso también muestra la efectividad respecto a los casos donde el decisor evaluó adecuadamente. En el resto de los casos (15) se obtiene que CA evalúa la competencia con un valor distinto al del decisor, este resultado valida el comportamiento diferente del algoritmo a partir de analizar información con un alto grado de incertidumbre.

3.4 Resultados experimentales en la recuperación de los casos

Se diseñó un experimento para probar las mejoras del comportamiento del algoritmo propuesto en el proceso de recuperación de los casos más semejantes con respecto a otros métodos de recuperación existentes en cuanto a la eficiencia.

Datos empleados

En la realización del experimento se emplearon 22 casos de individuos con las competencias asociadas a un rol. La información referente a las evidencias fue recolectada del Paquete de Gestión de Proyectos GESPRO 12.05.

Criterios de evaluación

La evaluación se realizó teniendo en cuenta aspectos de complejidad computacional como tiempo de recuperación (TR) y la precisión (P).

Se toma como precisión un valor numérico que representa la diferencia de las distancias entre el caso más cercano y el caso encontrado. Siendo cero el mejor valor posible a obtener cuando coincide con el caso que tiene posición de cercanía 1.

Análisis de los resultados

En la Tabla 3.2 se muestran los resultados de la aplicación del algoritmo de recuperación KNN (Secuencial) y en la Tabla 3.3 se muestran los resultados arrojados por el método ANN (PF) para $K=10$, representando la cantidad de árboles por la que se divide el conjunto de casos y $\tau = 2$, la cantidad de hijos por nodo en el árbol. Resultados en negrita se encuentran los mejores resultados resueltos por ambos métodos.

Se parte de un escenario donde existen 164 casos evaluados. Aplicando el método KNN secuencial se obtienen los resultados que se muestran en la Tabla 3.2. Sin embargo aplicando la solución propuesta

utilizando el método ANN (PF) los tiempo de recuperación obtenidos disminuyen considerablemente respecto al algoritmo KNN (ver Tabla 3.4). En el Anexo IV se muestran los detalles del entorno de prueba utilizado.

Nota: En la Tabla (3.2 y 3.3) el tiempo de recuperación está expresado en milisegundos (ms).

Tabla 3.2. Resultados experimentales de la aplicación del método KNN (Secuencial).

Nro	Nuevo Caso	Caso más cercano	Distancia más cercana	Caso encontrado	Distancia encontrada	Posición de cercanía	Tiempo de recuperación	Precisión
1	12	103	0.129035761	103	0.129035761	1	16706	0
2	3	86	0.192587762	86	0.192587762	1	15392	0
3	21	79	0.325524193	79	0.325524193	1	15505	0
4	14	39	0.209973124	39	0.209973124	1	17755	0
5	6	404	0.423760736	404	0.423760736	1	15419	0
6	27	68	0.089050019	68	0.089050019	1	16186	0
7	42	30	0.158923694	30	0.158923694	1	17755	0
8	24	41	0.233459081	41	0.233459081	1	16052	0
9	23	13	0.142064903	13	0.142064903	1	16509	0
10	46	383	0.135785869	383	0.135785869	1	15353	0
11	25	128	0.398208628	128	0.398208628	1	16509	0
12	1	103	0.160052352	103	0.160052352	1	18316	0
13	45	94	0.193771877	94	0.193771877	1	16811	0
14	31	17	0.102546526	17	0.102546526	1	16094	0
15	28	86	0.218121953	86	0.218121953	1	18067	0
16	48	68	0.078986964	68	0.078986964	1	15775	0
17	29	97	0.096334302	97	0.096334302	1	15897	0
18	8	67	0.150326771	67	0.150326771	1	16513	0
19	27	68	0.089050019	68	0.089050019	1	16186	0
20	43	92	0.246333764	92	0.246333764	1	16067	0
21	2	4	0.335469403	4	0.335469403	1	15000	0
22	37	640	0.135197297	640	0.135197297	1	16821	0

Tabla 3.3. Resultados experimentales de la aplicación del método ANN (PF).

Nro	Nuevo Caso	Caso más cercano	Distancia más cercana	Caso encontrado	Distancia encontrada	Posición de cercanía	Tiempo de recuperación	Precisión
1	12	103	0.129035761	103	0.129035761	1	5347	0
2	3	86	0.192587762	142	0.208877346	2	4926	0.016289585
3	21	79	0.325524193	404	0.344723077	3	4423	0.019198885
4	14	39	0.209973124	39	0.209973124	1	4797	0
5	6	404	0.423760736	404	0.423760736	1	4500	0
6	27	68	0.089050019	68	0.089050019	1	5453	0
7	42	30	0.158923694	87	0.237055324	4	2017	0.07813163
8	24	41	0.233459081	106	0.24152974	2	4979	0.008070659
9	23	13	0.142064903	383	0.149876471	2	4599	0.007811568
10	46	383	0.135785869	383	0.135785869	1	4259	0
11	25	128	0.398208628	100	0.400302997	2	4593	0.002094369
12	1	103	0.160052352	103	0.160052352	1	5256	0
13	45	94	0.193771877	142	0.197717794	3	4840	0.003945916
14	31	17	0.102546526	17	0.102546526	1	5199	0
15	28	86	0.218121953	86	0.218121953	1	5272	0
16	48	68	0.078986964	68	0.078986964	1	5221	0
17	29	97	0.096334302	131	0.121499289	2	5179	0.045164987
18	8	67	0.150326771	99	0.151545406	2	4798	0.011218635
19	27	68	0.089050019	68	0.089050019	1	5197	0
20	43	92	0.246333764	96	0.284669632	2	5125	0.038335868
21	2	4	0.335469403	4	0.335469403	1	4229	0
22	37	640	0.135197297	383	0.160375636	4	4806	0.025178339

En el experimento para la comparación de las muestras se aplicó el método no paramétrico *Kruskal-Wallis Test*. Para el nivel de significación se aplicó el método de Monte Carlo con intervalos de confianza del 99%. Se considera significativa una significación menor de 0.05. En la ejecución del test se utilizó el paquete estadístico SPSS versión 13.0.

Se comparó las muestras obtenidas de aplicar los resultados de los algoritmos respecto a las variables TR y P. La Figura 3.2 se muestra los estadígrafos básicos obtenidos para cada una de las muestras.

Ranks

	Algoritmo	N	Mean Rank
TR	Secuencial(KNN)	22	33.50
	ANN-split-treeK10T2	22	11.50
	Total	44	
Precision	Secuencial(KNN)	22	21.39
	ANN-split-treeK10T2	22	23.61
	Total	44	

Figura 3.2. Resultados del *Kruskal-Wallis Test*.

En la Figura 3.3 se muestra como resultado, que se encontraron diferencias significativas respecto a la variable TR teniendo en cuenta que el nivel de significación reportado fue de 0.000 y no se encontraron diferencias significativas respecto a la variable P.

Test Statistics^{b,c}

			TR	Precision
Chi-Square			32.273	.331
df			1	1
Asymp. Sig.			.000	.565
Monte Carlo	Sig.		.000 ^a	.571 ^a
Sig.	99% Confidence	Lower Bound	.000	.559
	Interval	Upper Bound	.000	.584

a. Based on 10000 sampled tables with starting seed 1573343031.

b. Kruskal Wallis Test

c. Grouping Variable: Algoritmo

Figura 3.3. Resultados finales del Kruskal-Wallis Test.

Se realizó un análisis sobre el comportamiento de variable tiempo de respuesta (ver figura 3.6) y precisión (ver figura 3.7) con cada uno de los algoritmos presentados, se puede apreciar que ANN (PF) es más rápido que la búsqueda secuencial aplicando KNN, sin dejar de ofrecer casi el óptimo en exactitud en cuanto a la variable precisión.

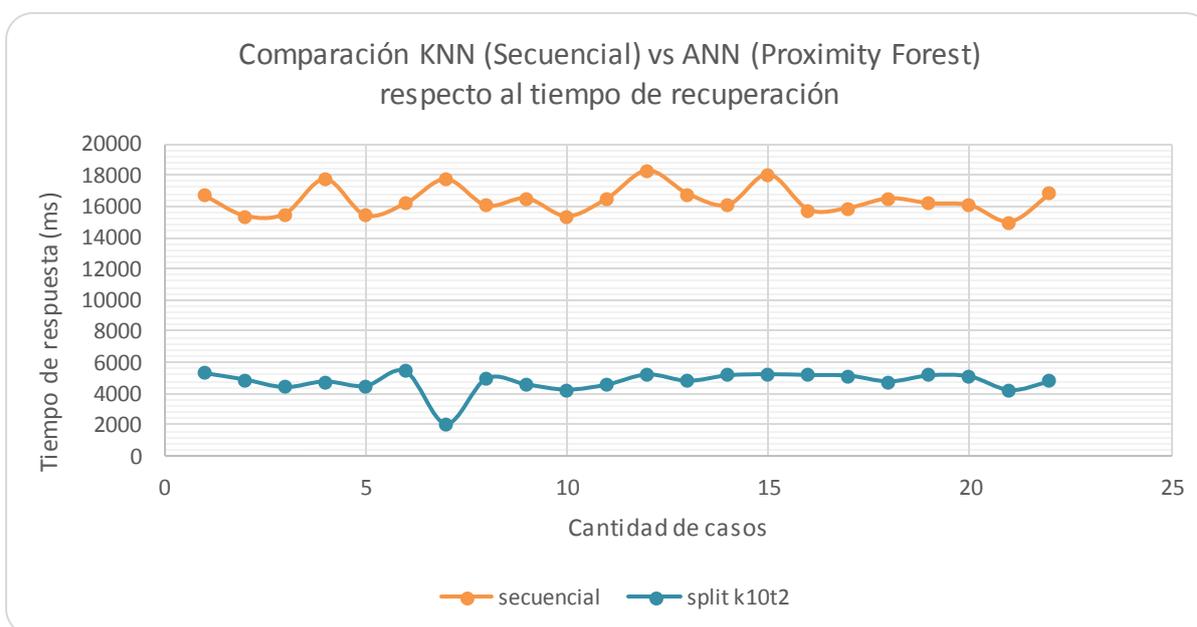


Figura 3.6. Análisis del comportamiento de los algoritmos respecto al tiempo de recuperación.

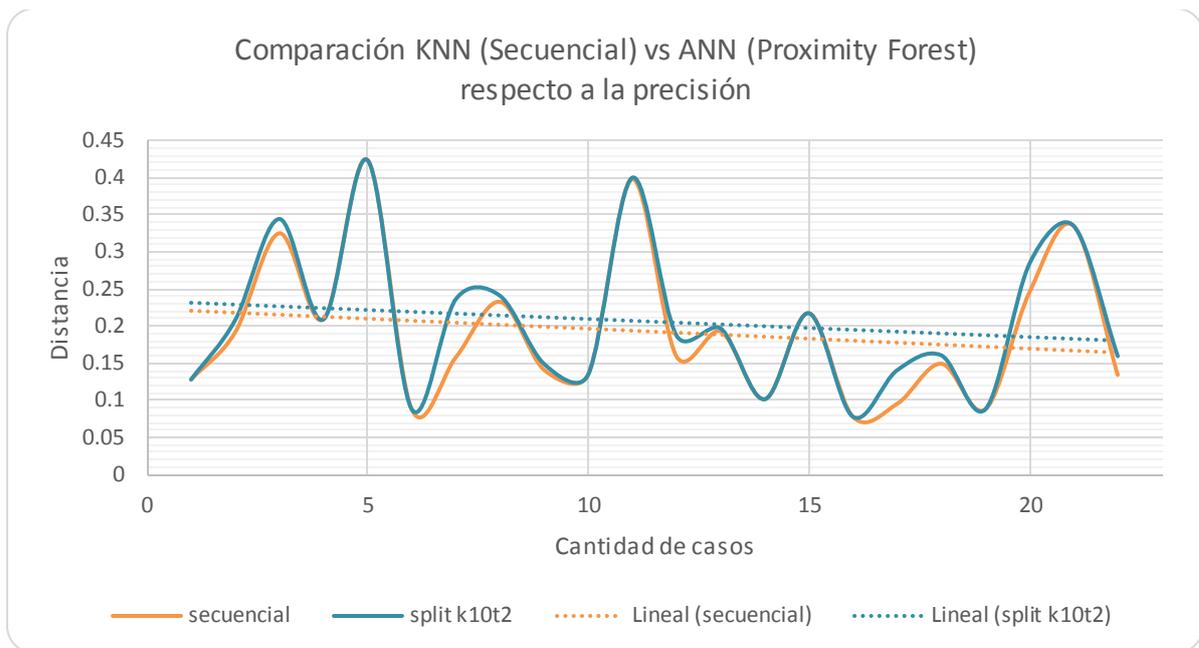


Figura 3.7. Análisis del comportamiento de los algoritmos respecto a la precisión.

En la Figura 3.7 se muestra un comportamiento estable del algoritmo propuesto que evidencia una alta capacidad de localizar soluciones relevantes de acuerdo al problema actual.

En la Figura 3.8 muestra que de los 22 casos que se le aplicaron los test estadísticos, el 90% de las respuestas localizan soluciones relevantes al problema actual comprendido entre las tres primeras posiciones de los diez casos más semejantes recuperados.

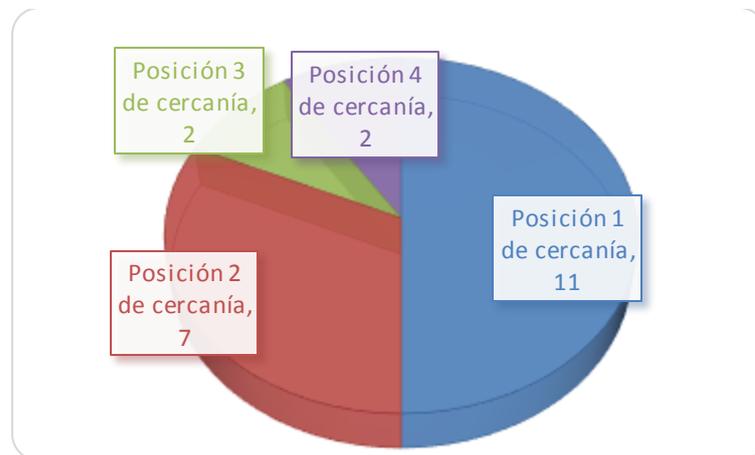


Figura 3.8. Capacidad de localizar soluciones relevantes al problema actual.

3.5 Validación estática

La validación estática se efectuó a partir de los usuarios de la propuesta. Para ello se aplicó la técnica de ladov (Kuzmina 1970), que permite el estudio del grado de satisfacción de los involucrados en un proceso o actividad objeto de análisis. A continuación se describe la aplicación de la técnica.

Para el desarrollo de esta técnica se aplicó una encuesta a un grupo de especialistas y/o decisores. El objetivo fundamental fue validar la fiabilidad, por ser este un aspecto susceptible de ser evaluado por los usuarios de la solución propuesta.

Las respuestas obtenidas permitieron conocer el grado de satisfacción en cuanto a:

- La fiabilidad en la toma de decisiones que se logra con el algoritmo propuesto.
- Las fortalezas del algoritmo que permiten elevar los niveles de fiabilidad.
- Los aspectos que consideren se debe incorporar para aumentar la fiabilidad.

La técnica de ladov constituye una vía indirecta para el estudio de la satisfacción, ya que los criterios que se utilizan se fundamentan en las relaciones que se establecen entre tres preguntas cerradas y dos abiertas. La encuesta aplicada (ver Anexo III) está conformada por siete preguntas que contempla las cinco preguntas anteriores más dos complementarias. A partir de estas preguntas se conforma el cuadro lógico de ladov que establece la relación entre las preguntas cerradas e indican la posición de cada sujeto en la escala de satisfacción. En la Tabla 3.5 se muestra el cuadro lógico de ladov y la escala de satisfacción empleada.

La escala de satisfacción está dada por los criterios:

1. Clara satisfacción.
2. Más satisfecho que insatisfecho.
3. No definida.
4. Más insatisfecho que satisfecho.
5. Clara insatisfacción.
6. Contradictoria.

Tabla 3.5. Cuadro lógico de ladov.

	1. ¿Considera usted que se debe continuar realizando la evaluación de las competencias de un individuo sin un algoritmo que permita el análisis de las evidencias previas y las evaluaciones de los expertos en procesos de anteriores?								
	No			No Sé			Sí		
3. ¿Satisface los niveles de fiabilidad esperado por usted la propuesta?	2. ¿Si UD. requiere elevar el nivel de fiabilidad en el proceso de evaluación de competencias de los individuos en el desempeño de los roles profesionales utilizaría usted el algoritmo propuesto?								
	Sí	No Sé	No	Sí	No Sé	No	Sí	No Sé	No
Muy satisfecho.	1	2	6	2	2	6	6	6	6
Parcialmente satisfecho.	2	2	3	2	3	3	6	3	6
Me es indiferente.	3	3	3	3	3	3	3	3	3
Más insatisfecho que satisfecho.	6	3	6	3	4	4	3	4	4
Para nada satisfecho.	6	6	6	6	4	4	6	4	6
No sé qué decir.	2	3	6	3	3	3	6	3	4

Para medir el grado de satisfacción se tomó una muestra de 22 especialistas en la UCI, teniendo en cuenta los años de experiencia vinculados a proyectos productivos, como jefe de proyecto o en cargos de dirección del área de producción, profesores de práctica profesional en el rol de Supervisores-Evaluadores y Tutores (SET). En la Figura 3.9 se muestra una representación de la composición de los encuestados.

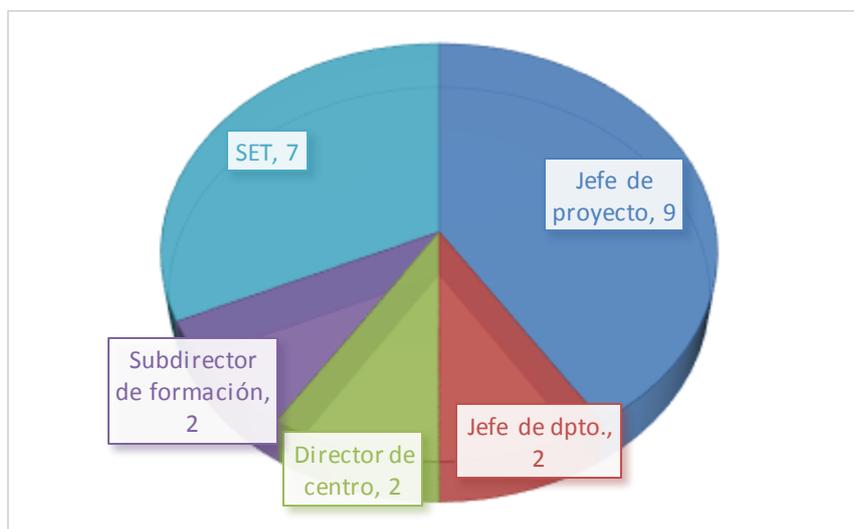


Figura 3.9. Composición de la muestra de los encuestados.

El resultado de la evaluación de la satisfacción individual se muestra en la Tabla 3.6. No hubo ninguna valoración entre: no definida, más insatisfecho que satisfecho, clara insatisfacción o contradictoria según la escala de satisfacción definida.

Tabla 3.6. Resultado de la satisfacción individual.

Nivel de satisfacción	Cantidad	%
Máximo de satisfacción.	18	81,82
Más satisfecho que insatisfecho.	4	18,18

Para obtener el índice de satisfacción grupal (ISG) se trabaja con los diferentes niveles de satisfacción que se expresan en la escala numérica que oscila entre +1 y -1 (ver Tabla 3.7).

Tabla 3.7. Escala numérica para el ISG.

+1	Máximo de satisfacción
0,5	Más satisfecho que insatisfecho
0	No definido y contradictorio
-0,5	Más insatisfecho que satisfecho
-1	Máxima insatisfacción

La satisfacción grupal se calcula por la siguiente fórmula definida en (3.2):

$$ISG = \frac{A(+1) + B(+0.5) + C(0) + D(-0.5) + E(-1)}{N} \quad (3.2)$$

Donde:

- A, B, C, D, E , representan el número de sujetos con índice individual 1; 2; 3 o 6; 4; 5.

- N representa el número total de sujetos del grupo.

Los valores que se encuentran comprendidos entre -1 y $-0,5$ indican insatisfacción; los comprendidos entre $-0,49$ y $+0,49$ evidencian contradicción y los que caen entre $0,5$ y 1 indican que existe satisfacción.

Para este caso el índice de satisfacción grupal se calcula como se muestra en (3.3):

$$ISG = ((18 + 0,5 * 4)) / 22 = 0,91 \quad (3.3)$$

Este resultado muestra el logro de una clara satisfacción y el reconocimiento del nivel de fiabilidad alcanzada sobre la mejora significativa que puede introducir el algoritmo propuesto en la UCI.

El lado *ov* contempla además dos preguntas complementarias de carácter abierto. En este caso fueron formuladas las siguientes:

- Pregunta 4. ¿Considera útil la posibilidad de predecir la evaluación de un individuo a partir de sus evidencias como criterio para la evaluación de las competencias asociadas a un rol? Argumente.
- Pregunta 7. ¿Qué elemento usted añadiría al algoritmo que se propone para aumentar su fiabilidad? Argumente.

Las preguntas abiertas son de mucha importancia ya que permiten profundizar en las causas que originan los diferentes niveles de satisfacción. Sobre las dos preguntas complementarias de carácter abierto los encuestados respondieron:

Pregunta 4

- Se considera muy útil la propuesta del algoritmo porque permite ahorrar tiempo en el proceso de evaluación.
- Permite realizar una evaluación de competencias basada en evidencias.
- Permite aprovechar el conocimiento existente en la organización.

Pregunta 7

- Se considera adecuado incluir el manejo de información heterogénea y el manejo de múltiples escalas lingüísticas.

En cuanto al resto de las preguntas hubo consenso en cuanto a la importancia de las técnicas y herramientas de ayuda a la toma de decisiones en las organizaciones y a la utilización del algoritmo para alcanzar una mayor agilidad en el proceso de evaluación de competencias.

La aplicación de la técnica de *ladov* ha aportado datos significativos respecto al grado de satisfacción de los decisores en el proceso de evaluación de las competencias profesionales de un individuo. Los criterios expresados en las preguntas abiertas fueron considerados para introducir mejoras.

3.6 Validación de la variable dependiente eficiencia en el proceso de evaluación

En este epígrafe se demostró que con la aplicación del algoritmo propuesto hubo una mejora significativa en la eficiencia del proceso de evaluación. Se toma como referencia el estudio realizado por (Rodríguez 2012) donde realiza una comparación de su modelo con el método de evaluación de 360 grados, utilizando una

muestra de 23 equipos de proyectos del Centro de Informatización de Entidades (CEIGE) de la UCI. En su propuesta la autora (Rodríguez 2012) concluye que para evaluar los 23 equipos utilizando el método de evaluación de 360 grados es necesario dedicar 13.28 horas y se necesita involucrar a 69 personas en el proceso. Sin embargo la implementación pesimista de su modelo requiere como máximo involucrar a 23 personas (una por equipo) en un tiempo promedio de 0.15 horas para evaluar los 23 equipos. Teniendo como referente la comparación anterior se aplicó el algoritmo a una muestra de 146 casos con un promedio de 19 competencias por cada caso, totalizando 2774 competencias. El tiempo total demorado fue de 0.083 horas, equivalente a 5 minutos. La base de caso contaba un total de 164 casos evaluados previamente por expertos. En el Anexo IV se muestran los detalles del entorno de prueba utilizado. Con la utilización del algoritmo propuesto se evidencia una reducción significativa del esfuerzo y el tiempo empleado en proceso de evaluación respecto a los dos métodos analizados. La propuesta de solución solamente requiere de un grupo de expertos que permitan corroborar la evaluación dada por el algoritmo. Para el caso en cuestión bastaría solamente precisar de dos expertos (ver Figura. 3.10).

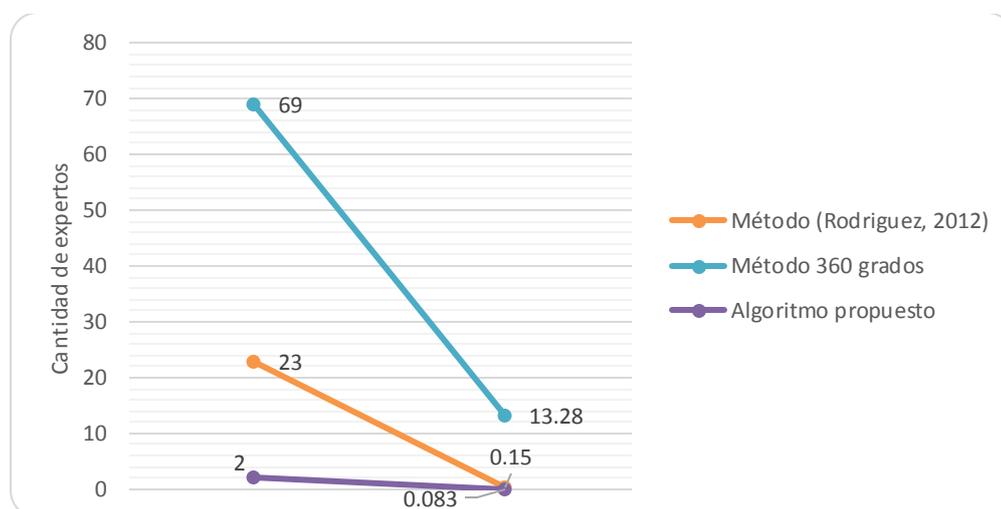


Figura 3.10. Comparación del algoritmo respecto a otros métodos de evaluación.

A partir del análisis realizado se concluye que el algoritmo propuesto disminuye significativamente el tiempo requerido para realizar la evaluación. Mostrando una reducción considerable del tiempo dedicado a evaluar y una disminución del esfuerzo de los expertos que intervienen en el proceso ya que la propuesta le brinda a los decisores una herramienta de trabajo que permite humanizar su trabajo.

3.7.1 Comparación de funcionalidades entre versiones de GESPRO

Se realizó una comparación de las funcionalidades principales de GRH de las versiones de GESPRO 12.05 y 13.05 con el objetivo de demostrar el porcentaje de las funcionalidades añadidas por la implementación del algoritmo (ver Tabla 3.8). Se toma como referencia el estudio inicial realizado por (Cuza 2013).

Tabla 3.8. Comparación de las principales funcionalidades de GRH en GESPRO 12.05 y GESPRO 13.05.

Nro.	Funcionalidades	GESPRO 12.05	GESPRO 13.05
1	Índice de desempeño RRHH	Sí	Sí
2	Datos miembros por proyectos	Sí	Sí
3	Experiencia productiva	Sí	Sí
4	Actividad de los usuarios	Sí	Sí
5	Miembros con pocas tareas	Sí	Sí
6	Control de asistencia	Sí	Sí
7	Aprovechamiento de los RRHH en el período	Sí	Sí
8	Resumen desempeño	Sí	Sí
9	Actividad de usuarios	Sí	Sí
10	Distribución por roles	Sí	Sí
11	Evaluación de desempeño	Sí	Sí
12	Competencias laborales	No	Sí
13	Estado real en las competencias por usuario	No	Sí
14	Actividad de formación	No	Sí
15	Estado ideal en las competencias por rol	No	Sí
16	Brechas entre el estado real por usuario e ideal por rol	No	Sí
17	Impacto de las Actividades de formación en las competencias	No	Sí
18	Costos de las actividades de formación	No	Sí
19	Plan de formación de un proyecto	No	Sí
20	Evaluación de competencias profesionales a partir de experiencias previas	No	Sí

De los datos expresados en la tabla anterior podemos concluir lo siguiente:

- El algoritmo implementado en GESPRO 13.05 aporta como nueva funcionalidad la representada con el número 20 resaltada en negrita, significando un 5 % del total; pero tiene una repercusión en las funcionalidades de la 11 a la 19.
- El aporte fundamental de la propuesta se basa en brindar una solución eficiente que permite agilizar el proceso de evaluación en cuanto a tiempo y a esfuerzo.
- Permitirá retroalimentar al resto de los procesos de GRH, además de diseñar planes de capacitación individuales, grupales o globales.

3.7 Análisis económico

El análisis económico de la propuesta se centra en determinar el costo de la implementación de la misma y además compararlo con el costo de aplicar el método de 360 grados. Se toma como referencia los datos del estudio realizado por (Rodríguez 2012). Se utilizó una ficha de costo donde se determinan los costos referidos a materiales utilizados y a mano de obra fundamentalmente. En el desarrollo del algoritmo propuesto intervinieron 3 especialistas: un analista, un desarrollador y un jefe de proyecto. Se concluye que para lograr el objetivo planteado en la investigación el costo total fue de **5,885.62 CUP** (ver Tabla 3.9).

Tabla 3.9 Ficha de costo de la solución.

CLIENTE: Laboratorio de Investigaciones de Gestión de Proyectos						Orden de trabajo: Nro. 1	
						Fecha de pedido: 26-04-2013	
PRODUCTO: Algoritmo basado en casos para la evaluación de competencias profesionales.						Fecha de inicio: 01-05-2013	
						Fecha de entrega: 21-10-2013	
						Fecha de terminación: 01-07-2013	
COSTOS DIRECTOS DE MATERIALES						COSTOS IND. DE PRODUCCIÓN	
Fecha	Tipo	Cantidad	UM	Precio	Importe	Tipo	Importe
	Papel	0.5	Resma	30.00	15.00	Depreciación de las PC	1,220.00
	Bolígrafos	3	Uno	1.00	3.00		
Total					18.00	Total	1,220.00
COSTO DE LA MANO DE OBRA						RESUMEN	
Fecha	Cargo	Cantidad	Mes	Salario	Importe		Importe
	Analista	1	1	555.00	555.00	Materiales directos	18.00
	Desarrollador	1	2	555.00	1,110.00	Mano de obra directa	4,647.62
	Jefe Proyecto	1	2	700.00	1,400.00	Costo directo	4,665.62
						Costos indirectos	1,220.00
		Tasa	Base				
	Vacaciones	9.09%	3,065.00		278.61	Costo total	5,885.62
	Impuesto Fuerza Trabajo	25%	3,343.61		835.90		
	Seguridad Social	14%	3,343.61		468.11	Unidades producidas	1
Total					4,647.62	Costo unitario	\$5,885.62

En la Tabla 3.9 se muestra que con la aplicación del método 360 grados se incurrió en un costo de 6,414.75 CUP, cifra que supera al costo de la implementación de la propuesta en 529.13 CUP.

Tabla 3.9 Ficha de costo de la aplicación del método 360 grados para evaluar 23 equipos.

COSTOS DIRECTOS DE MATERIALES						COSTOS IND. DE PRODUCCIÓN	
Fecha	Tipo	Cantidad	UM	Precio	Importe	Tipo	Importe
	Papel	1	Resma	30.00	30.00	Depreciación de las PC	1,220.00
	Bolígrafos	69	Uno	1.00	69.00		
Total					129.00	Total	1,220.00
COSTO DE LA MANO DE OBRA						RESUMEN	
Fecha	Cargo	Cantidad	Mes	Salario	Importe		Importe
	Jefe Proyecto	69	0.069	700.00	3,3400.00	Materiales directos	129.00
						Mano de obra directa	5,065.75
						Costo directo	5,194.75
						Costos indirectos	1,220.00
		Tasa	Base				
	Vacaciones	9.09%	3,340.75		303.67	Costo total	6,414.75
	Impuesto Fuerza Trabajo	25%	3,644.42		911.11		
	Seguridad Social	14%	3,644.42		510.22		
Total					5,489.19	Costo unitario	\$6,414.75

Se puede concluir que el costo de desarrollo de implementación del algoritmo es ligeramente inferior al método de evaluación integral, pero teniendo en cuenta la periodicidad del proceso de evaluación que puede ser mensual, trimestral, semestral o anual, el costo de evaluar con el algoritmo se traduce solamente a disponer de un grupo de expertos en dependencia de la cantidad de casos a validar. Sin embargo aplicando el método tradicional el costo y el esfuerzo siempre será superior teniendo en cuenta la cantidad de involucrados que pueden intervenir en el proceso.

Por ejemplo si se realiza el proceso de evaluación con una periodicidad mensual, con la aplicación del método 360 grados para el escenario lineal donde haya que evaluar 23 equipos de proyecto. En un año se necesitarían involucrar 828 personas y el tiempo promedio de evaluación sería de 159.36 horas lo que equivaldría a un costo total de 731,755.94 CUP.

Asumiendo que la composición de los 23 equipos es de 7 personas, la cantidad de personas a evaluar serían 161, lo que en un año representaría un total de 1932 personas. Si con el algoritmo propuesto para evaluar 146 casos se demora como promedio 5 minutos el equivalente a 0.083 horas, entonces para evaluar 1932 personas se demoraría 66.16 minutos, equivalente a 1.10 horas. Si se asume que dos expertos demoran 1 hora en validar todos los casos, entonces para validar los 1932 se necesitarían 27 expertos en un año y el tiempo que tardarían sería de 6.75 horas, asumiendo que dos expertos demoran solamente 0.5 hora en validar 146 casos. El tiempo total sería de 7.85 horas y representa la suma del tiempo demorado por los expertos más el tiempo de evaluación del algoritmo. Se puede concluir que el costo total asociado de aplicar la propuesta sería de 2,508.74 CUP, por lo que significaría un ahorro de 729,247.2 CUP en un año respecto al método tradicional de 360 grados.

3.8 Análisis social

La implementación del algoritmo basado en casos para evaluar competencias tiene como objetivo principal estimular la productividad, mejorar la eficiencia, eficacia y el rendimiento de los recursos humanos. Además conocer sus puntos débiles, presentar programas de mejora e identificar sus puntos fuertes y potenciarlos de manera que se logre un desempeño óptimo en el rol asumido.

Su importancia en el orden económico se encuentra reflejada en la implementación práctica de los lineamientos de la Política Económica y Social del Partido y la Revolución aprobados en el VI Congreso del PCC (PCC 2011), referidos a los lineamientos 8, 12 y 20 del capítulo I Modelo de Gestión Económica. Estos se encuentran relacionados con la elevación de la eficiencia, eficacia y el control en el empleo del personal, los recursos materiales y financieros, el fortalecimiento del sistema de control interno, para lograr los resultados esperados en cuanto al cumplimiento del plan con eficiencia, orden, disciplina y a que los ingresos de los trabajadores y sus jefes en las empresas estén vinculados a los resultados que se obtengan.

También tiene influencia en los lineamientos 132, 137, 138 del capítulo V Política de Ciencia, Tecnología e Innovación y Medio Ambiente. Se plantea prestar mayor atención en la formación y capacitación continua del personal. Por su parte también guarda relación con los lineamientos 169, 170, 171 y 172 del capítulo VI

Política Social teniendo en cuenta que abordan temas el proceso de reordenamiento laboral bajo el principio de idoneidad demostrada. Se plantea que los salarios garanticen que cada cual reciba según su trabajo, y el incremento de los salarios de manera gradual sea dirigido inicialmente a las actividades con resultados más eficientes. Además hace referencia a que se debe proyectar la formación de fuerza de trabajo calificada en correspondencia con las demandas actuales (PCC 2011).

Conclusiones del capítulo

Como resultados de la validación del algoritmo propuesto se obtuvo que:

- La implementación de la propuesta en el paquete de gestión de proyectos GESPRO 13.05 permitió la automatización de actividades que requieren de un elevado procesamiento de información y que este fuera realizado en menor tiempo y con mayor exactitud del que puede hacerse por las personas.
- La aplicación del cuasiexperimento permitió evaluar el algoritmo como clasificador y demostrar valores elevados de efectividad.
- Se demostró la capacidad del algoritmo de brindar criterios cercanos a los que toman los decisores.
- Se demostró la eficiencia respecto al tiempo de recuperación de los casos más semejantes de los resultados del algoritmo propuesto con la aplicación del test estadísticos de *Kruskal-Wallis*.
- La aplicación de ladov para la validación estática del modelo permitió corroborar la satisfacción con la propuesta por parte de los usuarios.
- La comparación realizada del algoritmo con otros métodos de evaluación muestra la superioridad en cuanto al tiempo y esfuerzo expresado en horas/hombres.
- La comparación realizada entre las versiones de GESPRO, permitió demostrar que aumentó el porcentaje de funcionalidades, posibilitando una mejora en la GRH.

CONCLUSIONES

- A partir de la sistematización de los referentes teóricos y los resultados del diagnóstico efectuado se confirma que los modelos y métodos existentes en la literatura presentan limitaciones, fundamentándose la necesidad de un nuevo algoritmo.
- Se desarrolló un algoritmo para evaluar de competencias profesionales que permite aprovechar el conocimiento existente en la organización y agiliza el proceso de evaluación de competencias profesionales con la aplicación técnicas de soft computing mediante la utilización del RBC y la CWW.
- El conjunto de métodos científicos utilizados para la validación de la propuesta permitieron comprobar la flexibilidad del algoritmo para brindar respuestas cercanas a los decisores, la mejora del tiempo de respuesta con respecto al método de indexación secuencial y la existencia de una alta satisfacción de los usuarios actuales y potenciales del algoritmo.

RECOMENDACIONES

Las principales recomendaciones derivadas del trabajo realizado son:

1. Extender su uso a la totalidad de los proyectos productivos de la UCI de manera que se amplíe la BC y se identifiquen nuevas áreas de mejora.
2. Incluir el manejo de información heterogénea y el manejo de múltiple escalas lingüísticas en los atributos de los casos.
3. Incorporar a la función de semejanza el cálculo del peso de los rasgos, atendiendo a diferentes criterios tales como: importancia a priori valorada por el usuario, método de los expertos, dispersión del rasgo o el carácter diferenciante del rasgo.
4. Añadir al a la posibilidad de otros tipos de aprendizaje, de modo que la base de casos se incremente a partir de nuevos casos resueltos de diversas formas.

BIBLIOGRAFÍA

- AAMODT, A. AND E. PLAZA Case-based reasoning. Foundational issues, methodological variations, and system approaches. *AI Communications*, 1994, 7(1), 39-59.
- ALLES, M. *Dirección estratégica de recursos humanos. Gestión por competencias*. Edtion ed. Garnica, Buenos Aires., 2000.
- ALLES, M. *Desarrollo del talento humano: basado en competencias*. Edtion ed.: Granica, 2005. ISBN 9789506411237.
- ÁLVAREZ, C. Modelo de Gestión de Recursos Humanos en el proyecto ERP Cuba. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos Universidad de las Ciencias Informáticas, 2012.
- ANDRÉ, M. Un modelo para la asignación de recursos humanos a equipos de proyectos de software. Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas Instituto Superior Politécnico "José Antonio Echeverría", 2009.
- ANDRÉS, R., J. L. GARCÍA-LAPRESTA AND L. MARTÍNEZ A multi-granular linguistic model for management decision-making in performance appraisal. *Soft Computing*, 2010/01/01 2010, 14(1), 21-34.
- ARAI, K. AND A. R. BARAKBAH Hierarchical K-means: an algorithm for centroids initialization for K-means. *Reports of the Faculty of Science and Engineering*, 2007, 36(1), 25-31.
- ARMSTRONG, M. *A handbook of human resource management practice*. Edtion ed.: Kogan Page Limited, 2003. ISBN 0749441054.
- ARYA, S., D. M. MOUNT, N. S. NETANYAHU, R. SILVERMAN, et al. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)*, 1998, 45(6), 891-923.
- AYALA, S. *Administración de los Recursos Humanos*. Edtion ed. Lima: Caballero Bustamante, 2004.
- BATCHELOR, B. G. *Pattern Recognition: Ideas in Practice*. Edtion ed. New York: Plenum Press, 1978. 71-72 p.
- BENTLEY, J. L. Multidimensional Binary Search Trees in Database Applications. *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, 1979, 5(4), 333-340.
- BERGMANN, R., J. L. KOLODNER AND E. PLAZA Representation in case-based reasoning. *The Knowledge Engineering Review*, 2005, 20(03), 209-213.
- BERN, M. Approximate closest-point queries in high dimensions. *Information Processing Letters*, 1993, 45(2), 95-99.
- BILES ET. AL., C. G., SCHULER, R. S. . *Audit Handbook of Human Resources Management Practices*. In *American Society for Personnel Administration*. 1986.
- BOHLANDER, G. W., S. SNELL AND A. SHERMAN *Managing Human Resources*. Edtion ed.: Mason: South-Western College., 2003.
- BONET, I. Modelo para la clasificación de secuencias, en problemas de la bioinformática, usando técnicas de inteligencia artificial. Tesis de doctorado Universidad Central "Marta Abreu" de Las Villas, 2008.
- BONISSONE ET. AL., P. P. Y. D., K.S. "Selecting Uncertainty Calculi and Granularity: An Experiment on Trading-off Precision and Complexity". In L.H.K.A.J.F. LEMMER. *Uncertain in Artificial Intelligence*,. North Holland, 1986, p. 217-247.
- BRIDGE, D., M. H. GÖKER, L. MCGINTY AND B. SMYTH Case-based recommender systems. *The Knowledge Engineering Review*, 2005, 20(03), 315-320.
- CALISOFT. Libro del Diagnóstico UCI - 2012. In. UCI, La Habana, 2012.
- CANÓS ET AL, L. C. E. T., LARA M. T., LIERN C. V., PÉREZ C. J. C. Modelos flexibles de selección de personal basados en la valoración de competencias 2008.
- CANÓS ET. AL., D., LOURDES, LIERN CARRIÓN, VICENTE. La agregación de información para la toma de decisiones en la empresa. In *XIV Jornadas de ASEPUMA y II Encuentro Internacional*. 2006.
- CANÓS ET. AL., L., T. CASASÚS, T. LARA, V. LIERN, J.C. PÉREZ. Un algoritmo fuzzy para la selección de personal basado en agregación de competencias. In *XV Jornadas de ASEPUMA y III Encuentro Internacional*. 2007.
- CANÓS, L., LIERN, V. Soft computing-based aggregation methods for human resource management. *European Journal of Operational Research*, 2008, 189.
- CARLSSON, C., R. FULLÉR Benchmarking and Linguistic Importance Weighted Aggregations. *Fuzzy Sets and Systems*, 1999.
- CELKO, J. *Joe Celko's Trees and hierarchies in SQL for smarties*. Edtion ed.: Elsevier, 2012. ISBN 0123877563.

CHARETTE, R. N. Why software fails. *IEEE Spectrum*, 2005, 42-49.

CHEN-TUNG CHEN, P.-F. P., WEI-ZHAN HUNG A two-phase fuzzy decision-making method based on multigranular linguistic assessment. *African Journal of Business Management*, 2012, 6.

CHEN-TUNG CHEN, W.-Z. H., HUI-LING CHENG APPLYING LINGUISTIC PROMETHEE METHOD IN INVESTMENT PORTFOLIO DECISION-MAKING. *International Journal of Electronic Business Management*, 2011, 9, 139-148.

CHEN, C.-T. AND W.-S. TAI. Measuring the intellectual capital performance based on 2-tuple fuzzy linguistic information. In *The 10th Annual Meeting of APDSI, Asia Pacific Region of Decision Sciences Institute*. 2005, vol. 20.

CHIAVENATO, I. *Gestión del Talento Humano*. Edtion ed. Bogotá. Colombia: Editorial Mc Graw Hill Interamericana S.A., 2002. ISBN 958-41-0288-5.

CHIAVENATO, I. *Administración de Recursos Humanos. El Capital Humano de las Organizaciones*. Edtion ed. México DF, México: McGraw-Hill, 2007.

CHONG, A. AND K. W. WONG 2008. On the Fuzzy Cognitive Map attractor distance. In *Proceedings of the IEEE Congress on Evolutionary Computation, Singapore 2008 IEEE*, 2652-2657.

CORMEN, T. H., C. STEIN, R. L. RIVEST AND C. E. LEISERSON. *Introduction to algorithms*. Edtion ed.: McGraw-Hill Higher Education, 2001. ISBN 0070131511.

COVER, T. AND P. HART Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 1967, 13(1), 21-27.

CUESTA, A. La gestión por competencias. 2001.

CUESTA, A. *Tecnología de Gestión de Recursos Humanos*. Edtion ed. La Habana: Editorial Academia, 2005a. ISBN 9592700532.

CUESTA, A. Metodología de Gestión por Competencias asumiendo la Norma Cubana sobre Gestión de Capital Humano. *Revista Brasileña de Gestión de Negocios*, 2010a, 13(40).

CUESTA, A. *Tecnología de gestión de recursos humanos*. Edtion ed. La Habana: Editorial Félix Varela y Academia, 2010b.

CUESTA, A. Evaluando desempeños: alineamiento estratégico y productividad. *Forum Empresarial*, 2012, 1(1), 1-30.

CUESTA, S. A. *Tecnología de Gestión de Recursos Humanos*. edited by E. ACADEMIA. Edtion ed. La Habana, 2005b. ISBN 9592700532.

CURTIS, B., W. HEFLEY AND S. MILLER *People CMM. A Framework for Human Capital Management*. edited by A. WESLEY. Edtion ed., 2009.

CUZA, B. Algoritmo para la elaboración de planes de formación profesional basado en competencias laborales para proyectos desarrolladores de software. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos Universidad de las Ciencias Informáticas, 2013.

DE ANDRÉS, R. Evaluación del Desempeño: Nuevos Enfoques desde las Teorías de Subconjuntos Difusos y de la Decisión Multi-criterio. Memoria presentada para optar al grado de Doctor con Mención Europea UNIVERSIDAD DE VALLADOLID, 2009.

DEMARCO, T. AND T. LISTER. *Peopleware: Productive Projects and Teams* [online]. 2da. [New York]: Dorset House Publishing Co., 1999.

DOUMPOS, M. A. C. Z. Preference disaggregation and statistical learning for multicriteria decision support: A review. *European Journal of Operational Research*, 2010, 209(3), 203-214.

DUBOIS, D., PRADE, HENRY *Fuzzy Sets and Systems. Theory and Application*. edited by I. ACADEMIC PRESS. Edtion ed., 1980.

DUCH, W. Similarity-based methods: a general framework for classification, approximation and association. *Control and Cybernetics*, 2000, 28(4).

DUTTA, S. AND P. P. BONISSONE *Integrating case based and rule based reasoning: The possibilistic connection*. Edtion ed.: INSEAD, 1990.

ESPINILLA, M., R. DE ANDRÉS, F. MARTÍNEZ AND L. MARTÍNEZ A 360-degree performance appraisal model dealing with heterogeneous information and dependent criteria. *Information Sciences*, 2012.

ESPINILLA, M., D. RUAN, J. LIU AND L. MARTÍNEZ. A heterogeneous evaluation model for assessing sustainable energy: a Belgian case study. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE, 2010, p. 1-8.

FARHAN, M., M. MANAF, N. M. NORWAWI AND M. H. A. WAHAB Faster Case Retrieval Using Hash Indexing Technique. *International Journal of Artificial Intelligence And Expert Systems (IJAE)*, 2011, 2(2), 81-95.

FAZHU, J., Y. XIUFENG AND S. KUN. Online Human Resource Management Personnel Performance Evaluation System of B2C E-commerce Businesses Based on Projection Pursuit Model. In *E-Business and E-Government (ICEE)*. 2010, p. 70-73.

FISHER, C. D., L. F. SCHOENFELDT AND J. B. SHAW *Human Resource Management*. Edtion ed. Boston: Houghton Mifflin Company, 2006.

FODOR, J. C. AND M. ROUBENS *Fuzzy preference modelling and multicriteria decision support*. Edtion ed.: Springer, 1994. ISBN 0792331168.

FRIEDMAN, J. H., J. L. BENTLEY AND R. A. FINKEL An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 1977, 3(3), 209-226.

GARCÍA, M. M. El empleo del razonamiento basado en casos en el desarrollo de Sistemas basados en el conocimiento para el diagnóstico. Trabajo de Tesis en opción al grado científico de Doctor en Ciencias Técnicas Universidad Central Marta Abreu, 1997.

GIONIS, A., P. INDYKY AND R. MOTWANIZ. Similarity Search in High Dimensions via Hashing. In *25th Conference Very Large Data Bases (VLDB)*. Edinburgh, Scotland, 1999.

GRAU, R., CORREA, C., ROJAS M. *Metodología de la investigación (Segunda Edición)*. Edtion ed.: UNIVERSIDAD DE IBAGUÉ CORUNIVERSITARIA, 2004.

GROUP, T. S. THE STANDISH GROUP REPORT. 2009.

GUTIÉRREZ, I. Un Modelo para la Toma de Decisiones usando Razonamiento Basado en Casos en condiciones de Incertidumbre. Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas Universidad Central Marta Abreu, 2003.

GUTIÉRREZ, I. AND R. E. BELLO. A Decision Case-Based System, That Reasons in Uncertainty Conditions. In M.T. ESCRIG, F. TOLEDO AND E. GOLOBARDES eds. *Topics in Artificial Intelligence*. Springer Berlin Heidelberg, 2002, vol. 2504, p. 54-63.

GUTIÉRREZ, I. AND R. E. BELLO Making decision in case-based systems using probabilities and rough sets. *Knowledge-Based Systems*, 2003, 16(4), 205-213.

GUTIÉRREZ, I., R. E. BELLO AND A. TELLERÍA Un sistema basado en casos para la toma de decisiones en condiciones de incertidumbre. *Revista Investigacion Operacional*, 2002, 23(2).

HAYES, J., A. ROSE-QUIRIE AND C. W. ALLINSON Senior managers' perceptions of the competencies they require for effective performance: implications for training and development. *Personnel Review*, 2000, 29(1), 92-105.

HERRERA, F., E. HERRERA-VIEDMA AND L. MARTÍNEZ A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *Fuzzy Systems, IEEE Transactions on*, 2008, 16(2), 354-370.

HERRERA, F., E. HERRERA-VIEDMA, L. MARTÍNEZ, F. MATA, et al. A multi-granular linguistic decision model for evaluating the quality of network services. In *Intelligent Sensory Evaluation*. Springer, 2004, p. 71-91.

HERRERA, F., HERRERA-VIEDMAN, E. Aggregation Operators for Linguistic Weighted Information. In *IEEE trans. on Systems, Man and Cybernetics - Part A: Systems and Humans*. 1997, vol. 27(4).

HERRERA, F. AND L. MARTÍNEZ A 2-tuple fuzzy linguistic representation model for computing with words. *Fuzzy Systems, IEEE Transactions on*, 2000, 8(6), 746-752.

HERRERA, F., MARTÍNEZ, L. A 2-tuple fuzzy linguistic representation model for computing with words. In *IEEE Transactions on Fuzzy Systems*. 2000, vol. 8(6), p. 746-752.

HERRERA, F. H.-V., E., AND MARTINEZ, L. . A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. In *IEEE Transactions on Fuzzy Systems*. 2008, vol. 16(2), p. 354-370.

JIMÉNEZ, A., S. RÍOS-INSUA AND A. MATEOS A decision support system for multiattribute utility evaluation based on imprecise assignments. *Decision Support Systems*, 2003, 36(1), 65-79.

JIN-HSIEN, W. AND H. JONGYUN A new version of 2-tuple fuzzy linguistic representation model for computing with words. *Fuzzy Systems, IEEE Transactions on*, 2006, 14(3), 435-445.

JING, H. Application of Fuzzy Data Mining Algorithm in Performance Evaluation of Human Resource. In *Computer Science-Technology and Applications, 2009. IFCSTA '09.*, 2009, vol. 1, p. 343-346.

KOLODNER, J. Making the Implicit Explicit: Clarifying the Principles of Case-Based Reasoning". In D.B. LEAKE. *Case-Based Reasoning: Experiences, Lessons, & Future Directions*. MIT Press, 1996.

KOLODNER, J. L. *Case-Based Reasoning*. Edtion ed., 1993.

KUZMINA, N. Metódicas investigativas de la actividad pedagógica [online]. 1970.

LAMY, J. B., A. ELLINI, J. NOBÉCOURT, A. VENOT, et al. Testing Methods for Decision Support Systems. In C.S. JAO ed. *Decision Support Systems*. InTech, 2010.

LAWRY, J. A methodology for computing with words. *International Journal of Approximate Reasoning*, 2001, 28(2–3), 51-89.

LI, Y., D. JIANG AND F. LI. The Application of Generating Fuzzy ID3 Algorithm in Performance Evaluation In *International Workshop on Information and Electronics Engineering (IWIEE)* Elsevier Ltd, 2012, vol. 29.

LIGEN, Y. AND D. ZHENLIN 2010. Research on the Evaluation of Enterprise's Core Competence Based on Unascertained Measure. In *Proceedings of the International Conference of Information Science and Management Engineering 2010* IEEE Computer Society.

LIU, X. AND H. LOU Parameterized Additive Neat OWA Operators with Different Orness Levels. *INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS*, 2006, 21, 1045–1072.

LUGO, J. A. Modelo para el control de la ejecución de proyectos basado en indicadores y lógica borrosa. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos Universidad de las Ciencias Informáticas, 2012.

MAHER, M. L. AND B. BALACHANDRAN. Flexible Retrieval Strategies for Case-Based Design. In J. GERO AND F. SUDWEEKS eds. *Artificial Intelligence in Design '94*. Springer Netherlands, 1994, p. 163-180.

MARTÍNEZ, L. Un nuevo modelo de representación de información lingüísticas basado en 2-tuplas para la agregación de preferencias lingüísticas Memoria para optar por el grado de Doctor en Informática Universidad de Granada, 1999a.

MARTÍNEZ, L. AND F. HERRERA An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges. *Information Sciences*, 2012, 207, 1-18.

MARTÍNEZ, L., HERRERA, F. An overview on the 2-tuple linguistic model for computing with words indecision making: Extensions, applications and challenges. In *Information Sciences*. 2012.

MARTÍNEZ, L., RUAN, D., HERRERA, F. Computing With Words in Decision Support Systems: An overview on Models and Application. *International Journal of Computational Intelligence Systems*, 2010, Vol. 3 No. 4.

MARTÍNEZ, L. L. Un nuevo modelo de representación de información lingüística basado en 2-tuplas para la agregación de preferencias lingüísticas. Universidad de Granada, 1999b.

MENDEL, J. M. Computing with words and its relationships with fuzzistics. In *Information Sciences*. 2007a, vol. 177(4).

MENDEL, J. M. Type-2 fuzzy sets and systems: an overview. *Computational Intelligence Magazine*, IEEE, 2007b, 2(1), 20-29.

MENDEL, J. M., ZADEH, LOTFI A., TRILLAS, ENRIC, YAGER, RONALD, LAWRY, JONATHAN, HAGRAS, HANI, GUADARRAMA, SERGIO. What Computing with Words Means to Me. In *IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE*. 2010.

MERIGÓ, L. J. M. Nuevas extensiones a los operadores OWA y su aplicación en los métodos de decisión. Universidad de Barcelona, 2008.

MING LI, Y. Z. The OWA-VIKOR method for multiple attributive group decision making in 2-tuple linguistic setting. *Journal of Convergence Information Technology*, 2012, 7.

MKAOUAR, R., E. MOULOUDI AND A. EL 2010. A competence based evaluation and selection problem by fuzzy linguistic computing. In *Proceedings of the International Symposium on Computational Intelligence and Design 2010* IEEE Computer Society.

MONDY, R. W. AND J. B. MONDY *Human Resource Management*. Edtion ed.: Prentice Hall, 2009. ISBN 9780136077282.

MONDY, W. AND R. NOE *Administración de Recursos Humanos*. Edtion ed.: Pearson Educación, 2005. ISBN 9789702606413.

MORALES, A. *Capital Humano. Hacia un sistema de gestión en la empresa cubana*. Edtion ed. La Habana: Política, 2009.

MUJA, M. AND D. G. LOWE. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In *VISAPP (1)*. 2009, p. 331-340.

MURPHY, K. R. AND J. N. CLEVELAND *Performance appraisal: An organizational perspective*. Edtion ed.: Allyn & Bacon, 1991. ISBN 0205123430.

O'HARA, S. AND B. A. DRAPER Are You Using the Right Approximate Nearest Neighbor Algorithm? IEEE Workshop on the Applications of Computer Vision (WACV), 2013.

OCN. Sistema de gestión INTEGRADA de los recursos humanos—VOCABULARIO: NC 3000: 2007 In. La Habana, Cuba: Oficina Nacional de Normalización (NC), 2007.

PACELLI, L. The Project Management: 18 Major Project Screw-Ups, and How to Cut Them Off at the Pass [online]. [New Jersey]: Prentice Hall, 2004.

PARK, C.-S. AND I. HAN A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. Expert Systems with Applications. Elsevier Science Ltd., 2002.

PAWLAK, Z. Rough Sets. International journal of Computer and Information Sciences, 1982, 11, 341356.

PCC. Lineamientos de la política económica y social del Partido y la Revolución. In. La Habana: VI Congreso del Partido Comunista de Cuba. Versiones Taquigráficas – Consejo de Estado, 2011.

PELAEZ, J. I., MESAS, A, LA RED, D.L Opinión de mayoría en toma de decisión en grupo mediante el operador QMA-OWA.

PÉREZ, L. Modelo para la evaluación por competencias en proyectos informáticos de la Universidad de las Ciencias Informáticas. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos Universidad de las Ciencias Informáticas, 2010.

PGDG. The PostgreSQL Global Development Group. PostgreSQL. 2013. Available from Internet:<<http://www.postgresql.org/>>.

PIÑERO, P. Y., S. TORRES, M. IZQUIERDO, J. LUGO, et al. GESPRO: Paquete para la gestión de proyectos. Nueva Empresa, 2013, 9(1).

PINO ET AL, M. A. M., SÁNCHEZ, M.C., PINO, M.L.Q. *Recursos Humanos*. edited by EDITEX. Edtion ed., 2008.

PIORUN, D. ¿Por qué fracasan los proyectos? . In.: Gerencia y Negocios en HispanoAmérica, 2003.

PMI *Guía de los Fundamentos para la Dirección de Proyectos (Guía del PMBOK)*. Edtion ed. Newtown Square, Pennsylvania 19073-3299 EE.UU. : Project Management Institute, Inc., 2008. ISBN 978-1-933890-72-2

PRAHALAD, C. AND G. HAMEL The core competence of the corporation. Boston (MA), 1990.

PRESSMAN, R. *Software Engineering: A Practitioner's Approach*. Edtion ed.: McGraw-Hill Science, 2004. , 2007.

RDCT. R Development Core Team. The R Project for Statistical Computing. 2013. Available from Internet:<<http://www.r-project.org/>>.

RODRÍGUEZ, E. Modelo para la evaluación de la composición de equipos de proyectos informáticos. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos Universidad de las Ciencias Informáticas, 2012.

ROUBENS, M. Fuzzy sets and decision analysis. Fuzzy Sets and Systems, 1997, 90(2), 199-206.

RUL ET AL *Administración de Recursos Humanos. 3ra Edición*. Edtion ed. Málaga: EDEA, 1996.

RUSPINI, E. H. Possibility as similarity: The semantics of fuzzy logic. In *Uncertainty in Artificial Intelligence*. 1990, vol. 6, p. 271.

RYAN, R. IT Project Management: Infamous Failures, Classic Mistakes, and Best Practices. MIS Quarterly Executive, 2007, 6(2).

SADIQ, R., M.J. RODRÍGUEZ, AND S. TEFAMARIAM. Integrating indicators for performance assessment of small water utilities using ordered weighted averaging (OWA) operators. In *Expert Systems with Applications*. 2010, vol. 37(7).

SAFARZADEGAN, G. S., HASSAN, SEBT MOHAMED, SHAHHOSSEINI, VAHID. Computing with words for hierarchical competency based selection of personel in construction companies. In *Applied Soft Computing*. 2011.

SÁNCHEZ, S. P. J. Modelos para la combinación de preferencias en toma de decisiones: Herramientas y aplicaciones. Universidad de Granada, 2009.

SEGARAN, T. Programming Collective Intelligence: Building Smart Web 2.0 Applications [online]. First Edition. O'Reilly Media, Inc., 2007.

SHIN, K.-S. AND I. HAN Case-based reasoning supported by genetic algorithms for corporate bond rating. Expert Systems with Applications, 1999, 16(2), 85-95.

SINGH, A. Architecture value mapping: using fuzzy cognitive maps as a reasoning mechanism for multi-criteria conceptual design evaluation. PhD Thesis Missouri University of Science and Technology, 2011.

SPSS-INC. SPSS® 13.0 Brief Guide [online]. 2004.

STANG, D. B. AND R. A. HANDLER. Magic Quadrant for Cloud-Based IT Project and Portfolio Management Services. I. GARTNER, 2013. G00247058.

TORRES, S. Modelo para la Gestión de Recursos Humanos en centro de desarrollo de sistemas de información. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos Universidad de las Ciencias Informáticas, 2011.

TUNING, P. Reflexiones y perspectivas de la Educación Superior en América Latina. Informe Final - Proyecto Tuning- América Latina 2004-2007. In. Universidad de Deusto: Bilbao, 2007.

TÜRKSEN, I. B. Meta-linguistic axioms as a foundation for computing with words. In *Information Sciences*. 2007, vol. 177(2), p. 332–359.

VAN SETTEN, M., M. VEENSTRA, A. NIJHOLT AND B. VAN DIJK. Case-based reasoning as a prediction strategy for hybrid recommender systems. In *Advances in Web Intelligence*. Springer, 2004, p. 13-22.

VERDECIA, E. Y. Metodología para la formación formativa de roles desde la práctica profesional. Tesis en opción al grado científico de Doctor en Ciencias Pedagógicas Universidad de las Ciencias Informáticas, 2011.

WANG, J.-H. AND J. HAO A new version of 2-tuple fuzzy linguistic representation model for computing with words. *Fuzzy Systems, IEEE Transactions on*, 2006, 14(3), 435-445.

WANG, X. AND Y. JIANG 2010. The Study of Enterprise Human Resource Evaluation Research Based on Neural Network. In *Proceedings of the Third International Symposium on Information Processing 2010* IEEE Computer Society.

WATSON, I. Is CBR a Technology or a Methodology? Tasks and Methods in Applied Artificial Intelligence, 1998, 525-534.

WAYNE ET AL., R. M., MONDY, J. B. *Human resource management*. Edtion ed. Upper Saddle River, N.J.:Prentice Hall, 2009. ISBN 9780136077282.

WESS, S., K.-D. ALTHOFF AND G. DERWAND. Using k-d trees to improve the retrieval step in case-based reasoning. In S. WESS, K.-D. ALTHOFF AND M. RICHTER eds. *Topics in Case-Based Reasoning*. Springer Berlin Heidelberg, 1994, vol. 837, p. 167-181.

WIERTMAN, M. J. *An Introduction to the Mathematics of Uncertainty*. edited by C. UNIVERSITY. Edtion ed., 2010.

WILKE, W. AND R. BERGMANN. Techniques and knowledge used for adaptation during case-based problem solving. In *Tasks and Methods in Applied Artificial Intelligence*. Springer, 1998, p. 497-506.

WILLIAMSON ET AL., C. *Gestión de Recursos Humanos*. Edtion ed., 2008.

XIAOFAN, C. AND W. FENGBIN. Application of Data Mining on Enterprise Human Resource Performance Management. In *Information Management, Innovation Management and Industrial Engineering (ICIII)*. 2010, vol. 2, p. 151-153.

YAGER, R. On ordered weighted averaging aggregation operators in multi-criteria decision making. In *IEEE Trans. On Systems, Man and Cybernetics*. 1988, vol. 18, p. 183-190.

YAGER, R. Families OWA operators. In *Fuzzy Sets and Systems*. 1993, vol. 59, p. 125-148.

ZADEH, L. Fuzzy Sets. *Information and Control*, 1965, 338-353.

ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning — I. *Information Sciences*, 1975a, 8(3), 199-249.

ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning — II. *Information Sciences*, 1975b, 8(4), 301-357.

ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning — III. *Information Sciences*, 1975c, 9(1), 43-80.

ZADEH, L. A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1978, 1(1), 3-28.

ZADEH, L. A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 1998, 2(1), 23-25.

ANEXOS

ANEXO I. Resumen comparativo de las herramientas de gestión de proyectos existentes

Tabla 1. Resumen comparativo de las herramientas de gestión de proyectos existentes.

No.	Herramientas	1. Soporte sobre la web	4. Tipo de licenciamiento	6. Entorno colaborativo	7. Seguimiento de problemas	8. Planificación	9. Gestión del portafolio de proyectos	10. Gestión de recursos	11. Gestión de documentos	12. Flujo de trabajo	13. Reportes y análisis	14. Gestión del presupuesto	15. Tiempo de seguimiento	16. Facturación	17. Manejo de incertidumbre	18. Planificación	19. Adquisición	20. Desarrollo	20.1 Evaluación de competencias	22. Gestión
1	2-plan	Si	Propietaria	Si	No	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	Si
2	Anyplan	No	Propietaria	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si	Si	No	No	Si
3	Assembla	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	No	Si	No	No	Si	Si	No	No	Si
4	AtTask	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si	Si	No	No	Si
5	Celoxis	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	Si	Si	No	No	Si
6	dotProject	Si	GPL	Si	Si	No	Si	Si	Si	No	Si	No	No	No	No	Si	Si	No	No	Si
7	Eclipse PPM software	Si	?	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	Si
8	EPM Live	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	Si	Si	No	No	Si
9	Endeavour Software Project Management	Si	GPL	Si	Si	Si	Si	Si	Si	No	No	No	No	No	No	Si	Si	No	No	Si
10	enQUIRE	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	Si	Si	No	No	No
11	FastTrack Schedule	No	Propietaria	Si	No	Si	Si	Si	No	Si	Si	No	No	No	No	Si	Si	No	No	Si
12	FusionForge	Si	GPL	Si	Si	Si	Si	No	Si	Si	No	No	Si	No	No	Si	Si	No	No	Si
13	Gemini	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	No	Si	No	No	Si	Si	No	No	Si
14	HP Project & Portfolio Software	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	No
15	InLoox	Si	Propietaria	Si	Si	Si	No	Si	Si	No	Si	Si	Si	Si	No	Si	Si	No	No	No
16	in-Step	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si	Si	No	No	Si
17	Microsoft Office Project Server	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si	Si	No	No	Si
18	Microsoft Project	No	Propietaria	No	No	Si	No	Si	No	No	Si	Si	Si	Si	No	Si	Si	No	No	Si
19	OpenERP	Si	Open source	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	Si	Si	No	No	Si
20	OpenProject	Si	GPL	Si	Si	Si	Si	No	Si	Si	No	No	No	No	No	No	Si	No	No	Si
21	Oracle Primavera EPPM (Primavera P6)	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	No
22	PlannerSuite	No	Propietaria	No	No	Si	Si	Si	Si	No	No	No	Si	Si	No	Si	Si	No	No	Si
23	ProjectManager.com	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	Si
24	Project.net	Si	GPL	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	Si
25	Project-Open	Si	Propietaria, GPL	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	Si
26	Redmine	Si	GPL	Si	Si	Si	Si	No	Si	Si	No	No	No	No	No	Si	Si	No	No	Si
27	SAP Business ByDesign	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	No	No	No	No	Si	Si	No	No	Si
28	VPMi	Si	Propietaria	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si	Si	Si	No	Si
29	GESPRO 13.05	Si	GPL	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	Si

ANEXO II. Información empleada en el cuasiexperimento

Tabla 2. Información para el cuasiexperimento.

Atributos del caso						Proceso de Evaluación	Resultado del Algoritmo
ID Caso	ID Competencia	CRE	CRP	CRC	CC	Evaluación	Evaluación
131	34	0.87	0.17	0.16	0.16	Medio	Medio
	32	0.87	0.17	0.16	0.15	Medio	Medio

13	26	0.08	0.001	0.13	0.13	Muy Bajo	Muy Bajo
	11	0.08	0.001	0.14	0.14	Muy Bajo	Muy Bajo

92	25	0.69	0.22	0.15	0.15	Medio	Medio
	8	0.69	0.22	0.11	0.11	Medio	Medio

137	39	0.99	0.61	0.10	0.10	Alto	Alto
	18	0.99	0.61	0.23	0.23	Alto	Alto

....

Tabla 3. Datos para determinar efectividad de CA.

ID Caso	ID Competencia	Clase Real	Clase según CA
131	34	Medio	Medio
	32	Medio	Medio

92	25	Medio	Medio
	8	Medio	Medio

....

Tabla 4. Datos para determinar efectividad de CD.

ID Caso	ID Competencia	Clase Real	Clase según CD
131	34	Medio	Medio
	32	Medio	Medio

92	25	Medio	Medio
	8	Medio	Medio

....

ANEXO III. Encuesta para la aplicación del método Iadov

Encuesta para la validación del algoritmo basado en casos para evaluar competencias profesionales.

Pregunta 1. ¿Considera usted que se debe continuar realizando la evaluación de las competencias de un individuo sin un algoritmo que permita el análisis de las evidencias previas y las evaluaciones de los expertos en procesos de anteriores?

Sí No No sé

Pregunta 2. ¿Si UD. requiere elevar el nivel de fiabilidad en el proceso de evaluación de competencias de los individuos en el desempeño de los roles profesionales utilizaría usted el algoritmo propuesto?

Sí No No sé

Pregunta 3. ¿Satisface la propuesta los niveles de fiabilidad esperado por usted?

Sí No No sé

Pregunta 4. ¿Qué añadiría usted al algoritmo propuesto para aumentar su fiabilidad? Argumente.

Pregunta 5. ¿El algoritmo propuesto contribuye a agilizar el proceso de evaluación de las competencias profesionales de un individuo?

Sí No No sé

Pregunta 6. ¿Satisface la propuesta los niveles de fiabilidad esperado por usted? Marque con una X cuál sería su respuesta.

Muy satisfecho.	
Parcialmente satisfecho.	
Me es indiferente.	
Más insatisfecho que satisfecho.	
Para nada satisfecho.	
No sé qué decir.	

Pregunta 7. ¿Cuáles considera son las principales fortalezas del algoritmo que contribuyen a la fiabilidad de las respuestas? Argumente.

ANEXO IV. Información del entorno de prueba

Tabla 5. Información del entorno de prueba utilizado.

Características	Descripción
Entorno	Virtualizado
Sistema operativo	Debian 7 64 bits
RAM	1 GB
Cantidad de procesadores	2
Cantidad de núcleos	4