

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 5. Laboratorio de Gestión de Proyectos

Facultad 1. Departamento de Técnicas de Programación



Procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos

Trabajo final presentado en opción al título de

Máster en Gestión de Proyectos Informáticos

Autor: Ing. Javier Heredia Ruiz

Tutor: Dra.C Surayne Torres López

La Habana, 1 de marzo de 2017

DEDICATORIA

A mi madre, que tanto esfuerzo ha hecho para que estuviera hoy aquí.

A la Revolución cubana y al Comandante en Jefe Fidel Castro Ruz que me dieron la posibilidad de hacerlo.

A mis amigos, sin quienes no hubiese podido llegar a esta meta.

A al claustro de la maestría de Gestión de Proyectos Informáticos en especial a Pedro y Surayne.

A todo el que me aconsejó cuando las ganas de abandonar el barco venían a mi mente.

A todos los que de una u otra manera contribuyeron a obtener este anhelado logro.

AGRADECIMIENTOS

Solo el que llega hasta aquí es capaz de imaginar la magnitud del esfuerzo realizado. Ha quedado atrás mucho estudio, trabajo y esfuerzo personal.

Quiero agradecer al colectivo de profesores de la maestría por sus enseñanzas que me guiaron hacia la meta, a todas las personas que han hecho posible la realización de esta tesis y que han tenido la gran paciencia de atenderme.

A mi tutora, que ha sido un ejemplo de dedicación y sabiduría.

A mis familiares, quienes me brindaron en todo momento su cooperación.

A todos, mis agradecimientos

El Autor.

DECLARACIÓN JURADA DE AUTORÍA Y AGRADECIMIENTOS

Declaro por este medio que yo Javier Heredia Ruiz, con carné de identidad 83110524587, soy el autor principal del trabajo final de maestría: “Procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos” 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 1 días del mes de marzo del año 2017.

Nombre y firma del maestrante

RESUMEN

En el desarrollo de proyectos, las instituciones emplean herramientas que contribuyan a la evaluación de competencias laborales. En ambientes que presentan incertidumbre en las evidencias del desempeño que tributan a la evaluación del personal en los proyectos de software, puede ser ineficiente el uso de esas herramientas según datos de encuestas realizadas. Las deficiencias en el manejo de las evidencias del desempeño durante el proceso de evaluación de competencias laborales en la ejecución de proyectos, afecta la objetividad del proceso de evaluación de competencias laborales. Una variante de solución es la introducción de técnicas de minería de textos como herramienta para la recuperación de información, haciendo especial énfasis en las funciones de similitud de textos. La presente investigación tiene como objetivo proponer un procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos que permita mejorar la eficiencia del proceso de evaluación de competencias laborales, a través de las fases y actividades que define. Los resultados de la aplicación del procedimiento se logran a partir de la ejecución de pruebas de Friedman para la comparación de tres roles aleatorios tomándose las evidencias del Sistema para la gestión de proyectos y posteriormente la realización de pruebas no paramétricas de Wilcoxon. La validación del procedimiento propuesto, en el módulo de Recursos humanos del Xedro-GESPRO 13.05, evidencia mejoras en el proceso de evaluación de competencias.

Palabras clave: Evaluación de competencias, funciones de similitud de textos, gestión de proyectos, minería de textos, recursos humanos.

Abstract

The deficiencies in the analysis of evidences that tax the evaluation of the personnel in the software projects, affects the efficiency with which the process of evaluation of labor competences is realized. From the analysis of the evidence and the evaluation process in the software development cycle, difficulties have been identified that undermine the veracity of the software. The main objective of this research is to propose a procedure for the use of functions that corroborate the relationship between the tasks and the labor competencies, as well as; the analysis of the evidences that tax the evaluation of the personnel in software projects and to improve the efficiency in the process of evaluation of labor competences. A study of the main bibliographies is carried out, allowing the identification of the criteria to be evaluated, obtaining the basis for the evaluation. An evaluation of the mining of texts as a tool for the discovery of patterns, with special emphasis on techniques and measures of text comparison. The measures of similarity were selected that showed the best resulted in the analysis of short texts. The results are obtained from the execution of Friedman tests for the comparison of three random roles and taking 35% of the evidences of these roles of the Project Management System and later nonparametric Wilcoxon tests were performed. The validation of the proposed procedure, adhered to the human resources module of Xedro_Gespro 13.05, evidences improvements in the competency evaluation process.

Keywords: *Competency assessment, features of similarity of text, human resources, project management, text mining.*

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA.....	9
1.1 Introducción	9
1.2 Análisis bibliométrico.....	9
1.3 Conceptualización de competencias	9
1.3.1 Análisis de evidencias del desempeño en la evaluación de competencias.....	10
1.3.3 Errores y deficiencias en el proceso de evaluación de competencias.....	12
1.3.4 El descubrimiento de conocimiento y la Minería de Datos	13
1.3.5 La minería de textos como herramienta para el descubrimiento de patrones	13
1.4 Selección de funciones de similitud para comparación de textos cortos	14
1.5 Conclusiones del capítulo	22
CAPÍTULO 2. PROCEDIMIENTO PARA LA EVALUACIÓN DE COMPETENCIAS LABORALES BASADO EN FUNCIONES DE SIMILITUD DE TEXTOS.....	24
2.1 Introducción	24
2.2 Procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos	24
2.3 Características generales del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos.....	25
2.3.1 Capacidades o competencias del personal ejecutor del procedimiento.....	26
2.3.2 Premisas para la aplicación de la propuesta de solución	26
2.5 Fases del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos.....	27
2.5 Fase 1: Procesamiento de las evidencias del desempeño	28
2.5.1 Establecer fecha de corte.....	28
2.5.2 Obtener las competencias	29
2.5.3 Obtener las evidencias del desempeño.....	30
2.5.4 Crear estructura	30
2.6 Fase 2: Análisis de las evidencias del desempeño	31
2.6.1 Selección y extracción de características	31
2.6.2 Reducción de dimensionalidad.....	32
2.6.3 Normalización de los datos.....	33
2.6.4 Limpieza de los datos	33
2.7 Fase 3: Obtención de similitud de las evidencias del desempeño	34
2.7.1 Calcular similitud entre las evidencias del desempeño	35
2.7.2 Comparar la similitud con el umbral	37

2.8 Fase 4: Agrupamiento de las similitudes de las evidencias del desempeño	39
2.9 Valoración de los resultados.....	40
2.10 Conclusiones del capítulo	40
CAPÍTULO 3. VALIDACIÓN DE LA PROPUESTA Y ANÁLISIS DE RESULTADOS	41
3.1 Introducción	41
3.2 Caracterización del entorno de aplicación	41
3.3 Ejecución del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos.....	42
3.3.1 Fase 1: Procesamiento de las evidencias del desempeño	42
3.3.2 Fase 2: Análisis de las evidencias del desempeño	43
3.3.3 Fase 3: Obtención de similitud de las evidencias del desempeño	45
3.3.4 Fase 4: Agrupamiento de las evidencias del desempeño	45
3.4 Comparación respecto a la eficiencia del procedimiento.....	49
3.4.1 Evaluación a través de Escalas Numéricas	49
3.4.2 Evaluación del procedimiento integrado al Modelo de evaluación de competencias a partir de evidencia utilizando Sistemas Basados en Conocimiento	50
3.4.3 Comparación de la eficiencia de las variantes para la ejecución del procedimiento	50
3.5 Valoración de los modelos.....	51
3.6 Análisis de la adaptabilidad del procedimiento.....	52
3.7 Valoración del impacto económico y social de la propuesta	53
3.8 Conclusiones del capítulo	55
CONCLUSIONES.....	56
RECOMENDACIONES	57
REFERENCIAS BIBLIOGRÁFICAS.....	58
ANEXOS.....	67

INTRODUCCIÓN

La definición de recursos humanos fue introducida a partir de los años ochenta impactando con cambios en el modo de considerar a las personas en las organizaciones. Este se concreta conceptualmente debido a que el personal que trabaja en una organización ha pasado de ser considerado un recurso más a ser una de las principales fuentes, posiblemente la principal, de ventajas competitivas de las empresas. Hoy día se hace alusión a capital intelectual, capital humano, potencial humano, talento humano, todos referidos al papel que tiene el hombre en la organización, surgiendo así, nuevas definiciones que dan respuesta a nuevas situaciones. El término competencias irrumpe el campo de la gestión de recursos humanos, y consecuentemente las disciplinas científicas que abordan este objeto (Gil, 2007).

Los elementos fundamentales que determinan el éxito de los proyectos dedicados al desarrollo de software, se encuentra el relacionado con los recursos humanos. La gestión del talento humano ha pasado a ser una función eminentemente estratégica y dinámica que proyecta como su principal clave de éxito el factor humano (Torres y otros, 2014). Por esa razón según el desempeño o comportamiento en las labores que realiza el recurso humano, será el resultado esperado o no de la organización.

Tomar en cuenta las competencias implica no solo atender a rasgos psicológicos de las personas, sino a comportamientos observables que son consecuencia de un conjunto de motivaciones, rasgos de personalidad, actitudes, valores, conocimientos, aptitudes y habilidades (Mitrani y otros, 1992). Asimismo, la evaluación de las personas no se habrá de apoyar exclusivamente en los exámenes y pruebas psicométricos, característicos del enfoque tradicional, sino que es necesario tener el recurso o técnica que permita tomar en consideración las experiencias laborales y los comportamientos exhibidos en el desempeño de un puesto de trabajo.

Los resultados dados por los sistemas de competencias han tenido logros muy efectivos tanto para la persona que desarrolla su trabajo como para que la empresa se encamine bien, hacia una meta firme (Muñoz, 2012). Aludiendo al tema de la evaluación en la gestión de competencias, Alles (Alles, 2006) refiere que: “Al implementar un sistema de gestión por competencias, las empresas se preguntan: ¿Cómo están mis trabajadores en relación con las competencias definidas? ¿Las cubren? ¿Deberé reemplazarlos? ¿Es posible desarrollar las competencias? Estas y otras interrogantes solo podrán responderse si la organización lleva a cabo la medición del nivel de desarrollo de las competencias de su personal” que en otras palabras se denomina evaluación de competencias.

La evaluación de competencias puede entenderse como un proceso por el cual se recoge información acerca de las competencias desarrolladas por un individuo y se comparan estas con el perfil de competencias requerido por un puesto de trabajo, de tal manera que pueda formularse un juicio de valor sobre el ajuste al mismo. Las competencias de los individuos son evaluadas antes de su incorporación a una organización, cuando se realiza la selección de personal, y además la evaluación de competencias estará presente 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 (Gil, 2007).

Hoy en Cuba, un país que tiene el objetivo de insertarse dentro de la alta competitividad empresarial, aplica la gestión por competencias y su evaluación en sus esferas productivas incluyendo la industria de desarrollo de software, en aras del incremento de la eficiencia y la eficacia en el desempeño laboral. La gestión por competencias, (y dentro de ella, como factor esencial la evaluación de las competencias) es una alternativa de reconocido éxito para alcanzar los objetivos de una empresa. Esta experiencia tiene un enfoque hacia las personas e incide de manera equilibrada en los procesos definidos dentro de la gestión de proyectos (Pérez, 2010).

Se identifica en todos los niveles que la evaluación de competencias presenta deficiencias que por consiguiente emanan problemas a la hora de emitir una valoración definitiva del proceso (Mano y Moro, 2009). Cuando la propia organización es capaz de asumir el rol de evaluador, pueden verse condicionados los resultados de la evaluación, debido a las relaciones que puedan existir entre los trabajadores de la entidad. Seguidamente no existe una garantía que las personas que se desempeñan como evaluadores estén capacitadas para desarrollar esa labor (Levy-Leboyer, 1997).

Por otro lado, además de una perspectiva amplia sobre el desempeño de los sujetos o sobre su potencial, la evaluación también requiere de experiencia, conocimiento profundo del puesto de trabajo para el que se realiza la evaluación y capacidad para el análisis y comprensión de la evaluación, que en ocasiones son elementos de los cuales carecen los evaluadores. Esto unido al engorroso proceso de analizar cada una de las evidencias, afectan la eficiencia con que se realiza el proceso de evaluación de competencias.

Conceptualizando el término de eficiencia, según la Norma Cubana NC 3000: 2007 (Oficina Nacional de Normalización, 2007) referente al Sistema de Gestión Integrada de Capital Humano — Vocabulario refiere que es la relación entre el resultado alcanzado y los recursos utilizados.¹ Existen entidades que se dedican total o parcialmente a la producción de software, donde un elemento de éxito lo constituye el análisis de la eficiencia en sus procesos.

La Universidad de las Ciencias Informáticas (UCI) es un centro de estudios que tiene la misión de producir software y brindar servicios informáticos. La UCI es considerada como un centro de estudio-producción, donde sus estudiantes, profesores y especialistas se integran a la producción de software nacional e internacional de manera directa e indirecta.

Por la prioridad que representa la producción de software cada vez de mayor calidad y que satisfaga las necesidades de los clientes es que se necesita gestionar de manera eficiente los recursos humanos en la

¹ NOTA: Así aparece en el apartado 3.2.15 de la NC ISO 9000:2005

Universidad. La evaluación del desempeño para la formación, la remuneración monetaria, el cambio o movimiento de puestos de trabajos se presenta indudablemente como factor clave en el camino que conduce al perfeccionamiento y al éxito.

Sin embargo, dentro de la gestión de los Recursos Humanos (RRHH) en la UCI el proceso de evaluación de competencias de sus profesionales no ha madurado lo suficiente. Las soluciones asociadas a la gestión de los recursos humanos ofrecidas por autores como (Aput, 2014), (Pérez, 2010), (Torres, 2011) (Rodríguez, 2012), (Álvarez, 2012) (Cuza, 2013), (Cervantes, 2015) profesores de la universidad, se enuncian a continuación:

- ✓ Definir indicadores para determinada área como el software multimedia.
- ✓ Evaluar competencias en los estudiantes, evidenciando claramente muchas lagunas que convergen en deficiencias múltiples a la hora de evaluar las competencias laborales de los profesionales.
- ✓ Definir y establecer los procesos de planificación, adquisición, desarrollo y gestión de los recursos humanos.
- ✓ Enfocarse a la gestión integrada del capital humano en un proyecto específico haciendo alusión a la evaluación de competencias.
- ✓ Evaluar las competencias en equipos de desarrollo desde el punto de vista genérico, en hacer un análisis de cómo elaborar los planes de formación profesional tomando la evaluación como una entrada.

Actualmente en la universidad se realizan procedimientos de evaluación de competencias. En el proceso se identifica un factor común: las evidencias. Según la Norma Cubana NC 3000: 2007 referente al Sistema de Gestión Integrada de Capital Humano — Vocabulario se refiere a evidencia objetiva a los datos que respaldan la existencia o veracidad de algo². Las evidencias constituyen el resultado de las tareas desarrolladas por el personal y avalan su desempeño en el proceso de evaluación de competencias (ver anexo 1). Son almacenadas en Xedro-GESPRO, una herramienta para la dirección integrada de proyectos desarrollada en la UCI y utilizada por varias organizaciones desarrolladoras de proyectos de diversa índole (Piñero y otros, 2014).

El evaluador debe analizar cada una de las evidencias del individuo, identificar cuáles tributan a las competencias que él debe poseer según su rol y definir el peso que tiene una determinada evidencia en una o varias competencias. Luego, realizar un análisis de estos elementos y como resultado, emitir una evaluación.

Una encuesta realizada arrojó que el 90% de los encuestados evidenciaron insatisfacción e inconformidad en el proceso de evaluación debido a que los juicios subjetivos de los evaluadores y la falta de unificación de criterios proporcionan lugar a resultados evaluativos muy diversos. Se ha dado el caso que dos profesionales que han tenido el mismo desempeño en un tiempo determinado, han obtenido evaluaciones diferentes. Estos resultados

² NOTA Así aparece en el apartado 3.8.1 de la NC ISO 9000:2005.

obtenidos fueron de la encuesta diseñada y aplicada por especialistas en el tema asociado a la identificación de la relación entre las tareas y las competencias laborales en los proyectos.

Se ha podido constatar que otro elemento que atenta con la veracidad de la evaluación es la correspondencia entre lo plasmado por el evaluador y el evaluado. En este aspecto el nivel de intrusión de los evaluadores es extremadamente alto, y existe el riesgo de acarrear errores en el proceso de evaluación. Los evaluadores no siempre cuentan con la capacitación suficiente para llevar a cabo el análisis de evidencias (Gil, 2007).

La pérdida de tiempo y esfuerzo de los evaluadores en esta tarea, mayor de lo planificado es otro de los problemas que se identificó en la encuesta. El análisis de las evidencias está caracterizado por una difícil identificación de la relación existente entre las evidencias y las competencias, y la asignación de un determinado peso, debido al número de evidencias. Todos estos inconvenientes traen consigo que en ocasiones los evaluadores en los proyectos no realicen el proceso de evaluación de competencias a los miembros de su equipo con la eficiencia y la eficacia requerida, resultando de manera global una ralentización de la evaluación e inconformidad con la misma, dificultándose dicho proceso tan vital en las organizaciones (Oficina Nacional de Normalización, 2007).

En la universidad hay muchos trabajadores vinculados a la producción de software que se enfrentan a este tipo de proceso evaluativo, lo cual puede acarrear como se evidenció en la encuesta realizada insatisfacciones, diversidad de criterios, falta de precisión y objetividad en los resultados. Las evidencias del desempeño de los miembros de los proyectos se encuentran en las tareas que los mismos realizan, en forma de texto y se encuentran almacenadas en la herramienta de gestión de proyecto. El proceso de evaluación de competencias está sujeto a la forma en que se determina la correlación y correspondencia entre las tareas y las competencias laborales. Por tanto, se necesita disponer de algún mecanismo que estandarice el proceso de evaluación de competencias laborales. Se plantea el siguiente **problema científico**: Las deficiencias existentes en el análisis de evidencias que tributan a la evaluación de competencias laborales, afecta la eficiencia con que se realiza el proceso de evaluación de competencias en los proyectos de software.

Objeto de investigación:

El proceso de evaluación de competencias en los proyectos de desarrollo de software.

Objetivo general:

Elaborar un procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos en las evidencias del desempeño durante el desarrollo de software que contribuya a mejorar la eficiencia del proceso de evaluación de competencias laborales.

Del mismo se derivan los siguientes **objetivos específicos**:

1. Elaborar el marco teórico de la investigación identificando las principales tendencias referentes al análisis de evidencias para la evaluación de competencias laborales.

2. Desarrollar un procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos que mejore la eficiencia en el análisis de evidencias en la evaluación de competencias laborales en los proyectos de software.

3. Validar el procedimiento propuesto a partir de la comprobación de la validez en un pre-experimento y su implementación en el Módulo de RRHH del paquete de gestión de proyectos Xedro-Gespro 13.05.

Campo de acción:

El campo de acción se enmarca en el proceso de análisis de evidencias que tributan a la evaluación del personal en los proyectos de software.

Tipo de investigación

Se define una **investigación de tipo correlacional**, en la cual no solo se especifican las características más significativas de las principales tendencias de evaluación de competencias laborales, así como, los métodos y funciones, constituyéndose así la base de la investigación, sino que también se manejan variables donde se identifican las relaciones que existen entre las mismas y sus valores a través de cálculos estadísticos.

Está guiada la investigación por la **hipótesis** de que:

El procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos se logrará mejorar la eficiencia del proceso de evaluación de competencias laborales en el desarrollo de software.

Operacionalización de las variables dependientes e independientes

A partir del objetivo y la hipótesis planteada se determina:

Variable Independiente:	Indicadores	Unidad de Medida (UM)
Procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos	Reusabilidad	Alta (3), Media (2), Baja (1)
	Aplicabilidad	Alta (3), Media (2), Baja (1)
	Precisión (es la fracción de evidencias recuperadas que son relevantes para la necesidad de información del usuario.)	%
	Recobrado (es la fracción de evidencias relevantes para una consulta que fueron recuperados.)	%
	Proposición de Fallo o Fall Out (es la proporción de evidencias no relevantes que son recuperados, fuera de todos los documentos relevantes disponibles.)	%

Variable Dependiente:	Indicadores	Unidad de Medida (UM)
Eficiencia en el proceso de evaluación de competencias laborales.	Tiempo dedicado al análisis de evidencias	(cantidad de evidencias analizadas/hora)
	Esfuerzo dedicado al análisis de las evidencias de una evaluación	(horas-hombre)

Para el desarrollo de la investigación se utilizaron varios **métodos de investigación**. El método sistémico a partir del análisis bibliográfico permitió obtener las principales tendencias en el área de recursos humanos referentes a la evaluación de competencias laborales, así como, los métodos y funciones para desarrollar sus procesos internos en proyectos de software. El método hipotético-deductivo permitió plantear objetivos específicos y una hipótesis a partir del problema, durante de la investigación fueron resueltos. La modelación permitió representar la propuesta de solución y otros conceptos, de manera que fuera fácil su comprensión. El histórico-lógico facilitó analizar funciones que mejoren el análisis de evidencias en la evaluación del personal en los proyectos de software.

La encuesta proveyó información sobre los criterios que tiene el personal involucrado, respecto al proceso de evaluación de competencias en los proyectos de software y la observación permitió realizar una percepción planificada y dirigida del fenómeno estudiado. El método pre-experimental permitió demostrar el vínculo entre los fenómenos objeto de investigación.

Muestreo

Población: La misma está constituida por las evidencias correspondientes a competencias laborales de los roles en proyectos de la UCI.

Técnica de muestreo: La muestra se tomó en forma dirigida no probabilística, seleccionándose tres roles aleatorios y tomándose las evidencias de estos roles del Sistema de Gestión de Proyectos (Xedro-GESPRO) de la UCI.

Diseño de investigación

El diseño de la investigación es pre-experimental siguiendo la estrategia de pre-prueba/post-prueba con un solo grupo.

Se analizará el comportamiento de las variables dependientes de la investigación a partir de la aplicación del procedimiento propuesto, estableciéndose las comparaciones pertinentes de acuerdo con los escenarios anterior y posterior respectivamente a la aplicación. El grupo de control estará constituido por una muestra de 3 bases de almacenamiento de evidencias de los centros de desarrollo de la facultad 1 en la UCI, sin utilizar el procedimiento, y el grupo experimental lo representa la misma muestra después de aplicado el procedimiento.

Análisis estadístico e instrumentos

Para realizar el análisis estadístico de los resultados se utilizarán pruebas de Friedman para la comparación de N muestras apareadas y posteriormente si se encuentran diferencias significativas, se realizarán sucesivas pruebas no paramétricas de Wilcoxon.

Como instrumentos para el análisis de la información se utilizarán los indicadores de calidad de los datos se utilizará el lenguaje y ambiente estadístico R y la herramienta SPSS.

Aportes prácticos de la investigación:

Procedimiento para realizar el análisis de evidencias del desempeño, el cual permite que la evaluación competencias laborales sea de forma semiautomática, disminuyendo el nivel de intrusión de los usuarios evaluadores y el esfuerzo y tiempo requerido para ello.

Listado de publicaciones, eventos y avales de la investigación.

Javier Heredia Ruiz, Propuesta que integra los sistemas de gestión bibliotecaria con las redes de telefonía móvil, La Habana, Cuba, Revista Ciencias de la Información, en el No. 1, Vol. 43, de 2012.

Javier Heredia Ruiz, Modelo de desarrollo de software y Guía de implantación en un centro productivo, Las Tunas, Cuba, ISBN 978-959-16-2427-7, 2015.

Javier Heredia Ruiz, Propuesta de política de salvaguarda de información para la Empresa de Mecanizado de Aluminio, Las Tunas, Cuba, ISBN 978-959-16-2427-7, 2015.

Javier Heredia Ruiz, Análisis de Fiabilidad del Sistema GESPRO v1.0 desarrollado en la Universidad de las Ciencias Informáticas (UCI), LACCEI 2015, Santo Domingo República Dominicana, SCOPUS y EBSO, 2015.

Javier Heredia Ruiz, Análisis de rendimiento de servidores en un proyecto de desarrollo de software, LACCEI 2015, Santo Domingo República Dominicana, SCOPUS y EBSO, 2015.

Javier Heredia Ruiz, Propuesta de proceso de selección del Capital Humano en el Centro de Desarrollo de Soluciones Integrales de Gestión de Entidades (CEIGE), LACCEI 2015, Santo Domingo República Dominicana, SCOPUS y EBSO, 2015.

Javier Heredia Ruiz, Surayne Torres López, Funciones para el análisis de textos cortos en proyectos de software, Informática 2016, La Habana, Cuba.

Javier Heredia Ruiz, Damián Cervantes Rodón, Los laboratorios virtuales, herramienta decisiva en el proceso de enseñanza-aprendizaje de la Ingeniería en Ciencias Informáticas, Informática 2016, La Habana, Cuba.

Javier Heredia Ruiz, Surayne Torres López, Funciones para el análisis de textos cortos en proyectos de software, revista Nueva Empresa, número especial, marzo de 2016, ISSN:1682-2455, RNSP:0472, La Habana, Cuba.

Estructura de la tesis:

La presente investigación consta de tres capítulos.

En el **primer capítulo** se elabora un marco teórico en torno a las principales tendencias de la evaluación por competencias dentro del área de gestión de recursos humanos, así como, los métodos basados en evidencias, consultando para ello la bibliografía especializada. Se realiza un análisis del entorno organizacional en cuestión y se analizan técnicas de minería de texto que resultan novedosas en esta área, específicamente las funciones de similitud de texto, para constituir la base de la solución de la investigación.

En el **segundo capítulo** se presentan los elementos que integran la propuesta de la investigación. Se parte de las características generales del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos. Se describen las fases por las que está compuesto, y las actividades establecidas en cada una de ellas, puntualizándose las entradas, salidas y herramientas y técnicas que utiliza. Finalmente, se realizan las conclusiones parciales.

En el **tercer capítulo** se muestra una síntesis de la aplicación del procedimiento propuesto, seguidamente se valida, luego del análisis bibliográfico para ello. Para este propósito se utilizaron funciones programadas en *AnalysisPro*, haciendo uso del paquete *pg_similarity*, con las tres variantes para el cálculo de la similitud entre los textos de las tareas de los miembros del proyecto y las competencias laborales. Se concluye con un análisis del impacto económico y social de la propuesta de solución.

Cuenta además la investigación con **conclusiones generales, recomendaciones, referencias bibliográficas y anexos.**

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

1.1 Introducción

En el presente capítulo se muestran los aspectos teóricos concernientes a las principales tendencias de la evaluación por competencias dentro del área de gestión de recursos humanos, así como, los métodos basados en evidencias, consultando para ello la bibliografía especializada. Se realiza un análisis del entorno organizacional en cuestión y se analizan técnicas de minería de datos y de textos que resultan novedosas en esta área, específicamente las funciones de similitud de texto, para constituir la base de la solución de la investigación.

1.2 Análisis bibliométrico

Para la contribución al soporte teórico de la investigación presentada se citaron o consultaron un total de 114 fuentes de información. Para utilizar la información contenida en estas fuentes se hizo uso de herramientas y estrategias de búsquedas en la Internet (ver anexo 2), así como, de bases de datos de publicaciones referenciadas en: Scielo, *Scopus*, *ScienceDirect*, *Springer*, ACM y IEEE. A continuación, se presenta el análisis bibliométrico de la misma.

Tabla 1. Análisis bibliométrico.

Bibliografía	Últimos 5 años	Años anteriores
Libros y monografías	7	23
Tesis de doctorados	5	4
Tesis de maestrías	9	2
Artículos en Revistas referenciadas en Web of Science, SCOPUS	20	15
Memorias de eventos	3	4
Artículos publicados en la web	4	13
Reportes técnicos y conferencias	1	1
Entrevistas personales	3	0
Total	52 (45.61%)	62 (54.38%)

1.3 Conceptualización de competencias

Boyatzis define competencias como: característica subyacente en una persona que está causalmente relacionada con un desempeño en un puesto de trabajo concreto y en una organización concreta (Boyatzis, 1982). Spencer y Spencer en su libro "*Competence at work, models for superior performance*" las definen como: Una característica subyacente de un individuo que está causalmente relacionado con un nivel estándar de efectividad y/o desempeño superior en un trabajo (Spencer, 1993). El contenido de este concepto coincide con el ofrecido por Boyatzis y los Spencer, privilegiando las cualidades humanas como causa del éxito en la actividad laboral (Zayas, 2010). Por su parte Luengas en su trabajo final de especialización la define como: la capacidad real que tiene una persona para aplicar conocimientos, habilidades y destrezas, valores y comportamientos, en el desempeño laboral, en diferentes contextos (Luengas, 2014).

En la presente investigación se adopta la definición de **competencia laboral** ofrecida por la Oficina Nacional de Normalización, aplicada en los entornos organizacionales y empresas del país. En la misma se plantea que competencia laboral es: “Un conjunto sinérgico de conocimientos, habilidades, experiencias, sentimientos, actitudes, motivaciones, características personales y valores, basado en la idoneidad demostrada, asociado a un desempeño superior del trabajador y de la organización, en correspondencia con las exigencias técnicas, productivas y de servicios. Es requerimiento esencial que esas competencias sean observables y que contribuyan al logro de los objetivos de la organización” (Oficina Nacional de Normalización, 2007:10).

El perfil de competencias en la investigación está asociado directamente con las actividades en los proyectos. La Dirección de los Centros productivos en la UCI tienen definidos los roles de desarrollo de software que trabajan en el proyecto y a partir de estos roles, está definido el perfil de competencias al que debe tributar cada uno de los ellos. Según Cuesta, para definir competencias laborales es de gran importancia identificar las dimensiones o pautas de conducta que las hacen observables. Se incluirá en esta investigación la definición de dimensiones de competencias laborales como complemento y descripción de las competencias (Cuesta, 2010).

1.3.1 Análisis de evidencias del desempeño en la evaluación de competencias.

La evaluación de competencias, cuando se enmarca en la evaluación del desempeño, supone una comparación entre las competencias demostradas por el individuo en su trabajo y las competencias exigidas por el Centro productivo para un desempeño exitoso en el puesto que ocupa. Así, en este contexto, son precisamente las evidencias que avalan el desempeño laboral en un período de tiempo determinado, y que son almacenadas de diferentes maneras para que puedan servir de entrada al proceso de evaluación de competencias.

En la industria de software y específicamente en el caso del desarrollo productivo, las actividades o tareas, constituyen las evidencias. Las mismas son guardadas en la herramienta destinada para gestionar todo lo concerniente a los proyectos. En la UCI las evidencias se almacenan en la herramienta Xedro-GESPRO constituyendo así, dichas evidencias la constancia del trabajo de cada individuo. Se considera la entrada fundamental para la evaluación de competencias, ya que las mismas constituyen textos cortos que serán utilizados para ser comparados y a partir de esto, obtener información relevante y útil para la evaluación de competencias laborales. Luego son comparadas con el perfil de competencias exigido para cada rol en dependencia del individuo que se esté evaluando.

Existen varios tipos de evidencias, entre las que se pueden citar las de: (Investigación, Gestión-Producción, Análisis-Producción, Desarrollo-Producción, Extensión y Postgrado). Para el caso de la investigación se tendrán en cuenta las de Desarrollo-Producción, revisando la relevancia que las mismas tendrían para el proceso de evaluación de competencias laborales en los proyectos, pues dicho proceso está sustentado sobre estas.

Este proceso de análisis de evidencias es un elemento común en la evaluación de competencias, independientemente de que se utilice una técnica u otra. A continuación, se mencionan diferentes técnicas de

evaluación de competencias solo se analizarán las técnicas que utilizan evidencias en forma de textos, debido a la relevancia que tienen en la investigación.

En la Tabla 1 se muestra la clasificación de los instrumentos que pueden emplearse asumiéndose como criterio el tipo de información sobre el que se centran. De esta manera se puede distinguir entre técnicas para valorar competencias que se apoyan prioritariamente en rasgos o características de las personas, técnicas que parten del comportamiento de las personas en sus puestos de trabajo, y técnicas que, pudiendo incluir la recogida de informaciones como las anteriores, integran estas con las valoraciones de los propios trabajadores o de otros miembros de la organización.

Tabla 2. Instrumentos para la evaluación de competencias laborales. Basado en las referencias del autor Gil Flores (Gil, 2007).

Fuentes de información	Instrumentos de evaluación
Experiencia práctica	• Lista de verificación • Sistemas de escalas para la observación • Incidentes críticos • Simulación y ejercicios prácticos
Características y experiencias del evaluado	• Test psicológicos • Recogida de información biográfica • Entrevista de evaluación • Portafolios
Valoración del evaluado o de otros miembros de la organización	• Auto-informe sobre comportamientos • Balance de competencias • Evaluación de 360 grados

De las técnicas relacionadas anteriormente que tienen la posibilidad de realizar el análisis de las evidencias obtenidas en el proceso productivo de los centros de desarrollo, las técnicas basadas en la Experiencia Práctica. El interés en este tipo de técnica está asociado, a que las mismas no interrumpen el cronograma del proyecto, y el objetivo es contribuir a la ejecución del mismo, no entorpecerlo. Por ello constituirán objeto de análisis en la investigación. Las técnicas basadas en la experiencia práctica se apoyan en el análisis del desempeño laboral de los trabajadores, tomando como fuente de información la experiencia de trabajo real o simulado. En este caso, el acceso a la información se realiza de modo directo, dado que el evaluador observa directamente el modo en que se comporta el sujeto evaluado.

Las técnicas basadas en características y experiencias del evaluado no se analizan debido, a que, en este caso el evaluador no accede directamente a la actuación del trabajador en su puesto, sino que valora esta a partir de sus características personales o a partir de la obtención de información mediada por los sujetos, acerca de su experiencia. En el caso de las técnicas basadas en valoraciones tampoco constituirán objeto de análisis en la investigación, debido a que, la información manejada en el caso de la investigación, las evidencias del desempeño no surgen directamente de las realizaciones prácticas de los sujetos, sino que implica recoger las valoraciones que diferentes personas hacen sobre el sujeto evaluado y por lo general se afecta la objetividad de

los juicios emitidos por los evaluadores y por ende la efectividad del proceso de evaluación de competencias laborales.

1.3.3 Errores y deficiencias en el proceso de evaluación de competencias

A pesar de la diversidad de técnicas existentes, presentan una deficiencia en común: el nivel de intrusión de la actividad humana es extremadamente alto, y encierra la dificultad que supone emitir juicios valorativos que se ajusten a las características y al mérito real de las personas evaluadas. Además de una perspectiva amplia sobre el desempeño de los sujetos o sobre su potencial, la evaluación requiere experiencia, conocimiento profundo del puesto de trabajo para el que se realiza la misma y capacidad para el análisis y comprensión de la situación evaluada. La utilización de evaluadores, puede llevar a que, la evaluación y el análisis de evidencias para emitir un juicio se vea afectada por ciertos errores, ver la Tabla 2.

Tabla 2. Errores comunes en el proceso de evaluación. Tomado de las referencias del autor Cuesta Santos (Cuesta, 2010).

Efecto halo	Hace referencia a la tendencia humana de establecer valoraciones globales sobre una persona basada en un solo rasgo o característica. Ejemplo: si una persona cae bien por su marcada sociabilidad, se puede tender a sobrevalorar su actuación positiva en general.
Tendencia central	Se refiere a la tendencia que presentan algunas personas a emitir calificaciones medias y rara vez extremas. Ejemplo: si la escala del indicado fuera de uno a diez, se tendería a puntuar un cinco.
Polaridad	En este caso se trata de emitir calificaciones extremas. Ejemplo: si el indicador fuera de la escala de uno a diez, se marcaría el uno o el dos, o, por el contrario, el nueve o el diez.
Proyección	Tendencia a proyectar aspectos positivos de uno mismo en la persona evaluada con la que uno se identifica, o a proyectar aspectos negativos de uno mismo en la persona con la cual no se identifica, o se rechaza.
Efecto recencia	Consiste en la tendencia de recordar aquello que acaba de ocurrir, olvidando o quedando relegado lo anterior. Este efecto hará que se evalúe positiva o negativamente a una persona por su conducta más reciente y no por su actuación en todo el período.
Efecto primacía	Efecto contrario al anterior, en el sentido a que hace referencia a que se recuerde mejor aquello que ha sucedido en primer lugar. Está relacionado con la fuerte latencia de las primeras impresiones negativas o positivas.

Todos estos elementos analizados constituyen errores en los que se puede incurrir fácilmente ya que por motivos psicológicos (por ejemplo) son usuales durante estas actividades. En la red de centros de la UCI tampoco se está exento de incurrir en estos errores, pues, para realizar el análisis de evidencias, los evaluadores deben revisar las tareas que cada individuo realiza en un período, e identificar según su opinión, cuáles evidencias corresponden con su perfil de competencias y cuáles no, para poder llevar a cabo la evaluación.

Cada evaluador a partir de las competencias laborales del individuo y las evidencias acumuladas en un período de tiempo, una vez determine si cada evidencia se relaciona o no con el perfil de competencias (según su criterio), aproxima de manera subjetiva si la evidencia en cuestión tiene o no poco peso en sus competencias, lo cual se ve reflejado al final en la evaluación del individuo y esta se ve afectada por dichas causas. Sin embargo, la explotación de técnicas de minería de datos (*Data Mining*) y de texto podría ser un área con un posible enfoque para crear alternativas de soluciones a las dificultades analizadas en la UCI.

1.3.4 El descubrimiento de conocimiento y la Minería de Datos

La minería de datos es una fase del descubrimiento de conocimiento en base de datos (KDD por sus siglas en inglés) que integra los métodos de aprendizaje y estadísticas para obtener hipótesis de patrones y modelos. El KDD es una técnica que combina recursos desarrollados en el área de la Inteligencia Artificial para extraer información importante de las bases de datos, y dentro del cual se referencia a la Minería de Datos, como un paso fundamental dentro del proceso (Ferruccio, 2004). La Minería de Datos consiste en aplicar análisis de datos y algoritmos de descubrimiento que producen una particular enumeración de patrones (o modelos) sobre los datos (Zapata y otros, 2014).

La limitante que existe es que las técnicas de minería de datos procesan información estructurada y, sin embargo, aproximadamente un 80% de la información se encuentra almacenada en forma textual no estructurada, de ahí que se desarrollen técnicas de minería de textos (*Text Mining*), que permitan encontrar patrones interesantes y útiles en información textual no estructurada (Durteler, 2001) (Arco, 2006), que apoyen la evaluación de competencias laborales en las distintas organizaciones.

1.3.5 La minería de textos como herramienta para el descubrimiento de patrones

La minería de texto es una de las áreas de investigación del procesamiento automático de textos. Ella se define como el proceso automático de descubrimiento de patrones interesantes en una colección de textos. Estos patrones no deben existir explícitamente en ningún texto de la colección, y deben surgir al relacionar el contenido de varios de ellos (Hearst, 1999) (Montes y Gómez, 2005) (Kodratoff, 2009).

La minería de textos es mucho más compleja que la minería de datos porque tiene que trabajar con datos textuales que son inherentemente no estructurados. Para lograr sus propósitos, la minería de textos necesita combinar varias técnicas, de ahí que sea un campo multidisciplinario que incluye la recuperación de información, el análisis de textos, la extracción de información, el agrupamiento, el resumen, la categorización, la clasificación, la visualización, la tecnología de bases de datos, el aprendizaje automático (Dixon, 1997) (Tan, 1999) (Arco, 2006).

La Búsqueda y Recuperación de Información en el campo de la Minería de Texto

El objetivo de recuperar esta información que se encuentra en textos y representarlos, es transformar un documento textual a un formato que sea adecuado como entrada a la aplicación de algoritmos (aprendizaje

automático, agrupamiento y clasificación) que permitan hacer minería de textos (Bordes y otros, 2011). Los problemas principales a la hora de recuperar información son provocados tanto por el uso del lenguaje natural como el silencio (debido a la sinonimia) y el ruido (debido a la polisemia). La sinonimia en este contexto se refiere a la semántica de los textos, es la relación de identidad o semejanza de significados entre determinadas palabras u oraciones. La polisemia por el contrario se presenta cuando una misma palabra o signo lingüístico tiene varias acepciones o significados (Bordes y otros, 2011).

Dentro de las relaciones semánticas entre términos, es necesario diferenciar entre “similitud” y “relación” semántica. Dos términos semánticamente similares son aquellos cuyo significado semántico es aproximado, mientras que dos términos relacionados pueden no ser semánticamente similares. La relación semántica entre conceptos es una noción más amplia que la similitud semántica, ya que esta se establece por la semejanza entre conceptos. Es por ello que la similitud semántica se entiende de forma generalizada, como el grado de proximidad taxonómica entre dos términos (Martínez, 2016).

Trabajos como: Clasificación de textos adaptada para conversión de texto en habla multidominio (Alias, 2006), *Clustering* iterativo de textos cortos con representaciones basadas en conceptos (Ingaramo, 2011) y Clasificación de Textos Cortos Usando Representaciones Distribucionales de los Términos (Jiménez y Balderas, 2012) han reconocido el grado de importancia y la complejidad que conlleva el procesamiento de textos cortos. Diferentes trabajos (Rosas *et. al.*, 2010) se han centrado en el análisis de las ventajas que se pueden obtener de la incorporación de información semántica a la representación de textos, donde se han obtenido resultados positivos en algunos casos.

1.4 Selección de funciones de similitud para comparación de textos cortos

Para la adecuación y comprensión de los textos al lenguaje de la computadora, se debe analizar la lingüística computacional. El agrupamiento es un procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios por lo general estarán asociados a la distancia o similitud que existe entre los vectores. Uno de los mayores problemas para llevar a cabo el agrupamiento, es encontrar una representación y función de similitud acorde para los textos que serán categorizados mediante un algoritmo de agrupamiento. La representación dependerá del objetivo de análisis que se esté intentando obtener. Cuando se refiere a una función adecuada, es referido al hecho de obtener adecuadamente la similitud entre textos, tarea de suma complejidad. Las funciones de similitud sirven para detectar que tan similares son dos contenidos y se materializan en algoritmos de computación que realizan esta tarea. Los investigadores han desarrollado múltiples funciones de este tipo y está ampliamente documentado que no todas son igual de eficaces ante la presencia de datos con las diferentes situaciones problemáticas (Elmagarmid, *et. al.*, 2007).

La selección de la función de similitud para trabajar en textos, debe ser de especial cuidado asumiendo que, de ello, dependerán los resultados. En relación con esto, debe ser minuciosa su selección para de esa manera garantizar que los grupos sean correctamente correlacionados y por consiguiente un adecuado agrupamiento.

Por ello es necesario realizar un análisis de los resultados que se obtienen con estas funciones y poder seleccionar las que mejores resultados ofrezcan específicamente para textos cortos.

Existen propuestas de funciones de similitud, las cuales pueden ser clasificadas en dos categorías: basadas en caracteres y basadas en palabras completas o *Tokens* (Elmagarmid *et. al.*, 2007.). Las primeras consideran cada cadena como una secuencia ininterrumpida de caracteres; las segundas como un conjunto de sub cadenas delimitadas por caracteres especiales como espacios en blanco, comas y puntos; consideran una cadena como un conjunto de *tokens* y calculan la similitud entre cada pareja de *tokens* mediante alguna función basada en caracteres.

A continuación, se enunciarán las principales funciones de similitud de textos que, según la bibliografía han arrojado buenos resultados. A su vez, se realizará el análisis de las mismas para poder obtener, asimismo, la(s) adecuada(s) para esta investigación. En la revisión de los diferentes métodos y vías en lo relativo a la comparación de cadenas de caracteres, se seguirá la organización propuesta por Naumann. Se puede apreciar de forma gráfica esta organización en la Figura 1.

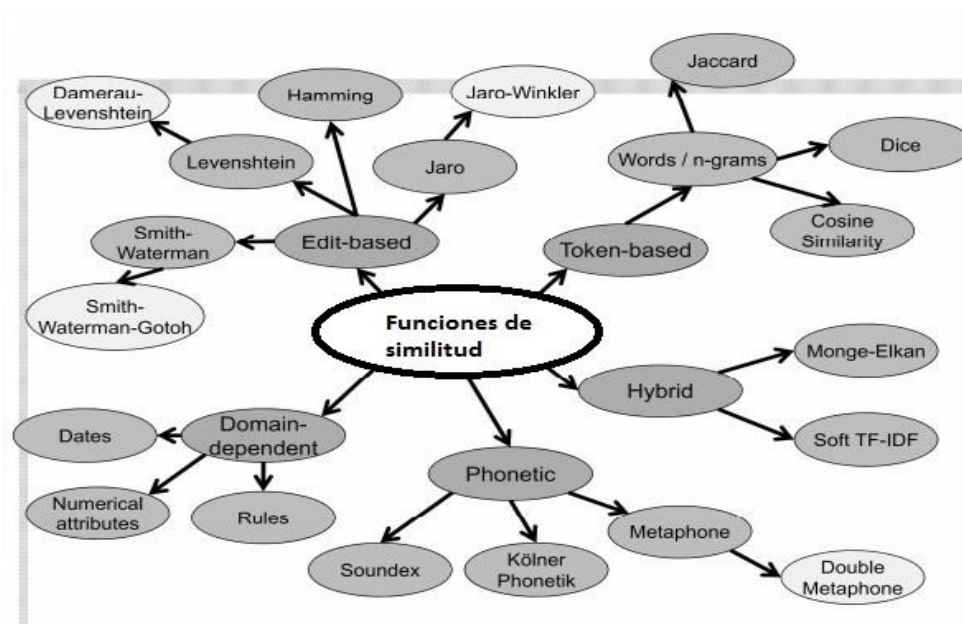


Figura 1. Organización de funciones de similitud para caracteres (Naumann, 2013).

Al referirse a la distancia que existe entre dos cadenas de textos, no se está indicando que exactamente sea la similitud entre las mismas, puesto que una magnitud puede calcularse a partir de la otra. Una distancia real $Z \in [0,1]$, donde el 0 indicará que las dos cadenas son idénticas y 1 que no presentan nada en común, lo que equivale a una similitud $S=1-Z$, que mientras mayor sea la distancia menor será la similitud y viceversa. En cambio, las funciones de similitud basadas en caracteres van a considerar cada cadena de caracteres como una secuencia ininterrumpida de estos y las funciones de similitud basadas en *tokens* suponen cada cadena como un conjunto de sub cadenas independientes por caracteres especiales, como por ejemplo, espacios en blanco, puntos, puntos y comas, o sea, como un conjunto de *tokens*, y van a calcular la similitud entre cada pareja de *tokens* mediante alguna función de similitud basada en caracteres. Las medidas o funciones a definir pueden ser

de distancia, similitud o disimilitud, basadas en caracteres o en tokens. En el caso de la investigación se tendrán en cuenta las basadas en similitud.

Distancia: Euclidiana, Métrica de Minkowski, Métrica Canberra y Hamming (esta última también considerada de similitud) Basándose en los criterios utilizados por (Arco, 2006) se detallan:

La Euclidiana considera que dos objetos son similares si la distancia entre ellos es cercana a cero y diferentes si la distancia entre ellos es cercana a uno. La distancia Euclidiana entre dos objetos O_i y O_j , descritos por k rasgos, es calculada como se muestra en la ecuación 1:

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

La métrica *Minkowski* calcula la distancia entre dos objetos O_i y O_j , descritos por k rasgos y donde se cumple que $\gamma \geq 1$, esta puede calcularse mediante la ecuación 2:

$$D(O_i, O_j) = \left(\sum_{h=1}^k |O_{ih} - O_{jh}|^\gamma \right)^{\frac{1}{\gamma}} \quad (2)$$

Por su parte, la métrica Canberra, se enfoca en el cálculo de dos objetos y a través de la ecuación 3 puede ser obtenida:

$$D(O_i, O_j) = \sum_{h=1}^k \frac{|O_{ih} - O_{jh}|}{(|O_{ih}| + |O_{jh}|)} \quad (3)$$

Otra forma utilizada de medir la distancia entre objetos es la distancia de *Hamming*. En su definición binaria a partir de la cantidad de atributos en que difieren dos objetos, como en sus variantes pesadas. En fuentes bibliográficas como (Naumann, 2013), (Arco, 2005) se presentan formas de medir la similitud entre objetos, considerando la medida como una relación asimétrica y teniendo en cuenta la probabilidad condicional de un objeto respecto al otro. La similitud entre dos objetos O_i y O_j se puede obtener calculando la correlación de Pearson entre ellos, a través de la ecuación 4:

$$D(O_i, O_j) = \frac{\sum_{h=1}^k (O_{ih} - \overline{atributo_h})(O_{jh} - \overline{atributo_h})}{\sqrt{\sum_{h=1}^k (O_{ih} - \overline{atributo_h})^2 \sum_{h=1}^k (O_{jh} - \overline{atributo_h})^2}} \quad (4)$$

En otro orden, se encuentran las funciones de similitud siguientes que también son utilizadas en el trabajo con textos cortos: *Jaro Distance*, *Jaro-Winkler Distance*, *Levenshtein Distance*, *Monge-Elkan Coefficient*, *Needleman-Wunsch Coefficient*, *Overlap Coefficient*, *Q-Gram Distance*, *Smith-Waterman Coefficient*, *Smith-Waterman-Gotoh Coefficient*, *Soundex Distance*.

Las funciones o métricas *Jaro* y *Jaro-Winkler* son utilizadas principalmente en el trabajo con cadenas de caracteres cortas (por ejemplo, los nombres propios o apellidos). La función *Jaro* determina la similitud entre dos secuencias de textos sobre la base de un carácter común. Esta métrica basa su funcionamiento en computar los caracteres comunes tales que el i-ésimo carácter sea igual al j-ésimo carácter donde el valor absoluto de i-j sea menor o igual que la mitad del mínimo entre las longitudes de los dos textos a comparar. Además de calcular el número de transposiciones que puede existir entre estos caracteres comunes, como la comparación fallida del i-ésimo de estos entre las dos cadenas. Esta técnica se utiliza principalmente en el área de detección de duplicados de textos de pequeña longitud. La distancia *Jaro* (d_j) de dos cadenas s_1 y s_2 es calculada a partir de la ecuación 5 como se muestra a continuación:

$$d_j = \left\{ \frac{1}{3} \left[\frac{m}{|S_1|} + \frac{m}{|S_2|} + \frac{m-t}{|m|} \right] \right\} \quad (5)$$

Donde $m = 0$; de otra manera si: m = número de caracteres que coinciden y t = la mitad del número de transposiciones, dos caracteres de las cadenas S_1 y S_2 , respectivamente, se consideran similares sólo si, son el mismo y no más allá de:

$$t = \left\lfloor \frac{\max(|S_1|, |S_2|)}{2} \right\rfloor - 1 \quad (6)$$

Se toma cada carácter de la cadena S_1 y se compara con cada uno de los caracteres que coincida en la cadena S_2 . El número de caracteres coincidentes (donde el orden de la secuencia es diferente) dividido por 2 definirá el número de transposiciones. Por ejemplo, en comparación *CRATE* con *TRACE*, solo R A E son los caracteres que coinciden, es decir $m=3$. Aunque 'CT' aparece en ambas cadenas, están más alejados que 1, es decir, $(5/2) - 1 = 1,5$. Por lo tanto, $t = 0$. En Dwane contra Duane las letras coincidentes ya están en el mismo orden, DANE, por lo que no son necesarias transposiciones. La distancia Jaro puede mostrar mejores resultados en caso de los métodos híbridos (Nitesh, 2015).

Winkler, 1990 propone una variante que asigna puntajes de similitud mayores a cadenas que comparten algún prefijo, basándose en un estudio realizado por *Pollock* y *Zamora* en 1984. *Cohen et. al.*, en 2003 proponen un modelo basado en distribuciones Gaussianas. *Yancey*, 2006 compara la eficacia de la similitud de Jaro y algunas de sus variantes. *Winkler*, ha propuesto algunas mejoras de este modelo, dicho proceso puede plantearse como sigue:

- 1) Se define un umbral real θ ,
- 2) se compara cada registro de R con el resto,
- 3) si la similitud entre una pareja de tuplas es mayor o igual que θ , se asumen duplicados; o sea, se consideran representaciones de una misma entidad real.

La distancia *Jaro-Winkler*, sin embargo, utiliza una escala p prefijo que emite puntuaciones mucho más adecuadas para las cadenas que coinciden desde el principio para un conjunto l longitud del prefijo. Dadas dos cadenas $S1$ y $S2$, su distancia *Jaro-Winkler* D_w se obtendría como se muestra en la ecuación 7:

$$d_w = d_j + (\ell p(1 - d_j)) \quad (7)$$

Donde:

d_j : distancia Jaro para las cadenas $S1$ y $S2$

l : longitud del prefijo común en el principio de la cadena, hasta un máximo de 4 caracteres

p : la longitud del prefijo común, p no debe exceder de 0,25, de lo contrario la distancia puede ser mayor que 1. El valor estándar para esta constante en la obra de Winkler es $p = 0,1$. (Serie Científica de la Universidad de las Ciencias Informáticas, 2015)

En otro orden, la distancia *Levenshtein* (LD) es considerada una métrica que permite medir el número de diferencias entre dos cadenas de caracteres. A continuación, la notación para LD :

$LD(s, t)$ donde: s = cadena de texto de origen y t = cadena de texto de destino

La distancia entre estas cadenas será medida por el número de inserción, eliminación o sustitución requerida para transformar la cadena de texto de origen (s) a la cadena de texto de destino (t). Una prueba del uso de este tipo de función se ejemplifica a continuación, a través de dos casos:

1. Si la cadena (s) fuente es "proyecto" y la cadena de destino (t) también es "proyecto", entonces $LD(s, t) = 0$, pues no se necesitan transformaciones. Las cadenas son idénticas entre sí.
2. Si la cadena (s) fuente original es "Paul" y la cadena de destino (t) es "Pual", entonces $LD(s, t) = 2$, porque habría dos sustituciones (cambios de "a" a "u" y "u" para "a") es necesario transformar la cadena original a la cadena de comparación.

La cuenta final de la distancia se convertirá en porcentaje mediante la ecuación 8 como se muestra a continuación (Chen, 2014):

$$Distance = (100 - total\ distance\ count) * 100 \quad (8)$$

El coeficiente de *Needleman-Wunsch* (NW) se utiliza comúnmente para la secuencia de alineación general y puntuación. Este algoritmo de coincidencia de cadenas, que incluye cuatro parámetros, encuentra la alineación óptima de dos secuencias de textos y calcula similitudes entre ellos. Los primeros dos parámetros en el algoritmo son dos cadenas, que lo ideal sería que fueran iguales. El tercer parámetro es una matriz de similitud, que muestra las relaciones entre cada carácter de las dos cadenas. La cuarta es una penalización por hueco que es un valor diseñado para acercarse en el marcador cuando las cadenas no coinciden (Güyer et. Al., 2015). La

puntuación más alta obtenida del algoritmo *NW* muestra la mejor similitud. Con el fin de hacer que las puntuaciones sean comparables entre sí, las puntuaciones deben ser normalizadas.

En cuanto al coeficiente *Overlap*, para su aplicación primero se debe pasar por un pre-procesamiento de las cadenas que van a ser comparadas, que se basa en su limpieza y normalización, eliminando palabras irrelevantes, además de acentos y puntuación. Este se enfoca en el cálculo como se muestra en la ecuación 9 del grado de superposición entre dos cadenas de caracteres. Realiza la intersección entre las cadenas que se van a comparar y divide sobre la cantidad de términos menor de las dos cadenas. Trata de mostrar que contenido de una cadena está dentro de otra.

$$\text{Overlap distance} = |Q \cap D| / \min(|Q|, |D|) \quad (9)$$

Donde Q y D representan las cadenas que se van a comparar. Este coeficiente indica en la escala entre 0 y 1 el nivel de similitud, donde 1 indica similitud y 0 disimilitud. (Tabares et. al., 2015)

Un q-gram, también llamado n-gram, es una subcadena de longitud q. El principio tras esta función de similitud es que, cuando dos cadenas son muy similares tienen muchos q-grams en común. Es común usar uni-grams (q = 1), bi-grams o di-grams (q = 2) y tri-grams (q = 3). Es posible agregar q – 1 ocurrencias de un carácter especial (no definido en el alfabeto Σ original) al principio y final de ambas cadenas. Esto llevará a un puntaje de similitud mayor entre cadenas que compartan algún prefijo o sufijo, aunque presenten diferencias en el medio (Amón y Echeverría, 2012). Mediante el siguiente caso se prueba el trabajo con esta función: la cadena " Peter y Velen " contiene los *bi-grams* "Ve", "el", "le", "en" y los *1-skip-2-grams* "VI", "ee", "ln". La distancia *q-grams* se obtiene mediante la tabulación de los *q-grams* que se producen en las dos cadenas y contando el número de *q-grams* que no se comparten entre las dos. Este formalmente puede ser denotado como sigue según se evidencia en la ecuación 10:

$$d_{q\text{gram}}(s, t; q) = \|v(s; q) - v(t; q)\|_1 = \sum_{i=1}^{|\Sigma|^q} |v_i(s; q) - v_i(t; q)| \quad (10)$$

Donde, $v(s; q)$ es un vector entero no negativo de la dimensión de ocurrencias de cada posible q-gramo en s. La fórmula (10) define la distancia q gramos entre dos cadenas s y t como la distancia L1 entre $v(s; q)$ y $v(t; q)$. Observe que sólo se necesita para almacenar y contar los que ocurren en realidad q-grams para evaluar la suma anteriormente (Van der Loo, 2014).

Monge Elkan, 1996 describe una similitud métrica basadas en conjuntos y plantea que es una variante de la métrica de Smith-Waterman como se ve en las ecuaciones 11 y 12 según corresponda. El alfabeto es letras mayúsculas y minúsculas, dígitos, punto, coma y todos los demás caracteres se ignoran (Cheatham, 2014).

$$\text{sim}_{\text{MongeElkan}}(x, y) = \frac{1}{|x|} \sum_{i=1}^{|x|} \max_{j=1, |y|} \text{sim}'(x[i], y[j]) \quad (11)$$

Donde: $|X|$ Es el número de tokens en x

Sim 'es la función de similitud interna (por ejemplo, LD)

Si las cadenas contienen sólo una palabra cada una, esta ecuación resultaría como sigue:

$$sim_{MongeElkan}(x, y) = sim'(x, y) \quad (12)$$

La similitud *Smith-Waterman* entre dos cadenas A y B resulta la máxima similitud entre una pareja (A', B'), sobre todas las posibles, tal que A' es subcadena de A y B' es subcadena de B. El modelo original define las mismas operaciones de la distancia de edición y, además, permite omitir cualquier número de caracteres al principio o al final de ambas cadenas (Smith y Waterman, 1981). Este aspecto permite identificar cadenas equivalentes con prefijos/sufijos que, al no tener valor semántico, pueden ser descartados (Amón *et. al.*, 2012). Un ejemplo del funcionamiento de este coeficiente es el siguiente: dos cadenas de textos, “Ing. Javier Heredia Ruiz” y “Javier Heredia Ruiz, Universidad de las Ciencias Informáticas” tendrían una similitud cercana a uno, pues semánticamente no significan lo mismo, aunque conceptualmente sí.

La distancia *Soundex* es considerada un algoritmo de codificación fonética, que se encarga de convertir un texto o palabra en código. El código obtenido por *Soundex* reside en sustituir las consonantes de la palabra afectada por un número; en caso de ser necesario se agregan ceros al final del código para conformar de esa manera un código de 4 dígitos. *Soundex* elige la clasificación de los caracteres con base, en el lugar de articulación en la lengua inglesa. Esta función a pesar de que es orientada a textos, sus limitaciones han sido ampliamente documentadas (Bhatti *et. al.*, 2014) y han dado lugar a varias mejoras, pero ninguna orientada hacia el idioma español (idioma en el cual se basa la investigación). Se evidencia que *Soundex* no está orientado al español ya que ni siquiera contempla el juego de caracteres (“ñ”, “ll”). Asimismo, la dependencia de la letra inicial, la agrupación por punto de articulación del idioma inglés y el estar limitado a cuatro caracteres implica que no es eficiente para detectar errores ortográficos comunes en el idioma español (Amón y Moreno, 2014).

Las funciones analizadas se utilizan para el análisis de similitud o distancia entre cadenas de textos y específicamente textos cortos, presentan limitantes que impiden obtener la similitud entre las tareas asignadas a los especialistas de un proyecto de software y las competencias previamente definidas en el mismo para cada rol. Ambos elementos constituyen las evidencias en estos proyectos y son guardadas en la herramienta que los gestiona (en el caso de la UCI es el Xedro-Gespro), en forma de texto cortos. De ahí, que se necesiten funciones de este tipo que garanticen con el menor margen de error la similitud entre este tipo de textos.

Con respecto a las medidas o funciones de similitud existentes, asociadas al agrupamiento de textos cortos se evidencia en menor medida en el ámbito asociado a la Gestión de proyectos informáticos en cuanto a su utilización y puesta en práctica. En este sentido, se puede asegurar que en la minería de texto el análisis del agrupamiento dependerá de la representación del texto y del peso que se le asigne a cada texto al caracterizarlo. Siguiendo esta línea, se han identificado tres funciones que han tenido los mejores resultados en el análisis de similitud o proximidad en textos cortos, estas son los coeficientes de proximidad o similitud *Dice*, *Coseno* y *Jaccard* (Tórres, 2015). Estos coeficientes pueden ser obtenidos por las ecuaciones 13 o 14, 15 y 16 como se muestra más adelante.

Seguidamente, se analiza el cálculo de la similitud S entre dos cadenas de textos D_i y D_j (en el caso de la investigación D_i y D_j constituirían las tareas asignadas a los integrantes de los proyectos y las competencias definidas para cada rol que desempeñe el integrante del proyecto respectivamente), basándose en los criterios utilizados por (Arco, 2006), y teniendo en cuenta los pesos asociados a los k términos que describen estos textos:

$$S(D_i, D_j) = \frac{2 \sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}{\sum_{h=1}^k peso_{ih}^2 + \sum_{h=1}^k peso_{jh}^2} \quad (13)$$

Dice

Si el peso de los términos es binario la ecuación se reduce a:

$$S(D_i, D_j) = \frac{2C}{A + B} \quad (14)$$

Donde C es el número de términos que D_i y D_j tienen en común, y A y B son el número de términos de D_i y D_j respectivamente.

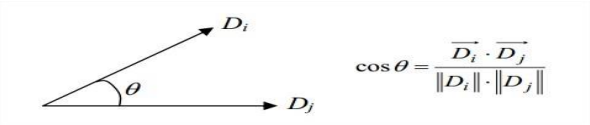
$$S(D_i, D_j) = \frac{\sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}{\sum_{h=1}^k peso_{ih}^2 + \sum_{h=1}^k peso_{jh}^2 - \sum_{h=1}^k (peso_{ih} \cdot peso_{jh})} \quad (15)$$

Jaccard

$$S(D_i, D_j) = \frac{\sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}{\sqrt{\sum_{h=1}^k peso_{ih}^2} \cdot \sqrt{\sum_{h=1}^k peso_{jh}^2}} \quad (16)$$

Coseno

Esta última función de proximidad, se utiliza en gran medida para determinar la similitud entre textos y su fundamento está asociado al coseno del ángulo que existe entre ellos. Si es relativamente cercano a uno el coseno del ángulo, se consideran textos similares, y si es cercano a cero, se consideran los textos diferentes. Seguidamente, se muestra otra variante de forma gráfica, donde los textos se representan como vectores, así como, la expresión que describe el coseno del ángulo entre ellos:



$$\cos \theta = \frac{\overrightarrow{D_i} \cdot \overrightarrow{D_j}}{\|D_i\| \cdot \|D_j\|} \quad (17)$$

Esta función de similitud aporta una buena medida del “parecido” de dos vectores en un espacio multidimensional, dicho espacio puede describir las características de los usuarios, tales como, las palabras

clave. La similitud entre estos elementos se mide computando el coseno entre el ángulo de estos dos vectores. Se plantea en la bibliografía (Torres, 2015), (Amón *et. al.*, 2012) que esta función es la más utilizada por los desarrolladores para el trabajado con textos cortos. La medida de similitud del Coseno equivale al cálculo del producto escalar de dos vectores de documentos (A y B) dividido por la raíz cuadrada de la sumatoria de los componentes del vector A multiplicado por la raíz cuadrada de la sumatoria de los componentes del vector B. Si no existe coincidencia alguna entre los componentes, la similitud de los vectores será cero ya que el producto escalar será cero (situación muy frecuente en la realidad pues los vectores pueden llegar a tener disímiles componentes y se puede dar el caso de la no coincidencia con mayor frecuencia de lo que se puede esperar). De igual forma resulta lógico deducir que la similitud máxima solo se puede obtener cuando todos los componentes de los vectores son similares, en este caso la función del coseno obtiene su máximo valor, la unidad, o sea uno.

Se considera cada elemento como un vector dentro de un espacio vectorial de m dimensiones y se calcula la similitud con el coseno del ángulo que forman como se muestra en la Figura 2. Es decir, dos vectores x_1 , x_2 serán más similares mientras más pequeño sea el ángulo que se forma entre ellos y viceversa.

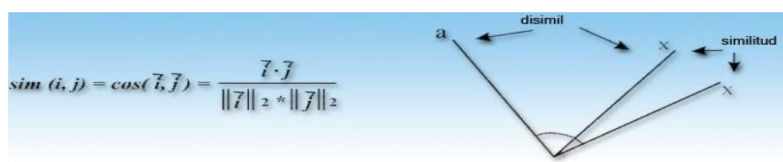


Figura 2. Cálculo de la similitud con el coseno del ángulo.

Estas tres funciones, coeficientes o métricas son aplicables directamente a vectores que representan textos de un mismo idioma, pero, para el caso de vectores que corresponden a distintos idiomas, es necesario realizar previamente un proceso de traducción (lo que no constituye objetivo de esta investigación).

Una vez analizadas y probadas las funciones de similitud para el trabajo con textos, se identificaron las más adecuadas de acuerdo a los resultados obtenidas por ellas. Para el caso de este trabajo se selecciona para la investigación los coeficientes Dice, Coseno y Jaccard, pues son las que una vez probadas y fundamentadas por la bibliografía, han arrojado los mejores resultados en cuanto a precisión e incidencia mínima de errores en la obtención de la similitud de textos cortos. Asociado a ello y al dominio del problema de la investigación constituyen la posible solución al mismo y de contribución a la eficiencia del proceso de evaluación de competencias laborales en el ámbito de la Gestión de proyectos en la UCI.

Solo resta obtener el cómo proceder a su utilización de forma adecuada en la producción de software de la Red de centros productivos. De ahí que se proponga un procedimiento de guía para trabajar con las funciones seleccionadas, bajo las condiciones del objeto de estudio de la presente investigación.

1.5 Conclusiones del capítulo

En el capítulo se analizan los elementos teóricos fundamentales asociados al objeto de estudio, el proceso de la evaluación de competencias laborales, las técnicas de minería de datos y de textos como alternativa a la

solución del problema planteado. Se exponen los fundamentos esenciales de las funciones de similitud de textos, dando cumplimiento al primer objetivo específico de la investigación, identificándose las siguientes conclusiones de su resultado y discusión:

En el proceso de evaluación de competencias, el análisis de evidencias es un proceso común que se lleva a cabo independientemente de las técnicas de evaluación que se utilicen; identificándose para la investigación las técnicas de evaluación de competencias laborales que analicen evidencias.

A partir del análisis de las técnicas de minería de datos, y dentro de ellas las asociadas a la minería de textos, se identificaron alternativas poco exploradas en el campo de la automatización de procesos para realizar análisis de evidencias en la evaluación de competencias laborales en la Gestión de Proyectos Informáticos.

El análisis de los referentes asociados a las funciones de similitud entre textos y específicamente textos cortos permitió identificar las que arrojaron los mejores resultados según el objeto de estudio de la investigación, en este caso fueron *Dice*, *Jaccard* y *Coseno*.

Según las ideas anteriores, se concluye la necesidad de proponer un procedimiento que sirva de guía para utilizar las funciones de similitud entre textos cortos identificadas de manera que tribute al proceso de evaluación de competencias laborales.

CAPÍTULO 2. PROCEDIMIENTO PARA LA EVALUACIÓN DE COMPETENCIAS LABORALES BASADO EN FUNCIONES DE SIMILITUD DE TEXTOS

2.1 Introducción

En el presente capítulo se define el procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos que permite mejorar la eficiencia del proceso de evaluación de competencias laborales. Se tratan y describen las características generales del procedimiento, las competencias o capacidades que debe tener el personal ejecutor de este, así como, pre-condiciones o premisas a tener en cuenta para la aplicación del mismo. Seguidamente, se detallan las fases, actividades y elementos principales que componen el procedimiento. Finalmente se realizan las conclusiones parciales.

2.2 Procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos

La permanente evolución existente en el ámbito tecnológico asociado a delimitar la evaluación del personal, aumenta la necesidad de gestionar eficientemente los elementos que pueden constituir las evidencias en las organizaciones. Proveer al personal de metodologías, procedimientos, técnicas y herramientas que contribuyan de forma positiva al desarrollo de esta actividad constituye un salto cualitativo en el desarrollo de un proyecto exitoso.

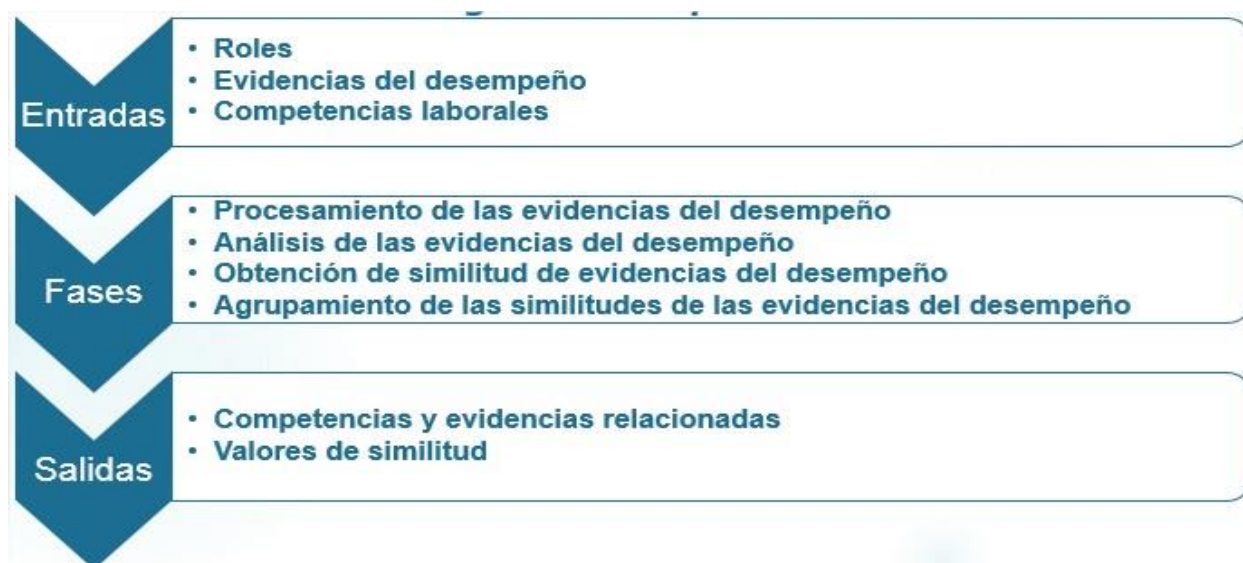


Figura 3. Estructuración del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos. Fuente: Elaboración propia.

En la Figura 3 se observa la vista general del procedimiento propuesto, en la cual se muestra la interacción entre cada uno de sus componentes estructurales: las entradas, las fases y las salidas. Una vez obtenido el resultado del procedimiento, el mismo será utilizado en aras de alcanzar mayor eficiencia en el proceso de evaluación de competencias laborales en los proyectos.

Las entradas (roles, evidencias del desempeño y las competencias laborales): constituyen los elementos de relevancia ya que es necesaria la asociación entre los roles y las competencias laborales que serán evaluadas posteriormente. Las evidencias por la identificación de su almacenamiento (documentos, bases de datos y/o herramientas de gestión de proyectos), para su procesamiento y utilización en el proceso.

Las fases (Procesamiento de las evidencias del desempeño, análisis de las evidencias del desempeño, obtención de similitud y agrupamiento): definen cada uno de los pasos a seguir para la obtención de los grupos de evidencias y competencias correlacionadas. En cada fase se encuentran las actividades más importantes, las que contribuyen a obtener los artefactos de salidas del procedimiento. Se describirán las fases a partir del epígrafe 2.4.

Las salidas (Competencias y tareas relacionadas y Valor de similitud): mostrarán el grupo de elementos similares necesarios para ser utilizados en el proceso de evaluación de competencias laborales en proyectos de software.

2.3 Características generales del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos

Las relaciones de similitud constituyen herramientas que son utilizadas en problemas de evaluación de competencias laborales en diversas áreas, entre ellas las asociadas a la gestión. Por ello, la gestión de proyectos no se encuentra exenta a su utilización. El procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos se centra en la explicación de cómo proceder, para la clasificación de los grupos de evidencias que se corresponden con las competencias definidas por la organización.

El procedimiento cuenta con las siguientes características:

- Se basa en las evidencias del desempeño del personal de la organización. Estas no son más que las tareas que realizan los RRHH durante la ejecución de un proyecto que avalan su desempeño laboral (Torres, 2015). Teniendo a su vez que, dichas evidencias se encuentran almacenadas en textos cortos y cada una cuenta con sus propias particularidades.
- Se basa en funciones de similitud para corroborar la relación entre las competencias y las evidencias del personal de los proyectos de software, almacenadas en herramientas para gestión de proyectos.
- Puede ser aplicado en el período designado por la organización ya que esta es la encargada de definir cuál será el intervalo de tiempo a evaluar el desempeño de sus RRHH; el que puede ser mensual, trimestral o anual. Para ello se tienen en cuenta elementos como: La cantidad de evidencias asociadas a un rol, el almacenado en las herramientas de gestión de proyectos, los objetivos de la organización y el período definido.
- Podrá ser ejecutado por el personal del proyecto que tenga dominio en el trabajo con las evidencias almacenadas.

2.3.1 Capacidades o competencias del personal ejecutor del procedimiento

Es de importancia desglosar de manera clara las responsabilidades necesarias y específicas de las personas que intervienen en el procedimiento para el análisis de evidencias del desempeño en la evaluación de competencias laborales. El personal encargado tiene que, conocer el proceso de almacenaje de la información que constituiría evidencias. En este acápite se describen los roles que se han sido identificados para ejecutar las actividades definidas en cada una de las fases.

Seguidamente se describen los roles, indicando sus principales responsabilidades:

- Líder de proyecto: es la autoridad en la organización y de inmediato interés en que se realice el procedimiento para el análisis de evidencias del desempeño en la evaluación de competencias laborales para mejorar la eficiencia.
- Rol genérico: especialista de la organización encargado de las actividades del procedimiento, desde la nueva solicitud de aplicación por el líder del proyecto hasta su culminación.

2.3.2 Premisas para la aplicación de la propuesta de solución

El procedimiento propuesto puede comenzar a utilizarse en cualquier etapa de la vida de un proyecto de software, aunque se recomienda su utilización desde etapas tempranas. Esto permitirá dar un completo seguimiento a los miembros del proyecto y orientar la gestión de las competencias laborales al logro de los objetivos de la organización, desde el planteamiento del alcance del proyecto. A continuación, se describe y fundamenta a manera de pre-condiciones los componentes y elementos a tener en cuenta.

Para llevar a cabo cualquiera de las tareas, actividades y acciones en el área de la Gestión de proyectos, es necesario que existan premisas o elementos de entradas. Dichos elementos constituirán la garantía de la ejecución exitosa de las mismas. En el caso de la propuesta de solución de la investigación se tienen en cuenta las siguientes:

- Los puestos de trabajo (PT) o roles definidos para el proyecto. Los roles son necesarios ya que a ellos se asocian las competencias laborales que serán evaluadas. (Torres y otros, 2014).
- Los elementos que constituyen evidencias del desempeño de los miembros de los equipos de desarrollo de proyectos.

Para la comprensión y entendimiento de qué elementos constituyen evidencias en el ámbito de gestión de proyectos, es importante referirse al concepto de proyecto y determinar cuáles son sus invariantes. Un proyecto se presenta como un esfuerzo temporal para lograr un objetivo bien determinado, estructurado en un conjunto de tareas organizadas lógicamente, con recursos limitados y un balance entre costo, tiempo y calidad (PMI, 2013), (Cuesta, 2012). Por tanto, en los proyectos son invariantes la existencia de fechas de inicio y fin del proyecto, y la existencia de tareas como forma de organización del trabajo de los recursos humanos involucrados en los mismos. Las tareas realizadas por los miembros de un proyecto,

constituyen evidencias de su desempeño y están compuestas por los siguientes elementos que se almacenan en forma de texto (PMI, 2013), (Piñeiro y otros, 2013):

- ✓ Título: describe las acciones que deben realizarse.
 - ✓ Responsable: recurso humano que debe realizar las acciones que incluye la tarea.
 - ✓ Tiempo estimado: tiempo en el que el recurso humano debe terminar la tarea.
 - ✓ Tiempo real dedicado: tiempo que dedicó el recurso humano para terminar la tarea.
 - ✓ Fecha de inicio: fecha en la que debe comenzarse a realizar.
 - ✓ Fecha de cierre planificada: fecha estimada para terminar la tarea.
 - ✓ Fecha real de cierre: fecha en que realmente se terminó la tarea, evidencia la eficiencia con que se realizó la tarea.
 - ✓ Evaluación: dada por el evaluador o jefe de proyecto, evidencia la eficacia con que se desempeñó el miembro del proyecto que la realizó ya que es una muestra de la calidad con que se concluyó.
- Competencias laborales establecidas para los roles en el proyecto: capacidad efectiva para llevar a cabo exitosamente una tarea previamente asignada.
 - Umbral: representa α -corte para determinar la posible relación tareas-competencias. Este umbral se utiliza para establecer a partir de qué valor de similitud entre las competencias laborales y las tareas que constituyen evidencias, se va a considerar que esas tareas contribuyen o muestran la competencia.
 - Documentación del proyecto: artefacto que recoge el negocio de la organización, identificándose las informaciones que son guardadas con el objetivo de identificar las de relevancia para el proceso de evaluación de las competencias.

Aunque en el desarrollo de un proyecto se generen otras evidencias, estas varían de un tipo de proyecto a otro; incluso en un mismo proyecto cambian si varía la metodología con que se desarrolla. Por estas razones estas evidencias no son genéricas para todos los proyectos.

La identificación de dónde se recogen y almacenan las tareas de los miembros de los equipos de los proyectos, es la entrada fundamental para el procedimiento. De igual forma deben estar definidas actividades cuyos resultados constituyen elementos primordiales para el procedimiento propuesto, las que, sin lugar a dudas, serán ejecutadas por el personal que represente un rol de los establecidos para ello.

2.5 Fases del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos

El procedimiento está conformado por las siguientes fases:

1. Procesamiento de las evidencias del desempeño: se obtienen las evidencias del desempeño de las herramientas de gestión de proyectos los elementos que constituyen evidencias del desempeño laboral. Se encarga de llevar la información no estructurada, proveniente de las diversas fuentes de información,

a una representación estructurada que permita a las restantes fases la correcta interpretación de la misma.

2. Análisis de las evidencias del desempeño: las evidencias son analizadas en búsqueda de mejoras en el rendimiento de la clasificación, garantizando que las mismas se encuentren libres de errores.
3. Obtención de similitud de evidencias del desempeño: se proponen las funciones de similitud de textos para ser utilizadas en el proceso de evaluación de competencias laborales, en correspondencia con el umbral determinado en cada caso.
4. Agrupamiento de las similitudes de las evidencias del desempeño: se conforman los grupos de evidencias que guardan relación entre sí en correspondencia con las competencias y el umbral establecido por la organización.

2.5 Fase 1: Procesamiento de las evidencias del desempeño

La fase de procesamiento de las evidencias del desempeño como se aprecia en la Figura 4, constituirá la base para la ejecución con éxito del procedimiento. Se encargará de, la información no estructurada obtenida de las diversas fuentes de información, llevarla a una representación estructurada que permita a las restantes fases poder interpretarla y a su vez, hacer uso de ellas.



Figura 4. Actividades de la Fase 1: Procesamiento de las evidencias del desempeño. Fuente: elaboración propia.

Seguidamente se describen cada una de las actividades, especificándose las entradas, técnicas y herramientas a utilizar y las salidas.

2.5.1 Establecer fecha de corte

Rol: Líder del proyecto

Entradas:

- Documentación del proyecto asociada al negocio del mismo: permite identificar las características e informaciones generales del proyecto que son guardadas con el objetivo de identificar las de relevancia para el proceso.

Salidas:

- Período en el que se escogerán las evidencias del desempeño a relacionar con las competencias definidas por la organización.

Herramientas y tecnologías:

Técnicas de recopilación de información: Para el levantamiento de la información necesaria y delimitar el período del proceso de evaluación de competencias, como es el caso de: tormentas de ideas, entrevistas a las personas con mayor conocimiento sobre el proceso.

Revisión documental y funcional: resultaría efectivo revisar la documentación sobre la información de procesos anteriores y sobre la base estimar el período adecuado.

Descripción:

La fecha de corte es la actividad inicial para llevar a cabo el procedimiento. Esta fecha es la que recoge el período para seleccionar las tareas asignadas a las personas del proyecto en un intervalo determinado. Dicho intervalo lo debe seleccionar el líder del proyecto teniendo en cuenta que es el máximo responsable de llevar a cabo el proceso de evaluación de competencias laborales de sus trabajadores. Por ello, este debe a través de las técnicas y herramientas seleccionadas para la información obtenida delimitar el período (anual, trimestral o mensual) a evaluar.

2.5.2 Obtener las competencias

Rol: Líder del proyecto y rol genérico

Entradas:

- Los puestos de trabajo (PT) o roles definidos para el proyecto: son necesarios ya que a ellos se asocian las competencias laborales definidas por la organización.

Salidas:

- Las competencias laborales y su estado ideal esperado para cada rol o puesto trabajo.

Herramientas y tecnologías: Para la obtención de las competencias laborales previamente definidas por la organización, se utilizan herramientas como: Microsoft Excel o Libre Office que permiten almacenar y editar a la vez, en caso de ser necesario. Para la utilización de las competencias por métodos o funciones automáticas se utilizan gestores de base de datos como: Postgresql 9.3.

Descripción:

Es necesario para la ejecución de esta actividad la obtención de las competencias laborales definidas en el proyecto. En este punto es decisivo el desempeño del líder de proyecto, pues es el encargado de obtener de los

medios donde se encuentren almacenadas en forma de texto las competencias laborales asociadas a los roles correspondientes e insertarla en alguna base de datos. De esta forma el rol genérico podría acceder a estas y analizarlas posteriormente de forma automática.

2.5.3 Obtener las evidencias del desempeño

Rol: Líder de proyecto y rol genérico

Entradas:

- Las tareas realizadas por los miembros de un proyecto que se almacenan en forma de texto en herramientas de gestión de proyectos.

Salidas:

- Las evidencias del desempeño de los miembros de los equipos de desarrollo del proyecto que se encuentran en las tareas asignadas a cada rol.

Herramientas y tecnologías: Para la obtención de las tareas que constituyen evidencia del desempeño se utilizan las herramientas para gestionar proyectos como es el caso de: *Xedro_Gespro*, *GanttProject*, *iceScrum*, *Project-Open*, *Planisware 5*, *Redmine*, *PlannerSuite* y otras.

Descripción:

El líder de proyecto asigna las tareas a cada uno de los miembros del proyecto y desde la herramienta para la gestión de proyecto. Desde dicha herramienta cada integrante del proyecto la cumple total, parcial o la incumple. El rol genérico a través de la propia herramienta, accede y obtiene el conjunto de tareas de un miembro del proyecto. Estas tareas constituirán las evidencias del desempeño. Se toman solamente las que tributan a las tareas que realizan los profesionales en los proyectos de desarrollo de software, debido a que el análisis de información está enfocado a las evidencias del desempeño que tributan a las competencias laborales.

2.5.4 Crear estructura

Rol: Rol genérico

Entradas:

- Las competencias laborales y su estado ideal esperado para cada rol o puesto trabajo.
- Las evidencias del desempeño de los miembros de los equipos de desarrollo del proyecto que se encuentran en las tareas asignadas a cada rol.

Salidas:

- Estructura para almacenar los valores finales del procedimiento.

Herramientas y tecnologías: tablas en Microsoft Excel, una base de datos en Postgresql en aras de que se almacenen los resultados obtenidos en las fases siguientes.

Descripción:

En esta actividad el rol genérico creará una estructura con las características necesarias para el almacenado de los resultados que se obtendrán de la ejecución del procedimiento. De esta estructura serán tomados de forma automática estos resultados que contribuirán de manera efectiva al proceso de evaluación de competencias laborales en proyectos de software, donde la premisa fundamental sea el análisis de evidencias del desempeño. La estructura puede ser creada según la organización, desde tablas en Microsoft Excel hasta una base de datos en Postgresql.

2.6 Fase 2: Análisis de las evidencias del desempeño

La fase de análisis de las evidencias del desempeño como se muestra en la Figura 5, que son manejadas por los sistemas automáticos en forma vectorial para buscar mejoras en la clasificación y garantizar que las mismas no contengan errores.



Figura 5. Actividades de la Fase 2: Extracción de características. Fuente: elaboración propia.

Se describe a continuación, cada una de las actividades, especificándose las entradas, técnicas y herramientas a utilizar y las salidas.

2.6.1 Selección y extracción de características

Rol: Rol genérico

Entradas:

- Las evidencias del desempeño de los miembros de los equipos de desarrollo del proyecto que se encuentran en las tareas asignadas a cada rol y que cuentan con una estructura determinada.

Salidas:

- Conjunto de términos indexados de características adecuadas vectorialmente de las evidencias del desempeño que pueden constituir el vocabulario.
- Base de datos de características de las evidencias del desempeño en forma vectorial.

Técnicas y herramientas:

En la selección efectiva de características, existen dos técnicas importantes: una función de evaluación que puede ser tipo filtro (se basan en métodos estadísticos que emplean técnicas de análisis univariado o multivariado) o *wrapper* (emplean como función de evaluación un algoritmo de clasificación para eliminar las características menos influyentes) y un procedimiento iterativo de búsqueda completa o heurística (encuentra el subconjunto óptimo de características, de acuerdo a la función de evaluación empleada).

Descripción:

En la actividad de selección y extracción de características el rol genérico parte de una secuencia de *tokens* (pequeños grupos de datos que representan un conjunto de información mayor previamente establecida), obtenida de la fase anterior y específicamente de las herramientas de gestión de proyectos y produce una secuencia de términos indexados basados en esos tokens. La forma en que esa secuencia de términos indexados es utilizada depende del procesamiento posterior que se le vaya a realizar al documento y si el vocabulario ha sido construido o no.

Para extraer las características de estos *token* serán utilizados procedimientos de esta índole con el fin de buscar mejorar el rendimiento del agrupamiento.

2.6.2 Reducción de dimensionalidad

Rol: Rol genérico

Entradas:

- Conjunto de términos indexados de características adecuadas vectorialmente de las evidencias del desempeño que pueden constituir el vocabulario.
- Base de datos de características de las evidencias del desempeño en forma vectorial.

Salidas:

- Subconjunto menor de términos indexados.
- El vocabulario que denota el número de términos indexados que permanecen para el posterior agrupamiento.

Técnicas y herramientas:

Técnicas de selección de rasgos y técnicas basadas en la reparametrización con el objetivo de controlar la dimensionalidad de los términos. Ejemplo: Selección de rasgos, umbral de frecuencia de documentos, homogeneidad ortográfica y uso de tesauros.

Descripción:

Una vez que el rol genérico identifica que el conjunto de términos indexados de características adecuadas vectorialmente de las evidencias del desempeño que pueden constituir el vocabulario es usualmente grande, se enfoca en reducir esa dimensión. El objetivo de la reducción de dimensionalidad es disminuir el número de rasgos que son finalmente usados para representar las evidencias del desempeño. Por ejemplo, para el agrupamiento posterior, ese conjunto de términos resultante debe ser lo más acertado posible. Además, aunque una mayor cantidad de rasgos se puede asumir como más información, existen rasgos que son irrelevantes y provocan la obtención de resultados no deseados.

2.6.3 Normalización de los datos

Rol: Rol genérico

Entradas:

- Subconjunto menor de términos indexados.
- El vocabulario que denota el número de términos indexados que permanecen para el posterior agrupamiento.

Salidas:

- Subconjunto de términos indexados normalizados.

Técnicas y herramientas:

- Normalización coseno para normalizar los vectores pesados.

Descripción:

Teniendo en cuenta que trabajar con las frecuencias absolutas de los términos en la representación Modelo de Espacios Vectoriales (VSM) no arroja buenos resultados en el agrupamiento. Normalizar el vector de pesos de un documento determinado permitirá no preocuparse por la variedad de longitudes que estos puedan tener. Por su generalidad, los vectores pesados son normalizados siguiendo la normalización coseno (también considerada como función de similitud). El rol genérico puede tener en cuenta términos probabilísticos, los que comúnmente son normalizados a través de, la suma de los componentes de un vector pesado del documento.

2.6.4 Limpieza de los datos

Rol: Rol genérico

Entradas:

- Subconjunto de términos indexados normalizados.

Salidas:

- Base de datos de características de los términos indexados normalizados libre de errores y listos para ser sometida al proceso de obtención de similitud.

Técnicas y herramientas:

- Utilización de los métodos para la limpieza: los cuales serán ejecutados en las bases de datos de los proyectos.
- Herramientas que tienen implementado funciones de limpieza de datos como el *AnalysisPro*.

Descripción:

La limpieza de los datos es la actividad que como se sugiere en (Escobar, 2016) permite al rol genérico eliminar los errores en la entrada de los términos, la estandarización de términos y la corrección de estos en el procedimiento. De esta forma se garantiza que se utilicen en la comparación, solo los elementos que estarán vinculados a los comportamientos identificativos de las evidencias del desempeño y no preposiciones y artículos.

Estas acciones se llevan a cabo a través de funciones de limpieza de datos integradas al paquete *AnalysisPro*, que realiza de manera automática la limpieza de los datos antes de realizar el análisis de las mismas.

2.7 Fase 3: Obtención de similitud de las evidencias del desempeño

La fase obtención de similitud permitirá determinar la relación, de las competencias y las tareas en proyectos de desarrollo de software, que no son más que las evidencias del desempeño. En la investigación como se pudo constatar en el capítulo 1 disímiles funciones dedicadas a la similitud entre textos, donde en algunos casos arrojaban buenos y no tan buenos resultados, y otras dependían de factores externos para lograr ese fin. A raíz de ello, se proponen para el procedimiento, tres de las funciones que mejores resultados arrojan para la comparación de la similitud que puede existir entre las tareas y las competencias, las de *Dice*, *Jaccard* y *Coseno*, Figura 6. Estas funciones según la bibliografía estudiada, han arrojado los mejores resultados en el proceso de agrupamiento de textos.

Para el trabajo con estas funciones de similitud se utilizará la notación y simbología de las fórmulas propuestas 12 y 13, en caso que se cumpla la condición de que los pesos sean binarios, 14 y 15.

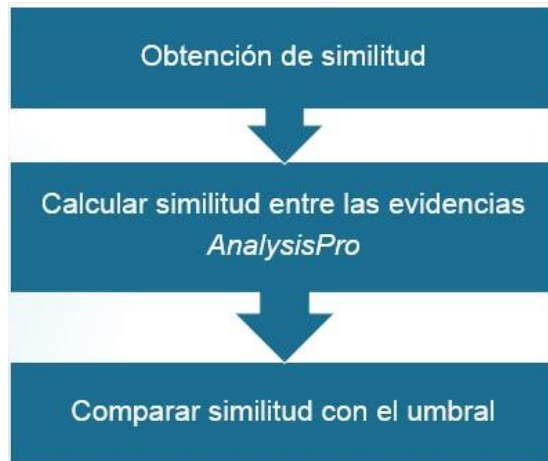


Figura 6. Actividades de la Fase 3: Obtención de similitud. Fuente: elaboración propia.

Se detallan a continuación, cada una de las actividades, especificándose las entradas, técnicas y herramientas a utilizar y las salidas.

2.7.1 Calcular similitud entre las evidencias del desempeño

Rol: Rol genérico

Entradas:

- Base de datos de las evidencias del desempeño en forma vectorial libre errores y listas para ser sometida al proceso de obtención de similitud.
- Funciones de similitud Dice, Jaccard y Coseno: seleccionadas luego del estudio que arrojó que con las mismas se obtienen los mejores resultados de similitud entre textos cortos.
- Puesto de trabajo, rol o roles de la persona en el proyecto.
- Umbral: se representa α -corte para determinar la posible relación tareas-competencias. Este umbral se utiliza para establecer a partir de qué valor de similitud entre las competencias laborales y las tareas que constituyen evidencias del desempeño, se va a considerar que esas tareas contribuyen o muestran la competencia.

Salidas:

- Conjunto de evidencias del desempeño en el período determinado por la organización para obtener la evaluación de las competencias laborales.
- El umbral determinado: encargado de controlar la flexibilidad del conjunto de resultados.
- Valor de similitud de las evidencias del desempeño.

Herramientas y tecnologías:

- Se utiliza el criterio de expertos. Método para la concordancia entre expertos. Se propone emplear un modelo de computación con palabras basado en 2-tuplas.
- AnalysisPro (Lugo, *et. al.*, 2013), haciendo uso del paquete pg_similarity, extensión para soportar consultas de similitud en *PostgreSQL*, con cada una de las variantes, ver Tabla 3.
- Análisis de la particularidad de cada función de similitud: teniendo en cuenta los errores identificados, el estudio de cada función, velando que el valor de los datos no sea afectado.

Descripción:

Escoger el umbral constituye un factor determinante y de éxito en la aplicación del procedimiento, se enfoca en la especificación de las fechas de inicio y fin del período en que se tomarán las evidencias del desempeño para el análisis. El período de evaluación y su duración dependerá de las características de la organización que desarrolle el proyecto. Los períodos pueden tener por ejemplo una duración de uno, tres, seis meses y anuales.

Se selecciona el valor del umbral α -corte a partir de criterios de expertos. Método para la concordancia entre expertos. Se propone emplear un modelo de computación con palabras basado en 2-tuplas. El modelo de representación lingüístico basado en 2-tuplas fue presentado en (Herrera *et. al.*, 2001) para mejorar los problemas de pérdida de información en los procesos de computación con palabras. El rol genérico es el encargado de llevar a cabo esta actividad.

El modelado lingüístico difuso basado en 2-tuplas (Herrera *et. al.*, 2001) permite reducir la pérdida de información que habitualmente se produce en los modelados lingüísticos difusos clásicos. Esta pérdida de información, que provoca una falta de precisión en los resultados, se debe al propio modelo de representación puesto que opera con valores discretos sobre un universo de discurso continuo. La principal ventaja del modelo computacional lingüístico basado en 2-tuplas, es que permite realizar procesos de computación con palabras de forma sencilla y, además, sin pérdida de información, puesto que utiliza un modelo continuo de representación de la información (Cervantes, 2015).

El cálculo de la similitud entre las evidencias del desempeño constituye lo esencial de la propuesta de solución debido a que, la similitud entre textos y puntualmente entre textos cortos como es el caso de la investigación constituye una solución para relacionar las competencias laborales con las tareas del personal en un proyecto de software, de forma que contribuya a la evaluación de competencias laborales. A su vez, aporta en gran medida al proceso de evaluación de competencias laborales en los proyectos de software que utilicen herramientas de gestión para almacenar las evidencias del desempeño de sus profesionales en forma de textos.

Tabla 3. Funciones de similitud Dice, Jaccard y Coseno en la extensión *pg_similarity*.

Función de similitud	Función	Parámetros
<i>Coseno</i>	<i>cosine(text, text) returns float8</i>	<i>pg_similarity.cosine_tokenizer (enum)</i> <i>pg_similarity.cosine_threshold (float8)</i>
<i>Jaccard</i>	<i>jaccard(text, text) returns float8</i>	<i>pg_similarity.jaccard_tokenizer (enum)</i> <i>pg_similarity.jaccard_threshold (float8)</i>
<i>Dice</i>	<i>dice(text, text) returns float8</i>	<i>pg_similarity.dice_tokenizer (enum)</i> <i>pg_similarity.dice_threshold (float8)</i>

Estas funciones son ejecutadas de forma automática a partir de la planificación del proceso de evaluación de competencias en el proyecto. Todas las funciones de similitud devuelven valores normalizados en lugar de puntuaciones de similitud ilimitadas. Las funciones de distancia devuelven puntuaciones no negativas no limitadas. En el caso de la investigación se basa solamente en las funciones asociadas a similitud y no de distancia.

2.7.2 Comparar la similitud con el umbral

Rol: Rol genérico

Entradas:

- Valor de similitud de las evidencias del desempeño.
- Umbral definido por la organización.
- Estructura para almacenar los valores finales del procedimiento.
- Relación tarea/competencias laborales (evidencias) de los miembros de los equipos de desarrollo en el período determinado por la organización, almacenada en la estructura creada en la fase inicial.

Salidas:

- Relación tarea/competencias laborales (evidencias) de los miembros de los equipos de desarrollo en el período determinado por la organización, almacenada en la estructura creada en la fase inicial.
- Almacenaje de relaciones entre tareas y competencias en el umbral definido, en la estructura previamente definida.

Herramientas y tecnologías:

- Implementación de los métodos: mediante el uso de las tecnologías escogidas realizar la implementación de las funciones seleccionadas. Uso del paquete *pg_similarity*, extensión para soportar consultas de similitud en *PostgreSQL*, con cada una de las variantes.
- Implementación de los métodos: mediante el uso de las tecnologías escogidas realizar la implementación de los métodos seleccionados.
- Estructura de tablas en Microsoft Excel o base de datos en Postgresql permiten almacenar los resultados obtenidos.

Descripción:

El rol genérico compara la similitud obtenida en la actividad anterior con el umbral establecido por el líder del proyecto. Esto permite corroborar si la evidencia está relacionada con la competencia en cuestión y se procede al guardado del valor de similitud.

Una vez ejecutado este paso se establece si la tarea está relacionada o no con la competencia, de acuerdo a los niveles de similitud calculados el apartado anterior y comparándolo luego con el umbral definido, encargado de controlar la flexibilidad del conjunto de resultados. Estos valores son utilizados por los operadores para emparejar cadenas. Para cada par de cadenas, si el valor calculado (usando la función de similitud correspondiente) es mayor o igual al valor de umbral, hay una coincidencia.

Comparar la similitud de la tarea con el umbral.

$$S(D_i, D_j) \geq \alpha\text{-corte (18)}$$

Donde $S(D_i, D_j)$ representa el grado de similitud entre la competencia D_i y la tarea D_j , obtenida de una de las variantes ofrecidas en las funciones de similitud propuestas anteriormente e incluso de las tres:

Dice

$$S(D_i, D_j) = \frac{2 \sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}{\sum_{h=1}^k peso_{ih}^2 + \sum_{h=1}^k peso_{jh}^2}$$

Jaccard

$$S(D_i, D_j) = \frac{\sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}{\sum_{h=1}^k peso_{ih}^2 + \sum_{h=1}^k peso_{jh}^2 - \sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}$$

Coseno

$$S(D_i, D_j) = \frac{\sum_{h=1}^k (peso_{ih} \cdot peso_{jh})}{\sqrt{\sum_{h=1}^k peso_{ih}^2 \cdot \sum_{h=1}^k peso_{jh}^2}}$$

Si se cumple la relación planteada en esta ecuación (18), entonces, se puede asegurar que la tarea está relacionada con la competencia en cuestión y se procede al guardado del valor de la similitud.

Posteriormente se almacena el valor de similitud entre las competencias y las tareas mayores que el umbral definido en forma de vectores en una base de datos con la denominación siguiente: $(IDMiembro_n, IDTarea_j, IDCompi, Competencia_j, Tarea_i, S(D_i, D_j) Jaccard, S(D_i, D_j) Coseno \text{ y } S(D_i, D_j) Dice$. Permitiendo así, almacenar los grupos de relaciones entre tareas y competencias en un umbral definido.

Pueden ser guardados los tres valores de similitud calculados con las ecuaciones 12, 14 y 15, o sencillamente uno de ellos. Valores $S(D_i, D_j)$ en el rango del α -corte establecido y 1, o sea, los valores de similitud en el umbral definido para ello. Dichos valores mostrarán la fortaleza que existe en la relación entre las tareas y las competencias laborales definidas para uno o varios roles en los que trabajan los miembros de los proyectos de software.

2.8 Fase 4: Agrupamiento de las similitudes de las evidencias del desempeño

Rol: Rol genérico

Entradas:

- Almacenaje de relaciones entre tareas y competencias en el umbral definido, en la estructura previamente definida.
- Algoritmos de agrupamiento según selección de la organización para el trabajo.

Salidas:

- Grupos de evidencias que guardan estrecha relación en el proceso de desarrollo de software según un rol o PT y un período establecido, que contribuye al proceso de evaluación de competencias laborales.
- Grupos de competencias.
- Los grupos de valores de similitud.

Herramientas y tecnologías:

- Algoritmos de agrupamientos seleccionados por la organización, que sean utilizados en la minería de textos y específicamente en el agrupado de textos cortos como vectores. Ejemplo de ellos: *Simultaneous Keyword Identification and Clustering of Text Documents (SKWIC)*, *Simultaneous Keyword Identification and Fuzzy Clustering of Text Documents* y *Extended Star*.

Implementación de los métodos de agrupamiento: mediante el uso de las tecnologías escogidas realizar la implementación de los métodos seleccionados.

Descripción:

El rol genérico en la fase asociada directamente al agrupamiento, conforma los grupos en sí, según su relación. El agrupamiento puede efectuarse de maneras diversas, todo ello dependiendo del algoritmo de agrupamiento

seleccionado para esto y de las propiedades del mismo. Entre los métodos de agrupamiento que se usan con frecuencia en la minería de textos, se encuentran:

- *Simultaneous Keyword Identification and Clustering of Text Documents* (SKWIC) que realiza un análisis de clusters duro y determinista.
- *Simultaneous Keyword Identification and Fuzzy Clustering of Text Documents* (FuzzySKWIC) que utiliza una técnica de análisis de clúster borrosa.
- *Extended Star* que aplica una técnica de análisis de clúster dura y con solapamiento.

Los tres algoritmos tienen en común que parten de una representación VSM del *corpus* textual y devuelven una colección de *clusters*. La estructura de los *clusters* varía en dependencia de los algoritmos de agrupamiento.

2.9 Valoración de los resultados

A partir de los resultados obtenidos con el procedimiento para la evaluación de competencias laborales basados en funciones de similitud de textos, se puede realizar un posterior agrupamiento de la correlación existente entre las tareas y las competencias laborales definidas, sin mediar la subjetividad del factor humano. De esta forma se puede realizar un correcto proceso de evaluación de competencias laborales en proyectos de software. La obtención de forma simultánea de los grupos de textos al agruparlos es de gran utilidad. Estos textos son separados en categorías según su relevancia y se genera de manera automática la descripción compacta de cada grupo en términos no sólo de los valores de los atributos, sino también de su relevancia y correspondencia.

2.10 Conclusiones del capítulo

Se desarrolla el procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos, para la obtención de la correlación existente entre las tareas del personal del equipo de desarrollo de software y las competencias laborales definidas por el proyecto en un determinado período de tiempo. De todo ello, se obtiene como salida la relación entre las evidencias del desempeño, que serán la antesala para un posterior agrupamiento que contribuirá a la evaluación de competencias laborales.

A diferencia de soluciones consultadas en la bibliografía, la propuesta se basa en las evidencias del desempeño almacenadas en herramientas de gestión de proyectos, en períodos de tiempo.

Se pretende con la propuesta de solución mejorar la obtención de la relación entre las evidencias del desempeño en los proyectos de manera que contribuya a la eficiencia del proceso de evaluación de competencias laborales como apoyo a la evaluación de competencias laborales en organizaciones con mejora continua.

CAPÍTULO 3. VALIDACIÓN DE LA PROPUESTA Y ANÁLISIS DE RESULTADOS

3.1 Introducción

En el presente capítulo se tiene como principal objetivo validar el procedimiento, a partir de su aplicación en herramientas para la gestión de proyectos. Para este propósito primeramente se lleva a cabo una caracterización del entorno de aplicación del procedimiento y una síntesis de la aplicación inicial. Seguidamente se hace uso de un pre-experimento, donde se realiza una comparación atendiendo a las variables de la investigación, con las tres variantes para el cálculo de la similitud entre textos de las evidencias del desempeño. Se concluye con un análisis del impacto económico y social del desarrollo y utilización de la propuesta solución.

3.2 Caracterización del entorno de aplicación

La propuesta está orientada a la ejecución de un procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos durante la gestión de proyectos, de manera que se reutilicen las experiencias de los expertos. Se enfoca específicamente en la relación existente entre toda la evidencia recogida en la herramienta para gestión de proyectos, para luego, analizarlas y obtener las que se consideren relevantes, posteriormente agruparlas y tener la correspondencia. Se contribuye de esa forma al proceso de evaluación de competencias laborales.

El cálculo de la similitud, a través de los coeficientes *Dice*, *Jaccard* y *Coseno*, se realiza a partir de las evidencias del desempeño arrojadas por el sistema de gestión de información del proyecto. En caso de la UCI se toman del Xedro-Gespro las evidencias del desempeño correspondientes al personal de los proyectos de software. Estas son recopiladas por el rol genérico, en un período, que para esta validación se sigue el período de evaluación de la universidad trimestral, determinado por el líder del proyecto. Los proyectos utilizados para la aplicación de la propuesta de solución corresponden a centros de desarrollo de software de la propia institución.

A las evidencias del desempeño almacenadas en el Xedro_Gespro se les calcula la similitud mediante funciones implementadas en un mecanismo automatizado en el paquete *pg_similarity* para *AnalysisPro* donde, a través de una función son introducidos los datos y posteriormente se obtiene el resultado de manera automática y el que se guarda en la estructura definida para ello. De esta forma se cuenta con la relación entre las evidencias del desempeño de manera actualizada y en el tiempo requerido, con un mínimo de esfuerzo y de errores en las operaciones de cálculo.

Para garantizar el procedimiento sobre las evidencias del desempeño existentes en la herramienta, partiendo de un enfoque supervisado, se evalúa por un grupo de expertos. Todo ello con el fin de contar con un procedimiento acorde al nivel de madurez de la gestión en la organización. En este caso se realiza la evaluación por los expertos con período trimestral.

Para la validación de la presente investigación se seleccionó la realización de un pre-experimento como vía para comparar una muestra de un grupo de 1051 tareas de miembros de los equipos y las competencias laborales

definidas para estos en el desarrollo de proyectos de software, de cinco proyectos de los centros de producción de la UCI. Seguidamente, se describe la ejecución del procedimiento.

3.3 Ejecución del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos

En consecuencia, con la muestra seleccionada se inicia la ejecución del procedimiento. A continuación, se muestra una valoración de los resultados obtenidos por cada una de las fases del mismo.

La red de centros de la UCI está conformada por un total de 14 centros de desarrollo distribuidos indistintamente por las cinco facultades con que cuenta la universidad. Debido a la cantidad de usuarios que colaboran y desarrollan un conjunto de soluciones informáticas dentro de los centros, fue necesaria la introducción de una herramienta de gestión de proyectos para la ayuda en la toma de decisiones a diferentes niveles: nivel de persona, nivel de proyecto, nivel de centro de producción y nivel alta gerencia. En respuesta a dicha limitante es que surge la herramienta para gestionar proyectos Xedro-Gespro.

Esta herramienta es utilizada en los 14 centros para controlar, planificar y gestionar el desarrollo productivo. Entre las áreas que gestiona se encuentra la relacionada a los recursos humanos. En consonancia con esta área es almacenado un gran volumen de información asociado al comportamiento de cada uno de los integrantes del proyecto, teniendo en cuenta a su vez, el cumplimiento de cada una de las tareas o actividades asignadas.

3.3.1 Fase 1: Procesamiento de las evidencias del desempeño

Un análisis preliminar realizado a la muestra seleccionada evidenció que la primera limitante se encontraba en que la información asociada al desempeño de los integrantes del proyecto (las tareas) era almacenada de manera no estructurada en la herramienta para la gestión de proyectos y por ende no se podía trabajar de manera automática con ellas. Por lo que es identificado que, en el proceso de evaluación de competencias, por la particularidad de realizarse por personas, se hacía un análisis muy subjetivo que suele introducir errores en el resultado del mismo.

Como resultado de la ejecución de esta fase inicial, se obtuvo del Xedro_Gespro los elementos que constituyen evidencias del desempeño laboral y las competencias laborales a comparar. Para el logro de ello primeramente se estableció la fecha de corte, la cual delimita el período en el cual se escogerán las tareas a evaluar del Xedro_Gespro, decisión tomada por el líder del proyecto que en este caso fue trimestral. Estas tareas se encontraban en forma de texto no estructurado, por lo que se identificó la necesidad de trabajar con una representación estructurada que permita a las actividades de las restantes fases la correcta interpretación de la misma. Se seleccionan de Microsoft Excel las competencias laborales asociadas a cada uno de los roles involucrados en el proceso. Una vez seleccionadas fueron almacenadas en una base de datos donde, a través del gestor de base de datos *Postgresql* en su versión 9.3 se pudo trabajar con las mismas. Por último, se creó la estructura que constituyó una base de datos como medio para almacenar el resultado final del procedimiento y

de ahí, poder extraer los grupos de evidencias del desempeño y tareas relacionadas para el proceso de evaluación de competencias laborales en los proyectos de software en la UCI.

3.3.2 Fase 2: Análisis de las evidencias del desempeño

En esta fase se llevó a cabo el análisis de las evidencias del desempeño, de manera estructurada para seleccionar y extraer las más adecuadas para el proceso posterior de obtención de similitud. Para ello se hizo uso de un procedimiento iterativo de búsqueda completa el cual permitió encontrar el subconjunto de características adecuadas. Todo ello de acuerdo a la función de evaluación de tipo filtro basada en métodos estadísticos con técnicas de análisis multivariado, luego de la realización del procedimiento de evaluación sobre todo el conjunto de evidencias del desempeño seleccionadas en el trimestre.

A partir de la obtención de las evidencias del desempeño relevantes, se llevó a cabo la reducción de la dimensionalidad con el objetivo de disminuir el número de características que son finalmente usadas para representar las evidencias del desempeño. En este caso la técnica de reparametrización selección de rasgos y dentro de ellas, la selección de los términos con mayores valores de calidad en la agrupación, se logró identificar aquellos términos que caracterizan cada agrupación. Seguidamente con el fin de no tener en cuenta las diferentes longitudes que pueden tener los vectores se llevó a cabo la normalización de las evidencias dimensionadas, haciendo uso de la técnica de normalización coseno.

Se trabajó en garantizar que las evidencias dimensionadas y normalizadas a su vez, no contengan errores. Con anterioridad, se identificó que el principal problema radicaba en la realización del proceso de evaluación de competencias, por la particularidad de realizarse por personas, se hacía un análisis muy subjetivo y que solía introducir errores en el resultado del mismo. Teniendo en cuenta la subjetividad con que se realiza el proceso se hace un análisis de la información utilizada en el mismo. De aquí que se identificaran los siguientes errores (Escobar, 2016):

Tareas incompletas:

- 1) Tareas que no tienen asignada una fecha de inicio.
- 2) Tareas que no tienen asignada una fecha de fin.
- 3) Tareas que no están asignadas a ningún usuario.

Tareas incorrectas:

- 4) Tareas con fecha final menor que la fecha inicial.
- 5) Tareas con valor de tiempo estimado igual a cero.
- 6) Tareas duplicadas.

Tareas que presentan inconsistencias:

7) Nombres de tareas muy generales que no permiten exactamente saber que se va a realizar.

8) Nombres de las tareas que indican un problema y no una tarea.

9) Tareas que son no conformidades.

10) Utilización de términos no estandarizados en la descripción de las tareas. Ejemplo:

a. Utilización para las tareas de diseño de interfaz de usuario: realizar diseño, diseñar.

b. Incorrecta clasificación de las tareas de Desarrollo-Producción. Clasificadas como tareas que realmente son de gestión o formación.

El tratamiento de los errores identificados se plantean dos estrategias: la eliminación de los errores y la sustitución de cadenas para los problemas de estandarización. En correspondencia a esto se propone erradicar los errores de datos incompletos y datos incorrectos según Escobar, 2016 mediante la implementación de consultas SQL como se describe a continuación (Escobar, 2016).

Mediante la técnica de imputación de datos:

- Error 1: si la fecha de inicio es nula, se realiza la imputación del dato mediante la asignación del valor de la fecha de fin de la tarea.
- Error 2: si la fecha de fin es nula, se le asigna el día después de la fecha de inicio de la tarea.
- Error 3: si la tarea no tiene asignada ningún usuario, se le asigna como responsable el usuario que la creó.
- Error 4: si la fecha de fin es menor que la fecha inicial, se le asigna como fecha final a la tarea la misma fecha de inicio.
- Error 5: cuando el tiempo estimado de la tarea es nulo o cero, se calcula la cantidad de días entre la fecha de inicio y de fin, y se multiplica por 6, que es la cantidad de horas estimadas que se trabaja por día.
- Error 6: para la erradicación de los duplicados se hizo uso de la Distancia de Jaro (ESTOQUE y BARÓ, 2013), la cual evalúa que tan iguales son dos cadenas de caracteres. Se valida primeramente que la tarea esté asignada a la misma persona y tenga la misma fecha de inicio y de fin, luego se calcula la distancia de Jaro para el nombre de la tarea, y si esta es mayor que 85 se considera que las cadenas son semejantes y se procede a la eliminación de una de ellas.

Por su parte, los errores de las tareas inconsistentes identificados debido a la complejidad de solución que presentan no van a ser erradicados en esta iteración del procedimiento, proponiéndose mejoras en la fase 4 que ayuden a mitigar la presencia de estos errores.

Como resultado de la ejecución de esta fase, se obtiene la base de datos de características de las evidencias del desempeño en forma vectorial libre de errores y listas para ser sometida al proceso de obtención de similitud. De esta forma se garantiza que las mismas no contengan elementos que puedan afectar el proceso.

3.3.3 Fase 3: Obtención de similitud de las evidencias del desempeño

Una vez identificados los métodos que permitirán erradicar los errores encontrados en la fase anterior se procedió a la siguiente: Obtención de similitud. Como primera actividad de esta fase se hace uso del umbral definido en la fase inicial, factor determinante y de éxito en la aplicación del procedimiento ya que, se enfocó en la especificación del período en que se tomarán las evidencias del desempeño para el análisis. Con el uso de una función implementada en *plpgsql* incluida en el *AnalysisPro* para cada método, permitió la introducción de los parámetros necesarios en cada caso. Estos parámetros la función la toma de las herramientas donde están almacenados, en este caso en Microsoft Excel y seguidamente se obtuvieron los valores de similitud.

Una vez obtenidas las similitudes fueron asociadas al período establecido en la configuración del sistema, que como se planteó anteriormente fue trimestral, en la implementación realizada para Xedro-Gespro 13.05 utilizándose las variantes de las funciones de similitud *Jaccard*, *Dice* y Coseno. Para la automatización de esta actividad se utilizaron las funciones de similitud entre textos incluidas en el paquete *pg_similarity* para *AnalysisPro* (Taveira, 2012).

3.3.4 Fase 4: Agrupamiento de las evidencias del desempeño

En el caso de la ejecución de la cuarta fase correspondiente al agrupamiento, como se ha dicho anteriormente, se formaron los grupos de competencias y tareas relacionadas, además de los valores de similitud obtenidos. Estos elementos fueron agrupados en grupos afines en la estructura establecida en la fase inicial que en este caso fue una base de datos relacional. Como la tarea del agrupamiento puede efectuarse de maneras diversas, se seleccionaron los algoritmos: *Simultaneous Keyword Identification and Fuzzy Clustering of Text Documents* y *Extended Star*, ambos que parten de una representación VSM del corpus textual y devuelven el conjunto de evidencias del desempeño que guardan relación entre sí.

Utilizando las fórmulas definidas en el capítulo anterior para el cálculo de las funciones en este procedimiento y con su implementación en *AnalysisPro* se permitió la obtención de los resultados respecto a la relación tareas - competencias laborales. Estas funciones se ejecutan automáticamente. Se obtiene de esta forma la caracterización de las competencias laborales, ver anexo 3.

Se realizó un pre experimento de los resultados de los cálculos de las tres funciones propuestas, con el objetivo de establecer cuál ofrece los mejores resultados y utilizarla en la aplicación de la relación entre las evidencias del desempeño en un modelo posterior. Se tomó como muestra para el pre-experimento un grupo de 1051 tareas de miembros de los equipos y las competencias laborales definidas para estos en el desarrollo de proyectos de software. Se les aplicaron las funciones para la limpieza de los datos (Escobar, 2016) y posteriormente se les calculó el valor de similitud entre tareas y competencias sin establecer un umbral. De lo

que se obtuvo un vector $(IDMiembro_n, IDTarea_j, IDCompi, Competencia_j, Tarea_i, S(D_i, D_j) \text{ Jaccard}, S(D_i, D_j) \text{ Coseno y } S(D_i, D_j) \text{ Dice}$. Pueden ser guardados los tres valores de similitud calculados con las ecuaciones 13, 15 y 16, o sencillamente uno de ellos; donde los valores de la similitud se encuentran en el intervalo [0-1].

Se aplicó primero la prueba de *Shapiro-Wilk* para comprobar la normalidad de los datos, los resultados de esta prueba mostraron que los datos de similitud de las competencias y las tareas no seguían una distribución normal, al obtenerse valores de significancia menores de 0.05, Tabla 4. Posteriormente se aplicó la prueba no paramétrica de Friedman que arrojó la existencia de diferencias significativas entre los resultados de las funciones de similitud.

Para determinar si existen diferencias significativas o no, se aplica el test no paramétrico de *Wilcoxon* para dos muestras relacionadas, aplicando el método de Monte Carlo con un 99% de intervalo de confianza. Los resultados de la aplicación de este test dos a dos al procedimiento propuesto, arrojó que la función de similitud Coseno ofrecía los mejores resultados. Para la realización de estas pruebas se hizo uso de las funciones implementadas en lenguaje PLR integrando las funcionalidades del paquete *Stats* (Hollander, 2014) de R con *PostgreSQL*. Estas funciones se encuentran integradas al paquete *AnalysisPro*.

Grupo- 1: Coseno, Dice

Grupo- 2: Jaccard

Tabla 4. Normalización de los datos.

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Jaccard	.207	1651	.000	.795	1651	.000
Cosine	.186	1651	.000	.824	1651	.000
Dice	.194	1651	.000	.845	1651	.000

La variante de la función de similitud de Coseno estima los valores de similitud más altos, es decir es más flexible para establecer la relación entre las competencias laborales y las tareas que constituyen evidencias del desempeño en el proyecto. Teniendo en cuenta que, aunque se realizan acciones de limpieza y de selección de los *tokens* más relevantes, las tareas y las competencias son definidas por personas con diferentes grados de conocimiento y formas de dicción, es necesario contar con más flexibilidad para establecer la similitud entre estos cuerpos textuales.

Posteriormente las mismas tareas y competencias fueron entregadas a un grupo de expertos en el desarrollo de proyectos de software a los cuales se les solicitó que evaluaran la similitud en valores de (R-Relacionada, D-Duda y N-No relacionada), aplicando el Modelo lingüístico de 2-tuplas para encontrar consenso. Se aplicó otra ronda para las tareas que quedaron con dudas, definiendo si estaban relacionadas o no.

Como los datos no siguen una distribución normal, se aplicó la prueba no paramétrica de Friedman para las muestras relacionadas. Los resultados mostraron diferencias significativas entre los resultados como se muestra en la Tabla 5.

Tabla 5. Pruebas no paramétricas de Friedman

Friedman

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Cosine	1651	.210770	.0722726	.1048	.5774	.166667	.188982	.223607
Dice	1651	.202372	.0678384	.0870	.5714	.153846	.181818	.222222

Por consiguiente, al encontrarse diferencias significativas entre los resultados, se aplicaron pruebas de *Wilcoxon* para dos muestras relacionadas, a través de las funciones Dice y Coseno, lo que mostró que entre los resultados de estas muestras no existen diferencias significativas, como se aprecia en la Tabla 6.

Tabla 6. Pruebas de Wilcoxon.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
Dice - Cosine	Negative Ranks	1422(a)	711.50	1011.00
	Positive Ranks	0(b)	.00	.00
	Ties	229(c)		
	Total	1651		

a Dice < Cosine

b Dice > Cosine

c Dice = Cosine

Para ambas pruebas (*Friedman* y *Wilcoxon*) se aplicó el método de simulación Monte Carlo, ver Tabla 7, con un 99% de intervalo de confianza.

Tabla 7. Método de simulación Monte Carlo.

Test Statistics (b, c)

	Dice - Cosine
Z	-32.676(a)
Asymp. Sig. (2-tailed)	.000
Monte Carlo Sig. (2-tailed)	.000
99% Confidence Interval	.000
Lower Bound	.000
Upper Bound	.000
Monte Carlo Sig. (1-tailed)	.000
99% Confidence Interval	.000
Lower Bound	.000
Upper Bound	.000

Para la realización de estas pruebas se implementaron funciones en lenguaje PLR integrando las funcionalidades del paquete *Stats* de R con *PostgreSql*. Las mismas se integraron al paquete *AnalysisPro*.

Tabla 8. Proceso de comparación de las tareas y las competencias usando las funciones de similitud seleccionadas.

issues_cp text	dimension_cp text	dice_cp double precis	cosine_cp double precis	jaccard_cp double precis
actualizar cambios modelo base datos	llevar una bitacora actividades realizadas base datos	0.333333	0.338062	0.2
actualizar cambios modelo base datos	documentar base datos	0.5	0.516398	0.333333
actualizar cambios modelo base datos	verificar controlar estado uso reglas integridad base datos	0.307692	0.316228	0.181818
actualizar cambios modelo base datos	controlar objetos logicos crean actualizan base datos	0.222222	0.248069	0.125
actualizar cambios modelo base datos	administrar politicas seguridad auditoria firewall bases	0.166667	0.169031	0.090909
actualizar cambios modelo base datos	implementar diferentes politicas seguridad base datos	0.222222	0.248069	0.125
actualizar cambios modelo base datos	activar auditorias especificas actividades base datos	0.307692	0.316228	0.181818
actualizar cambios modelo base datos	actualizar forma periodica estadisticas bd alimentar modelo	0.142857	0.149071	0.076923
actualizar cambios modelo base datos	configurar servidores aplicaciones conectarse base datos	0.266667	0.282843	0.153846
actualizar cambios modelo base datos	realizar reorganizaciones base datos mejorar rendimiento	0.307692	0.316228	0.181818
actualizar cambios modelo base datos	analizar rendimiento base datos realiza afinamiento serv	0.266667	0.282843	0.153846
actualizar cambios modelo base datos	levantar indicadores metricas conocer comportamiento base	0.307692	0.316228	0.181818
propuesta inicio montaje modelo datos	recuperar ordenar restringir datos utilizando sentencia	0.166667	0.169031	0.090909
propuesta inicio montaje modelo datos	manipular datos tablas utilizando sentencias insert update	0.153846	0.158114	0.083333
propuesta inicio montaje modelo datos	visualizar datos varias tablas utilizando join inner join	0.125	0.13484	0.066667
propuesta inicio montaje modelo datos	instalar parches service pack base datos	0.181818	0.182574	0.1

Como se puede apreciar en la Tabla 8 queda evidenciado como las tareas realizadas por los miembros de los equipos de desarrollo que tienen alguna relación con una determinada competencia arrojan resultados estadísticos mayores que 0 al aplicar las funciones de similitud (Dice, Coseno y Jaccard), pudiéndose identificar esta relación de manera simple y rápida que anteriormente estaba condicionada a la subjetividad del evaluador, lo cual significa un ahorro de tiempo y esfuerzo significativo para los miembros de proyecto que necesitan la información para apoyar a la evaluación de competencias laborales.

En aras de garantizar mejoras en el trabajo con las técnicas, herramientas y el procedimiento, se identifican las siguientes, con el fin de mantener el nivel de eficiencia obtenido.

Mejoras a las herramientas de gestión de proyectos:

- Las interfaces donde se asignan las tareas deben ser modificadas con el fin de eliminar errores de edición y de completitud de los datos. Se deberán establecer campos obligatorios a llenar como: usuario a asignar la tarea, período a evaluar y duración.
- Estandarizar los términos asociados a las tareas para evitar los errores de edición.
- Incorporación de técnicas de generación semiautomática de tareas a partir de procesos de gestión preestablecidos, por ejemplo: a partir de requisitos, riesgos y otros.
- Incluir técnicas de validaciones de las tareas asignadas, para garantizar por ejemplo que: en el período insertado la fecha de fin debe ser igual o mayor que la fecha de inicio.

Mejoras a la organización:

- Capacitar al personal que interactuará con la herramienta, a través de cursos de entrenamiento u otros.

Una vez culminado el procedimiento se genera el portafolio del mismo, donde se guardarán los artefactos realizados a lo largo del proceso. Este portafolio constituirá la base para futuros procesos de obtención de similitud de textos.

3.4 Comparación respecto a la eficiencia del procedimiento

Para valorar la eficiencia del procedimiento propuesto se tuvo que integrar al Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento. Se seleccionó el método de Escalas Numéricas ya que también se basa en evidencias, aunque no propone indicadores (de Andrés, 2009) (Pérez, 2010) (Palmero y otros, 2012).

Los métodos basados en los criterios de expertos aumentan subjetividad a partir de la percepción humana. Los diagnósticos, dependiendo de la calidad de las preguntas realizadas, pueden ofrecer una imagen acertada de las competencias laborales, pero no muestran el desempeño y evalúan solo un momento. La recolección de evidencias ofrece una imagen más objetiva del desempeño de los miembros de los equipos, mientras la utilización de un cuadro de mando integral, ayuda a la visualización del estado de las competencias laborales. Se destacan en lo referente al método utilizado para la evaluación las propuestas de (Cuesta, 2011), (Hernández y otros, 2009) y la realizada en esta investigación.

Debido a las condiciones planteadas con anterioridad en la investigación los equipos de trabajo donde se aplicó el procedimiento propuesto, no eran los mismos, ni presentaban la misma cantidad de miembros. Sin embargo, el análisis estuvo encaminado a conocer la eficiencia de la aplicación del mismo, no a comparar la evaluación dada por los distintos métodos y/o modelos a un mismo grupo de personas.

3.4.1 Evaluación a través de Escalas Numéricas

El motivo de selección del método de Escalas Numéricas para la aplicación del procedimiento propuesto está dado por la coincidencia de basar la evaluación del personal utilizando evidencias. Este fue aplicado a miembros de proyectos del centro CIDI, en la Facultad 1 de la Universidad de las Ciencias Informáticas. Para ello, se escogieron 37 personas que pertenecían a proyectos comunes que tuvieran asignados al menos dos de los roles establecidos en la universidad para el desarrollo de software, para los cuales están definidas las competencias laborales a mostrar. Añadido a ello, se seleccionó para la evaluación un especialista que contara con más de cinco años de experiencia en el desarrollo de software y al menos tres vinculado en procesos de evaluación de los recursos humanos. En el Anexo 10 se muestra un resumen de la aplicación del método.

Una vez aplicado el procedimiento y en consulta con el especialista se pudo constatar que es un trabajo que manualmente sería muy tedioso e inoperativo. El tiempo que se demora por persona y de tareas es como promedio de tiempo de 15 y 23 minutos respectivamente, evaluando un solo rol. A medida que aumentan las tareas el tiempo crece considerablemente, pues hay que chequear parámetro por parámetro de evaluación según la tarea (Sucede de igual manera si el especialista se desempeña en más de un rol).

Para realizar un mejor trabajo se debería tener en cuenta la variedad de valores que pueden tener las tareas: complejidad, tiempo planificado y ejecutado, completada y evaluada. Ello complejizaría aún más el tiempo de evaluación por especialista.

3.4.2 Evaluación del procedimiento integrado al Modelo de evaluación de competencias a partir de evidencia utilizando Sistemas Basados en Conocimiento

Por su parte, con el Modelo de evaluación de competencias a partir de evidencia utilizando Sistemas Basados en Conocimiento el cual tiene integrado el procedimiento propuesto, siguiendo la misma muestra y área de aplicación como el caso de Escalas numéricas, se identifican dos momentos fundamentales:

Primer momento: cuando se emplean expertos utilizando el modelo de representación lingüística 2-tuplas para la creación de la base de conocimiento. Esta actividad puede realizarse solo una vez al comenzar a utilizar el modelo o cuando existan cambios en las competencias laborales de la organización. La cantidad de expertos utilizados para lograr el consenso de la pre-evaluación inicial puede variar de acuerdo a la cantidad de roles y de competencias laborales definidas en el proyecto. Con tres expertos se puede obtener una evaluación satisfactoria al contar con la caracterización de las competencias.

El segundo momento se realiza durante el proceso de obtención de la similitud entre las tareas asociadas a los roles, de acuerdo a la frecuencia planificada en el proyecto u organización, por lo que, se realizará periódicamente. En este momento la necesidad de RRHH sería solamente de una persona por proyecto para que compruebe si el resultado ofrecido por el sistema es el esperado para cada situación.

3.4.3 Comparación de la eficiencia de las variantes para la ejecución del procedimiento

La capacidad de poder obtener un determinado efecto que generalmente es deseado contando con el mínimo de recursos disponibles y en el menor tiempo posible, no es más que la eficiencia. Como se planteó para evaluar la eficiencia de la propuesta se comparan los resultados obtenidos de la implementación del procedimiento en el Modelo de evaluación de competencias a partir de evidencia utilizando Sistemas Basados en Conocimiento y el método de Escalas Numéricas a miembros con los mismos roles de proyectos del centro CIDI de la Facultad 1 de la UCI.

En la Figura 7 se muestra una comparación de los procesos de evaluación de competencias laborales, respecto a la cantidad de personas o expertos evaluadores requeridos en un proyecto en un año, infiriendo que se realizan trimestralmente (recordar que la organización define el período), con la implementación del procedimiento propuesto. Se tuvo en cuenta para la selección de los RRHH a utilizar en cada método, la necesidad requerida desde el inicio hasta la culminación del mismo. Por otra parte, el período escogido fue de un año delimitándose por trimestres, según el período de evaluación en la UCI.



Figura 7. Personal requerido para la aplicación del procedimiento con el método Escalas Numéricas y el Modelo de evaluación de competencias a partir de evidencia utilizando Sistemas Basados en Conocimiento.

Fuente: elaboración propia.

Si bien el método de Escalas Numéricas, mantiene un número menor de personas para la ejecución, se observó que requiere más tiempo, el cual puede aumentar significativamente de acuerdo a la cantidad de evidencias del desempeño a calcular similitud. Por su parte, el Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento, a pesar de requerir más personal en el momento inicial, luego solo requiere uno, aunque la cantidad de evidencias del desempeño aumente no lo afecta en cuanto al tiempo, ya que el procesamiento de las mismas se hace de forma automática y fuera del horario del horario laboral.

La cantidad de RRHH requeridos para llevar a cabo el procedimiento con el Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento, es necesario utilizar una persona por proyecto que corrobore el resultado obtenido en el sistema, lo que implicaría a 200 personas, tomando como ejemplo una organización con dimensiones como la UCI. Si se toma que los equipos son de siete personas y que para cada proyecto la persona ocupa 0.5 horas, la cantidad de horas total para realizar el procedimiento sería de 100 horas para toda la UCI.

El Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento, enfoca su utilización en organizaciones orientadas a proyectos, que deben cumplir con cronogramas específicos y donde el análisis de las evidencias del desempeño es una tarea de apoyo que necesita ser realizada con la eficiencia necesaria. Teniendo en cuenta, además, que métodos donde se toman valoraciones de todos los miembros del equipo pueden ofrecer informaciones que debido a la subjetividad e influencia de la opinión de líderes y directivos no sean las adecuadas.

3.5 Valoración de los modelos

Los resultados obtenidos en la aplicación del procedimiento para la evaluación de competencias laborales basado en funciones de similitud de textos por el método de Escalas Numéricas y el Modelo de Evaluación de competencias a partir de evidencia utilizando sistemas basado en conocimiento; permitió evidenciar que el

modelo con los mejores resultados fue este último. Todo ello se puede afirmar teniendo en cuenta la cantidad de personas involucradas para la ejecución del mismo y el tiempo requerido para la realización del proceso. Por lo tanto, con este modelo se facilita la automatización del proceso de evaluación de competencias laborales de las evidencias del personal de proyecto de software, utilizando funciones de similitud de textos, evitando la utilización de técnicas invasivas para estimar el desempeño que atenten con la veracidad y objetividad de tan importante proceso.

3.6 Análisis de la adaptabilidad del procedimiento

Es necesario que el procedimiento se pueda extender a otras áreas u organizaciones en los cuales también se realice de alguna u otra manera procesos de análisis de evidencias o tareas de sus recursos humanos. El análisis de la adaptabilidad de la propuesta estará en función de la generalidad que debe presentar para poder aplicarse a otros entornos no solo de gestión de proyectos, a otros modelos y en evaluar la extensibilidad que debe mostrar para valores añadidos en un futuro.

Para la obtención del procedimiento se tuvo en cuenta el estudio bibliográfico realizado de varias funciones de similitud de textos, identificándose las que coincidían específicamente con los textos cortos. Las competencias y tareas se guardan en forma de textos en herramientas para gestión de proyectos y son de vital importancia para la realización del mismo. En función de este análisis, se obtuvo que no cumplían las características adecuadas para utilizarlas en este caso y otras que suplen las deficiencias encontradas en este estudio, conformándose la propuesta de solución. La relación de las herramientas para la gestión de proyectos que se basan en la asignación de tareas a los roles de los proyectos se pueden apreciar en el Anexo 4.

En el Anexo 4 se recogen las herramientas destinadas a gestionar proyectos que cuentan con la particularidad de que asignan tareas, actividades, peticiones a los roles de los proyectos donde son utilizadas. Haciendo alusión a lo planteado como premisa para la ejecución de este procedimiento, se basa en las tareas del personal de los proyectos. Estas tareas constituyen las evidencias del desempeño, por ende, se puede asegurar que el procedimiento puede ser implementado en cualquiera de las organizaciones que utilice alguna de las herramientas anteriormente enunciadas.

De las fases propuestas, aunque son específicas del procedimiento propuesto, el 75% mantiene un nivel de generalización alto. En ellas se describen que se debe realizar, dejando un margen para que la organización o proyectos según sus potencialidades determinen de qué forma realizar el procedimiento. El otro 25 % que lo engloba la fase de procesamiento de las evidencias del desempeño presenta un nivel medio, debido a que la organización puede determinar el uso de una herramienta para realizar el proceso. En este caso en específico, se puede ejecutar el resto del procedimiento, lo que no se implementarían métodos, solo se configuraría la herramienta, o podría darse el caso que se haga uso de ambos.

En el caso de la propuesta no se propuso el uso de herramientas debido a que, ello incorpora un intermediario entre el sistema para la gestión de proyecto y la realización del procedimiento, lo que conlleva a caer nuevamente en subjetividades y más trabajo por parte de los RRHH de la organización destinado a ello. La

extensibilidad del procedimiento se refiere a la flexibilidad que debe presentar el procedimiento, que permita un crecimiento en las actividades que se realizan. Se realizó un análisis del nivel de generalización de las fases propuestas, lo cual está ligado a su nivel de flexibilidad, demostrando que el procedimiento puede ser aplicable no solo a entornos de gestión de proyectos informáticos, sino a otros donde se trabaje y procese el desempeño del personal como es el caso de: áreas asociadas a la investigación, ciencia, educación, medicina y otros.

Es meritorio especificar que pueden incluirse nuevas actividades a este procedimiento siempre y cuando apoyen el objetivo por el cual se realiza, quedando a decisión de la organización la inclusión de nuevas acciones e instrumentos en caso de ser necesarios. Según lo analizado anteriormente, el nivel de generalización de las fases propuestas permite la introducción de nuevos cambios sin que, por ello, se vea afectada la ejecución del procedimiento, por lo que se puede afirmar que el nivel de adaptabilidad del procedimiento propuesto es alto.

3.7 Valoración del impacto económico y social de la propuesta

La propuesta de solución para la obtención de la similitud entre las evidencias del desempeño se integró al Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento, el cual está siendo utilizado en la herramienta de gestión de proyectos Xedro-GESPRO 13.05 (Piñero y otros, 2013) en el módulo de gestión de recursos humanos.

Actualmente la herramienta Xedro-GESPRO 13.05 se utiliza por la red de centros productivos de la UCI, en la Empresa Mixta Guardián del Alba, en la Empresa de Desarrollo de Tecnologías para la Defensa (XETID) y en COPEXTEL División Tecnostar (Ver Anexo 5, Anexo 6, Anexo 7, Anexo 8). Para el caso específico de la UCI, con la aplicación del procedimiento propuesto integrado al Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento se benefician alrededor de 14 centros de desarrollo de software, en los cuales se gestiona un promedio anual de 200 proyectos. (Basado en datos del uso histórico de Xedro-GESPRO desde junio del año 2011 hasta mayo del 2015)

Para realizar una valoración del impacto económico de la propuesta en una organización como la UCI, se tendrá en cuenta la comparación de la eficiencia de las distintas variantes propuestas, donde se establece la cantidad de personas y horas necesarias para realizar del procedimiento utilizando el método de Escalas numéricas y el Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento. Además, se toma como tarifa horaria la de un especialista, que se obtiene a partir del fondo salarial del trabajador y la Resolución 8 del 2005. El costo de realizar el procedimiento con cada uno de los métodos sería el que se muestra en la Tabla 9, el cual se obtendría de multiplicar la cantidad de personal requerido por la cantidad de horas requeridas por el salario mensual. Este análisis económico se refiere a un solo una iteración del procedimiento, que puede realizarse en una organización como la UCI cuatro veces al año, por lo que estos valores de costo se multiplicarían. Además, existe el hecho de que, todas las actividades del personal se verían afectadas en los cronogramas de sus proyectos por intervenir en el procedimiento.

Tabla 9. Análisis del costo de realización del procedimiento.

Método/Modelo de evaluación de competencias laborales	Personal Requerido	Horas requeridas	Escala salarial	Costo de realizar el procedimiento
Escalas numéricas	2	2	\$4.90	\$470.4
Modelo de evaluación de competencias a partir de evidencias utilizando Sistemas Basados en Conocimiento	3	2	\$4.90	\$705.6

Antes de la aplicación de la propuesta, la herramienta Xedro_Gespro no contaba con funcionalidades que facilitaran la evaluación de las competencias laborales y la toma de decisiones asociada a esta. Su aplicación facilita el cumplimiento del requisito 4.5.4 de la Norma Cubana NC 3001:2007 que establece la planificación, ejecución, control de la capacitación y desarrollo del capital humano (ONN, 2007).

La propuesta contribuye a la formación en gestión de proyectos y la gestión del conocimiento, al impartirse en cursos de la Maestría en Gestión de Proyectos Informáticos coordinada por el Laboratorio de Investigaciones en Gestión de Proyectos de la UCI (UCI EVA Postgrado, 2015), tales como: Curso Básico de Gestión de Proyectos, Gestión de Recursos Humanos, Herramientas para la Gestión de Proyectos y Dirección Integrada de Proyectos.

La utilización de la propuesta de solución contribuye a la implementación práctica de lineamientos aprobados en el VI Congreso del PCC (PCC, 2011) como: el 8, referido al incremento de facultades a las direcciones de las entidades, que estará asociado a la elevación de su responsabilidad sobre la eficiencia, eficacia y el control en el empleo del personal; el 12 relacionado con fortalecer el sistema de control interno de las entidades, para lograr los resultados esperados en cuanto al cumplimiento de su plan con eficiencia; y el 20, referente a que los ingresos de los trabajadores y sus jefes en las empresas estarán vinculados a los resultados que se obtengan.

Sin lugar a dudas, constituye un elemento ventajoso para la aplicación de este procedimiento, el uso del software libre como elemento primordial para lograr la soberanía tecnológica, lo cual ayuda a garantizar el desarrollo nacional de manera integral y sustentable. El entorno computacional y funcionalidades desarrolladas sobre tecnologías de software de código abierto promueven y reflejan estos preceptos. Esto representa las siguientes ventajas para GESPRO: dominio pleno de sus funcionalidades; detección de posibles errores y correcciones a tiempo, mejora continua sobre la base del desarrollo colaborativo.

3.8 Conclusiones del capítulo

La utilización del Modelo de evaluación de competencias a partir de evidencias utilizando sistemas basado en casos demostró mejores resultados para la obtención de la similitud de las evidencias del desempeño que los restantes modelos estudiados.

La aplicación del procedimiento a otro contexto demostró su adaptabilidad y pertinencia en herramientas informáticas para la gestión de proyectos como Xedro-GESPRO considerándose un aporte novedoso, que permite elevar la eficiencia, competitividad y valor agregado en este tipo de herramientas.

La propuesta de solución contribuirá a la eficiencia del proceso de evaluación de competencias laborales al ser puesto a disposición de una amplia gama de usuarios de la herramienta Xedro-GESPRO y por ende apoyará a la toma de decisiones en las organizaciones.

CONCLUSIONES

El estudio de los referentes teóricos y metodológicos de la investigación permitió elaborar el marco teórico relacionado con el análisis de las evidencias del desempeño laboral, donde se identificaron las funciones de similitud Dice, Jaccard y Coseno como las de mejores resultados en el análisis de evidencias.

En el procedimiento se describen los elementos y premisas necesarios para su aplicación, demostrando su flexibilidad y adaptabilidad.

La aplicación de las pruebas no paramétricas de Friedman y el Test de Wilcoxon demostraron la mejora en la eficiencia del proceso de evaluación de competencias laborales.

La utilización del procedimiento evidenció ahorro de tiempo y esfuerzo para los gestores de proyecto que necesitan la información como apoyo a la evaluación de competencias laborales.

La aplicación del procedimiento apoya la estrategia planteada en los lineamientos de la política económica y social del partido y la revolución, al apostar por el mejoramiento de la industria de desarrollo de software y la superación de sus recursos humanos.

RECOMENDACIONES

A pesar de los resultados obtenidos con la propuesta de no englobar todas las posibilidades de investigación al campo de acción definido en la investigación, por lo que se enuncian las siguientes recomendaciones:

1. Evaluar la aplicación del procedimiento propuesto a otros elementos de producción orientados a proyectos, que no solo sean las evidencias del desempeño y las competencias laborales.
2. Tener en cuenta para futuras investigaciones las medidas de validez interna teniendo en cuenta que capturan propiedades estructurales en la distribución de los datos y suelen basarse en conceptos tales como cohesión y separación.

REFERENCIAS BIBLIOGRÁFICAS

- ALIAS, Francesc, et al.** Clasificación de textos adaptada para conversión de texto en habla multidominio. *Procesamiento del Lenguaje Natural*, 2006, vol. 37, p. 267-274.
- ALLES, Martha Alicia.** Dirección Estratégica de Recursos Humanos. Gestión por Competencias. Buenos Aires: Ediciones Granica S.A., 2006, ISBN: 978-950-641-477-1.
- ALMUNIA, Pablo.** IEDGE – Dirección de Proyectos Informáticos. [En línea] 2011. [Citado el: 13 de febrero de 2013.] <http://blog.iedge.eu/tecnologia-sistemas-informacion/direccion-siti/pablo-almunia-direccion-de-proyectos-informaticos/>.
- Amón, Iván, and Claudia Jiménez.** Funciones de similitud sobre cadenas de texto: una comparación basada en la naturaleza de los datos. *CONF-IRM Proceedings*, 2010, Vol. 58.
- Amón, Iván, y Echeverria, Jaime.** ALGORITMO FONÉTICO PARA DETECCIÓN DE CADENAS DE TEXTO DUPLICADAS EN EL IDIOMA ESPAÑOL. *Ingenierías Universidad de Medellín*, Print version ISSN: 1692-3324. *Rev. ing. univ. Medellín*, 2012, vol.11 no.20.
- André, M.** Un modelo para la asignación de recursos humanos a equipos de proyectos de software. Tesis doctoral, Instituto Superior Politécnico “José Antonio Echeverría”, Facultad de Ingeniería Informática, La Habana, 2009.
- ARCO, Leticia, et. al.** Agrupamiento de documentos textuales mediante métodos concatenados. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, 2006, vol. 10, no 30, p. 43-53.
- Baeza Yates, Ricardo. Berthier., Robeiro Neto.** *Modern Information Retrieval. An Introduction to Information Retrieval.* Cambridge University Press, Ed Greengrass. *Information Retrieval: A Survey*, 1998.
- BECK, Kent, et. al.** Manifiesto Ágil. *Versión obtenida el*, 2001, vol. 22, p. 07-15.
- BERRY, M.** *Survey of Text Mining. Clustering, Classification, and Retrieval.* s.l. Springer-VerlaG, 2004. ISBN: 0-387-95563-1.
- Bordes, F. J. V., Javier, F., Oliván, S., & Antonio, J.** Evaluación del funcionamiento y recuperación de información textual de los principales motores de búsqueda y metabuscadores genéricos de la World Wide Web. 2011.
- BOXALL, P., PURCELL, J. y WRIGHT, P. M.** *Human Resource Management: Scope, Analysis and Significance. The handbook of Human Resource Management.* Oxford University Press, 2007.
- BOYATZIS, R. E.** *The competent manager: A model for effective performance.* John Wiley & Sons, 1982.

- BROOKS, Fred.** *The mythical man-month: essays on software engineering*. s.l.: Addison-Wesley, 1975. ISBN: 0-201-00650-2.
- CASTILLO ARREDONDO, Santiago.** *Compromisos de la Evaluación Educativa*. Madrid: Prentice Hall, 2002. ISBN: 84-205-3562-1.
- Cervantes Rodón, Damián.** Método de evaluación de composición de equipos de proyectos de desarrollo de software, Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos. Tesis de maestría. Universidad de las Ciencias Informáticas, La Habana, Cuba, 2015.
- CHARETTE, Robert N.** Why software fails [software failure]. *Ieee Spectrum*, 2005, vol. 42, no 9, p. 42-49.
- COCKBURN, A y HIGHSMITH, J.** Agile Software Development: The people factor: *IEEE Computer*, 2001, Vol. 34. ISSN: 0018-9162.
- COCKBURN, Alistair; HIGHSMITH, Jim.** Agile software development, the people factor. *Computer*, 2002, vol. 34, no 11, p. 131-133.
- COTTERELL, Mike y HUGHES, Bob.** *Software Project Management*. Berkshire, England: McGraw Hill, 1999. ISBN: 007-709505-7.
- CUESTA SANTOS, Armando.** Metodología de Gestión por Competencias asumiendo NC sobre Gestión del Capital Humano. *FECAP*, 2011, vol. 13. ISSN: 1806-4892.
- Cuesta, A.** Metodología de Gestión por Competencias Asumiendo la Norma Cubana de Gestión de Capital Humano. *RBGR-Revista Brasileira de Gestao de Negócios*, 2011, vol. 13, nº 40, pp. 300-311, ISSN: 1806-4892.
- CUESTA SANTOS, Armando.** *Tecnología de Gestión de Recursos Humanos*. La Habana, Cuba: s.n., 2010. S.N.
- Cuesta, A.** Evaluando desempeños: alineamiento estratégico y productividad, *Forum Empresarial*, vol. 17, nº 1, pp. 1-30, 2012. ISSN: 1541-8561.
- CURTIS, B. et. al.** A Field Study of the Software Design Process for Large Systems: *IEEE Trans. Software Engineering*, 1988, Vols. SE-31.
- CURTIS, B., et. al.** People Management Capability Maturity Model. Software Engineering Institute, 1994.
- CHEATHAM, Michelle; HITZLER, Pascal.** The properties of property alignment. En *Proceedings of the 9th International Conference on Ontology Matching-Volume 1317*. CEUR-WS.org, 2014. p. 13-24.
- CHEN, Yoke Yie; YONG, Suet-Peng; ISHAK, Adzlan.** Email hoax detection system using Levenshtein distance method. *Journal of computers*, 2014, vol. 9, no 2, p. 441-446.
- de Andrés, R.** Evaluación de desempeño: Nuevos enfoques desde las teorías de subconjuntos difusos y de la decisión multi-criterio. Tesis doctoral. Universidad de Valladolid, Valladolid, 2009.

- DÍAZ VILLEGAS, Jorge.** *Oportunidades para una Simbiosis Investigación-Industria en la Producción de Software Regional*. Temuco, s.n., 2010.
- DIXON, M.** An overview of document mining technology, 1997. URL citeseer.ist.psu.edu/dixon97overview.html.
- DIXON, M.** An Overview of Knowledge Discovery in Texts: A Definition and Applications Proc. of the 11th International Symposium on Foundations of Intelligent Systems (ISMIS-99), 1999.
- DÜRSTELER, J. C.** Minería de Texto: La revista digital de InfoVis.net., 2001.
- Elmagarmid, A.K., Ipeirotis, P.G. and Verykios, V.S.** Duplicate Record Detection; A Survey. IEEE Transactions on Knowledge and Data Engineering, 2007, vol.19, no.1, pp.1-16.
- ERNST & YOUNG, Consultores.** *Gestión por Competencias. Manual del Director de Recursos Humanos*. s.l Disponible en: <http://formacionhumana.files.wordpress.com/2008/06/competencias.pdf>, 2008.
- ERRECALDE, Marcelo Luis; INGARAMO, Diego; ROSSO, Paolo.** Proximity estimation and hardness of short-text corpora. En *2008 19th International Workshop on Database and Expert Systems Applications*. IEEE, p. 15-19.
- ESTOQUE, E. and L. BARÓ.** Implementación de algoritmos para la limpieza de datos. Tesis de grado, Universidad de las Ciencias Informáticas. La Habana, Cuba, 2013.
- FERNÁNDEZ GONZÁLEZ, A., et. al.** De las capacidades a las competencias: una reflexión teórica desde la psicología. Revista Varona, 2003, no 36-37, p. 22-25.
- Fernández Luna, J. M.** Modelos de Recuperación de Información basados en Redes de Creencia, 2013.
- FERRUCCIO, y otros.** Minería de Datos (1º CONTECSI Congreso Internacional de Gestión de la Tecnología y Sistemas de la Información): Sao Paulo, s.n., 2004.
- FUNG, Benjamin Chin Ming.** *Hierarchical document clustering using frequent itemsets*. Tesis Doctoral. Simon Fraser University, 2002.
- FRANKES, J., NAKHAEIZADEH, G. y RENZ, I.** *Advances in Soft Computing. Text Mining. Theoretical Aspects and Applications*. s.l Physica-Verlag, 2003, ISBN: 3-7908-0041-4.
- FRANKES, W. B. y BAEZA-YATES, R.** *Information Retrieval. Data Structures & Algorithm*. S.l Prentice-Hall PTR, 1992, ISBN: 0-13-463837-9.
- Gan, Guojun, Ma Chaoqun, y Wu Jianhong.** *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM, 2007, Páginas 3-65, ISBN: 978-0-898716-23-8.
- GIL FLORES, Javier.** Universidad de Sevilla, España. La evaluación de competencias laborales: s.n., 2007, Vol. XXI, ISSN: 1139-613X.

- GÜYER, Tolga; ATASOY, Bilal; SOMYÜREK, Sibel.** *Measuring disorientation based on the Needleman-Wunsch algorithm. The International Review of Research in Open and Distributed Learning*, 2015, vol. 16, no 2.
- HEARST, Marti A.** *Untangling text data mining. En Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics*, 1999, p. 3-10.
- Hernández, V., Miranda, Y., Quintana, L., Mederos, R. y Toyos, A.** Evaluación de las competencias laborales y su influencia en la calidad del servicio del hospital Dr. Mario Muñoz Monroy,» *Revista de Ingeniería Industrial*, 2009, vol. 30, nº 1, ISSN: 0258-5960.
- HERRERA, Francisco; MARTÍNEZ, Luis.** A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 2001, vol. 31, no 2, p. 227-234.
- Hollander, M., Wolfe, D. A. y Chicken, E.** *Nonparametric Statistical Methods*, Tercera ed., Hoboken: John Wiley & Sons, Inc., 2014, ISBN: 978-0-470-38737-5.
- INGARAMO, Diego Alejandro, et. al.** "Clustering iterativo de textos cortos con representaciones basadas en conceptos". *Procesamiento del Lenguaje Natural*. 2011, ISSN: 1135-5948, N. 46, pp. 19-26
- Iraurgi, I., Martínez-Pampliega, A., Iriarte, L., & Sanz, M.** Modelo cognitivo-contextual del conflicto interparental y la adaptación de los hijos. *Canales de psicología*, 2011, 27(2), 562-573, SSN edición impresa: 0212-9728, ISSN edición web (<http://revistas.um.es/analesps>): 1695-2294
- JACKSON, P. y MOULINIER, I.** *Natural Language Processing for Online Applications. Text Retrieval, Extraction and Categorization*. s.l John Benjamins Publishing Company, 2002, ISBN: 902724988.
- JACOBSON, Ivar, BOOCH, Grady y RUMBAUGH, James.** *El Proceso Unificado de Desarrollo de Software*. Madrid: Addison Wesley. Pearson Educación S.A, 2000, ISBN: 84-036-2.
- JIMÉNEZ, Juan Manuel Cabrera; BALDERAS, Hugo Jair Escalante.** *Clasificación de Textos Cortos Usando Representaciones Distribucionales de los Términos*, 2012.
- KODRATOFF.** 2009. Knowdad de Carnegie Mellon. 2009. Disponible en: <http://www.sei.cmu.edu/reports/09tr003.pdf>.
- Kuna, H., Rey, M., Martini, E., Rambo, A., & Podkowa, L.** Avances en el Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación. *Revista Latinoamericana de Ingeniería de Software*, 2015, 3(1), 47-55.
- Lanquillon, C.** *Enhancing Text Classification to Improve Information Filtering*, Tesis doctoral, Otto-von-Guericke-Universität Magdeburg, Universitätsbibliothek, 2001.

López, J.A., Barahona, N.A., Estrada, M., Favela, M.A y Cuen, F.A. Evaluación de competencias del profesional de enfermería que labora en hospitales públicos y privados. *Rev. Enfermería Actual en Costa Rica*, 2014, 27, 1--15. DOI: <http://dx.doi.org/10.15517/revenf.v0i27.16057>

López, B. Limpieza de datos: Reemplazo de valores ausentes y estandarización. Tesis doctoral, Universidad Central "Marta Abreu" de las Villas, Santa Clara, 2011.

LUGO, J. A. Modelo para el control de la ejecución de proyectos basado en indicadores y lógica borrosa. Tesis de maestría, Laboratorio de Gestión de Proyecto, Facultad 5, Universidad de las Ciencias Informáticas, La Habana, Cuba, 2012.

Lugo, J. A., Torres, S., Piñero, P. Y. y Delgado, R. Control de la ejecución de proyectos basado en indicadores y lógica borrosa. *Iberoamerican Journal of Project Management*, 2013, vol. 4, nº 1, pp. 15-31, ISSN: 2346-9161.

Lugo, J. A., Torres, S., Piñero, P. Y. y Delgado, R. Cálculo automático de indicadores como técnica de apoyo para la decisión en el Sistema Nacional de Salud,» *Revista Cubana de Información en Ciencias de la Salud*, 2013, vol. 24, nº 4, pp. 15-35, ISSN: 2307-2113.

LENGNICK HALL, M. L. et. al. Strategic human resource management: The evolution of the field. *Human Resource Management Review*, 2009, vol. 19, no 2, p. 64-85.

LEVY-LEBOYER, Claude; PRIETO, José M. Gestión de las competencias: Como analizarlas, como evaluarlas, como desarrollarlas. *Gestión*, 1997, ISBN: 84-8088-178-X.

LEZCANO, R. D. Minería de Datos. [En línea] 2002. [Citado el: 21 de 3 de 2013.] <http://www.google.com.cu/depar/areas/informatica/SistemasOperativos/MineriaDatosLezcano>

Mano González, Marta de la; Moro Cabero, Manuela. "La evaluación por competencias: propuesta de un sistema de medida para el grado en Información y Documentación". *BiD: textos universitaris de biblioteconomia i documentació*, 2009, núm. 23, <<http://bid.ub.edu/23/delamano2.htm>>. DOI: <http://dx.doi.org/10.1344/105.000001504> [Consulta: 06-06-2016].

MANRÍQUEZ PANTOJA, Luis. ¿Evaluación en competencias? *Estudios pedagógicos (Valdivia)*, 2012, vol. 38, no 1, p. 353-366.

Mark P.J. van der Loo. The stringdist Package for ApproximateString Matching. *The R Journal*, 2014, vol. 6/1, June, ISSN: 2073-4859.

MARTÍNEZ COMECHE, Juan Antonio. Los modelos clásicos de Recuperación de información y su vigencia. En III Seminario Hispano-mexicano de investigación en Bibliotecología y Documentación, 29-31 marzo 2006, México DF. (In Press).

- MARTÍNEZ, Israel Alonso.** Análisis y optimización del recurso UMLS en la recuperación de información biomédica mediante métricas de similitud semántica. Tesis Doctoral, Universidad Pontificia Comillas, 2016.
- MCCLELLAND, David C.** Testing for competence rather than for" intelligence." *American psychologist*, 1973, vol. 28, no 1, p. 1.
- MONTES Y GÓMEZ, Manuel.** Minería de Texto empleando la Semejanza entre Estructuras Semánticas. Tesis Doctoral, Computación y Sistemas, México DF., 2005, Vol. 9. ISSN: 1405-5546.
- Naumann, F.** Similarity measures. *Information Systems*, Jun 2013.
- NEIL THOMPSON, Donaval.** *Proyectos Informáticos: Fracasos y Lecciones Aprendidas. Revista de Derecho y Tecnologías de la Información*, 2006, vol. 4.
- Nitesh Pradhan Manasi Gyanchandani, PhD Rajesh Wadhvani, PhD.** A Review on Text Similarity Technique used in IR and its Application (0975 – 8887). 2015, vol 120, No.9.
- OFICINA NACIONAL DE NORMALIZACIÓN, NC 3000.** *Sistema de Gestión Integrada de Capital Humano. Vocabulario.* La Habana, Cuban National Bureau of Standards, 2007.
- ONN.** Norma Cubana NC 3000:2007. Oficina Nacional de Normalización, La Habana, 2007.
- PACELLI, Lonnie.** *Grandes errores en la Gestión de Proyectos.* s.l Financial Times Prentice Hall, 2004, ISBN: 0131490478.
- Palmero, M., Padrón, J. Rizo, L. Tejo, Y. Rodríguez y C. A. Torre.** Procedimiento para evaluar el impacto de la capacitación en la Empresa Constructora Integral de Mayabeque, *Ingeniería Industrial*, 2012, vol. XXXIII, nº 3, pp. 215-225, ISSN: 1815-5936.
- MORENO PARRA, JHOSEP DAVID y YARA PEÑALOZA, JONATHAN ANTONIO.** Introduccion a la inteligencia de negocios con énfasis en una selección de algoritmos de minería de datos, Universidad Tecnologica de Pereira, facultad de Ingenierías de sistemas y ciencias de la computación, Pereira, 2014.
- PÉREZ QUINTERO, LISSET.** Modelo para la evaluación por competencias en proyectos informáticos de la Universidad de las Ciencias Informáticas. Tesis de maestría, Universidad de las Ciencias Informáticas, La Habana, 2010.
- PCC.** Lineamientos de la Política Económica y Social del Partido y la Revolución. Partido Comunista de Cuba, La Habana, 2011.
- PMI.** Guía de los Fundamentos de la Dirección de Proyectos, Pennsylvania, EEUU: PMI Publications, 2013, ISBN: 1-930699-73-5.
- Piñero, P. Y. y otros.** GESPRO Paquete para la gestión de proyectos. Nueva Empresa, 2013, vol. 9, nº 1, pp. 45-53, ISSN: 1682-2455.

- PIORUN, Daniel.** ¿Por qué fracasan los proyectos? Parte II. Gerencia.com. [Versión electrónica]. Obtenida el, 2003, vol. 29.
- PRESSMAN, Roger S.** *Ingeniería de Software. Un enfoque práctico.* Mc Graw Hill, 2004, ISBN: 84-481-3214-9.
- PROJECT MANAGEMENT INSTITUTE, Inc.** *Guía de los Fundamentos de la Dirección de Proyectos. Tercera Edición. (Guía del PMBOK).* s.n., 2004, ISBN: 1-930699-73-5.
- APUT, Rolando Quintana.** Indicadores para evaluar las competencias del líder de proyecto en el equipo de desarrollo de software multimedia. Avanzada Científica, 2014, vol. 17, no 1, p. 1-15.
- REEL, John. S.** Critical Success Factors in Software: IEEE Software, 1999, vol. 16 Issue 3, Page 18-23.
- LÓPEZ, Carmen Torres; GARCÍA, Leticia Arco.** Representación textual en espacios vectoriales semánticos. Revista Cubana de Ciencias Informáticas, 2016, vol. 10, no 2, p. 148-180.
- López Salazar, Ricardo y Velázquez-Pompeyo, René Ileana.** Industria electrónica, investigación y desarrollo (I+D) y competencias profesionales. Estrategias para el escalamiento industrial. Estudio de caso en Mérida, Yucatán Telos. Universidad Privada Dr. Rafael Bellosillo Chacín, Maracaibo, Venezuela. 2015, vol. 17, núm. 3, pp. 377-397.
- LYLE, M.; SIGNE, M.** Competence at work: Models for superior performance. Amerika: John Willey & Sons Inc, 1993.
- RICHARD, Diego.** Management: MEDWAVE. Santiago de Chile. 2005. Disponible en: <http://www.mednet.cl/link.cgi/Medwave/Enfermeria/enfermeriaquirurgica05/1/2686>.
- Rodríguez, E.** Modelo para la evaluación de la composición de equipos de proyectos informáticos. Tesis de maestría. Universidad de las Ciencias Informáticas, La Habana, 2012.
- Robertson, S. E.** The probability ranking principle in IR. *Journal of Documentation*. Sparck Jones, K. Search term relevance weighting given little relevant information. *Journal of Documentation*, 35(1): 30-48; Croft, W. B.; Harper, D. J. Using probabilistic models of document retrieval without relevance information. *Journal of Documentation*, 1979, 35(4):285-295.
- RODRÍGUEZ MORENO, María Luisa.** *Evaluación, balance y formación de competencias laborales: propuestas para mejorar la calidad en la formación profesional y en el mundo del trabajo.* s.l LAERTES, 2006, ISBN: 8475845762.
- Rodríguez, H., Guerrero, A., Pierra, A. y Tamayo, S. E.** El movimiento del software libre en Cuba. Revista Cubana de Ciencias Informáticas, 2009, vol. 3, nº 3-4, pp. 5-11, ISSN: 1994-1536.
- ROSAS, María V.; ERRECALDE, Marcelo L.; ROSSO, Paolo.** Un análisis comparativo de estrategias para la categorización semántica de textos cortos. Procesamiento del lenguaje natural, 2010, vol. 44, p. 11-18. ISSN: 1135-5948.

Ruiz Rangel, Jonathan. Minería de datos como soporte a la toma de decisiones empresariales en una arquitectura SOA, Barranquilla, Ed. Coruniamericana, 2013, vol. I, p. 25-34.

Salton, G.; McGill, M. J. *Introduction to modern information retrieval.* New York: McGraw-Hill. Meadow, Ch. T.; Boyce, B. R.; Kraft, D. H. Text information retrieval systems. Toronto: Academic Press, 1992, Salton, G.; Wong, A.; Yang, C. S. A vector space model for automatic indexing. Communications of the ACM, 1975, 18(11): 613-620.

Samaniego Lema, Y. M., & Duran Mendoza, L. G. Implementación de un sistema de control en el área de consulta externa de un hospital para mejorar su desempeño. 2015.

Serie Científica de la Universidad de las Ciencias Informáticas. 2015, vol. 8, No. 1, octubre, ISSN: 2306-2495 | RNPS: 2343 <http://publicaciones.uci.cu> Pág. 120-132

SOMMERVILLE, Ian. *Ingeniería de Software.* Addison Wesley Pearson Educación S.A, 2005, ISBN: 84-7829-074-5.

TAN, Ah-Hwee, et. al. Text mining: The state of the art and the challenges. En Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases. 1999, p. 65-70.

Taveira, E. Introduction · Installation · Functions and Operators · Examples · License. 16 junio 2012, [En línea]. Available: <http://pgsimilarity.projects.pgfoundry.org/>.

TENBRINK, Terry D. *Evaluación: Guía Práctica para Profesores.* Madrid, España: Narcea S.A., 2006, ISBN: 84-277-0464-X.

THE STANDISH GROUP. *The chaos manifesto the standish group.* [En línea] 2012. [Citado el: 23 de enero de 2013.]

Torres, S. y colectivo de autores. Gestión de recursos humanos para centros de desarrollo de sistemas de información, Ciencias de la Información, 2014, vol. 45, nº 2, pp. 3-8, ISSN: 0864-4659.

UCI EVA Postgrado. Entorno Virtual de Aprendizaje. Formación Postgraduada. 2015. [En línea]. Available: <http://evapostgrado.uci.cu/course/category.php?id=15>. [Último acceso: 10 febrero 2016]

ULRICH, D. <http://www.aedipe.es>. [En línea] 2006. [Citado el: 14 de abril de 2013.] <http://www.aedipe.es/documentos/HRarchitect.doc>.

Valentina Tabares Morales, Néstor d. duque Méndez, Paula A. Rodríguez Marín, Julián Moreno Cadavid, Demetrio A. Ovalle Carranza. FROAC: Una Iniciativa Colombiana para la Integración de Repositorios de Objetos de Aprendizaje. Universidad nacional de Colombia, Colombia, Campus Virtuales, 2015, 01. IV.

VELASCO, Calle, et. al. Modelo basado en técnicas de procesamiento de lenguaje natural para extraer y anotar información de publicaciones científicas. Tesis Doctoral, ETSI_Informática, 2014.

ZAYAS AGUERO, Pedro Manuel. Las competencias: Una visión teórico-metodológica. Universidad de Holguín, Cuba, 2010. ISSN: 1696-8360. <http://www.eumed.net/ce/2010a/pmza.htm>.

ANEXOS

Anexo 1 Gestión de las tareas de los miembros de los equipos en el Xedro-GESPOR 13.05

The screenshot shows the Xedro-GESPOR 13.05 software interface. The top navigation bar includes: GESPRO, Proyectos, Planificación, Ejecución, Contratación, Configuración, and Ayuda. Below this is a sub-navigation bar with: Vistazo, Planificación, Ejecución, Contratación, Gestión documental, and Configuración. On the left, there is a sidebar with 'Peticiones' and a list of tasks. The main table displays task details.

	ESTADOS	ASUNTO	AUTOR	ASIGNADO A	FECHA	FECHA CUMPLIMIENTO	% REALIZADO
7362	Nueva	Realización del contrato	Rachel Olivera	Rachel Olivera	26/11/2013	29/11/2013	
7361	Nueva	Dictamen de la comisión de evaluación	Rachel Olivera	José Alejandro Lugo	25/11/2013	25/11/2013	
7360	Nueva	Realización de la aprobación del proyecto	Rachel Olivera	José Alejandro Lugo	25/11/2013	29/11/2013	
7359	Nueva	Elaboración del proyecto técnico	Rachel Olivera	José Alejandro Lugo	09/11/2013	23/11/2013	

Anexo 2 Ejemplo de estrategia de búsqueda utilizada en la investigación

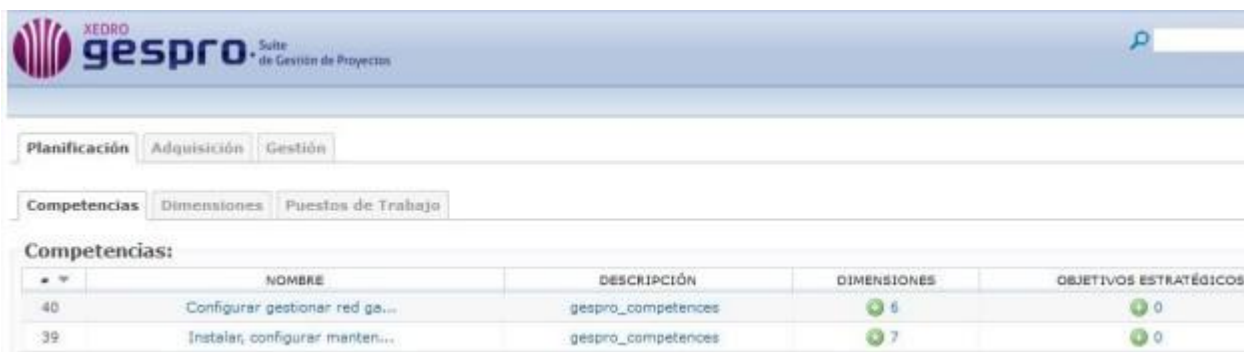
Ejemplo de estrategia de búsqueda utilizada en la investigación.

Necesidad: utilización de funciones de similitud de textos en el desarrollo de software.

	1	2	3
Concepto	Funciones de similitud	Desarrollo	Software
Variantes	Coeficientes Medidas Coefficients	Desarrollar Elaborar Develop Factory	Producto de software Aplicaciones Applications
Sinónimos	Ecuaciones - Computers Similarity functions Measures of similarity	Construir- build Producir- manufacture	Apps Programas- computadoras Computer programs

Cobertura: 2015 - 2017 **Idioma:** Inglés y Español **Fuente:** EBSCO

Anexo 3 Gestión de las competencias laborales en el Xedro-Gespro 13.05



The screenshot shows the Xedro-Gespro Suite de Gestión de Proyectos interface. The top navigation bar includes 'Planificación', 'Adquisición', and 'Gestión'. Below this, there are tabs for 'Competencias', 'Dimensiones', and 'Puestos de Trabajo'. The 'Competencias' tab is active, displaying a table with the following data:

	NOMBRE	DESCRIPCIÓN	DIMENSIONES	OBJETIVOS ESTRATÉGICOS
40	Configurar gestionar red ga...	gespro_competences	6	0
39	Instalar, configurar manten...	gespro_competences	7	0

Anexo 4 Relación de herramientas de gestión de proyectos que basan su funcionamiento en evidencias o tareas

No	Herramientas para gestionar proyectos	Asignan tareas a los roles en los proyectos	
		Si	No
1	OpenProj	x	-
2	OmniPlan	x	-
3	ProjectSim	x	-
4	Microsoft Project	x	-
5	Contactizer Pro	x	-
6	LisaProject	x	-
7	FastTrackShedule	x	-
8	MacProject	x	-
9	24SevenOffice	x	-
10	Project KickStart	x	-
11	OpenWorkbench	x	-
12	Rachota	x	-
13	03spaces	x	-
14	Collabtive	x	-
15	KPlato	x	-
16	KKOOP	x	-
17	dotProject	x	-
18	TaskJuggler	x	-
19	Mingle	x	-

20	Jira	x	-
21	Huddle	x	-
22	Wrike	x	-
23	PlannerSuite	x	-
24	codeBeamer	x	-
25	Basecamp	x	-
26	InLoox	x	-
27	Hyperoffice	x	-
28	Kayako helpdesk software	x	-
29	PRISM	x	-
30	CentralDesktop	x	-
31	KForge	x	-
32	WorkPLAN Enterprise	x	-
33	Ubidesk	x	-
34	Teamcenter	x	-
35	ProjectSpaces	x	-
36	Glasscubes	x	-
37	MicroPlanner X-Pent	x	-
38	FogBugz	x	-
39	Feng Office Community Edition	x	-
40	Projectplace	x	-
41	Gemini	x	-
42	Zoho Projects	x	-
43	Php GroupWare	x	-
44	Smartsheet	x	-
45	PSNext	x	-
46	Doliban ERP/CRM	x	-
47	Assembla	x	-

48	eGroupWare	x	-
49	TrackerSuite.NET	x	-
50	Tenrox	x	-
51	Teamwork PM	x	-
52	Severa	x	-
53	Redmine	x	-
54	ProjeKtron BCS	x	-
55	Primavera Project Planner	x	-
56	Planisware 5	x	-
57	Oracle Project Portfolio Management	x	-
58	Microsoft Office Project Server	x	-
59	Merlin	x	-
60	Journyx	x	-
61	Genius inside	x	-
62	Easy Project. NET	x	-
63	Endeavor Software Project Managent	x	-
64	Clarizen	x	-
65	AtTask	x	-
66	Project-Open	x	-
67	Project. NET	x	-
68	Daptiv	x	-

Anexo 5 Caracterización de las competencias laborales

GESPRO	Proyectos	Planificación	Ejecución	Contratación	Configuración	Ayuda	Conectado como admin	Mi cuenta	Terminar sesión
 gespro Suite de Gestión de Proyectos							Ir al proyecto...		
Competencias Dimensiones Puestos de Trabajo Evaluación Competencia Competencia por Puesto Trabajo Evaluación de Desempeño									
Filtros									
Evaluación Competencia:									
	USUARIOS	COMPETENCIA	EFICACIA (IRHE)	EFICIENCIA (IRHF)	EFFECTIVIDAD (IRHC)	COSINE (IRHCC)	EVAL. SISTEMA	EVAL. EXPERTO	
351	Indira Rodríguez	Ejecutar auditorías configu...	1.0	1.0	1.98	0.17	Media	Ninguno	Evaluar Borrar
350	Javier Menéndez	Ejecutar auditorías configu...	1.0	2.0	4.71	0.14	Media	Ninguno	Evaluar Borrar
349	Rosel Sosa	Ejecutar auditorías configu...	0.86	2.0	1.7	0.11	Muy alta	Ninguno	Evaluar Borrar
348	Marielis Izquierdo	Ejecutar auditorías configu...	0.24	2.0	1.14	0.19	Media	Ninguno	Evaluar Borrar
347	Surayne Torres	Ejecutar auditorías configu...	0.96	2.0	2.91	0.14	Muy alta	Ninguno	Evaluar Borrar
346	Isuel Méndez	Ejecutar auditorías configu...	0.48	2.0	0.17	0.14	Baja	Ninguno	Evaluar Borrar

Anexo 6 Acta de aceptación del Xedro-Gespro 13.05 en la empresa Xetid



Acta de aceptación

ACTA DE ACEPTACIÓN

En cumplimiento del Convenio de colaboración XETID y en función de la ejecución del proyecto: Personalización de GESPRO para el XETID, se hace entrega de los productos que se relacionan a continuación:

- Instalación en sitio del producto GESPRO.

La Parte Cliente, luego de haber revisado el producto de trabajo determina que se acepta el producto porque es de interés para la empresa y ha mantenido un adecuado nivel de funcionalidad desde el año 2010 en nuestra organización, y se ha mantenido una buena atención por los desarrolladores de la solución a las solicitudes de cambio y nuevos requerimientos asociados a los procesos de la empresa, que han sido presentados por los especialistas de XETID. Además nos mantenemos interesados en la continuidad del trabajo conjunto.



Entrega Departamento Gestión de Proyectos
COAE
Dr.C Pedro Yobanis Piñero Pérez
Cargo: Jefe de departamento de gestión de Proyectos





Recibe Dirección de Producción XETID
TCor. Rolando Ramirez Concepción

Cargo: Director de Producción

Representante Parte Suministradora (Universidad de las Ciencias Informáticas, Centro de Consultoría y Desarrollo de Arquitecturas Empresariales, Departamento de Investigaciones en Gestión de Proyectos)

Anexo 7 Aval de instalación del Xedro-Gespro 13.05 en la empresa Guardián del Alba



Caracas, 25 de diciembre de 2013

Empresa Socialista de Capital Mixto Guardián del Alba

A: Rectora de la UCI

A: Quién pueda Interesar

Estimada Rectora

Sirva la presente para avalar el impacto para nuestra organización de la instalación del sistema GESPRO 13.05 en la misma.

Primero: Notificamos que el sistema GESPRO 13.05 ha sido instalado exitosamente en nuestra organización, y se encuentra visible en la red PDVSA bajo la dirección URL: <http://gesproguardiandelalba.pdvs.com>.

Segundo: El monto estimado del servicio profesional de instalación del sistema y su soporte técnico, calculado en función de la tarifa establecida de \$40 USD la hora del profesional, asciende a la suma de \$ 78400 USD. Y ha incluido:

- Diagnóstico y consultoría a la organización en el tema de gestión de proyectos.
- Instalación de la solución informática GESPRO 13.05 bajo licencia GNU GPL 2.0 y soporte técnico.
- Cursos de entrenamiento en el uso del sistema a todas las sedes de la organización.
- Curso básico de gestión de proyecto en todas las sedes de la organización.
- Apoyo a la toma de decisiones de la empresa con las funcionalidades del sistema basadas en la sumariazación lingüística de los datos.



Tercero: El monto estimado de sistemas similares en el mercado es superior, elemento que indica un aporte en la sustitución de importaciones para nuestra empresa.

- Microsoft Project Server, \$ 239199 (Sistema privativo)
- VPMI Enterprise \$ 152400 (Sistema privativo)

Cuarto: La solución instalada tiene un aporte a la soberanía tecnológica de la Empresa, la corporación PDVSA y del pueblo Venezolano, es una solución basada en software libre y licencia GNU GPL 2 que cumple con lo establecido en el Decreto N° 3.390, de diciembre de 2004.

Quinto: El servicio profesional tiene un impacto social que se refleja en la cantidad de profesionales venezolanos formados en gestión de proyectos. En total recibieron formación básica en gestión de proyectos 96 profesionales, distribuidos en los estados de: Distrito Capital Caracas, Miranda, Mérida, Zulia, Anzoátegui y Monagas.

Fraternalmente,

Rafael Román Morales
Gerente General Empresa Socialista de Capital Mixto Guardián del Alba



Anexo 8 Acta de aceptación del Xedro-Gespro 13.05 en la empresa Copextel

ACTA DE ACEPTACIÓN

En cumplimiento del **Convenio de colaboración** el Departamento de Gestión de Proyectos de CDAE y en función de la ejecución del proyecto: **GESPRO COPEXTEL**, se hace entrega del producto que se relaciona a continuación:

Lista de productos que serán aceptados:

- Especificación de requisitos de software
- GESPRO Transaccional Versión: 13.05.00 Descripción: Es una herramienta de Gestión de proyectos que permite el manejo de los proyectos de una sola entidad. La gestión de proyectos es la disciplina que se encarga de organizar y administrar recursos de manera tal que se pueda culminar todo el trabajo requerido en el proyecto dentro del alcance, el tiempo, y coste definidos. El uso de herramientas para la gestión de proyectos se ha ido generalizando al paso del tiempo con el objetivo de facilitar el trabajo de los especialistas

Artefactos:

- Paquete GESPRO Transaccional

Entrega El Grupo de desarrollo de activos de GESPRO

Nombre y apellidos: Ing. Ilana Pérez Pupo

Cargo: Jefe del proyecto GESPRO COPEXTEL

Firma: 

Recibe COPEXTEL S.A.

Nombre y apellidos: Jesús Castellanos Hernández

Cargo: Gerente General de la División Tecnostar Informática y Comunicaciones

Firma: 

1

Anexo 9 Dictamen junta de acreditación de la maestría de Gestión de Proyectos como Maestría de Excelencia



Junta de Acreditación Nacional

La Habana, 21 de enero de 2015
"Año 57 de la Revolución"

DICTAMEN

DE LA JUNTA DE ACREDITACIÓN NACIONAL SOBRE EL PROCESO DE EVALUACIÓN Y ACREDITACIÓN A LA MAESTRÍA EN GESTIÓN DE PROYECTOS INFORMÁTICOS PERTENECIENTE A LA UNIVERSIDAD DE CIENCIAS INFORMÁTICAS.

La Junta de Acreditación Nacional (JAN), reunida el 21 de enero del 2015 elaboró el siguiente dictamen sobre la maestría de referencia.

La ejecución de este programa de Maestría en una universidad con un modelo docente-productivo ha permitido la obtención y generalización de soluciones que se aplican en centros de desarrollo de la propia UCI, en la administración de procesos universitarios y en otras entidades productivas.

La creación de herramientas tecnológicas como GESPRO, para la gestión de proyectos elaborada en software libre, garantiza independencia tecnológica. Su actualización sistemática es un resultado directo de la introducción del programa en diferentes procesos productivos con gran aceptación. GESPRO se utiliza, a su vez, como medio de enseñanza e instrumento de trabajo en la propia maestría.

Como resultado de las investigaciones realizadas en el marco del programa de maestría, se han desarrollado software que se emplean en la medicina, la educación, la gestión empresarial, el manejo de distintos tipos de recursos y la gobernabilidad institucional en entidades del MIC (ALBET, CALISOFT, COPEXTEL, DESOFT, SOFTEL), del MINFAR (XETID), del MININT (DTS) y la empresa mixta cubano-venezolana GUARDIÁN DEL ALBA.

Los egresados han dado su aporte al trabajo que se ejecuta para la certificación de procesos de la UCI a través del estándar internacional integrado de madurez y capacidad de organizaciones (CMMI, nivel 2).

Las líneas de investigación en las que se trabaja en el programa están en correspondencia con las de la Universidad. Se destaca la existencia de un grupo científico (Grupo de Investigación en Gestión de Proyectos) en el que participan los miembros del claustro, los

Anexo 10 Resumen de aplicación del procedimiento utilizando Escalas numéricas

No	Cantidad de Tareas	Tipo Tarea Dominante	Tiempo Prom x Tarea Horas	Rol	Rol Admin. BD	Rol Diseñador BD	T Total Eval min
1	53	Baja	24	Diseñador BD	0	8	0:19:32
2	2	Baja	9	Diseñador BD	0	0	0:01:20
3	15	Media	21	Diseñador BD	0	11	0:05:40
4	5	Media	6	Diseñador BD	0	2	0:01:40
5	79	Baja	26	Diseñador BD	10	10	0:27:42
6	32	Baja	43	Administrador de BD	9	0	0:11:31
7	1	Baja	6	Administrador de BD	0	0	0:02:06
8	10	Alta	52	Diseñador BD	0	6	0:03:48
9	112	Media	6	Administrador de BD	15	7	0:39:44
10	7	Alta	19	Diseñador BD	0	0	0:04:11
11	8	Baja	43	Administrador de BD	0	1	0:02:16
12	86	Alta	30	Diseñador BD	0	13	0:30:27
13	10	Baja	16	Administrador de BD	0	6	0:03:07
14	6	Baja	45	Diseñador BD	0	5	0:03:33
15	37	Media	17	Administrador de BD	8	2	0:11:53
16	23	Media	2	Diseñador BD	2	6	0:19:43
17	3	Alta	3	Diseñador BD	0	4	0:02:50
18	117	Media	4	Diseñador de BD	5	11	0:37:17
19	10	Media	7	Diseñador BD	0	2	0:03:39
20	2	Media	30	Diseñador de BD			0:01:23
21	9	Baja	24	Administrador de BD	1	1	0:05:31
22	26	Baja	47	Diseñador de BD	0	0	0:08:54
23	10	Baja	57	Diseñador BD	2	5	0:13:19
24	3	Media	5	Diseñador BD	0	1	0:03:06

25	46	Media	4	Administrador de BD	5	14	0:24:41
26	2	Media	10	Diseñador BD	0	3	0:01:57
27	5	Alta	120	Diseñador de BD	0	5	0:04:34
28	4	Baja	7	Diseñador BD	0	2	0:04:03
29	10	Baja	20	Diseñador de BD	0	2	0:09:21
30	56	Baja	13	Administrador de BD DBA	18	7	0:20:42
31	8	Alta	26	Diseñador BD	0	6	0:05:10
32	34	Baja	23	Diseñador BD	0	11	0:14:29
33	296	Baja	4	Diseñador BD	2	13	0:44:48
34	7	Media	3	Diseñador de BD	0	3	0:05:25
35	26	Media	41	Diseñador BD	0	8	0:12:15
36	12	Baja	28	Diseñador de BD	0	4	0:07:57
37	11	Media	8	Diseñador BD	0	8	0:09:04
Total tareas analizadas			Tiempo total requerido para la evaluación				7:08:38
		1183					Horas