



Proceso de limpieza de datos y extensión del Repositorio para Investigaciones en Gestión de Proyectos

Techniques for cleaning data and extension in the building of Repositories for Research in Project Management

Claudia Celeste Rivero Hechavarría¹

Iliana Pérez Pupo²

Pedro Y. Piñero Pérez³

Lainier A. PiedraDieguez⁴

Arisney Figueredo Ramos⁵

Eliuvis Matos Matos⁶

Randy H. Bartumeu Huergo⁷

1 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

2 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

3 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

4 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

5 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

6 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

7 Departamento de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.



Resumen

Los repositorios de datos permiten la compartición, reutilización y localización de los datos para el aprendizaje y descubrimiento del nuevo conocimiento dentro de las organizaciones. Los datos deben estar limpios y en el formato adecuado para ser analizados. El presente trabajo tiene como objetivo presentar las técnicas empleadas en el proceso de limpieza de datos, en la construcción del Repositorio para Investigaciones en Gestión de Proyectos. En este escenario, son utilizadas entre otras, las técnicas de seudonimización y anonimización. Se realiza el diagnóstico de las técnicas por año y los resultados que se obtuvieron respectivamente. Se presenta la caracterización de siete nuevas bases de datos incluidas en el repositorio para fortalecer el desarrollo de investigaciones en gestión de proyectos. Las siete nuevas bases de datos contienen información relacionada con la evaluación de indicadores de cliente, programa y organización. Incluyen, además, información de otros indicadores asociados a los recursos humanos de organizaciones y proyectos, así como también indicadores de proyectos y las competencias de usuarios. La aplicación de las técnicas empleadas permitió obtener el Repositorio para Investigaciones en Gestión de Proyectos con datos limpios para el intercambio de la información con la comunidad científica internacional.

Palabras clave: Descubrimiento de Conocimiento, Gestión de Proyectos, Intercambio de información, Repositorio de Datos, Técnicas de limpieza de datos.

Abstract

Data repositories allow the sharing, reuse and localization of data for learning and discovery of new knowledge within organizations. The data must be clean and in the ideal format to be analyzed. The present work aims to present the techniques used in the process of data cleansing, in the construction of the Repository for Research in Project Management. In this scenario, pseudonymization and anonymization techniques are used among others. The diagnosis of the techniques is made per year and the results obtained respectively. The characterization of seven new databases included in the repository is presented to strengthen the development of research in project management. The seven new databases contain information related to the evaluation of client, program and organization indicators. They also include information on other indicators associated with the human resources of organizations and projects, as well as project indicators and user competencies. The application of the techniques used allowed to obtain the Repository for Research in Project Management with clean data for the exchange of information with the international scientific community.

Keywords: Knowledge Discovery, Project Management, Information Exchange, Data Repository, Data Cleaning Techniques.



Introducción

La ingeniería del software constituye una de las industrias emergentes que mayor crecimiento e impacto ha experimentado en años recientes. El desarrollo de los proyectos de esta industria respecto a la calidad y las buenas practicas está marcado por la existencia de numerosos estándares y libros desarrollados por diversos autores, entre los que se destacan, entre otros, el PMBOK (PMI, 2017), la ISO 21500 (ISO, 2012), CMMI (CMMI, 2010) e Ingeniería de Software. Un enfoque práctico (Presman, 2010).

A pesar de la gran cantidad de bibliografía que recoge buenas prácticas persisten dificultades como se refleja en los Reportes CHAOS del *Standish Group*. En los reportes realizados del 2011 y hasta el 2015 se señala que el porcentaje de los proyectos exitosos oscila entre los proyectos discutidos y los proyectos fracasados. En el 2015, el 19% y el 52% fueron los porcentajes de los proyectos fracasados y los proyectos discutidos respectivamente. Los proyectos exitosos representaron solamente el 29% (*The Standish Group*, 2016). En muchos de los casos las deficiencias se presentan por la falta de conocimiento de buenas prácticas para el control y seguimiento de proyectos (Lugo, 2015).

Una encuesta realizada a 45 profesionales de doce organizaciones de software, permitió confirmar, además, que los problemas asociados con factores humanos constituyen un factor de incidencia a considerar en el resultado de los proyectos (Torres, 2015).

En este escenario aprender de la experiencia de proyectos anteriores, ayudaría a disminuir el elevado número de proyectos cancelados o atrasados. Entre las técnicas que pueden ser aplicadas en este escenario se destaca la minería de datos.

Entendemos en el contexto de esta investigación que la minería de datos constituye un campo interdisciplinario donde convergen técnicas estadísticas, matemáticas, inteligencia artificial, el reconocimiento de patrones y bases de datos para el descubrimiento de conocimiento no trivial que reside de manera implícita en los datos (Gorunescu, 2011).

En particular, en la gestión de proyectos los datos generalmente provienen de sistemas de información y de bases de datos de proyectos terminados. Pero constituye una problemática el acceso y la organización de los datos porque con frecuencia los datos de proyectos no están públicos o se encuentran publicados de forma parcial y desordenada. Por otra parte, la calidad de la información, incluso en bases de datos internacionales de proyectos, no es suficiente para el desarrollo de las investigaciones (Aguilar, 2016). Además, un estudio realizado en (Santiesteban, A. M., et al., 2016) detectó que los repositorios de datos existentes no cubren las necesidades de investigaciones en la temática de gestión de proyectos.

En el Departamento de Investigaciones en Gestión de Proyectos de la Universidad de las Ciencias Informáticas (UCI), se trabaja en función de continuar desarrollando una plataforma para el desarrollo de investigaciones en gestión de proyectos. Para ello cuenta con un conjunto de investigadores experimentados, un programa de formación de masters en Gestión de Proyectos desde el 2006 y un sistema de información para la dirección integrada de proyectos GESPRO (Piñero, P., et al., 2014). A partir del conjunto de investigaciones desarrolladas por este departamento y como parte de este trabajo se ha logrado construir un repositorio de bases de datos para el desarrollo de investigaciones en gestión de proyectos. Los datos en dicho repositorio son provistos por expertos a partir de diferentes investigaciones rea-



lizadas como parte de tesis de maestría y doctorado que han empleado el sistema de información GESPRO (Piñero, P., et al., 2014).

En este trabajo, en particular, se presentan las técnicas para el tratamiento de errores que se emplearon en el desarrollo de los procesos de limpieza y mejora de la calidad de los datos contenidos en el repositorio. Se incluyen mejoras y extensiones al repositorio propuesto en (Santiesteban, A. M., et al., 2016). Además, se presenta la caracterización de las siete nuevas bases de datos incluidas en el mismo para fortalecer el desarrollo de investigaciones en gestión de proyectos.

El trabajo se encuentra estructurado de la siguiente forma. En la primera sección se presenta el proceso de limpieza de las bases de datos empleado en la construcción del repositorio. En la segunda sección se presenta el análisis de resultados, finalmente se presentan las conclusiones y la bibliografía.

Proceso de limpieza de las bases de datos del repositorio.

La construcción del Repositorio para Investigaciones en Gestión de Proyectos tiene sus inicios en el 2011. Se decide desarrollar el sistema de información en gestión de proyectos tanto para apoyar la innovación y la toma de decisiones en el proceso como para fomentar las investigaciones que permitieran elevar los resultados exitosos en la ejecución de proyectos.

Como parte del desarrollo del repositorio se lleva a cabo un proceso de limpieza de las bases de datos que se incluyen en el mismo. El proceso de limpieza y construcción del repositorio se concibe alineado completamente con los procesos de minería de datos que se muestran en la Ilustración I, y que se describen a continuación:

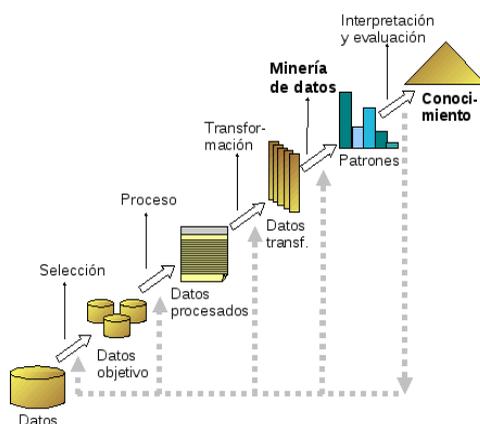


Figura 1: Procesos que intervienen en la minería de datos.

1. Selección: tiene como objetivo la caracterización y selección de las fuentes de datos.
2. Preprocesamiento y transformación: se aplican técnicas como limpieza de datos, la integración y transformación de datos, la reducción de datos y la selección de atributos.

3. Aplicar las técnicas de minería: es el proceso del cálculo de resúmenes y valores derivados. En esta etapa se
4. perfeccionan constantemente las técnicas y algoritmos que se encargan de extraer y representar el conocimiento de forma adecuada para la toma de decisiones. Se combinan técnicas potenciando las ventajas de cada una y atenuando sus debilidades.
5. Interpretación y evaluación: en este paso se procede al análisis de los resultados descubiertos. Incluye a su vez la resolución de posibles inconsistencias con otros conocimientos anteriores a la investigación.

El alcance de esta investigación cubre los procesos de: selección, preprocesamiento y transformación, no obstante, el “Paso 3 Aplicación de técnicas de minería de datos”, se explica en el análisis de resultados a partir del análisis de las investigaciones desarrolladas con el uso del repositorio.

Paso 1: Selección. Extensión del repositorio.

En el proceso de selección descrito en este trabajo se tomaron: 11 bases de datos contenidas en la primera versión del repositorio (Santiesteban, A. M., et al., 2016) y un conjunto de datos obtenidos en investigaciones desarrolladas entre el 2015 y el 2017. Estos datos corresponden a las temáticas relacionadas en la primera versión del repositorio ampliando y a siete nuevas temáticas, a partir de las cuales se extiende el repositorio y se crea la versión 17.10 del repositorio. A continuación, en la Tabla 1 se listan las dieciocho bases de datos seleccionadas para conformar el repositorio y que se emplearán en la etapa de preprocesamiento y transformación.

Tabla 1: Bases de Datos del Repositorio para Investigaciones en Gestión de Proyectos.

Bases de datos contenidas en el repositorio.	Bases de datos Incorporadas en la versión 17.10.
<ol style="list-style-type: none"> 1. Asignación de Recursos y Tareas. 2. Evaluación de Proyectos. 3. Evaluación Borrosa de Proyectos. 4. Gestión de Proyectos. 5. Gestión de Tareas. 6. Gestión de Recursos Humanos y Materiales. 7. Gestión de Riesgos. 8. Gestión de Cambios. 9. Gestión de Desviaciones. 10. Gestión de Requerimientos. 	<ol style="list-style-type: none"> 11. Gestión de Indicadores. 12. Base de Datos Evaluación de Proyectos Hard. 13. Base de Datos Gestión de Competencias de Usuarios. 14. Base de Datos Gestión de Indicadores del Cliente. 15. Base de Datos Gestión de Indicadores de la Organización. 16. Base de Datos Gestión de Indicadores de Programas. 17. Base de Datos Gestión de los Indicadores de los Recursos Humanos del Proyecto. 18. Base de Datos Gestión de los Indicadores de los Recursos Humanos de la Organización.



Caracterización las siete bases de datos incluidas Base de datos “Evaluación de Proyectos Hard”

La base de datos (BD) Evaluación de Proyectos Hard, contiene información a partir del año 2015. Está basada en 7 indicadores con valores lingüísticos que evalúan a los proyectos.

Tabla 2: Características de la Base de datos “Evaluación de Proyecto Hard”.

Indicadores		Totales (10)	Indicadores que relaciona la BD
Cantidad de Atributos	Entero	2	<ol style="list-style-type: none"> Índice de la calidad del dato. Índice de la ejecución. Índice del Rendimiento de la ejecución. Índice de Rendimiento de la Eficacia. Índice de Rendimiento de la Planificación. Índice de Rendimiento de los Recursos Humanos. Productividad.
	Texto	8	
Cantidad de Proyectos, de acuerdo a los indicadores que se relacionan en la base de datos, evaluados de	Bien	408	
	Regular	272	
	Mal	510	

Base de datos “Gestión de Competencias de Usuarios”.

La BD Gestión de Competencias de Usuarios obtenida a partir de la investigación (Torres, 2015) contiene información a partir del año 2015. Relaciona los usuarios y las competencias que estos presentan. Está basada, además, en 4 indicadores que arrojan una evaluación del usuario, la cual se almacena con términos lingüísticos.

Tabla 3: Características de la Base de Datos Gestión de Competencias de Usuarios.

Indicadores		Totales (8)	Indicadores que relaciona la BD
Cantidad de Atributos	Decimal	1	<ol style="list-style-type: none"> Índice de Rendimiento del Recurso Humano con respecto a la Eficacia. Índice de Rendimiento del Recurso Humano con respecto a la Eficiencia. Índice de Rendimiento del Recurso Humano con respecto a la Efectividad. Índice de Rendimiento del Recurso Humano con respecto al Cospine.
	Texto	4	
	Entero	1	
	Fecha	2	

Base de datos “Gestión de Indicadores del Cliente”.

La BD Gestión de Indicadores del Cliente, contiene información a partir del año 2015. Relaciona 12 indicadores orientados al cliente.



Tabla 4: Características de las bases de datos Gestión de Indicadores del Cliente, Gestión de Indicadores de la Organización y Gestión de Indicadores de Programas.

Indicadores		Totales (16)	Indicadores que relaciona la BD	
Cantidad de Atributos	Decimal	12	1. Índice de la calidad del dato 2. Índice de la ejecución 3. Índice del Rendimiento de la ejecución 4. Índice de Rendimiento de la Eficacia 5. Índice de Rendimiento de la Planificación 6. Índice de Rendimiento de la Logística. 7. Índice de Rendimiento de los Recursos Humanos.	8. Índice de Rendimiento de Costos 9. Plan. 10. Costo Planificado. 11. Valor Ganado. 12. Costo Real.
	Texto	1		
	Entero	1		
	Fecha	2		

Base de datos “Gestión de Indicadores de la Organización”.

La BD Gestión de Indicadores de la organización, contiene información a partir del año 2015. Relaciona 12 indicadores orientados a la organización, los cuales están expuestos en la Tabla 4. Esta BD también cuenta con los mismos atributos expuestos en la Tabla 4.

Base de datos “Gestión de Indicadores de Programas”.

La BD Gestión de Indicadores de Programas, contiene información a partir del año 2015. Relaciona 12 indicadores orientados a los programas, los cuales están expuestos en la Tabla 4. Esta BD también cuenta con los mismos atributos expuestos en Tabla 4.

Base de datos “Gestión de los Indicadores de los Recursos Humanos del Proyecto”.

La BD Gestión de Indicadores de los Recursos Humanos del Proyecto obtenida a partir de la investigación (Torres, 2011) contiene información a partir del año 2015. Relaciona 4 indicadores que arrojan una evaluación cualitativa y cuantitativa del recurso humano por proyectos.

Tabla 5: Características de la BD Gestión de Indicadores de los Recursos Humanos del Proyecto.

Indicadores		Totales (9)	Indicadores que relaciona la BD
Cantidad de Atributos	Decimal	4	1. Índice de Correlación del Recurso Humano con respecto al Trabajo. 2. Índice de Rendimiento del Recurso Humano con respecto a la Eficiencia. 3. Índice de Rendimiento del Recurso Humano con respecto a su Aprovechamiento. 4. Índice de Rendimiento del Recurso Humano con respecto a la Eficiencia.
	Texto	3	
	Fecha	2	

Base de datos “Gestión de los Indicadores de los Recursos Humanos de la Organización”.

La BD Gestión de Indicadores de los Recursos Humanos de la Organización contiene información a partir del año 2015. Relaciona 4 indicadores que arrojan una evaluación cualitativa y cuantitativa de los recursos humanos que almacena.

Tabla 6: Características de la BD Gestión de Indicadores de los Recursos Humanos de la Organización.

Indicadores		Totales (9)	Indicadores que relaciona la BD
Cantidad de Atributos	Decimal	4	Los indicadores que se exponen en la Tabla 5.
	Texto	2	
	Fecha	2	

Paso 2: Preprocesamiento y transformación. Técnicas de limpieza de datos.

Este proceso tiene como objetivo elevar la calidad de las bases de datos. Entre los atributos de calidad que se persiguen en este proceso se encuentran: confidencialidad de la información, datos correctamente escritos en el idioma de cada base de datos, niveles adecuados de incompletitud de información en cada uno del registro de datos (inferior al 10% del total de información), la no existencia de registros completamente vacíos sin información, la eliminación de datos anómalos en las bases de datos que no estén diseñadas para el tratamiento de datos anómalos, el cubrimiento de las áreas de conocimiento de la gestión proyectos, la estandarización de los términos en las bases de datos.

Los procesos de preprocesamiento y transformación de los datos se realizan en todas las bases de datos del repositorio, es decir, tanto en las contenidas como en las incorporadas en la presente investigación. En la Tabla 7 se muestra un diagnóstico desde la fecha en que se inició el repositorio hasta la fecha. El diagnóstico refleja los problemas y soluciones que se identificaron y aplicaron respectivamente en las bases de datos del repositorio y el sistema de información que provee los datos para el mismo. También se relacionan en la Tabla 7 las técnicas de limpieza de datos que se realizaron en las bases de datos que se obtuvieron en los períodos que se señalan.

Tabla 7: Diagnóstico de la construcción del Repositorio para Investigaciones en Gestión de Proyectos.

Período 2011 - 2012	Período 2015-2018
Problemas: ✖ Datos incompletos. ✖ Falta de estándares en la representación de los datos. Soluciones: <ul style="list-style-type: none"> • Mejoras en el Sistema de Información. • Uso de Listas. • Estandarizaciones. Creación de campos obligatorios. Técnicas de limpieza de datos: <ul style="list-style-type: none"> • Imputación de datos incompletos (Juicio de Expertos). • Eliminación de datos muy incompletos. • Verificación por expertos de los datos de proyectos. 	Existen datos más limpios y consistentes. Problema: ✖ Se afecta la integridad y confidencialidad de los datos. Solución: <ul style="list-style-type: none"> • Se aplican técnicas para conservar la integridad y confidencialidad de los datos. Datos limpios. Técnicas de limpieza de datos: <ul style="list-style-type: none"> • Anonimización. • Seudonimización. • Recopilación.

A partir de los problemas del período 2015-2018 que se ponen en la Tabla 7, se decide emplear las técnicas de Anonimización, Seudonimización y Recopilación. A continuación, como parte de este trabajo se explican las mismas.

La técnica de la anonimización consiste en la eliminación de tuplas o columnas. En las bases de datos del repositorio existían columnas que no arrojaban información relevante para investigaciones o tuplas repetidas, las que se deciden eliminar. Así mismo, se relacionaban columnas que contenían información confidencial pero no relevante para las investigaciones en gestión de proyectos, por tanto, son eliminadas con el objetivo de proteger la confidencialidad de los datos contenidos.

La técnica de la seudonimización consiste en el renombramiento de valores tales como el nombre de personas. En las tablas de las bases de datos del repositorio se empleó la misma para renombrar nombres de proyectos, entidades implicadas, empresas y personas.

La recopilación de los datos consiste en el empleo de técnicas y herramientas como las entrevistas, la encuesta, el cuestionario, la observación, entre otras, para obtener el conjunto de datos para el repositorio.

Para aplicar las técnicas antes mencionadas se emplean los lenguajes de programación SQL y PL/PgSQL y el gestor de bases de datos PostgreSQL. Con el objetivo de elevar la calidad en la construcción de las bases de datos para investigaciones se realiza un proceso manual de verificación de los datos luego de aplicadas las técnicas. Este proceso afecta los costos de obtención de las bases de datos, pero mejora los atributos de calidad requeridos.

Resultados y discusión

Repositorio para Investigaciones en Gestión de Proyectos.

A partir de la aplicación de las técnicas empleadas como parte del proceso de limpieza de datos realizado se obtuvo la versión 17.10 del Repositorio para Investigaciones en Gestión de Proyectos, que cuenta con dieciocho bases de datos independientes. (Ver Tabla 8).

Tabla 8: Características Generales de las BD del Repositorio.

Base de Datos	Tipo de Datos	Cantidad de Atributos	Cantidad de Instancias	Cantidad de Proyectos
Asignación de Recursos y Tareas	Mezclado	38	9 569	88
Evaluación de Proyectos	Mezclado	84	8 430	667
Evaluación Borrosa de Proyectos	Mezclado	19	8 430	667
Gestión de Proyectos	Mezclado	49	136	134
Gestión de Tareas.	Mezclado	26	12 563	103
Gestión de Recursos Humanos y Materiales	Mezclado	28	612	35
Gestión de Riesgos	Mezclado	19	2 379	345
Gestión de Cambios	Mezclado	20	593	44
Gestión de Desviaciones	Mezclado	10	2 506	330
Gestión de Requerimientos	Mezclado	17	1 652	36
Gestión de Indicadores	Mezclado	93	34 810	996
Evaluación de Proyectos Hard	Mezclado	19	8 430	667
Gestión de Competencias de Usuarios	Mezclado	11	10 452	0
Gestión de Indicadores del Cliente	Mezclado	16	387	0
Gestión de Indicadores de la Organización	Mezclado	16	87	0
Gestión de Indicadores de Programas	Mezclado	16	87	0
Gestión de los Indicadores de los Recursos Humanos del Proyecto	Mezclado	9	6 539	84
Gestión de los Indicadores de los Recursos Humanos de la Organización	Mezclado	8	3 149	0

En la Tabla 9 se muestra, además, un breve análisis en el que se manifiesta la relación de las bases de datos del repositorio con 9 áreas de conocimiento de la gestión de proyecto. Para la presente investigación no se obtuvieron datos correspondientes al área de la Gestión de las Comunicaciones. En próximos trabajos serán incorporados al repositorio.



Tabla 9: Relación entre las bases de datos del repositorio y las áreas de conocimiento de la gestión de proyectos.

<u>Base de Datos</u>	<u>G. de Alcance</u>	<u>G. de Tiempo</u>	<u>G. de los Costos</u>	<u>G. de la Calidad</u>	<u>G. de los RR.HH</u>	<u>G. de los Riesgos</u>	<u>G. de Adquisiciones</u>	<u>G. de Interesados</u>	<u>G. de Integración</u>
Asignación de Recursos y Tareas	x	x	x				x		x
<u>Evaluación de Proyectos</u>		x	x		x				x
<u>Evaluación Borrosa de Proyectos</u>		x	x		x				x
<u>Gestión de Proyectos</u>	x	x	x		x			x	
<u>Gestión de Tareas</u>	x	x	x		x				

Tabla 10: Relación entre las bases de datos del repositorio y las áreas de conocimiento de la gestión de proyectos.

Materiales									
Gestión de Riesgos			x			x	x		x
Gestión de Cambios		x	x		x				
Gestión de Desviaciones		x	x				x		
Gestión de Requerimientos		x	x		x				
Gestión de Indicadores		x	x	x	x	x		x	x
Evaluación de Proyectos Hard			x	x	x				x
Gestión de Competencias de Usuarios				x		x			
Gestión de Indicadores del Cliente				x		x		x	
Gestión de Indicadores de la Organización				x	x	x			x

Gestión de Indicadores de Programas			x	x	x				x
Gestión de los Indicadores de los Recursos Humanos del Proyecto					x				x
Gestión de los Indicadores de los Recursos Humanos de la Organización					x				x

Uso y Explotación del Repositorio para Investigaciones en Gestión de Proyectos.

El uso y explotación contribuye al aumento de las investigaciones asociadas a la Gestión de Proyectos. Como muestra de ellos son las tesis de maestría y doctorados que se desarrollan actualmente y las realizadas haciendo uso del mismo. Estas últimas suman un total de 13 investigaciones (Peña, 2017), (Castro, 2016), (Jiménez, 2017), (Pestano, 2011), (Torres, 2011), (Vázquez, 2011), (Lugo, 2012), (Piñera, 2013), (Pacheco, 2014), (Rodríguez, 2015), (Bermúdez, 2015), (Lugo, 2015), (Torres, 2015) además de 11 publicaciones (Castro, Pérez, Piñero P. R., Piñero P. Y., et. at., 2016), (Castro, Pérez, Piñero P. Y., et. at., 2016), (Castro, et a., 2016), (García, et al., (2016), (Torres, Castro, Pérez, Piñero P. Y., Diéguez, 2016), (Sosa, Pérez, García, Peñaherrera, Piñero P. Y., 2016), (Castro, Pérez, Piñero P. Y., Martínez, Cruz, 2016), (Boaventura, Peña, Verdecia, Fustiel, 2016), (Guerrero, Jiménez, León, Piñero P. Y., Romillo, 2016), (Guerrero, Pérez, Ventura, Morell, Piñero P. Y., 2017), (Guerrero, Morell, Noaman, Ventura, 2016) . A continuación, en la Figura 1 se expresan estos resultados.

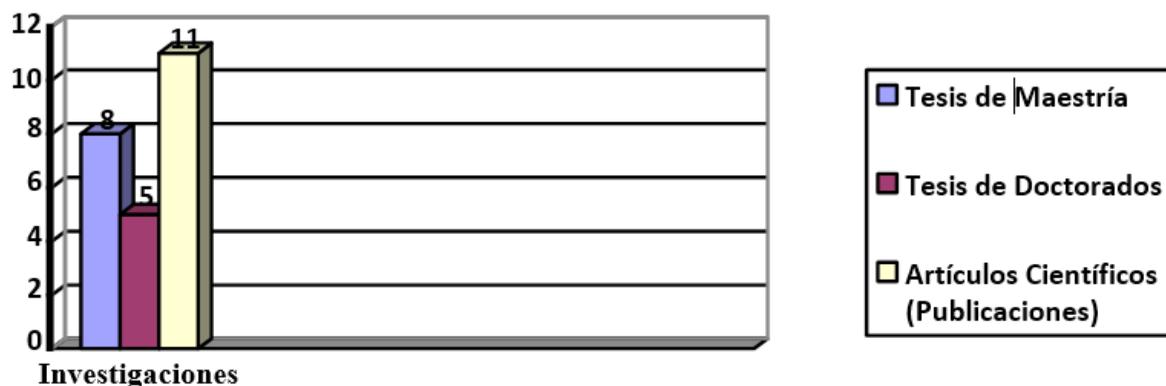


Figura 1: Investigaciones realizadas empleando el repositorio.

Conclusiones

La extensión del Repositorio para Investigaciones en la Gestión de Proyectos, basada en siete nuevas bases de datos, permitió alcanzar un total de dieciocho bases de datos, lo que permite el fortalecimiento de futuras investigaciones. Con la inserción de datos reales actualizados de proyectos, las bases de datos del repositorio fueron actualizadas, lo que permite obtener mayor cantidad de datos para el desarrollo de investigaciones.

El proceso de limpieza realizado a las bases de datos del repositorio, permitió la mejora y calidad de los datos contenidos en el mismo. Esto constituye un impacto positivo en los resultados de las investigaciones que se realicen con los datos.

Con la realización del proceso de limpieza, se fortalece el empleo del repositorio mediante la publicación de las bases de datos para la comunidad científica internacional.

Referencias

- PMI (2017), Project Management Body of Knowledge (PMBOK), 6th ed., Pennsylvania: PMI Publications.
- ISO, ISO 21500 (2012): Guidance on Project Management, International Organization for Standardization, Ginebra.
- CMMI (2010): CMMI para Desarrollo, Versión 1.3. Mejora de los procesos para el desarrollo de Mejores productos y servicios. Software Engineering Institute. EE. UU.
- Presman PhD, R. S. (2010): Ingeniería de Software. Un enfoque práctico, 7ma ed., México.
- The Standish Group (2016), The CHAOS Manifesto. Think Big, Act Small, Boston: The Standish Group.
- Lugo, A. (2015): Modelo para el control de la ejecución de proyectos basado en softcomputing (Tesis de doctorado), Universidad de las Ciencias Informáticas, La Habana.
- Torres, S. (2015): Modelo de evaluación de Competencias a partir de Evidencias durante la gestión de Proyectos (Tesis de doctorado), Universidad de las Ciencias Informáticas, La Habana.
- Gorunescu, F. (2011): Data Mining, Concepts, Models and Techniques. Intelligent Systems Reference Library, Volume 12. University of South Australia.
- Aguilar PhD, J. (2016): Panel “Aplicaciones de la Inteligencia Artificial y el reconocimiento de Patrones a la Gestión de Proyectos”, II Taller Internacional de Gestión de Proyectos.
- Piñero, P., Torres, S., Izquierdo, M., Lugo, A., Rizo, J., Abelardo, F. N., Santiesteban, A. M., Pérez, A. D., Pérez, I., Torres, K. M., Machado, R., Tasé, M., Fernández, A., Delgado, R (2014). GESPRO. Paquete para la Gestión de Proyectos. La Habana, Cuba: UCI.
- Santiesteban, A. M., Castro, F., Rojas, C., Rivero C., Piñero, P., Pérez, I., Rodríguez, O., Acuña, H.. (2016):” Repositorio para el desarrollo de investigaciones en Gestión de Proyectos”, II Taller internacional de Gestión de Proyectos, II Conferencia Científica Internacional UCIENCIA, ISBN 978-959-286-054-4.



- Pompa, M.: Procedimiento de limpieza de datos en el GESPRO (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana, 2015.
- Peña, M. (2017): Modelo para la evaluación de la factibilidad de proyectos basado en técnicas de soft-computing (Tesis de Doctorado), Universidad de las Ciencias Informáticas, La Habana.
- Castro, F. (2016): Un modelo para el aseguramiento de ingresos basado en técnicas de soft computing (Tesis de Doctorado), Universidad de las Ciencias Informáticas, La Habana.
- Jiménez, G. (2017): Modelo de control de proyectos basado en el análisis geo-referencial para la ayuda a la toma de decisiones (Tesis de Doctorado), Universidad de las Ciencias Informáticas, La Habana.
- Pestano, H. (2011): Propuesta de modelo de desarrollo para líneas de productos de software en centros de producción (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Torres, S. (2011): Modelo para la Gestión de Recursos Humanos en Centros de Desarrollo de Sistemas de Información (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Vázquez, (2011): Definición de una arquitectura de referencia para una línea de productos de software (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Lugo, A. (2012): Modelo para el control de la ejecución de proyectos basado en indicadores y lógica borrosa (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Piñera, (2013): Algoritmo para la sumarización lingüística de datos a partir de reglas de asociación (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Pacheco, (2014): Descubrimiento de conocimiento a partir de la relación rasgos de la personalidad-rendimiento laboral en proyectos informáticos (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana. Rodríguez, (2015): Aprendizaje de reglas difusas para la evaluación de proyectos basado en técnicas de clasificación (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Bermudez, (2015): Sistema basado en técnicas de soft computing para la evaluación de la ejecución de proyectos (Tesis de Maestría), Universidad de las Ciencias Informáticas, La Habana.
- Castro, G. F.; Pérez, I., Piñero, P. R., Piñero, P. Y., et. at. (2016). Plataforma para aseguramiento de ingresos, aplicación en gestión de proyectos y telcos. IV Taller Internacional Las TIC en la Gestión de las Organizaciones, Informática 2016.
- Castro, G. F.; Pérez, I., Piñero, P. Y., et. at. (2016). PRODanalysis, un Sistema para el Aseguramiento de Ingresos Basado en Minería de Outliers. INNOVA Research Journal, Vol 1, No. 7, 18-36. ISSN 2477-9024. Disponible en: <http://www.journaluidegye.com/magazine/index.php/innova/article/view/34>.
- Castro, G.F., Pérez, I., Piñero, P. Y., García, R. (2016). Método para el aseguramiento de ingresos en entornos de desarrollo de software. Revista Cubana de Ciencias Informáticas Vol. 10, No. Especial UCIENCIA, ISSN: 2227-1899 | RNPS: 2301 <http://rcci.uci.cu> Pág. 43-57. Disponible en: <http://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1463>.
- García, R., Pérez, I., Villavicencio, N., Piñero, P. Y., Beovides, S. (2016). Experiencias usando algoritmos genéticos en la planificación de proyectos. Revista Cubana de Ciencias Informáticas Vol. 10, No. Especial UCIENCIA, ISSN: 2227-1899 | RNPS: 2301 <http://rcci.uci.cu> Pág. 71-86. Disponible en: <http://rcci.uci.cu>



uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1462.

- Torres, S., G.F. Castro, Pérez, I., Piñero, P. Y., Diéguez, L.P. (2016) Rough Sets for Human Resource Competence Evaluation and Experiences. Applied Mathematics, 2016, 7, 1317-1325. <http://dx.doi.org/10.4236/am.2016.712116>
- Sosa, R., Pérez, I., García, R., Peñaherrera, E., Piñero, P. Y. (2016). Ecosistema de Software GES-PRO-16.05 para la Gestión de Proyectos. Revista Cubana de Ciencias Informáticas Vol. 10, No. Especial UCIENCIA, ISSN:2227-1899 | RNPS: 2301 <http://rcci.uci.cu>. Pág. 239-251. Disponible en: <http://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1475>.
- Castro, G.F., Pérez, I., Piñero, P. Y., Martínez, N., Cruz, Y. (2016). Aplicación de la minería de datos anómalos en organizaciones orientadas a proyectos. Revista Cubana de Ciencias Informáticas Vol. 10, No. Especial UCIENCIA, ISSN: 2227-1899 | RNPS: 2301 <http://rcci.uci.cu>. Pág. 195-209. Disponible en:<http://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1456>.
- Boaventura José, C., Peña Herrera, E., Verdecia Vicet, P., & Fustiel Álvarez, Y. (2016). Elección entre una metodología ágil y tradicional basado en técnicas de soft computing. Revista Cubana de Ciencias Informáticas,10, 145-158.
- Guerrero Enamorado, A., Jiménez Moya, G., León Companioni, A., Piñero Pérez, P. Y., Romillo, A., (2016). SIGESPRO: Sistemas de Información Geográfica para controlar proyectos.Vol. 10, Núm. 2: Abril-Junio. Guerrero Enamorado, A., Pérez Pupo, I., Ventura, S., Morell, C., & Piñero Pérez, P. Y. (2017). Evaluación de proyectos usando sistemas basados en algoritmos genéticos de aprendizaje de reglas. Revista Cubana de Ciencias Informáticas, 11(4).
- Guerrero-Enamorado, A., Morell C., Amin Y. Noaman & Sebastián Ventura (2016) An Algorithm Evaluation for Discovering Classification Rules with Gene Expression Programming, International Journal of Computational Intelligence Systems, 9:2263-280, DOI: 10.1080/18756891.2016.1150000

