Herramienta para la Automatización de la Revisión Sistemática

Edgar Felipe Uribe Rayas Centro de Investigación en Matemáticas, CIMAT A.C. Unidad Zacatecas edgar.uribe@cimat.mx

Jorge Alexis Márquez del Real Centro de Investigación en Matemáticas, CIMAT A.C. Unidad Zacatecas jorge.marquez@cimat.mx

J. Guadalupe Uribe Dévora Centro de Investigación en Matemáticas, CIMAT A.C. Unidad Zacatecas guadalupe.uribe@cimat.mx

Claudia Valtierra Alvarado Centro de Investigación en Matemáticas, CIMAT A.C. Unidad Zacatecas claudia.valtierra@cimat.mx

Jezreel Mejia Miranda Centro de Investigación en Matemáticas, CIMAT A.C. Unidad Zacatecas jmejia@cimat.mx Resumen: La realización de una revisión sistemática provee a los investigadores las bases para obtener conocimiento científico mediante la identificación de artículos científicos relevantes, los cuales contengan las tendencias en investigación y experimentación acerca de un tema en específico. Existen varios estudios acerca de cómo implementar una revisión sistemática, sin embargo, estas investigaciones solo indican los pasos que conforma el protocolo de la revisión sistemática, sin proporcionar alguna herramienta o aplicación de software que soporte la automatización de este protocolo, dando como resultado que los investigadores dediquen una buena cantidad de tiempo en la detección y búsqueda de artículos científicos relevantes de manera manual. Por lo tanto, en este artículo se presenta la propuesta de una herramienta para la automatización del protocolo de la Revisión sistemática, con la finalidad de reducir el tiempo en el desarrollo del protocolo de la Revisión Sistemática.

Palabras clave: Revisión Sistemática, Protocolo, Herramientas.

A Tool to Automatize the Systematic Review

Abstract: Systematic Reviews provide to the researchers the base for obtaining scientific knowledge through of the identification of relevant scientific articles, which contain research tendencies and experimentation about a specific topic. There are several studies about how to implement a systematic review, however these studies only indicate the systematic review protocol and do not provide some software tool or application to support the automation of this protocol, giving as a result that the researchers expend large amounts of time to search and detect relevant scientific articles manually. Therefore, in this paper is proposed a tool to automatize the systematic reviews protocol, in order to decrease the time to perform the systematic review protocol.

Keywords: Systematic Review, Protocol, Tools.

1. Introducción

La inclusión de revisiones sistemáticas en la ingeniería de software provee un medio efectivo para la identificación y evaluación de la literatura e investigaciones existente en un dominio específico. Esta actividad es considerada como el primer paso a llevar a cabo en la obtención de conocimiento, entendimiento y delimitación de un tema de investigación

científico(Mian, Conte, Natali, Biolchini, & Travassos, n.d.). Varios análisis sobre revisiones sistemáticas y su aplicación a la ingeniería del software, han sido realizados (Group, 2007; Jørgensen & Shepperd, 2007; Kitchenham & Dybå, 2004; Mian et al., n.d.). Estos trabajos analizan las actividades que se realizan en una revisión sistemática y proponen nuevos pasos o actividades para llevar a cabo este protocolo, sin embargo, y a pesar de sentar las bases para el establecimiento del protocolo de la revisión sistemática, hasta el momento se han detectado pocas herramientas existentes, que de manera eficiente soporten la automatización de las actividades de una revisión sistemática de manera completa. La mayoría de herramientas como Mendeley (Ltd, 2013), RefWorks (RefWorks, 2013), JabRef (Team, 2013), EndNote (Thomson, 2013), que sirven de apoyo al investigador se enfocan principalmente en la elaboración de citas bibliográficas, administración de estudios encontrados y en la interacción con algunas bibliotecas digitales.

Por lo tanto, el objetivo principal de esta investigación es presentar el desarrollo de una herramienta o aplicación de software que permite la automatización del protocolo que sigue una revisión sistemática. El artículo está estructurado de la siguiente manera: en la sección 2, se presenta las fases y las actividades del protocolo de la revisión sistemática; en la sección 3 se analizan herramientas existentes para establecer una comparativa y detectar mejoras para ser incluidas en la herramienta propuesta; en la sección 4, se presenta la propuesta de la herramienta para la revisión sistemática; en la sección 5, se presenta el caso de estudio en los que se aplica la herramienta y finalmente, la sección 6 se presentan las Conclusiones y trabajos futuro posibles.

2. Protocolo de la Revisión Sistemática

De acuerdo a la definición de Kitchenham (Group, 2007), la revisión sistemática es un método que permite a los especialistas obtener resultados relevantes y cuantificables, esto lleva a la identificación, selección y producción de evidencias basadas en la investigación de un tópico en particular (Mian et al., n.d.). Para establecer las fases y pasos que debería contener la herramienta que se propone, fue necesario realizar un análisis de diferentes autores acerca de cómo implementar una revisión sistemática. Pino(Espa, 2006) señala las actividades que se deben realizar para llevar a cabo una revisión sistemática. Kitchenham, Dyba y Jorgensen (Kitchenham & Dybå, 2004) señalan la importancia que adquiere la práctica de revisiones sistemática en la ingeniería del software y explican de manera general como se implementa una revisión, de donde se puede extraer el método a implementar en la herramienta. Un ejemplo de selección de estudios primarios, y la importancia de la estadística se muestra también en el trabajo realizado por Jorgensen y Shepperd (Jørgensen & Shepperd, 2007). Un enfoque de aplicación de las revisiones sistemáticas y los pasos a seguir en la Ingeniería del Software es presentado por Mian, Conte, Natali, Biolchini y Travassos (Mian et al., n.d.), establece el proceso de las revisiones sistemáticas adecuado a la ingeniería del software.

De la investigación realizada se obtuvieron las fases principales y los pasos que estas deben integrar para una revisión sistemática, las cuales son las siguientes:

- Planificación de la revisión: la primer fase, en esta etapa se identifica la investigación, se establecen los objetivos y se realizan las siguientes actividades:
 - Elección de nombre de la investigación.
 - Formulación de las cadenas de búsqueda.

- Selección de fuentes de búsqueda de estudios y establecer los criterios de selección de estudios primarios y secundarios.
- Realizar la revisión sistemática: en esta segunda fase se ejecutan las siguientes actividades:
 - Ejecución de la búsqueda.
 - Evaluación de la calidad de los estudios.
 - Revisión de los estudios seleccionados.
 - Extracción de la información.
 - Documentos de extracción de datos.
 - Ejecución de la extracción.
- Reporte de la revisión: tercera y última fase en la que se resume y analizan los resultados utilizando métodos estadísticos, las actividades que se contemplan son:
 - Cálculo estadístico.
 - Presentación de resultados.

3. Herramientas existentes

Una vez establecidas las fases del protocolo y las actividades contenidas en cada una de ellas, se realizó un estudio sobre las herramientas existentes que favorecen la ejecución de una revisión sistemática. Se encontró que si bien existen aplicaciones que brindan soporte en algunos aspectos, éstas no cubren completamente el proceso que conlleva la realización de una revisión sistemática, entre las herramientas que se encontraron se citan las siguientes:

 Mendeley (Ltd, 2013), es una herramienta para la elaboración de citas bibliográficas, control de estudios agregados de manera manual y cuenta con un plugin llamado webimporter, el cual se utiliza para agregar la funcionalidad de descargas de la aplicación a los distintos navegadores web, sin embargo, las búsquedas se siguen realizando de manera manual, lo que

- conlleva a que los investigadores sigan dedicando un considerable lapso del tiempo de la revisión sistemática a la búsqueda de artículos.
- RefWorks (RefWorks, 2013), es una aplicación online de paga, permite un periodo de prueba de 30 días gratuitos. La herramienta funciona como un gestor en línea de artículos y referencias, artículos pero de forma manual, es decir el investigador debe tener los artículos o estudios digitales en su computadora e importarlos a la aplicación que se mueve sobre la plataforma web, en palabras sencillas RefWorks sirve para creación de citas bibliográficas y administración de los artículos.
- JabRef (Team, 2013) es una aplicación open source que permite entre otras características, al igual que las dos anteriores herramientas, la administración de artículos y referencias, así como la inclusión de éstas en investigaciones.
- EndNote (Thomson, 2013) es una herramienta para administración y publicación de bibliografías, lo interesante de la herramienta es que permite aplicar hasta 5000 estilos de citas bibliográficas.

Cabe decir que las 4 aplicaciones permiten la creación de una base de datos de artículos alojados en la computadora de los investigadores, además, permiten la búsqueda de datos de un artículo en internet, pero con el fin de crear citas bibliográficas, y aunque son de una gran utilidad no brindan el soporte necesario para la realización de una revisión sistemática.

Como resultado del análisis, la Tabla 1 muestra que los pasos identificados del protocolo de la revisión sistemática no son incluidos en las herramientas analizadas.

Herramienta Elementos	Mendeley	RefWorks	JabRef	EndNote
Citas bibliográficas	Si	Si	Si	Si
Búsquedas en internet	Si	Si	Si	Si
Gestión de estudios	Si	Si	Si	Si
Asignación de Prioridad	No	No	No	No
Búsquedas en bibliotecas digitales	Si	Si	No	Si
Extracción automática de información	Si	Si	Si	Si
Elaboración de Plantillas	No	No	No	No
Elaboración de gráficos para análisis	No	No	No	No
Selección de estudios primarios.	No	No	No	No

Tabla 1. Comparación de funciones de herramientas similares.

Como puede observarse, no existe alguna herramienta que cubra y automatice las actividades identificadas para el protocolo de la revisión sistemática, y que en el desarrollo de la herramienta deben ser establecidos como funcionalidad principal.

4. Propuesta de Herramienta para la Revisión Sistemática

En esta sección se describen los módulos que integran la herramienta propuesta. Para facilitar el uso de la herramienta, ésta se divide en módulos de fácil acceso y comprensión, para proporcionar a los investigadores un ambiente amigable y práctico. A continuación se describen los módulos de la

aplicación, los principales son mostrados en la Figura 1, y en la Figura 2, y están basados en la identificación de las fases y actividades obtenidas en la sección 2.

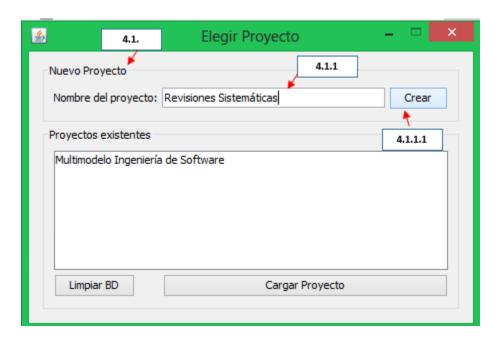


Figura 1. Módulo de Soporte a la Planificación de la Revisión.

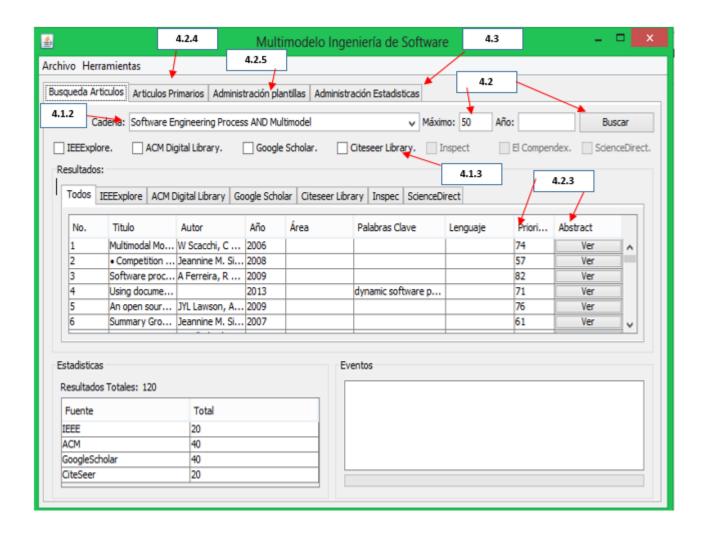


Figura 2. Principales Módulos de la Herramienta para la Automatización del Protocolo de la Revisión Sistemática.

4.1 Planificación de la revisión.

La herramienta cubre esta fase del protocolo mediante los siguientes módulos:

4.1.1 Elección de nombre de la investigación:

se establece el entorno para realizar esta función mediante la ventana de nuevo proyecto que permite la creación de un proyecto de investigación o revisión sistemática, a continuación se agrega una breve descripción de este módulo.

4.1.1.1 Creación de Un proyecto

Al crear un nuevo proyecto se le asigna un nombre de acuerdo al tópico de la investigación que se realiza, este proyecto de manera automática, se agrega a la base de datos de la aplicación, todas las búsquedas, estudios, archivos, plantillas, estadísticas y descargas se almacenen para su posterior uso en la base de datos. La principal función de esté módulo es proveer orden en la administración de estudios y revisiones, al tener ordenados en la base de datos los proyectos y sus elementos, simultáneamente se crea una estructura de carpetas que permite tener ordenados artículos, plantillas y estadísticas de cada estudio permitiendo un perfecto control sobre la ubicación de cada elemento de la investigación.

4.1.2 Formulación de las cadenas de búsqueda:

El motor de búsqueda que se implementa en la herramienta está basado en los criterios utilizados por las principales bibliotecas digitales, esta actividad se realiza en base al módulo de la herramienta descrito enseguida.

4.1.2.1 Módulo de búsqueda de artículos.

Permite la entrada de cadenas de búsqueda formadas por los investigadores, las cadenas pueden estar formadas por los principales tópicos de una investigación, luego el sistema las convierte a expresiones regulares para realizar la extracción de estudios en las fuentes, los datos por los que se puede buscar artículos o formar cadenas de búsqueda son: a) Titulo, b) Palabras

principales del tópico de la investigación, c) Tópicos enlazados con conectores lógicos booleanos AND, OR y NOT, d) Autor, e) Sinónimos y abreviaturas.

4.1.3 Selección de fuentes de búsqueda de estudios y establecer los criterios de selección de estudios primarios y secundarios:

La herramienta cuenta con un módulo de búsqueda de estudios el cual permite la elección de bibliotecas digitales, el módulo que realiza esta función es el siguiente:

4.1.3.1 Bibliotecas Digitales:

Este módulo incrustado en la pantalla de búsqueda permite selección y la extracción de estudios de las siguientes fuentes: 1) IEEXplore, 2) ACM Digital Library, 3) Citeseer Library, 4) Inspect, 5) ElCompendex, 6) ScienceDirect, 7) Scholar Google.

4.1.3.2 El criterio de selección de estudios y proceso de selección de estudios.

El sistema favorece la selección de estudios, y considera los siguientes factores: a) Año de publicación, b) Incidencia de autores en el tema de investigación, c) Incidencia de palabras clave dentro del abstract de cada documento.

4.2 Realizar la revisión sistemática

La herramienta permite la automatización de esta segunda fase mediante un conjunto de módulos que realizan las funciones requeridas y que se describen a continuación:

4.2.1 Número máximo de estudios a buscar por biblioteca.

Esta opción permite establecer un máximo de estudios ordenados por prioridad que se buscaran y mostraran en cada librería.

4.2.2 Ejecución de la búsqueda.

Cuando se han introducido la cadena de búsqueda, se eligió un máximo de estudios y se han elegido las fuentes para extracción de los estudios, la herramienta cuenta con un botón que inicia la búsqueda en las fuentes seleccionadas este botón también permite detener una búsqueda e iniciar una nueva con distintos parámetros.

4.2.3 Calidad y revisión de los estudios seleccionados.

La evaluación de los estudios se lleva a cabo en base a la priorización de estudios primarios, los módulos que realizan la actividad relacionada a la calidad de los estudios son:

4.2.3.1 Resultados:

En esta pantalla la herramienta permite observar los resultados emanados de una búsqueda, los resultados se muestran:

- Todos: muestra todos los resultados encontrados.
- Por biblioteca Digital: mediante pestañas que muestran los resultados encontrados en una biblioteca en específico.

Los campos que se muestran de cada estudio cuando las bibliotecas y los documentos permiten extraerlos son: a) Título, b) Autor, c) Año, d) Área, e) Palabras clave, f) Lenguaje, g) Prioridad, h) Abstract.

4.2.3.2 Evaluación de la calidad de los estudios:

El sistema permite mostrar el resumen (Abstract) de cada documento mediante la opción ver para constatar de manera visual, si el estudio corresponde a la prioridad asignada. La selección de estudios primarios basada en la prioridad del documento encontrado y la evaluación mediante la lectura y la visualización del abstract, si estos dos criterios coinciden a la vista del investigador el estudio se puede considerar primario o secundario en caso de no cumplir con las expectativas del proyecto de la revisión.

4.2.4 Extracción de la información.

El módulo base que realiza las funciones de extracción y síntesis de la información de los estudios de esta actividad es el de administración de artículos primarios que se describe enseguida.

4.2.4.1 Módulo de administración de artículos primarios:

La herramienta permite ordenar y mantener un registro de los artículos primarios del proyecto, el ordenamiento se puede mostrar por los siguientes criterios:

- Por archivo asignado.
- Sin archivo asignado.
- Por plantilla.

Identificado un estudio como primario, esta funcionalidad permite dos opciones:

- Descargar el artículo de forma automática.
- Agregar artículo de forma manual.

4.2.5 Documentos de extracción de datos y ejecución de la extracción.

herramienta ofrece un sistema de creación y administración de plantillas estandarizadas para reportar la información de estudios, se toman en cuenta para la conformación de una plantilla los siguientes puntos:

- Una plantilla puede contener la información de un documento o de varios documentos.
- Los campos de la plantilla son personalizados, es decir el investigador los puede definir.
- El sistema de Administración de plantillas permite la creación, modificación y eliminación de plantillas.
- El sistema de Administración de plantillas permite la visualización, modificación, eliminación y exportación de los artículos que conforman una plantilla.

4.3 Reporte de la revisión

Para esta tercera fase, la herramienta permite la elaboración de reportes en base a las plantillas en conjunto con el módulo de análisis que permite realizar gráficas estadísticas. El objetivo es proveer a los investigadores un sistema que les permita el análisis de los estudios que han seleccionado y obtener

resultados en base a las plantillas de reporte que se han creado. Para agregar una gráfica primero se selecciona la plantilla de la cual se va a graficar el criterio u criterios que se van a evaluar, por defecto se selecciona la opción general y en este caso se grafican los elementos de todas las plantillas existentes, también es posible seleccionar solo una plantilla, automáticamente se muestran los campos generales de las plantillas, aunque también se puede seleccionar de manera manual los criterios personalizados en cada plantilla. La visualización de las gráficas se realiza en el mismo módulo mediante la opción ver y para el control de documentación de la investigación también se pueden exportar a un documento .docx, al igual que los reportes de las plantillas.

4.3.1 Sistema de referencias

El sistema cuenta con un módulo de adición de citas bibliográficas, las cuales se integran a las plantillas de reporte o si se desea se pueden importar desde el menú herramientas lo que permite exportar las citas bibliográficas en varios formatos, entre los cuales se encuentran entre otros:

- IEEE
- AMERICAN MEDICAL ASOCIATION (AMA).
- Hardvard Reference format 1 (autor-fecha).
- Nature.
- American Sociological Association.

5. Caso de estudio

Actualmente la herramienta se aplica para realizar la revisión sistemática en la investigación de 4 tesis de estudiantes en el Centro de Investigación en Matemáticas CIMAT Unidad Zacatecas, con lo cual se espera tener una visión del comportamiento de la herramienta. Hasta el momento se han introducido

cadenas de búsqueda en la herramienta para obtener archivos de las diversas fuentes. Con ésta se obtiene un número de estudios, asi como, sus datos y archivos y una prioridad asignada según el criterio de la herramienta, en contraste para comprobar su efectividad se introduce la misma cadena en cada una de las bibliotecas digitales en donde se encontraron los estudios en las diferentes bases de datos que maneja la herramienta y se ha constatado la efectividad pues se logran detectar los mismos estudios y con similar prioridad lo que indica que los criterios aplicados son correctos en un alto porcentaje. Cabe resaltar, que las busquedas manuales que se han realizado en forma manual se han reducido hasta en un 70% de cuando se hacía manualmente, debido a la automatización de las búsquedas, descargas y sistema de monitoreo que permite la herramienta.

6. Conclusiones y Trabajos Futuros

El uso de la herramienta agiliza el proceso de realización de una revisión sistemática, hasta este momento con las pruebas que se han realizado, las búsquedas con la herramienta proveen resultados factibles y acordes con las búsquedas manuales. La herramienta brinda un soporte idóneo para la elaboración de reportes, se han buscado herramientas similares pero ninguna realiza las funciones que la aplicación objeto de esta investigación ejecuta. En conclusión el equipo de trabajo de la investigación de desarrollo de la herramienta espera colaborar con este elemento en la investigación para facilitar el trabajo de los investigadores y mejorar el uso de fuentes literarias, final el criterio del investigador aunque Como trabajo futuros implica probar la aplicación en más casos de estudio para constatar su efectividad y tener un rango más amplio de resultados del uso de ésta, se planea de acuerdo a los resultados obtenidos una nueva versión que pueda incluir mejoras y correcciones que se puedan detectar, ya está en proceso del desarrollo de un plugin que permita integrar la herramienta con los distintos procesadores de texto para ampliar la funcionalidad y uso de la herramienta, aunque en un principio sea sólo en la inserción de citas bibliográficas, al final se buscará una mayor interacción con herramientas para brindar al investigador un marco de trabajo ideal.

Referencias

Espa, R. (2006). Española de Calidad e Ingeniería del Software.

Group, S. E. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering.

Jørgensen, M., & Shepperd, M. (2007). A Systematic Review of Software Development Cost Estimation Studies, 33(1), 33–53.

Kitchenham, B. A., & Dybå, T. (2004). Evidence-based Software Engineering.

Ltd, M. (2013). Mendeley Ltd. Retrieved November 01, 2013, from http://www.mendeley.com/

Mian, P., Conte, T., Natali, A., Biolchini, J., & Travassos, G. (n.d.). A Systematic Review Process for Software Engineering.

RefWorks. (2013). Refworks. Retrieved October 30, 2013, from http://www.refworks.com/

Team, J. D. (2013). JabRef[desktop]. Retrieved November 02, 2013, from http://jabref.sf.net

Thomson, R. (2013). EndNote[desktop]. Reuters Thomson. Retrieved from http://endnote.com

Notas biográficas:

Edgar Felipe Uribe Rayas es Ingeniero en Sistemas Computacionales, egresado de la Universidad Politécnica de Zacatecas (UPZ), actualmente estudia la Maestría en Ingeniería del Software en el Centro de Investigación en Mátemáticas (CIMAT) Unidad Zacatecas. Su interés es el desarrollo de

software basado en procesos, modelos de calidad, la seguridad informática y la arquitectura de software, ha desarrollado varios sistemas de software y sitios web mediante el uso de varios lenguajes de programación.

Jorge Alexis Márquez del Real es Ingeniero en Sistemas Computacionales, egresado de la Universidad Politécnica de Zacatecas (UPZ), actualmente estudia la Maestría en Ingeniería del Software en el Centro de Investigación en Mátemáticas (CIMAT) Unidad Zacatecas. Su interés es la y arquitectura de

software, la aplicación de modelos matemáticos a los procesos de software, ha desarrollado varios sistemas de software y sitios web mediante el dominio de varios lenguajes de programación.

J. Guadalupe Uribe Dévora es Ingeniero en Sistemas Computacionales, egresado de la Universidad Politécnica de Zacatecas (UPZ), actualmente estudia la Maestría en Ingeniería del Software en el Centro de Investigación en Mátemáticas (CIMAT) Unidad Zacatecas. Su interés es el desarrollo de

software basado en procesos, la arquitectura de software y los modelos de calidad, ha desarrollado varios sistemas de software y sitios web mediante el dominio de varios lenguajes de programación.

Claudia Valtierra Alvarado es una Ingeniero en Sistemas Computacionales, egresada de la Universidad Politécnica de Zacatecas (UPZ), actualmente estudia la Maestría en Ingeniería del Software en el Centro de Investigación en Matemáticas (CIMAT) Unidad Zacatecas. Su interés son los procesos de

software, el estudio de la caracterización de organizaciones de desarrollo y el diseño web, ha desarrollado varios sistemas de software y sitios web mediante el dominio de varios lenguajes de programación.

Jezreel Mejia Miranda es doctor en Informática por la Universidad Politécnica de Madrid (UPM), España, donde se le concedió la nota máxima, Cum Laude, y mención de "Doctorado Europeo". Realizó una estancia de investigación para obtener el doctorado europeo en la Universidad Fernando Pessoa en

Oporto, Portugal. Previamente, en el Instituto Tecnológico de Orizaba, Veracruz, cursó la maestría en Ciencias de la Computación y la licenciatura en Informática. Es miembro del grupo de investigación Cátedra de Mejora de Procesos Software en el Espacio Iberoamericano (MPSEI), donde participa en proyectos internacionales de investigación con entidades educativas y de gobierno (Instituto Tecnológico de Orizaba; Instituto Regional de Zacatecas; Facultad de Informática de la UPM) y de vinculación con la industria (clúster de empresas de desarrollo de software en Zacatecas). Asimismo, es miembro del comité científico de diversos congresos internacionales como: CISTI (2009-2013), CERMA (2009-2013), del Coloquio de Investigación Multidisciplinaria del Instituto Tecnológico de Orizaba (2011) y del Infonor Chile 2012 y de la revista

internacional RISTI (2010-2013). Ha publicado diversos artículos técnicos en temas relacionados con la gestión de proyectos, entornos multi-modelo, modelos y estándares de calidad y temas relacionados en entornos outsourcing. También ha participado en proyectos de la empresa multinacional everis consulting. Actualmente, el Dr. Jezreel Mejía Miranda es investigador del Centro de Investigación en Matemáticas, A.C. (Cimat), Unidad Zacatecas, en el área de Ingeniería de Software. También forma parte del equipo oficial de traducción al español del libro CMMI-DEV v1.2 y 1.3, versiones reconocidas por el prestigioso Software Engineering Institute (SEI) de la Carnegie Mellon University. Como investigador, sus áreas de interés son: entornos multimodelo, gestión de proyectos software, modelos y estándares de calidad (CMMI, ISO, TSP, PSP, etc.), metodologías ágiles, métricas, mejora de procesos en entornos outsourcing y entornos de desarrollo tradicional. Cuenta con certificación en CMMI e ISO 20000.



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 2.5 México.