

<b>Descripción del anteproyecto</b> <b>Licenciatura en Ingeniería de Software</b> <b>Facultad de Estadística e Informática</b>	<b>Junio</b> <b>2024</b>
--	-----------------------------

Xalapa, Veracruz, a 7 de Junio de 2024

**PROYECTO DE TITULACIÓN PARA REGISTRO.**

Cuerpo Académico	Ingeniería y Tecnología de Software
Nombre del proyecto de Investigación VINCULACIÓN/PLADEA-FEI	Ingeniería de Software e Inteligencia Artificial
LGAC que alimenta	LGAC 1. Gestión, modelado y desarrollo de Software
Línea de Investigación	Administración de proyectos
Duración Aproximada	12 meses
Modalidad de Trabajo Recepcional	Tesis
Nombre del trabajo recepcional	Comparación de técnicas de clustering para el reportes de fallos.
Requisitos	Capacidad de abstracción, lectura de artículos en idioma inglés, interés en el área de inteligencia artificial.
<b>RESPONSABLE DEL TRABAJO RECEPCIONAL.</b>	
Director	Dr. Oscar Alonso Ramírez
Codirector	Dr. Jorge Octavio Ocharán Hernández
Alumnos Participantes	uno
<b>DESCRIPCIÓN DEL PROYECTO DE INVESTIGACIÓN</b>	
<p>Muchos de los factores que impactan en el tiempo y confiabilidad de un proyecto de Software, pueden deberse al factor humano que pudieran ser solventados desarrollando muchas tareas de manera automática y tomando decisiones autónomas, haciendo más eficiente la labor de un Ingeniero de Software y evitando el sesgo del humano. Por otro lado, la Inteligencia Artificial carece hoy en día de metodologías de desarrollo de software que permitan la construcción de sistemas de cómputo de calidad, y que se adapten al tipo de sistemas que permitan la experimentación e investigación en el área de Inteligencia Artificial. Por lo tanto, surge la necesidad de colaborar entre ambas disciplinas para fortalecer los resultados de ambas áreas de investigación, aportando las fortalezas de cada una en la otra.</p>	

**Es por ello que el presente proyecto busca desarrollar colaboración entre la Ingeniería de Software y la Inteligencia Artificial, para contribuir al desarrollo de ambas disciplinas, mediante la aplicación de técnicas de Inteligencia Artificial que aporten soluciones a problemas de procesos y del producto de software, así como la aplicación de estrategias, métodos y procesos que soporten a la investigación, desarrollo, y experimentación en el ámbito de la inteligencia Artificial.**

### **DESCRIPCIÓN DEL TRABAJO RECEPCIONAL.**

Una buena atención a los reportes de fallos es una actividad que ayuda a garantizar la calidad de los productos de software (Tran et al., 2019). Regularmente, habrá una persona encargada de analizar cada reporte y de decidir cómo se atenderá el reporte. Sin embargo, cuando la cantidad de estos reportes incrementa se pueden presentar problemas para atenderlos de la manera adecuada.

Uno de los grandes problemas en el manejo de reportes de fallos, son los reportes duplicados. (Gupta & Gupta, 2021). Estos reportes duplicados llevan al desperdicio de tiempo y recursos, por lo cual han surgido trabajos que buscan identificar los reportes duplicados para mejorar el manejo de los reportes como por ejemplo el trabajo de (Kukkar et al., 2020). Otra alternativa, ha sido la implementación de algún tipo de clasificación (Otoom et al., 2019) o de clustering (Hammad et al., 2018) para los reportes de fallos y así identificar reportes similares y ello permita asignar los recursos de una manera más eficiente. Sin embargo, el número de clases que se utiliza puede ser muy limitado para identificar distintos tipos de reportes y no todas las técnicas de clustering presentan flexibilidad al número de clusters a utilizar.

Debido a lo anterior, se propone realizar una revisión sistemática de la literatura para analizar trabajos relacionados. Se propone como términos tentativos de búsqueda “clustering”, “bug report”, “bug similarity”. Se propone investigar y analizar cuáles son las características que otros trabajos analizan en los reportes de fallo con el fin de clasificarlos, así como de los algoritmos o técnicas para extraer dichas características y el pre-procesamiento que se realiza de los atributos. Se deberá buscar trabajos que realicen técnicas de clustering para los reportes de fallos y analizar los resultados reportados para poder determinar cuáles técnicas se utilizaran para la comparación. También se deberá buscar repositorios públicos de reportes de fallos.

El objetivo de este trabajo es realizar una comparación entre distintos métodos de clustering aplicados a reportes de fallos que permitan identificar reportes similares. La selección de cuales técnicas se van a comparar y sobre qué tipo de reportes se trabajará, se decidirá después del análisis de los resultados de la revisión sistemática.

### **RESULTADOS ESPERADOS.**

Reporte de una revisión sistemática de la literatura  
Borrador de un artículo  
Documento de tesis

**BIBLIOGRAFÍA RECOMENDADA.**

Gupta, S., & Gupta, S. K. (2021). A systematic study of duplicate bug report detection. *International Journal of Advanced Computer Science and Applications*, 12(1).

Hammad, M., Alzyoudi, R., & Otoom, A. F. (2018). Automatic clustering of bug reports. *International Journal of Advanced Computer Research*, 8(39), 313-323.

Kukkar, A., Mohana, R., Kumar, Y., Nayyar, A., Bilal, M., & Kwak, K. S. (2020). Duplicate bug report detection and classification system based on deep learning technique. *IEEE Access*, 8, 200749-200763.

Otoom, A. F., Al-jdaeh, S., & Hammad, M. (2019, August). Automated classification of software bug reports. In *proceedings of the 9th international conference on information communication and management* (pp. 17-21).

Tran, H. M., Le, S. T., Nguyen, S. V., & Ho, P. T. (2020). An analysis of software bug reports using machine learning techniques. *SN Computer Science*, 1(1), 4.

Luaphol, B., Srikudkao, B., Polpinij, J., & Kaenampornpan, M. (2018, October). Assembling relevant bug report using the constraint-based k-means clustering. In *2018 International conference on information technology (InCIT)* (pp. 1-6). IEEE.

Rocha, T. M., & Carvalho, A. L. D. C. (2021). Siameseqat: A semantic context-based duplicate bug report detection using replicated cluster information. *IEEE Access*, 9, 44610-44630.

<p>_____</p> <p><b><u>Dr. Oscar Alonso Ramírez</u></b></p> <p>_____</p>	<p>_____</p> <p>_____</p>
<p><b>Nombre y Firma del Director del Trabajo</b></p>	<p><b>Nombre y Firma del Codirector del Trabajo</b></p>
<p><b>Vo. Bo.</b></p> <p>_____</p> <p><b>Dr. Ángel Juan Sánchez García</b> <b>Responsable de CA-ITS</b></p>	<p><b>Vo. Bo.</b></p> <p>_____</p> <p><b>Dr. Jorge Octavio Ocharán Hernández</b> <b>Coordinación de Academia de</b> <b>Experiencia Recepcional</b></p>

**NOTAS:**

- 1) Casos excepcionales serán evaluados por la Academia de ER.
- 2) Tratando de un CA externo a la Licenciatura en Ingeniería de Software, el proyecto deberá llevar el aval de los CA de la misma que se asocie con el tema.
- 3) El Vo. Bo. del Responsable de CA se obtiene en la reunión de cada CA, donde se presentan los temas del mismo para su aprobación.
- 4) El Vo. Bo. de la Coordinación de ER se obtiene en una reunión de la academia que se programa para ello.