PROGRAMA DE INCENTIVOS PARA LA FORMACIÓN DE DOCENTES-INVESTIGADORES

Nombre del programa de posgrado:

Maestría en Informática con Énfasis en Innovación e Investigación.

Categorización PRONII: **No**

Nombre de la Institución: Universidad Católica Nuestra Señora de la Asunción

Vinculación a Proyectos I+D: **No**

Nombre del beneficiario: **Néstor Fabián Riveros Godoy**

Vinculación docencia, tutoría o centro de investigación: **Dr. Carlos Rodríguez**

- Publicaciones realizadas durante el programa:

An Early Alert System for Software Vulnerabilities based on Vulnerability Repositories and Social Networks." The 47th Latin American Computing Conference (CLEI 2021), October 2021, San Jose, Costa Rica.

- Presentación en el evento SegurInfo Paraguay 2021

Título de tesis:

Alertas tempranas sobre vulnerabilidades del software a partir de Twitter y fuentes CVE, con análisis de avisos del día cero para vulnerabilidades no documentadas

RESUMEN

La cantidad de información existente sobre vulnerabilidades del software, la diversidad y heterogeneidad de las fuentes de consultas, y la poca conciencia de los usuarios en cuanto a la importancia de las actualizaciones de seguridad del software, hacen que el riesgo de ataques cibernéticos sea muy alto. En este trabajo, se aborda la problemática de las alertas tempranas sobre vulnerabilidades del software aprovechando información existente en redes sociales y registros oficiales de vulnerabilidades. Para ello, se propone la creación de una solución basada en la recuperación automatizada de información sobre vulnerabilidades a partir de dichas fuentes utilizando como marco las preferencias del usuario, el etiquetado inteligente de vulnerabilidades y la disponibilización de dicha información a usuarios finales interesados en vulnerabilidades.

OBJETIVOS

La tesis aborda los problemas planteados, mediante una solución que permite generar alertas tempranas sobre vulnerabilidades del software. Tomamos como base las fuentes de datos obtenidas de registros oficiales de vulnerabilidades y, en forma complementaria, información sobre vulnerabilidades extraída de las redes sociales. La primera fuente, provee información formalmente reportada a organizaciones encargadas de gestionar la información sobre vulnerabilidades (p.ej., CVE), mientras que la segunda permite la identificación de potenciales vulnerabilidades del día cero. Para el efecto, combinamos técnicas de recuperación de la información, NLP (del inglés, Natural Language Processing) y etiquetado automático e inteligente de vulnerabilidades. El empleo de estas técnicas permite identificar, extraer y clasificar vulnerabilidades en forma conveniente, de manera a permitir una búsqueda y descubrimiento más efectivos de vulnerabilidades que sean de interés para los usuarios.

Los objetivos específicos que abordamos son los siguientes:

1. Automatización de: la extracción, el análisis y la clasificación de la información sobre nuevas vulnerabilidades, utilizando la red social Twitter como sensor de avisos y las fuentes de CVE como referencias de contenido documentado.

2. Utilizamos la red social Twitter para descubrir avisos de alertas no documentadas y que se generan día a día como vulnerabilidades del día cero.

3. Abordamos la problemática de la heterogeneidad de los términos utilizados en el ámbito del software y las vulnerabilidades (p.ej., el CMS Wordpress es típicamente también referido como WP), para así facilitar la búsqueda de información relativa a las mismas.

4. Diseñamos una solución centrada en el usuario, cuya experiencia de uso sea lo más sencilla posible, para que se pueda etiquetar preferencias específicas que nos ayuden a conocer la realidad tecnológica referente al software de interés, para proveer alertas sobre vulnerabilidades que afecten exclusivamente a un entorno de software definido.

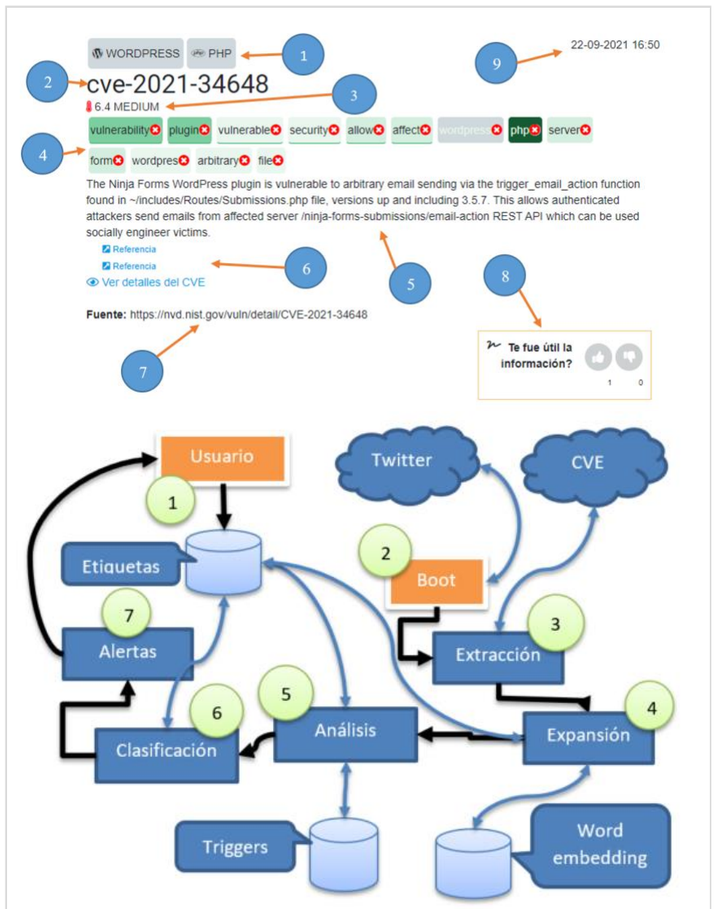


Gráfico que representa los principales componentes del sistema

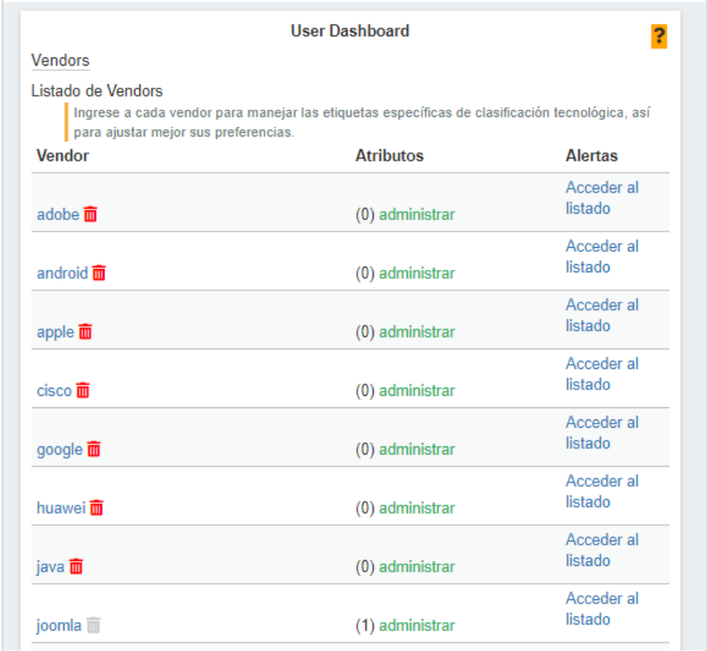
APORTES DE LA INVESTIGACIÓN

El sistema permite proveer información útil y relevante sobre vulnerabilidades del software en el menor tiempo posible y con el menor esfuerzo de parte del usuario, para con esa información obrar en consecuencia de acuerdo con el nivel de gravedad de la vulnerabilidad detectada automáticamente.

ACTIVIDADES REALIZADAS

Se desarrolló completamente una aplicación online que se puso a disposición de los usuarios y se realizaron dos validaciones con usuarios representativos y con perfiles técnicos del área de ciberseguridad.

La primera respecto a la usabilidad de la herramienta, y la segunda para validar la calidad de los resultados obtenidos, para ello se contó con la participación y las pruebas de personal del MITIC.



Ejemplo de tecnologías analizadas

Fotografía en las instalaciones del MITIC, el uso de la herramienta en pantalla gigante.

Ejemplo de una vulnerabilidad detectada en la herramienta

RESULTADOS OBTENIDOS

En la validación de usabilidad surgieron cuestiones que se fueron mejorando en cada iteración. En la segunda validación los resultados demostraron que la herramienta era capaz de detectar y clasificar automáticamente las vulnerabilidades, y proveer al usuario en el tiempo oportuno los avisos, incluyendo el nivel de gravedad del fallo en el caso de las vulnerabilidades documentadas.

CONCLUSIÓN

Inicialmente queríamos que la herramienta sea utilizada por usuarios de todos los niveles de conocimiento, pero se concluye que esta herramienta está enfocada únicamente a usuarios con conocimientos técnicos del área de ciberseguridad. Respondemos a la pregunta de investigación con la siguiente afirmación: efectivamente sí se puede clasificar la información, nuestra herramienta logra descubrir y clasificar vulnerabilidades del software, y provee la alerta temprana al usuario final en tiempo y forma, para de esta manera lograr, no solo un mayor conocimiento de lo que afecta a su entorno tecnológico real, sino en el tiempo justo, para de esa manera, obrar en consecuencia.

VISIÓN Y PLANES FUTUROS

Proponemos incorporar las mejoras sugeridas por los usuarios. Se evidenció un consumo importante de procesador respecto a cálculos y procesamiento, especialmente a nivel Base de Datos, el prototipo se desarrolló y se puso en funcionamiento asumiendo solamente algunas tecnologías importantes y referenciales del mercado tecnológico paraguayo, por lo tanto, faltan incorporarse otras más y realizar el aprendizaje de texto embebido de más tecnologías, además, se debe depurar aún más las semillas descubiertas y curadas. También se deben incorporar más usuarios referentes del área de ciber tecnologías de las Redes Sociales, en este caso Twitter, según pudimos averiguar entre nuestros usuarios de prueba, serían aproximadamente 200 usuarios en el idioma inglés y unos pocos en el idioma español.

“Este programa de posgrado fue cofinanciado por el Consejo Nacional de Ciencia y Tecnología - CONACYT con recursos del FEEI”