Universidad Católica
"Nuestra Señora de la Asunción"
Facultad de Ciencias y Tecnología
Departamento de Ingeniería Electrónica e Informática

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

Autor: Néstor Fabián Riveros Godoy

Tutor: Carlos Rodríguez

Proyecto Final de Máster

para optar por el título de:

Mágister en Informática con Énfasis en Investigación e Innovación

Asunción-Paraguay

Octubre 2021

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, proponemos 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. Estudios de usuarios demuestran la viabilidad de la propuesta como herramienta útil para la gestión de alertas tempranas sobre vulnerabilidades del software.

Palabras claves: vulnerabilidad del software, etiquetado inteligente, alertas tempranas, procesamiento de lenguaje natural.

Agradecimientos

"El tiempo es la cosa más valiosa que una persona puede gastar"

Theophrastus

Agradecimiento a mí familia por su inmenso apoyo. A los docentes por su paciencia, capacidad y compresión. Y un especial agradecimiento a mi tutor, el Dr. Carlos Rodríguez, por ser un guía inconmensurable en este trabajo. Fueron extensas horas de reuniones y debates que enriquecieron el proceso de este trabajo para llegar al objetivo.

Néstor Fabian R.G.

Con el apoyo de:

Fondo para la Excelencia de la Educación y la Investigación

"La Maestría en Informática con Énfasis en Investigación e Innovación, Código POSG17-93 es cofinanciada por el Consejo Nacional de Ciencia y Tecnología – CONACYT, con recursos del FEEI"

Institución ejecutora del programa: Universidad Católica "Nuestra Señora de la Asunción"

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

De un tiempo a esta parte, las empresas han aumentado su dependencia del *software*, el cual ayuda a mejorar la eficiencia de sus procesos. Esto, sin embargo, las lleva a ser más vulnerables al riesgo de sufrir ataques cibernéticos que podrían resultar en graves daños y pérdidas económicas [7]. En este contexto, el aumento del home office a causa del COVID-19[10] (*El COVID se refiere a la pandemia del 2020 hasta la fecha de este documento, ocurrida por la enfermedad respiratoria muy contagiosa causada por el virus SARS-CoV-2*), trae consigo un nuevo desafío al expandirse la conectividad fuera de las empresas, pues los sistemas de *software* no actualizados con los últimos parches de seguridad, las contraseñas débiles, la falta de protección contra virus y malware, y la ausencia de políticas de seguridad informáticas claras, hacen que los dispositivos y sistemas de *software* utilizados para el efecto sean especialmente vulnerables y podrían, por lo tanto, exponer a las empresas a algún tipo de ciberataque[1]. Cuando las empresas crearon sus sistemas tecnológicos, asumieron que la mayoría de los empleados estarían dentro de su perímetro de defensa; pero estos modelos de control con el Covid-19 y la cuarentena cambiaron sustancialmente.

En la Tabla 1.1, presentamos el análisis y una breve explicación de nuestras fuentes de consultas en línea para acreditar las mismas. La finalidad de esta tabla es validar la información que utlilizamos y sustentar la importancia de las fuentes consultadas.

Para comprender el impacto de la ciberseguridad y sus potenciales inconvenientes, nos pareció importante comentar sobre una vulnerabilidad destacada del año 2018, conocida como EternalBlue[2], el cual causó graves problemas en el protocolo Server Message Block (SMB) de Microsoft. Uno de los ataques más famosos de su uso fue el ***Wanna Cry***[3], el cual es un ransomware, que a su vez es un malware que básicamente encripta o bloquea el acceso a los datos del sistema que afecta y pide rescate a la

[1] https://www.rdstation.com/mx/blog/home-office/
[2] https://es.wikipedia.org/wiki/EternalBlue
[3] https://www.kaspersky.es/resource-center/threats/ransomware-wannacry

Proveedor	Descripción
iProUP	Es un portal de noticias del mundo digital. Es el primer sitio argentino en abarcar temáticas como startups, innovación, criptomonedas, incubadoras, fintech y blockchain. Tiene una antigüedad de más de 3 años en el mercado digital, y con más de 24mil referencias al sitio según estadísticas públicas de Google.
Cybersecurity Ventures	Es el investigador líder mundial y página número uno de consulta sobre temas referentes a la economía cibernética global. Además es una fuente confiable de datos, cifras y estadísticas de ciberseguridad. Tiene más de 6 años en el mercado digital, y con más de 40mil referencias hacia su sitio según estadísticas públicas de Google.
Panda Security	Es una empresa española fundada en 1990, especializada en la creación de soluciones de seguridad informática. Centrada inicialmente en la creación de un programa antivirus, la compañía ha ampliado sus objetivos expandiendo su línea de negocio hacia los servicios de ciberseguridad avanzada con tecnologías para la prevención del cibercrimen.
Kapervsky	Fundada en Rusia en 1997. Es una compañía internacional dedicada a la seguridad informática con presencia en aproximadamente 195 países del mundo.
ESET NOD32	Es un programa de antivirus desarrollado por la empresa ESET, de origen eslovaco, su data de fundación es 1987.
WeLiveSecurity	El cual es una creación de ESET –experimentados investigadores con profundo conocimiento de las últimas amenazas y tendencias en materia de seguridad de la información. Es una fuente editorial de noticias de seguridad en Internet, opiniones y análisis, cubriendo alertas y ofreciendo tutoriales, videos y podcasts. El sitio busca satisfacer a todos los niveles de conocimiento, desde programadores aguerridos hasta personas buscando consejos básicos para asegurar su información en forma efectiva, según Google existen más de 400mil referencias al sitio.

Tabla 1.1: Listado de sitios online consultados

víctima para recuperar su información, el mismo afectó a más de 300.000 empresas en todo el mundo y generó costes totales de alrededor de 4 mil millones de dólares en pérdidas. Cabe destacar que esto se podría haber evitado, pues al momento del ataque ya existían los parches de seguridad para dichas vulnerabilidades, los cuales fueron publicados meses antes de los incidentes[1].

Microsoft, proveedor de dicho protocolo, en ese entonces deslindó responsabilidades, alegando que los parches estuvieron disponibles en tiempo y que fueron las empresas las responsables por no actualizar su *software*[2]. Con esta anécdota, se puede constatar, que las consecuencias de no aplicar los parches de seguridad pueden ser catastróficas. En este contexto, en un estudio realizado por Beattie et al.[6] se analiza la problemática de las actualizaciones del *software* y cuándo aplicar los parches de seguridad al sistema. El artículo explica la disyuntiva entre la aplicación de una actualización de seguridad que pudiera no ser aún estable y el riesgo de que el *software* sea comprometido debido a la espera prolongada por una actualización estable.

Un dato no menos importante de mencionar respecto al ransomware ***Wanna Cry***, según el informe denominado "*Estado de la Ciberseguridad en Paraguay - Año 2020*"[3] publicado por el **Cert-Py** (Centro de Respuestas ante Incidentes Cibernéticos del Paraguay) hasta hoy en día, *Wanna Cry* sigue siendo el ransomware con mayor actividad y con más detecciones hechas por *Kaspersky*, tanto a nivel global como también específicamente en el Paraguay. Por lo tanto, se puede comprender, que este *malware* aún sigue muy vigente y las alertas de seguridad respecto a este malware no cesan.

Según un artículo publicado por *Cyber Security Ventures*[4], se estima que para el 2025, el coste del cibercrimen alcanzará los 10.5 billones de dólares anuales a nivel mundial. Además, según *Welive security*[5], las vulnerabilidades en el *software* y *hardware* de los productos tecnológicos son elementos, que con frecuencia, se identifican en los incidentes de seguridad. Tales vulnerabilidades son aprovechadas por los exploits, que son códigos especialmente diseñados para explotar dichas fallas.

En otro artículo online, publicado por *Panda Security*, se menciona que las empresas **no tienen bien establecidas las políticas de parchado del *software***[6], por lo que no siempre es prioritaria la búsqueda diaria de vulnerabilidades del *software* dentro de las empresas. En otras palabras, hay poca conciencia sobre la amenaza que representan las vulnerabilidades del *software* y la importancia de las actualizaciones de seguridad. Por otro lado, el reporte publicado por el *Banco Interamericano de Desarrollo* (BID), denominado *"Riesgos, avances y el camino a seguir en América Latina y el Caribe"* [7],

[1] https://medium.com/@rootedshell/ms17-010-eternalblue-57692e719768

[2] https://www.adslzone.net/2017/05/13/microsoft-culpa-las-empresas-del-ataque-son-responsables-por-no-actualizar-windows/

[3] https://bit.ly/2YiGAyq

[4] https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/

[5] https://www.welivesecurity.com/la-es/2020/01/30/vulnerabilidades-exploits-reportados-2019/

[6] https://www.pandasecurity.com/es/mediacenter/seguridad/consecuencias-no-aplicar-parches/

[7] https://bit.ly/3uujbpn

se menciona el importante crecimiento en la región del interés sobre aspectos relacionados a la seguridad del *software* y el cibercrimen, lo cual podría interpretarse como un aumento (*aunque no lo suficiente*) en la consciencia de las personas y empresas sobre las consecuencias en torno a estas cuestiones.

Ciertamente, el aspecto de la seguridad del *software* se ha vuelto un desafío mayor en los últimos años, ya que la cantidad de vulnerabilidades, y ataques siguen creciendo a un ritmo constante, las contramedidas no son suficientes. En este contexto, las fuentes de información de ciberseguridad son elementos fundamentales en la operación diaria de los profesionales de la seguridad. Entre las fuentes más importantes se encuentran los **CVE** (del inglés *Common Vulnerabilities and Exposures*), mediante los cuales se documentan las vulnerabilidades del *software* oficialmente reportadas por profesionales de la seguridad (*p.ej., hackers éticos*).

En la Figura 1.1[1], se puede observar un gráfico estadístico de la cantidad de vulnerabilidades documentadas del *software*, reportadas desde el 2010 hasta el 2020 a nivel mundial. Esta figura denota una clara tendencia hacia un aumento de las vulnerabilidades reportadas anualmente, en particular, durante los últimos años. Adicionalmente a las fuentes oficiales sobre vulnerabilidades (*p.ej., CVE*), existe otra categoría de vulnerabilidades conocidas como vulnerabilidades del día cero (del inglés, zero-day o 0-day vulnerability). Éstas se refieren a vulnerabilidades que son desconocidas por personas u organizaciones a quienes afectan y quienes deberían estar interesadas en mitigarlas.

Por otro lado, las últimas noticias a nivel mundial hablan de que el 2020 no solo estuvo marcado por la pandemia, sino por sofisticados ataques del cibercrimen, entre estos artículos destacados encontramos uno que analiza *"los 10 peores ataques registrados en el 2020"*[2], y otro cuyo título es *"Los ciberataques que marcaron el 2020"*[3]. En los cuales se evidencia el nivel de exposición en el que se encuentran las empresas, y como dato no menor se menciona que el 70 % de los ataques fueron por motivos económicos y el restante con fines de espionaje.

A todo esto hay que considerar que hoy en día, las empresas deben redoblar esfuerzos para garantizar la protección de datos de los usuarios en los sistemas, pues existen las reglamentaciones globales denominadas *Reglamentaciones Generales de Protección de Datos*[4] o directamente conocidas por sus siglas RGPD, y más específicamente el punto focal se refiere al impacto que podría tener la pérdida total o parcial de los datos personales en un sistema comprometido, por ello, es aún más importante tener alertas tempranas de vulnerabilidades del *software* para obrar en consecuencia, y de esa manera mitigar algún impacto negativo que podría ocurrir.

[1] https://bit.ly/3AT2tTb
[2] https://www.muycomputer.com/2020/12/30/ciberseguridad-en-2020/
[3] https://www.itmastersmag.com/noticias-analisis/los-ciberataques-que-marcaron-el-2020/
[4] https://bit.ly/3kWBfFO

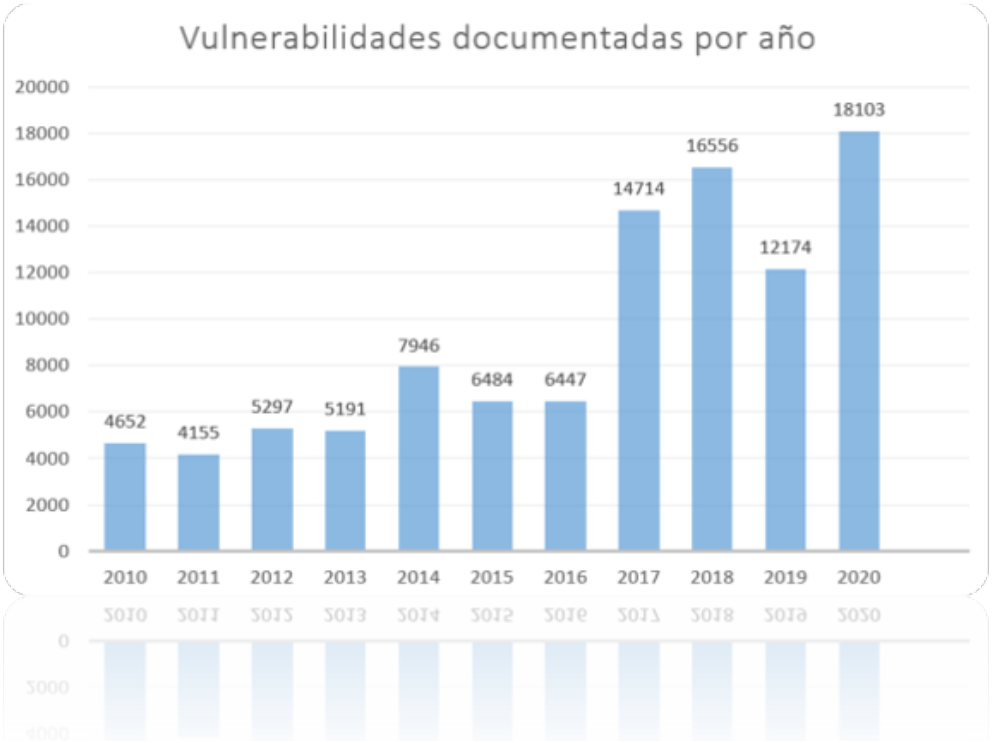

Figura 1.1: Vulnerabilidades reportadas desde el 2010 hasta el 2020

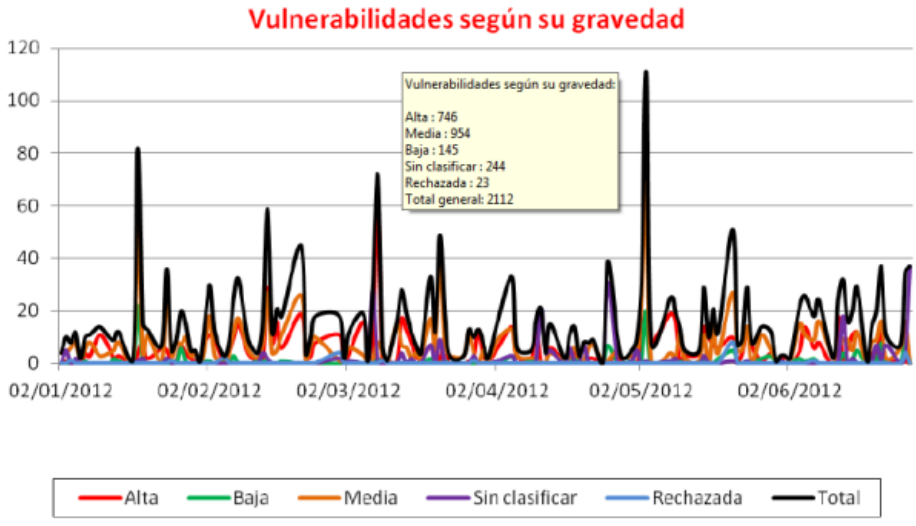

Figura 1.2: Vulnerabilidades por gravedad

1.1. La problemática

En el proceso de descubrimiento de vulnerabilidades documentadas del *software*, hemos identificado una serie de operativas comunes al momento de detección del fallo en un sistema informático, se enfatiza que lo primero que debería hacerse es aplicar los parches de seguridad y revisiones sugeridas según los CVE correspondientes. Ahora bien, muchas vulnerabilidades pueden tener un nivel de gravedad bajo por lo que se podría retrasar la atención; en la Figura 1.2[1] se puede observar un gráfico estadístico de los niveles de gravedad del fallo. En cambio, para aquellas vulnerabilidades críticas, se deben tomar las acciones pertinentes con la mayor inmediatez posible. Sin embargo, según una publicación realizada por el sitio *Welive Security*, *"la excesiva cantidad de alertas produce fatiga de seguridad*" [2], también encontramos una publicación interesante respecto al tema de fatiga en las empresas [3] o esta breve explicación en Wikipedia [4], **lo cual podría acarrear que aquellas vulnerabilidades verdaderamente críticas pasen desapercibidas**. Este inconveniente es análogo a las fatigas por alarmas presentes, típicamente, en contextos relativos a las emergencias médicas.

Independientemente de las fuentes consultadas, el descubrimiento de nuevas vulnerabilidades requiere típicamente la inversión de mucho tiempo y esfuerzo, optando típicamente por una o más de las siguientes opciones: (i) conocer la estructura del buscador de cada repositorio (p.ej., CVE o NVD (del inglés, *National Vulnerability Database*[5])) con el fin de poder encontrar información útil; (ii) estar inscripto en listas de correos dedicadas a difundir información sobre vulnerabilidades; (iii) seguir a profesionales y expertos de seguridad fiables en las redes sociales. Sumado a estos esfuerzos, se debe tener en cuenta, además, que los desarrolladores de *software* y otros profesional de las Tecnologías de Información y Comunicaciones (TIC) no siempre comprenden en profundidad los aspectos relativos a la ciberseguridad,[6] lo cual podría exacerbar la dificultad de acceso a la información relativa a vulnerabilidades del *software*.

Según el informe publicado por SANS Institute en el año 2015,*"Muchos expertos en seguridad no comprenden el desarrollo de software y la mayoría de los desarrolladores de software no comprenden la seguridad"* [7], no es un informe menor pues se evidencia aún más la problemática en el sector, a sabiendas que la primera línea de desarrollo debería comprender mejor estas cuestiones y estar más informados al respecto, y las alertas tempranas, con los niveles de gravedad del fallo, podrían motivar a estar más atentos. En este contexto, no solo un especialista de seguridad debe estar

[1] https://bit.ly/3AYwlxw
[2] https://www.welivesecurity.com/la-es/2016/10/11/fatiga-de-seguridad-incidentes/
[3] https://www.cytomic.ai/es/tendencias/fatiga-de-alerta-empresas/
[4] https://bit.ly/3zB0IJ2
[5] https://nvd.nist.gov
[6] https://www.sans.org/blog/2015-state-of-application-security-closing-the-gap
[7] https://www.sans.org/blog/2015-state-of-application-security-closing-the-gap/

alertado oportunamente sobre temas relacionados a las vulnerabilidades, sino también los usuarios no técnicos, desarrolladores de *software*, dueños de empresas y gerentes de TI, entre otros.

Por todo lo anterior, hemos verificado, que estar informados en el menor tiempo y con la mayor precisión posible sobre los nuevos descubrimientos de vulnerabilidades, y que afecten a un entorno de *software* específico, conduciría a optimizar el conocimiento sobre la existencia de vulnerabilidades de interés para las empresas, y mejoraría la conciencia sobre la importancia de las actualizaciones de seguridad y las contramedidas de seguridad aplicadas a los sistemas de *software* utilizados.

1.2. Objetivos generales

En esta tesis proponemos abordar los problemas, indicados previamente, mediante la propuesta de una solución que permita 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.

1.3. Preguntas de investigación

Las preguntas de investigación a responder son:

1. P1: ¿Cuál es la problemática existente al momento de buscar y clasificar la información relevante sobre vulnerabilidades del *software*?

2. P2: ¿Es posible clasificar la información sobre vulnerabilidades del software para proveer al usuario alertas tempranas sobre problemáticas que pueden afectar a sus sistemas específicos?

1.4. Objetivos específicos

Los objetivos específicos que abordaremos son los siguientes:

1. Automatizar 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. Utilizar la red social *Twitter* para descubrir avisos de alertas no documentadas y que se generan día a día como vulnerabilidades del día cero.
3. Abordar 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. Proponer 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.

Para dar respuesta a estos objetivos, se pretende diseñar un mecanismo que ayude a realizar lo siguiente: (i) monitorear y descubrir publicaciones de vulnerabilidades documentadas y no documentadas, para clasificarla de acuerdo a la gravedad, (ii) crear etiquetas de clasificación para el contenido analizado, (iii) alertar al usuario respecto a las vulnerabilidades que puedan estar afectando a su entorno, y (iv) que la interfaz de usuario cumpla los principios del diseño centrado en las personas para que la experiencia de usuario sea lo más natural y fácil posible.

1.5. Estructura de la tesis

Este trabajo se compone de 4 partes siguientes distribuidas de la siguiente manera: En el capítulo 2 se presenta el marco teórico, con los trabajos relacionados y el estado del arte, en el capítulo 3 presentamos el análisis de la problemática así como los desafíos y requisitos que se desprenden de la misma. Seguidamente, en el capítulo 4, se encuentra la propuesta y la validación, y por último en el capítulo 5 exponemos los trabajos futuros y la conclusión.

Capítulo 2

Marco teórico

2.1. Trabajos Relacionados

A continuación, discutimos los trabajos relacionados categorizándolos de la siguiente manera: (i) repositorios y servicios en línea de vulnerabilidades; (ii) descubrimiento y extracción de la información sobre vulnerabilidades; (iii) gestión y etiquetado de la información; (iv) procesamiento de Language Natural.

2.1.1. Repositorio y Servicios en Línea de Vulnerabilidades

Entre los servicios en línea sobre seguridad informática encontramos a **HISPASEC**[1] el cual provee un mecanismo de suscripción a una lista de correos para el envío diario de avisos sobre vulnerabilidades informáticas, con el propósito de divulgar y concienciar a los usuarios sobre problemáticas del sector. Según nuestro análisis, la información proveída por el sitio es muy amplia y de diversidad de contenido, y no precisamente sobre temas específicos relativos a la realidad tecnológica del usuario, y como ya hemos evidenciado en el capítulo anterior, la saturación de información podría acarrear que aquellas críticas pasen desapercibidas. En la tabla 2.1 se pueden ver otras listas de correo que cumplen similar función.

En nuestra tesis, nuestro sistema genera avisos por correo electrónico, pero se filtra y envía sólo aquella información que se alinéa a los intereses del usuario final según sus preferencias o etiquetas tecnológicas pre definidas.

Google Hacking Database[2] es un repositorio muy importante en el ámbito de la seguridad informática, el cual provee un potente motor de búsqueda para el descubri-

[1] https://hispasec.com/es/

[2] https://www.exploit-db.com/google-hacking-database

Item	Fuente	Descripción
1	securityfocus.com	SecurityFocus es un portal de noticias sobre seguridad informática en línea y proveedor de servicios de seguridad de la información. Hogar de la conocida lista de correo de Bugtraq, los columnistas y escritores de SecurityFocus incluyeron al ex fiscal de delitos cibernéticos del Departamento de Justicia, Mark Rasch, y al pirata informático convertido en periodista Kevin Poulsen.
2	us-cert.cisa.gov	Como parte de su misión, CISA lidera el esfuerzo por mejorar la seguridad, la resistencia y la confiabilidad de la infraestructura de comunicaciones y ciberseguridad. Ofrece actualización de las últimas noticias e información gratuitas a través de las redes sociales, distribución de noticias y actualizaciones gratuitas por correo electrónico para ayudarlo a mantenerse informado.

Tabla 2.1: Otras listas de correo

miento de vulnerabilidades. Según nuestro análisis, la información contenida en este repositorio es muy relevante y a la vez de alta tecnicidad. Esto implica, que el usuario final, deberá de conocer muy bien la terminología y los aspectos técnicos entorno a las vulnerabilidades. Debido al tecnicismo en el ámbito de las vulnerabilidades, debe invertir una cantidad de tiempo y esfuerzo significativos para encontrar vulnerabilidades y parches de seguridad de su interés (p.ej., en correspondencia con su realidad tecnológica).

Cert Paraguay: [1] es una iniciativa pública paraguaya a incidentes de seguridad informáticos, entre otras cosas, día a día disponibilizan un listado de vulnerabilidades descubiertas del tipo CVE, aunque al momento de escritura del documento, vemos que la lista de alertas sobre vulnerabilidades no está actualizada, por lo que posiblemente no será una fuente de consulta primaria para el profesional de seguridad, y si pretende ser una fuente de consulta diaria se debería actualizar con más asiduidad la información. Otro de sus productos proveídos, también de acceso público y gratuito, es una suscripción a una lista de correos, que al igual que otras citadas con anterioridad, proveen información muy variada y sin posibilidad de categorizar la información que se desea.

En el ámbito de soluciones y servicios de información sobre vulnerabilidades integradas, hemos identificado a **TimeSys**[2] el cual provee un servicio de suscripción

[1] https://www.cert.gov.py/
[2] https://www.timesys.com

dirigido a profesionales del área de seguridad. El servicio se centra en el monitoreo de incidentes de seguridad enfocado exclusivamente a plataformas Linux. El servicio provee un conjunto de paneles (en inglés, *dashboards*) que permiten acceder a información sobre vulnerabilidades del tipo CVE. El sitio está muy bien logrado y con una importante contribución a nuestra idea respecto al concepto de monitoreo de las vulnerabilidades, aunque consideramos como dos grandes falencias, al menos comparando con nuestra propuesta, que es de pago y que se enfoca exclusivamente a plataformas Linux, dejando de lado diversidad de otras plataformas y tecnologías. Al ser una plataforma cerrada no tenemos forma de saber qué técnicas usan para el monitoreo de las alertas y ni siquiera sabemos si utilizan procesamiento de Lenguaje Natural, además al ser de pago, no pudimos probar si permiten el tagging de preferencias. Así que nuestro análisis al sitio, se basó exclusivamente al nivel de usabilidad y calidad en general de los servicios, no así al nivel de validar los resultados, por lo tanto, no podemos ni negar ni afirmar que sea un modelo operativo similar a nuestra propuesta a nivel técnico.

Hemos identificado también otras fuentes útiles de información sobre vulnerabilidades del *software*, estas son: GitHub,[1] Information Security Stack Exchange[2] y Stack Overflow.[3] Los mismos proveen referencias técnicas que contienen una base de conocimientos muy importante sobre vulnerabilidades del *software* y ciberseguridad en general. Según nuestro análisis, estos recursos son consultados posterior al descubrimiento de una vulnerabilidad y además utilizados, más que nadie, por expertos en el tema informático-técnico. Por lo tanto, la consulta a estas fuentes se encuentra casi al final del ciclo de vida del CVE. En nuestro trabajo, estas fuentes son consultadas para el entrenamiento de terminologías para nuestro texto embebido (*word embedding*).

Las redes sociales son también fuentes importantes de información sobre vulnerabilidades del *software* [24] y hoy en día, se han vuelto la primera fuente de consulta en la búsqueda de información de lo que ocurre globalmente. Por ejemplo, en *Twitter*[4] se pueden encontrar cuentas verificadas de especialistas de seguridad informática y que día a día comparten sus conocimientos técnicos en este ámbito, por lo que actuan de sensores cruciales para estar informados sobre nuevos CVEs, así como también sobre vulnerabilidades del día cero. Sin embargo, similarmente a lo que ocurre con las listas de correos, la información no está filtrada y por ende no siempre resulta de utilidad para el contexto tecnológico del usuario que la recibe.

Finalmente, se encuentran también las fuentes de información tradicionales tales como la NVD, OWASP[5] y Mitre,[6] las cuales mantienen repositorios sobre vulnerabilidades, estándares y esquema de clasificación de debilidades del software, entre otras informaciones de utilidad.

[1] https://github.com/
[2] https://security.stackexchange.com/
[3] https://stackoverflow.com/
[4] https://twitter.com/
[5] https://owasp.org/
[6] https://cve.mitre.org/

2.1.2. Extracción de la información sobre vulnerabilidades

Para extraer la información de *Twitter* y desde las fuentes de información que documentan las vulnerabilidades del *software*, rescatamos algunos estudios relevantes tales como el trabajo de Alqahtani et al. [4]. En el mismo, los autores abordan el enfoque de modelado semántico, que aprovecha las tecnologías Web para establecer vínculos de trazabilidad entre los repositorios de avisos de seguridad y otros repositorios de *software*. Exploramos este estudio para aprovechar la propuesta semántica mencionada y así optimizar nuestra extracción de la información. En el mismo contexto, identificamos varias investigaciones referentes a diversas técnicas propuestas para mejorar la extracción del contenido y el aprendizaje automatizado [12, 21, 19, 9]. La más importante, en el contexto de nuestro trabajo, es la propuesta de Arnav et al. [12], en la que se menciona la manera de extracción de la información sobre vulnerabilidades desde el repositorio NVD.

2.1.3. Gestión y Etiquetado de Información sobre Vulnerabilidades

En cuanto a la gestión de la información, la base de conocimientos existentes fue fundamental para abordar nuestro trabajo y tomamos como referencia la investigación [22], donde los autores abordan las fuentes de información de vulnerabilidades, la heterogeneidad de las mismas y la utilización de lenguaje natural para consultarlas. También, en el mismo contexto, en la propuesta de Atymtayeva et al. [5], se explora la construcción de una base de conocimientos para expertos en seguridad. Ambas propuestas tienen relación directa en la construcción de bases de conocimientos y que es fundamental para el descubrimiento de las vulnerabilidades. De estos trabajos rescatamos las referencias de fuentes y las técnicas de exploración de la información utilizadas. En estos trabajos se enfoncan en la creación de bases de datos en donde se busca facilitar la búsqueda de la información sobre vulnerabilidades conocidas. Esto implica que los usuarios finales deben dedicar tiempo y esfuerzo en la búsqueda proactiva de vulnerabilidades relacionadas a su entorno tecnológico.

El uso de información recolectada a partir de las redes sociales ha demostrado ser de gran utilidad en contextos variados tales como la detección y predicción de desastres naturales, eventos sociales globales, entre otros contextos. Por ejemplo, Sakaki et al. [25] exploran la clasificación semántica de tweets para alertas tempranas en el ámbito de los terremotos. Mientras que Alexander [1] propone el uso de las redes sociales como un termómetro de variedad de eventos globales y como mecanismo de gestión de desastres y crisis. Si bien estos trabajos han sido realizados en contextos diferentes a la gestión de información sobre vulnerabilidades, las problemáticas de base y mecanismos utilizados guardan (en esencia) una relación con la gestión de alertas en el ámbito de las vulnerabilidades del software.

Finalmente, en el ámbito del etiquetado para propósitos de gestión de la información, Vig et al. [27] proponen un mecanismo de etiquetado de documentos basados en técnicas de aprendizaje de máquina y VSM (del inlgés, *Vector Space Model*). Dicho mecanismo contribuye a la clasificación de contenido descubierto y una mejor presentación y filtrado de la información en base a las preferencias de los usuarios finales.

2.1.4. Procesamiento de Lenguaje Natural

El uso de técnicas de NLP ha tenido un fuerte repunte en los últimos años gracias a los importantes avances recientes en el área [16]. Por ejemplo, técnicas basadas en la representación distribuida de palabras [23] ha permitido la creación de una variedad de mecanismos de procesamiento de lenguage natural para propósitos de clasificación y compresión de lenguaje natural. *Word2vec* es una de tales técnicas [18]. El mismo utiliza un modelo de red neuronal para aprender asociaciones de palabras de un cuerpo de texto en un dominio, creando un VSM con la propiedad de que las palabras semánticamente relacionadas se encuentran cercanas unas a otras en el espacio vectorial.

En el ámbito de las vulnerabilidades del *software*, Mumtaz et al. [21] proponen la incrustación de palabras (en inglés, *word embedding*) de dominio específico utilizando como corpus fuentes de datos conteniendo información relativa a vulnerabilidades del *software*. Mokhov et al. [19], sin embargo, han empleado técnicas de NLP para el análisis de código fuente con el propósito de detectar debilidades y vulnerabilidades del *software* en la fase de diseño, ciertamente podrían detectarse fallos en la codificación aplicando patrones de búsqueda, esto sigue siendo no óptimo pues no todos los fallos son vulnerabilidades realmente y podría acarrear demasiados falsos positivos. Mientras que Khazaei et al. [14] han propuesto la predicción del puntaje basado en CVSS (del inglés, *Common Vulnerability Scoring System*[1]) utilizando técnicas de NLP sobre las descripciones de vulnerabilidades.

2.2. Conclusiones del Capítulo

En este capítulo se explicaron conceptos específicos, se vieron los trabajos relacionados y el estado del arte. En general hemos encontrado importantes estudios en el área, de los cuales rescatamos importantes temas para aprovecharla dentro de nuestra propuesta. Se enfatiza el tema referente a la cantidad de información existente y que es un problema en un ambiente real, por lo tanto, también se evidencia que es importante una solución más óptima en ese ámbito, y qué todo lo hecho hasta ahora es insuficiente como para encontrar el equilibrio entre información oportuna y eficacia.

[1] https://nvd.nist.gov/vuln-metrics/cvss

Capítulo 3

Búsqueda, Clasificación y Alerta sobre Vulnerabilidades del Software

En el presente capítulo, nos enfocamos en la pregunta de investigación *P1* y exploramos la problemática de la búsqueda y clasificación de información relevante sobre vulnerabilidades del software.

En la figura 3.1 presentamos los conceptos más importantes que forman parte de la problemática de las vulnerabilidades del software y que tomamos en consideración en nuestro análisis. Como se puede apreciar, el *software* se encuentra en una posición central de interés, seguido por las vulnerabilidades que pudieran afectarlo. A continuación, nos encontramos con los *exploits* que podrían aprovechar esas vulnerabilidades para provocar un comportanmiento no deseado o planificado. En un caso ideal, un usuario interesado (p.ej., un profesional de TI o experto en seguridad) alertado oportunamente sobre una vulnerabilidad del *software*, y que afecte a su entorno tecnológico, debería parchar o realizar las acciones de mitigación pertinentes para abordar el problema, y de esta manera, evitar algún tipo de ataque al software en cuestión. Como ya vimos en el capítulo anterior, es de vital importancia, y como primer paso hacia la prevención de la materialización de ataques, la creación de conciencia en los usuarios sobre la existencia de dichas vulnerabilidades. En una gran cantidad de situaciones, la información sobre vulnerabilidades descubiertas se encuentran públicamente disponibles, tanto en las redes sociales (p.ej., Twitter) como sitios espcializados (p.ej., repositorios de CVEs). La cantidad de información sobre vulnerabilidades podría, sin embargo, ser abrumadora, por un lado, y por otro, no siempre pertinente a la realidad tecnológica del usuario.

Como primer paso, entonces, necesitamos descubrir la información existente en los repositorios de vulnerabilidades (p.ej., CVE). Sin embargo, para poder enterarnos de una nueva vulnerabilidad, indefectiblemente, debemos tener una fuente de consulta primaria que nos provea información de manera oportuna. Una alternativa es acudir a

redes sociales como *Twitter*, en la que se pueden encontrar usuarios que activamente alertan a la comunidad sobre nuevas vulnerabilidades en tiempo (cuasi) real. Optar por esta alternativa presupone un número de desafíos, entre ellos: i) estar contínuamente atentos a lo que ocurre en *Twitter*; ii) una vez alertados de una nueva vulnerabilidad ir a la fuente original para interpretar la información propia y correctamente; y iii) avisar al usuario sobre una vulnerabilidad que se ajuste exclusivamente a su entorno tecnológico (para evitar introducir *ruido* en los avisos).

En un contexto como el descrito en el párrafo anterior, la información es muy variada y por ello necesitamos clasificarla correctamente. Para ello, es necesario que el usuario interesado defina sus preferencias tecnológicas, por ejemplo, a través de una interfaz de usuario intuitiva. Una vez expresadas las preferencias del usuario, es necesario realizar la categorización de la información (p.ej., mediante etiquetado o *tagging*), que sirve como resumen de las palabras más importantes dentro de esa vulnerabilidad y además alinearlas con las preferencias expresadas por el usuario. Paralelamente a la información oficial sobre vulnerabilidades en las redes sociales, se publican también aquellas vulnerabilidades aún no documentadas y conocidas comúnmente con vulnerabilidades del día cero (*0-day*), que también son de suma importancia y que deben ser tenidas en cuenta como parte de la concienciación de usuarios sobre la existencia de vulnerabilidades.

Otro aspecto importante a tener en cuenta es el aspecto relativo a la flexibilidad en el uso de terminologías dentro del dominio, es decir, términos técnicos similares que se refieren a una misma entidad. Por ejemplo, *WordPress* es también referido como *WP*, y *XSS* como *Cross Site Scripting*. El abordaje de este flexibilidad terminológica es de vital importancia para: i) facilitar la búsqueda y recuperación de información relativa a vulnerabilidades; ii) dar flexibilidad al usuario en el uso de terminilogías y jergas propias del dominio en la expresión de sus preferencias tecnológicas; y iii) proveer un etiquetado más rico semántica y terminológicamente.

En el siguiente capítulo presentamos el diseño de la propuesta, teniendo en cuenta estos desafíos, requerimientos y funcionalidades esperadas para el abordaje de la problemática expuesta en esta tesis.

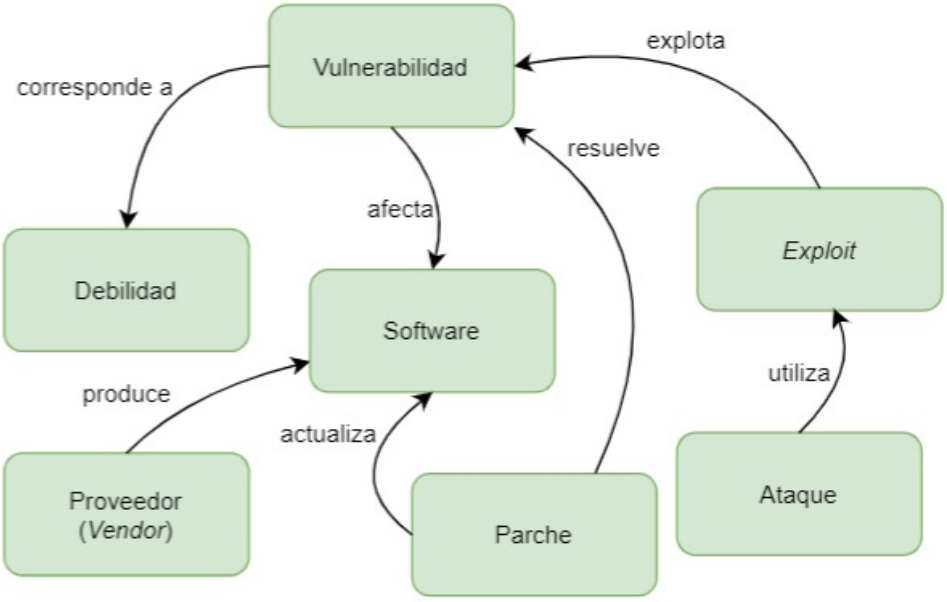

Figura 3.1: Conceptos claves en el ámbito de las vulnerabilidades del software

Capítulo 4

Diseño de la propuesta

En esta sección presentamos (i) el esquema de nuestra arquitectura y explicamos cada componente de la misma, (ii) la forma en el que realizamos la implementación, y (iii) los pasos que se hicieron para la validación de la solución propuesta.

En este contexto, el desafío inicial era identificar cuentas de usuarios de quién consumir información en *Twitter*, por lo tanto, analizamos la factibilidad con una prueba en campo, nos volcamos a una tarea diaria de búsqueda de información compartida en el microblog y lo qué es más, analizamos principalmente cómo se compartía la información sobre vulnerabilidades del *software*. Despúes de varios días de seguimiento, identificamos primariamente a unos pocos usuarios, para posteriormente ir incorporando a otros según unos criterios que se verán más adelante.

En nuestro proceso de análisis, en las redes sociales, descubrimos palabras comunes en los avisos de alertas y vulnerabilidades que se iban compartiendo, a estas palabras la denominamos *semillas* y usamos como punto de partida el estudio de Rupinder et al [13] para luego crear nuestra propia lista de aproximadamente 120 palabras *semillas*, pero a su vez, debíamos lidiar con la variación de dichas palabras en el idioma (*por ejemplo, podrían estar en plural, o conjugadas, etc.*), por ello nos propusimos hallar la manera de extraer la raíz de la misma para evitar este tipo de ambiguedades.

Una vez identificado a los usuarios confiables y, a su vez, como ya descubrimos las *semillas*, nos dimos cuenta que, además de solo aquellos usuarios confiables, habían *hashtag* referente a vulnerabilidades compartidas por otros usuarios de perfil técnico. Por lo que consideramos, no solo extraer información de aquellos usuarios identificados sino también realizar una búsqueda constante de terminos comunes. La API pública de *Twitter* permite ambas formas de uso, tanto extraer información por usuario como así también buscar palabras específicas, como una problemática específicamente del microblog *Twitter* es que se puede consumir solo 140 caractéres de cada tweet compartido.

Seguidamente vimos la necesidad de aplicar el *tagging* a las alertas descubiertas, para de esta manera, crear un resumen por cada una de ellas. En el caso de las vulnerabilidades documentadas del tipo CVE, esto sería realmente útil, pues, al contener normalmente mucha información, desplegaríamos en una primera presentación al usuario solo aquella relevante para su contexto, y dejaríamos información complementaria cuándo así el usuario tenga necesidad de ahondar. Ahora bien, con la información extraída de Twitter respecto a las vulnerabilidades no documentadas, esto podría ser un problema, pues al contener solamente 140 caracteres, realizar un resumen correctamente, se vuelve una tarea más dificil, y más aún al realizar la expansión de consulta con términos relacionados, pues al tener poca cantidad de palabras, se torna complejo realizarlo correctamente, de ahí que la información resumida no sería muy provechosa para el usuario final y el tagging generado podría contener mucho ruido para el contexto de preferencia.

Para el aprendizaje automatizado optamos por el *Word2vec*, que nos pareció uno de los algorítmos más optimos para esta propuesta, y sería esta la parte más innovadora del trabajo, para ello, analizamos la implementación en Python del método skip-gram, aunque la incorporación de nuevas palabras en una base de datos ya aprendida es una de las debilidades en dicha implementación, por ello, decidimos utilizar el mismo concepto para realizar una propia implementación en PHP con un componente distintivo, que es respecto a la incorporación dinámica del texto embebido, por lo tanto, con un aprendizaje constante e intentamos solucionar la problemática de la incorporación de nuevas palabras.

Como componente principal en nuestra propuesta y para poder clasificar la información de un contexto tecnológico específico, involucramos al usuario, quien deberá proveer al sistema sus etiquetas de acuerdo a su realidad tecnológica del software, de este modo, el sistema propuesto podrá realizar un match de acuerdo a las vulnerabilidades que se descubran dinámica y automáticamente, por lo que la información llegará al usuario y no será él quién tenga que salir a buscarla. Posteriormente, con la información descubierta, el usuario deberá aplicar los parches correctivos, las revisiones y las contramedidas de seguridad, pues la propuesta de tesis sólo contempla las alertas.

La idea central, en la cual se basa la solución propuesta en esta tesis, consiste en mantener al usuario al tanto de las vulnerabilidades existentes que afecten al entorno tecnológico de su interés. Abogamos por la posibilidad de que los usuarios puedan configurar preferencias sobre su entorno tecnológico, de modo a obtener alertas personalizadas. Para lograr este objetivo, nos basamos en técnicas de extracción de información Web y recuperación de la información a partir de redes sociales, utilizando técnicas de expansión de consultas (en inglés, query expansion). La presentación de la información es realizada utilizando mecanismos basados en *timelines* [3] junto con el etiquetado automático e inteligente [27] basado en técnicas de *word embeddings* [18], mientras que la entrega de alertas la realizamos utilizando el paradigma de *push notifications* [8].

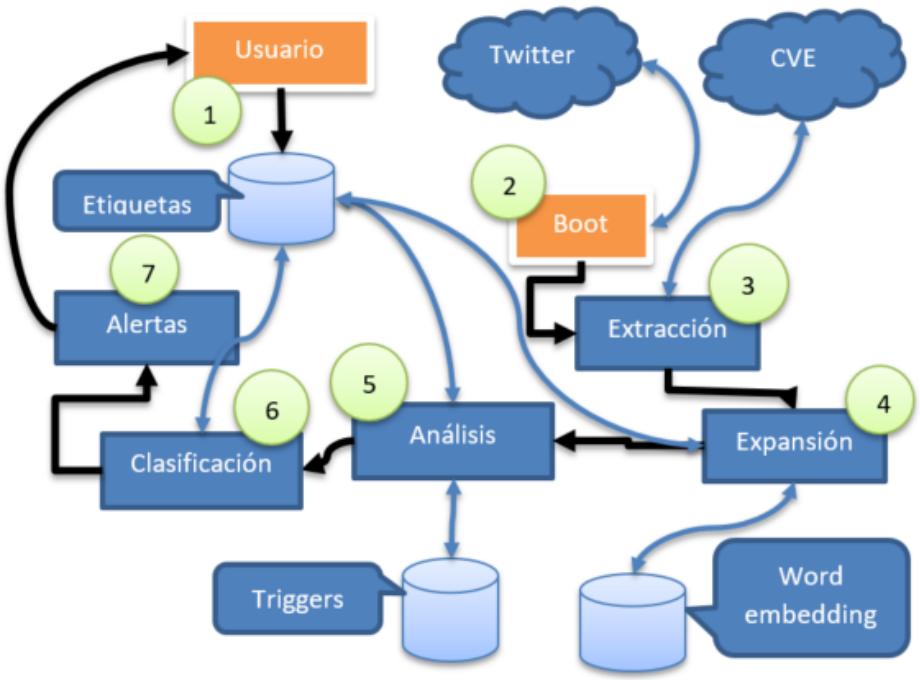

Figura 4.1: Diagrama conceptual del diseño de la propuesta desarrollada en el marco de este trabajo.

La Figura 4.1 presenta la arquitectura de la solución propuesta en esta tesis para el acceso a alertas tempranas sobre vulnerabilidades del software. Esta solución permite a los usuarios finales la definición de preferencias sobre los productos y *vendors* que se corresponden con el entorno tecnológico de interés del usuario. En base a dichas preferencias, nuestro sistema de alerta temprana permite identificar información relevante publicada en Twitter y fuentes oficiales de vulnerabilidades (CVE), y presentarlas convenientemente al usuario.

En la Figura 4.4, el *Usuario* ① define primeramente los productos (p.ej., SQL Server) y *vendors* (p.ej., Microsoft) de su interés en base a su entorno tecnológico utilizando etiquetas. Dichas preferencias son configuradas por el usuario utilizando la interfaz de usuario presentada en la Figura 4.5. Dichas etiquetas son utilizadas para la construcción de consultas (en inglés, *queries*) a ser utilizadas para la recuperación de información sobre vulnerabilidades. Debido a que tanto los productos como los *vendors* son expresados de múltiples formas (p.ej., "Wordpress" es también expresado mediante las siglas "WP" en la jerga de las TIC), es importante contar con la capacidad de recuperar información relevante inpedendientemente de la terminología utilizada para representar dicha información.

Para lidiar con esta variedad terminológica, el componente de *Expansión* ④ permite extender las etiquetas de preferencia del usuario mediante la utilización de técnicas

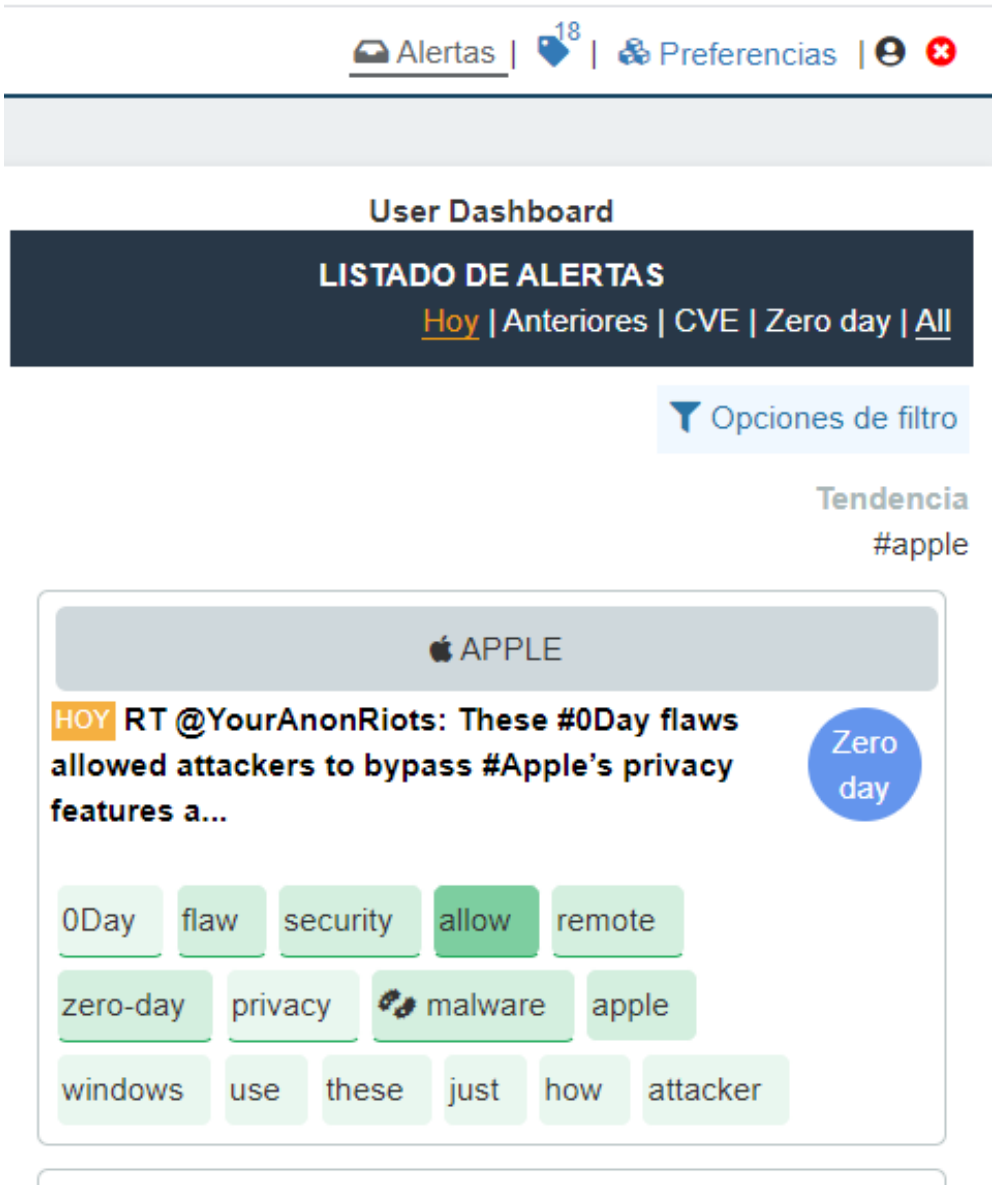

Figura 4.2: Pantalla inicial de la propuesta o user dashboard.

de *word wmbedding* [18], las cuales permiten la representación de palabras en un espacio vectorial de manera tal que la palabras semánticamente similares se encuentren cercanas unas a otras (en este trabajo, utilizamos *word embeddings* entrenados para el dominio de la informática y la ciberseguridad, como detallamos en la siguiente sección de este artículo). Por ejemplo, si el usuario selecciona la etiqueta "Microsoft Internet Explorer", el componente de expansión genera palabras adicionales y relacionadas para enriquecer la consulta, tales como "Internet Explorer", "Explorer" y "IE". La utilización de consultas expandidas de esta manera permite el aumento del *recall* al momento de recuperar información relacionada a las preferencias del usuario [17].

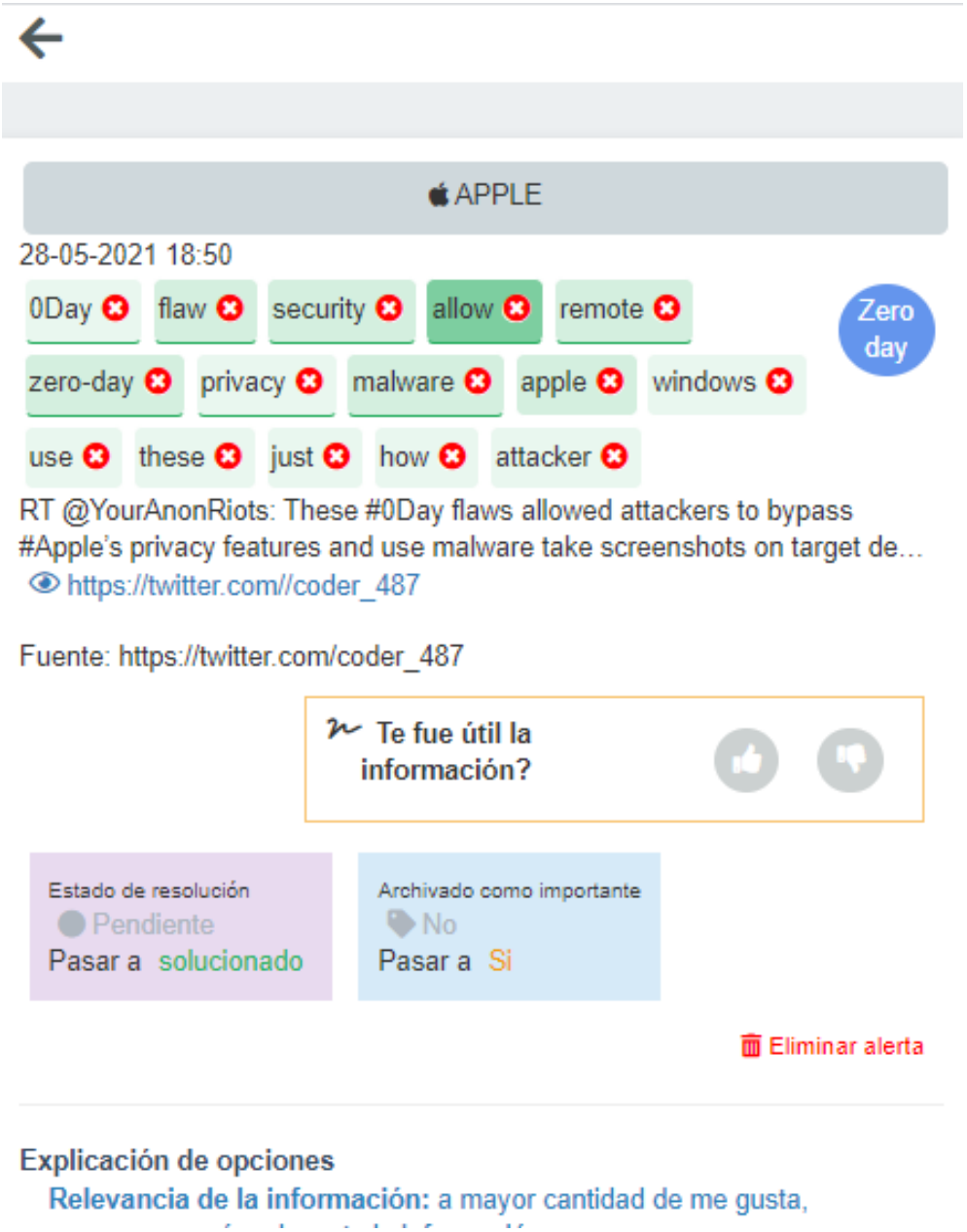

Figura 4.3: Pantalla de ampliación de alertas

Una vez expandidas las consultas de la manera indicada en el párrafo anterior, el componente de *Extracción* ③, en conjunción con el componente *Boot* ②, realizan las consultas pertinentes tanto a Twitter (mediante sus correspondientes API) como a CVE (utilizando ténicas de extracción de datos Web [9]). La información recuperada de estas fuentes es posteriormente procesada por los componentes de *Análisis* ⑤ y *Clasificación* ⑥, los cuales permiten categorizar y etiquetar la información recuperada para presentarlas al usuario final en forma de *Alertas* ⑦ personalizadas. Utilizamos el

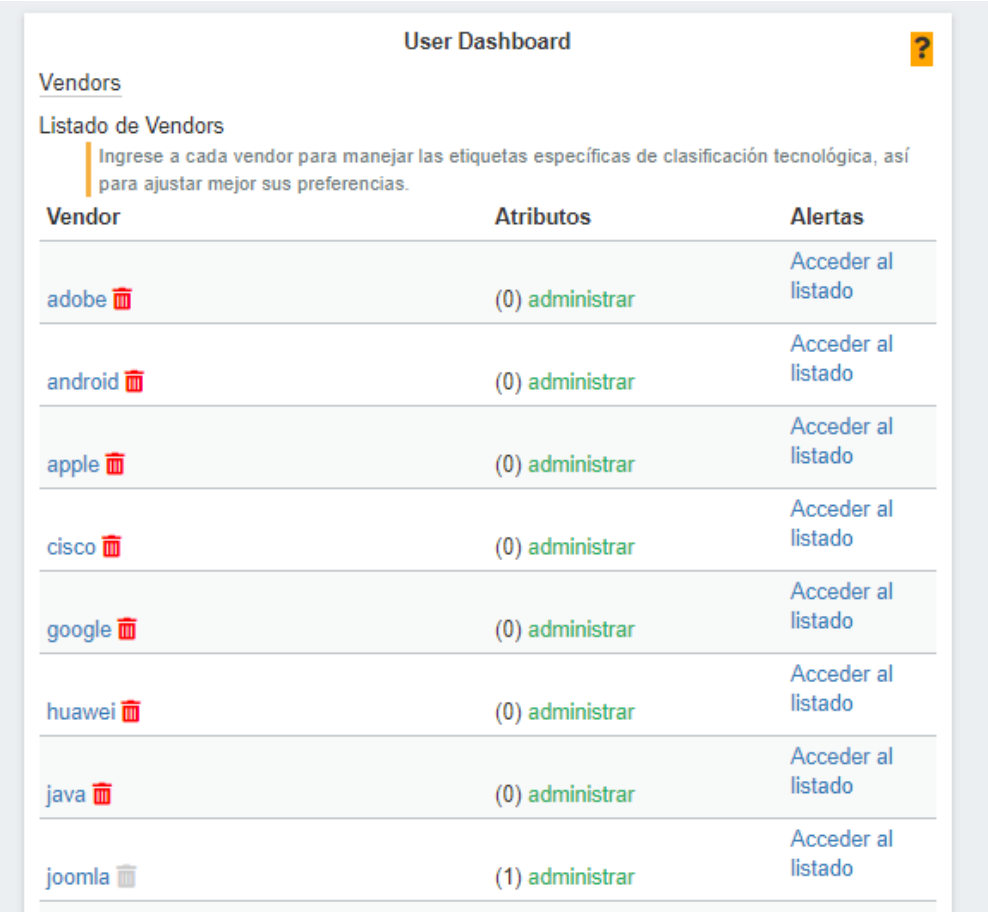

Figura 4.4: Pantalla de definición de preferencias de vendors

concepto de *etiquetado inteligente* [27], el cual es generado utilizando palabras semánticamente relacionadas en el espacio vectorial generado utilizando técnicas de *word embedding* [18]. Dichas alertas son presentadas al usuario utilizando el paradigma de *timelines* [3] (ver Figura 4.2). Hemos optado por esta opción debido a que el mismo es considerado apropiado para la presentación de información naturalmente asociada al aspecto temporal (en el dominio de la ciberseguridad, los reportes de vulnerabilidades están naturalmente asociados a una fecha). Adicionalmente, *timelines* es un paradigma de presentación de la información actualmente muy utilizado y comprendido por los usuarios, particularmente debido a la exposición frecuente de las personas a dicho paradigma en la utilización de aplicaciones para redes sociales (p.ej., Twitter, Facebook, Instagram).

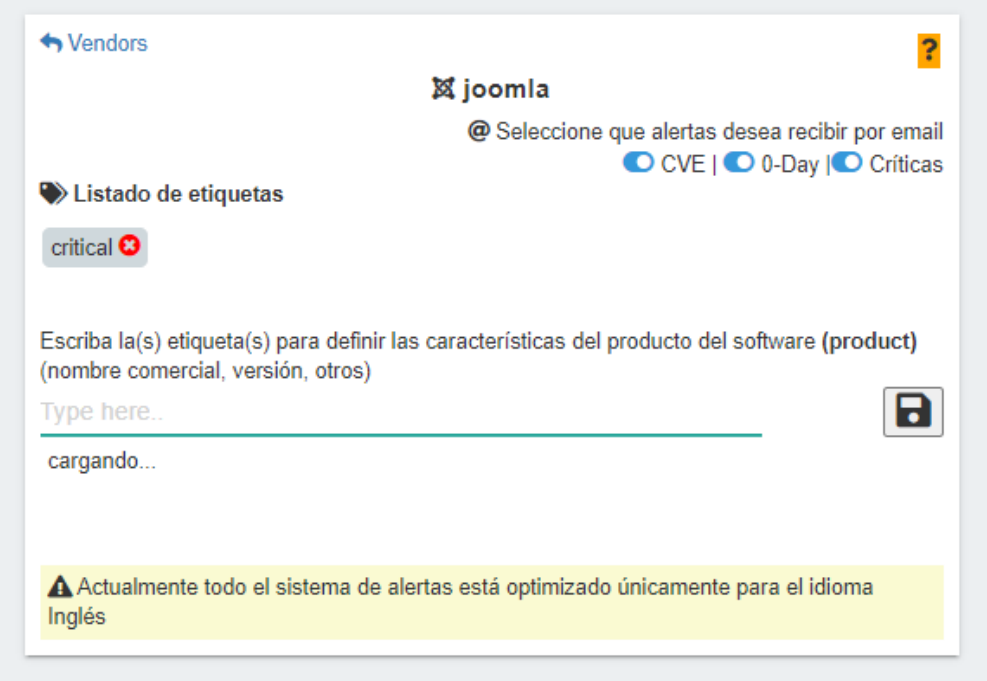

Figura 4.5: Pantalla definición de etiquetas de caracterización del entorno tecnológico

4.1. Implementación y Evaluación

La solución propuesta en esta tesis es implementada como una application Web utilizando un *stack* de tecnologías apropiadas para el efecto. En esta sección, proveemos los detalles tecnológicos de la implementación de nuestra propuesta.

4.1.1. Implementación de la Propuesta

Para la implementación del *back end* utilizamos PHP versión 7 y el motor de datos MySQL versión 5. El acceso a la información publicada en Twitter lo hacemos mediante sus API públicamente disponibles. Realizamos consultas a varias cuentas de usuarios verificados y cuyos avisos son considerados confiables. Recordamos que utilizamos esta red social como principal fuente de información sobre potenciales vulnerabilidades del día cero. En contrapartida, utilizamos los repositorios de NVD como principal fuente para la extracción de información sobre vulnerabilidades formalmente reportadas y documentadas. El acceso a dicha información la realizamos utilizando herramientas tradicionales de extración de datos Web [9], en particular, la librería Simple HTML DOM.[1]

Para propósitos de extracción de datos (particularmente, a partir de la red social

[1]https://simplehtmldom.sourceforge.io/

Twitter), contamos con una lista de más de 120 palabras semillas [13] que permiten detectar la presencia de información relacionada a vulnerabilidades. Ejemplo de estas palabras semillas incluyen “exploit”, “vulnerability”, “attack”, entre otras semillas. Adicionalmente, contamos también con palabras bloqueadas que frecuentemente resultan en falsos positivos [17] (p.ej., “podcast”, “seminar”, “webinar”, “article”). De manera a lidiar con las inflexiones del lenguaje (verbos conjugados, palabras en plural, gerundios, etc.), utilizamos la raíz de las palabras arriba mencionadas, basándonos en un recurso públicamente disponible del Diccionario de Cambridge.[1] La utilización de las palabras raíz nos permite aumentar el *recall* en la recuperación de información [17], lo cual contribuye a disminuir la omisión de información relevante relativa a las vulnerabilidades del software.

Como adelantabamos en la sección anterior, la expansión de consultas [13] es realizada utilizando técnicas de *word embedding*. En particular, para este trabajo, hemos producido nuestra propia implementación de la arquitectura de *word2vec* [18] en su variedad de *skip-gram* utilizando el lenguaje PHP y utilizando como corpus información textual del ámbito de la ciberseguridad. La utilización de *word embeddings* entrenados a partir de corpus específicos del dominio (en nuestro caso, ciberseguridad) han demostrado ser más performantes respecto a la utilización de corpus de índole general (p.ej., Wikipedia) [21, 20]. Esta implementación forma parte del componente de *Expansión* de la arquitectura presentada en la Figura 4.1.

En cuanto a la implementación del *front end*, hemos utilizado los estándares HTML5[2] y CSS.[3] Los componentes de la interfaz de usuario fueron modelados con la librería Bootstrap,[4] la cual nos ha permitido obtener una alta y estandarizada interactividad para la construcción de una interfaz basada en el concepto de Responsive Web. Los procesos asíncronos fueron implementados utilizando Ajax[5] y JQuery,[6] mientras que JSON fue utilizado como mecanismo principal de representación de datos para la comunicación vía API. Las Figuras 4.4, 4.5 y 4.3 son capturas de pantallas que ejemplifican las interfaces de usuarios utilizadas en nuestra solución para definición de preferencias del usuario, definición de etiquetas y presentación de alertas sobre vulnerabilidades, respectivamente.

4.1.2. Evaluación de usabilidad de la Propuesta

Hemos realizado la evaluación de nuestra propuesta para identificar los beneficios y limitaciones de la misma. Para el efecto, hemos llevado a cabo entrevistas contextuales

[1] https://dictionary.cambridge.org
[2] https://es.wikipedia.org/wiki/HTML5
[3] https://www.w3.org/wiki/Es/CSS
[4] https://getbootstrap.com/
[5] https://es.wikipedia.org/wiki/AJAX
[6] https://es.wikipedia.org/wiki/JQuery

[11] y pruebas de usabilidad [15] del sistema involucrando 5 personas (promedio de edad = 28 años) del ámbito de las TIC y la ciberseguridad. Este número de participantes es considerado aceptable para este tipo de estudios [15].

Para el momento de esta validación, comprendida entre el 21-02-2021 y el 30/04/2021, se habían descubierto un total de 2.070 (dos mil setenta) alertas, de las cuales 144 correspondían a aquellas documentadas oficialmente (CVE) y de gravedad crítica eran aproximadamente 30. Cabe recordar, que el sistema no pretende ser una Base de Datos de todas las vulnerabilidades propiamente, sino que está explíticamente enfocado a ser una Base de Datos de vulnerabilidades específicas de acuerdo a un contexto tecnológico definido, por ello, la cantidad total de alertas descubiertas en nuestro sistema, no tiene por que ajustarse a la cantidad total de alertas globales, sino que solo se descubren aquellas que *matchean* con preferencias (tipo tags) proporcionadas por el usuario dentro del sistema, en ese momento solo fueron definidos algunos pocos tags para la validación.

Los participantes fueron invitados y han proveído su consentimiento para participar en el estudio mediante correos electrónicos. Las sesiones con cada usuario fueron llevadas a cabo mediante video llamadas, en forma indepediente, con una duración de 1 hora. La participación fue voluntaria y no remunerada. El protocolo utilizado para el estudio fue el siguiente: (i) el investigador introduce el estudio y el propósito del mismo al participante; (ii) el investigador realiza una demostración de la herramienta para familiarizar al participante con la misma; (iii) el investigador provee una serie de tareas a ser realizada con la herramienta por el participante, quien las ejecuta bajo la observación del investigador; (iv) se lleva cabo una entrevista semi-estructurada para obtener una retroalimentación del participante. El audio de toda la experiencia es registrado por parte del investigador para propósitos de análisis posteriores.

La entrevista semi-structurada fue realizada en base a las siguientes temáticas: (i) preguntas generales sobre percepciones y sentimientos relativos a la experiencia; (ii) preguntas sobre las percepciones en cuanto a las tareas realizadas con la herramienta; (iii) preguntas sobre la herramienta en sí. A continuación, proveemos los resultados obtenidos por este estudio.

En la Tabla 4.1 se enumeran las 5 preguntas realizadas. En líneas generales, los usuarios declararon haber completado el estudio sin signos de agotamiento. Un total de 3 usuarios declaron que los ejercicios realizados en esta experiencia incidieron positivamente en sus habilidades y conocimientos sobre vulnerabilidades (p.ej., aprendieron nuevos conceptos), mientras que los restantes 2 mencionaron que la experiencia no tuvo incidencia en dichos aspectos, pero que la herramienta en sí podría ser de utilidad para sus actividades diarias.

Percepciones y sentimientos sobre la experiencia

Se pudo observar una lenta comprensión de las opciones/funcionalidades del sistema por parte de los participantes, pero a medida que iban interactuando con la herra-

Item	Cuestionario
1	¿Cómo te sientes ahora, al terminar este ejercicio? ¿Te sientes cansado, agotado, relajado?
2	¿Te parece que el ejercicio fue muy extenso y/o extenuante?
3	¿Consideras este ejercicio como algo útil para tu profesión / habilidades?
4	¿Te sentiste desconcertado y/o confundido durante la interacción con el sistema (por ejemplo, sin saber qué hacer para llevar a cabo el ejercicio)?
5	En líneas generales, ¿te sientes satisfecho con esta experiencia?. ¿Te sientes satisfecho con el ejercicio en sí?

Tabla 4.1: A. Preguntas generales sobre percepción/sentimientos relativos a la experiencia

mienta, se ha notado una rápida mejora en cuanto a la captación de la dinámica y las funcionalidades proveídas por la herramienta. En este sentido, un total de 3 participantes reconocieron explícitamente que todo sistema nuevo tiene su curva de aprendizaje, y por ende, es normal una cierta falta de compresión al inicio. En particular, un participante expresó cuanto sigue:

."El uso [correcto] de cualquier sistema implica [requiere] conocer el sistema, así que es normal perderse un poco con la herramienta en la primera sesión." (P3)

Los participantes P1 y P2 fueron los primeros en participar en el estudio, y en base a sus primeras experiencias y recomendaciones, el sistema fue corregido para las sesiones con los siguientes participantes. Por ejemplo, el participante P2 sugirió mejorar la iconografía de la herramienta para evitar confusiones. Como resultado de las correcciones, se ha notado una mejoría en la experiencia de los participantes P3, P4 y P5.

Percepciones sobre las tareas realizadas con la herramienta

En la Tabla 4.3 se enumeran las 10 preguntas realizadas. Los participantes fueron consultados si es que sintieron una falta de habilidades y/o conocimientos para realizar las tareas con la herramienta. Todos afirmaron que no experimentaron dicha falta debido a la familiaridad de los participantes con el dominio y con la dinámica de la herramienta en sí. En cuanto a lo último, consideramos que los usuarios no encontraron mayores dificultades debido a la familiaridad de los mismos con el paradigma de *timeline* (p.ej., debido a la familiaridad con aplicaciones de redes sociales como Facebook y Twitter), así como el etiquetado y filtrado de información.

Al ser abordados sobre la cercanía de las tareas realizadas con la herramienta respecto al día a día de sus actividades laborales (en TIC y ciberseguridad), todos coincidieron enfáticamente de que las mismas son efectivamente cercanas a dichas actividades, particularmente para especialistas del área de ciberseguridad.

Los participantes fueron además consultados sobre el nivel de conocimiento sobre

Item	Cuestionario
1	¿Cómo te sientes ahora, al terminar este ejercicio? ¿Te sientes cansado, agotado, relajado?
2	¿Qué te pareció el modo de listar las vulnerabilidades encontradas por el sistema?
3	¿Qué te pareció la funcionalidad de configurar tus preferencias para recibir alertas sobre vulnerabilidades?
4	¿Qué te pareció el etiquetado de vulnerabilidades en base a palabras claves?
5	Dadas las tareas realizadas durante el ejercicio, ¿qué te parece el sistema utilizado en relación a otros sistemas/aplicaciones/softwares que utilizado previamente para tareas similares?
6	¿Has notado en la interfaz del sistema algún elemento que no has visto nunca previamente?
7	¿Te has sentido obligado a comprender nuevos conceptos, elementos u operaciones con los cuales nunca antes has tenido que lidiar? Explica cuáles son y por qué.
8	¿Consideras que los procedimientos del sistema son intuitivos y similares a otros sistemas/aplicaciones/software para acceder a información relativa a vulnerabilidades? Favor, explica el por qué.
9	¿Qué te parece el mecanismo utilizado en el sistema? (Complicado / Fácil / Con muchos pasos)
10	¿Qué te parecieron las opciones complementarias respecto a marcar una alerta a solucionado o archivarlo para un seguimiento? ¿qué otras opciones más consideras necerias para un correcto seguimiento a la alerta?
11	¿Qué opinas respecto al sistema de valoración propuesto sobre las alertas ?
12	¿Has encontrado intuitiva las herramientas de filtros por etiqueta y gravedad del fallo?
13	¿Qué opinas sobre la posibilidad de eliminar etiquetas de las alertas? Entendiendo que eso ayudaría también a otros usuarios para identificar mejor las alertas, como así también otros usuarios que hagan lo mismo creará una mejor identificación para tus alertas. ¿Te parece una opción colaborativa interesante? ¿Puedes comentarnos mejor tu percepción a este respecto?
14	¿Te parecen útiles los avisos de vulnerabilidades por email? La cuales se filtraran teniendo en cuenta siempre las etiquetas de tus preferencias.
15	¿Recomendarías este sistema a otros/as? ¿Por qué?

Tabla 4.2: B. Preguntas sobre el sistema

ciberseguridad que se requeriría para poder utilizar el sistema de manera efectiva. Un total de 2 participantes coincidieron en que la herramienta no está orientada a usuarios sin conocimientos del dominio específico de la ciberseguridad. Los restantes observaron que podría ser de utilidad en caso de que la herramienta sea enriquecida con conceptos y definiciones fundamentales que permitan una mejor comprensión del dominio para usuarios no expertos. En este sentido, uno de los participantes afirmó:

"Para que usuarios no técnicos puedan usar la herramienta, creo que debería haber más definición de conceptos y una explicación más amplia de cuestiones relacionadas al ámbito de [la] seguridad [...]" (P5)

Al ser abordados sobre la utilidad de poder consultar y discernir entre vulnerabilidades formalmente reportadas (CVE) y aquellas del día cero, todos los participantes coincidieron en que es una funcionalidad útil. Por ejemplo, uno de los participantes agregó:

"[...] los puntos altos [altamente positivos] del sistema son los filtros y como filtro básico está la opción de listado por CVE y/o 0-day, útiles según la necesidad, [...] los usaría en mi día a día." (P4)

Percepciones sobre la herramienta

En la Tabla 4.2 se enumeran las 15 preguntas realizadas. Se destacó como *ventajosa* la utilización de la herramienta como apoyo para concentrar y tener organizados los avisos sobre las vulnerabilidades del software, en relación a otros mecanismos de acceso a información sobre vulnerabilidades (p.ej., buscadores y repositorios de vulnerabilidades), y como muy importante la *inmediatez* en la entrega de las alertas. Dos de los participantes proveyeron observaciones muy interesantes en ese sentido:

."En mi día a día utilizo muchas fuentes de información [sobre vulnerabilidades] y todo es muy abrumador, esta herramienta me facilitaría la vida en la obtención rápida de la información pues tener la información oportuna en el momento oportuno me da muchas ventajas técnicamente hablando." (P4)

."En mi experiencia, cuando recibíamos algún aviso o nos enterabamos por cualquier medio de alguna vulnerabilidad, en ese momento empezabamos a buscar en cualquier lugar [fuente], descentralizadamente, y en general, no siempre teníamos el tiempo suficiente para invertir en buscar información [sobre vulnerabilidades]. Normalmente no teníamos protocolos ni buenas fuentes de alertas." (P5)

En las primeras sesiones con los participantes P1 y P2, se presentaron dificultades en la compresión en cuanto al modo de listar las vulnerabilidades. Luego de ajustes realizados en base a las recomendaciones proveídas por estos dos participantes, las sesiones siguientes con los participantes restantes han demostrado mejoras en cuanto a la comprensión de la presentación (listado) de la información.

En cuanto a la configuración de las preferencias tecnológicas del usuario en base a etiquetas, las cuales permiten recibir altertas de vulnerabilidades en base a dichas

Item	Cuestionario
1	¿Te sentiste confiado en la realización del ejercicio?
2	¿Sentiste que en algún momento te faltaban ciertas habilidades/información/conocimiento para realizar el ejercicio? Favor, explícate.
3	¿Consideras que las tareas se acercan (o están alejadas) en relación a tu experiencia laboral en situaciones reales? Favor, explícate.
4	¿Qué tanto conocimiento sobre vulnerabilidades te parece que es necesario para poder utilizar este sistema? (mucho, normal, poco).
5	Dadas las tareas realizadas durante el ejercicio, ¿qué te parece el sistema utilizado en relación a otros sistemas/aplicaciones/softwares que utilizado previamente para tareas similares?
6	¿Sientes que has podido definir tus preferencias tecnológicas exitosamente? Favor, explícate.
7	¿Qué tan fácil te pareció la selección de vendors (proveedor del producto)? ¿Por qué?
8	¿Qué tan fácil te pareció la selección de etiquetas (palabras claves)? ¿Por qué?
9	¿Qué tan fácil te pareció navegar los resultados proveídos por el sistema? ¿Por qué?
10	¿Qué tan ÚTIL te pareció la posibilidad de seleccionar entre la posibilidad de elegir fuentes de datos oficiales (CVE) y 0-day (redes sociales)? ¿Por qué?

Tabla 4.3: C. Preguntas sobre las tareas

preferencias, la mayoría de los usuarios experimentaron dificultad en configurarlas. Sin embargo, les pareció acertada la decisión de permitir filtrar la información en base a las preferencias tecnológicas del usuario, de modo a poder alivianar la carga cognitiva que implica estar proactivamente buscando nuevas vulnerabilidades. En ese sentido, un participante afirmó:

"Partiendo de la experiencia con algunos repositorios de información sobre vulnerabilidades y foros, en los que la cantidad de información no relevante es muy grande, y que este sistema me permita aproximarme a mis preferencias tecnológicas y que me provea ya lo justo y necesario, me parece muy acertada y me permite relajarme y concentrarme en otras tareas." (P4)

Respecto al etiquetado automático de las vulnerabilidades, todos los usuarios coincidieron en que el mismo es realizado correctamente por la herramienta. El participante P2 resaltó que las etiquetas permiten comprender con un golpe de vista la categoría a la que pertenece la vulnerabilidad listada por la herramienta.

Cuando los participantes fueron consultados si es que han sido expuestos a conceptos o elementos que normalmente no tienen en cuenta en sus actividades diarias, 2 participantes resaltaron los conceptos de *niveles de gravedad* y *vulnerabilidades del día cero*. En particular, les pareció interesante que la herramienta contemple estos dos aspectos. Por otro lado, al consultarles sobre cómo compararían la herramienta en relación a soluciones similares, los 5 participantes declaron no conocer un sistema similar para el acceso a información sobre vulnerabilidades, enfatizando la relevancia de nuestra propuesta, y aclarando que sus fuentes primarias de información son los foros y repositorios de vulnerabilidades (los cuales no son directamente comparables a nuestro sistema de alertas tempranas de vulnerabilidades).

Al ser consultados sobre la funcionalidad de la herramienta para marcar una vulnerabilidad como solucionada o pendiente de seguimiento, todos los participantes coincidieron en afirmar que les resultaría muy útil. Incluso fueron más allá de esta funcionalidad y 4 participantes propusieron la posibilidad de compartir con la comunidad comentarios al respecto y la solución utilizada para abordar la vulnerabilidad. En este sentido, un participante sugirió cuanto sigue:

"Me parecen geniales y útiles las opciones complementarias [marcar una vulnerabilidad como solucionada o pendiente de ser solucionada], pero [además] me gustaría poder compartir con la comunidad comentarios y/o explicar los pasos que realicé en una vulnerabilidad solucionada." (P5)

Adicionalmente, un participante propuso la funcionalidad de poder asignar a otro usuario, mediante la herramienta, el trabajo de mitigar una vulnerabilidad:

"[...] me gustaría tener más información respecto al seguimiento en sí, y poder derivar a alguien el trabajo y saber cuándo lo solucionó." (P2)

En relación al uso del filtrado de información mediante etiquetas y nivel de grave-

dad, todos los participantes coincidieron en su relevancia y utilidad. Por ejemplo, un participante observó que:

"[...] puntos altos [altamente positivos] del sistema son las opciones de filtro, tanto para la búsqueda por etiquetas y más especialmente por grado de criticidad de la vulnerabilidad." (P4)

Dado que la herramienta permite el filtrado colaborativo de etiquetas utilizadas en las vulnerabilidades, permitiendo agregar/eliminar etiquetas, hemos también consultado al respecto a los usuarios. Todos los usuarios reconocieron su potencial utilidad, siempre y cuando sea correctamente utilizado. Además, todos expresaron su preocupación de un uso malintencionado (o por desconocimiento) de esta funcionalidad. Este es un problema ampliamente reconocido en sistemas de *crowdsourcing* [2, 26]. Por ejemplo, un participante afirmó:

"[...] particularmente me preocupa un tanto que cualquiera pueda eliminar las etiquetas, más que nada por que podría eliminarse alguna útil por desconocimiento o malintencionadamente, pero usándola bien me parece una buena opción colaborativa en la que nos beneficiemos todos en la comunidad. Podría haber una especie de usuarios colaboradores y que no todos puedan tener la opción de borrar [etiquetas] ." (P5)

La propuesta presentada arriba por el participante P5 se encuentra alineada con mecanismos utilizados en sistemas de *crowdsourcing* existentes. En particular, con la comunidad de preguntas y respuestas Stackoverflow,[1] donde usuarios con cierta reputación ganada en la comunidad acceden a permisos especiales de curación de contenido en la plataforma.

Finalmente, al consultar a los participantes si recomendaría a otros la utilización de la herramienta, todos coincidieron que sí la recomendarían. Uno de los participantes agregó:

"[...] porque es útil y fácil de usar para gente que esté en el ámbito de las vulnerabilidades, porque facilita tener a mano la información relevante. Rescatando todas sus funcionalidades: definición de preferencias, tagging de la información, filtros y avisos por email me parece una herramienta genial." (P1)

4.2. Evaluación de validación de resultados

Posterior a la primera validación, que se detalló en el apartado anterior, sugieron varios aspectos a mejorar y varios puntos críticos programáticos en el prototipo, ciertamente se hicieron ajustes menores en muchos aspectos, se destacan 2 en particular, (i) aquella referente al filtrado de la información descubierta, con especial énfasis en

[1] https://stackoverflow.com

la curación de *"las semillas"* y ajustes en el módulo de aprendizaje automátizado. Y (ii) se realizaron ajustes respecto al envío personalizado de alertas por e-mail, el cual no estaba filtrando correctamente la información a enviar al usuario, por consiguiente, generaba mucho ruido en la bandeja del usuario y al límite de la fatiga de alertas.

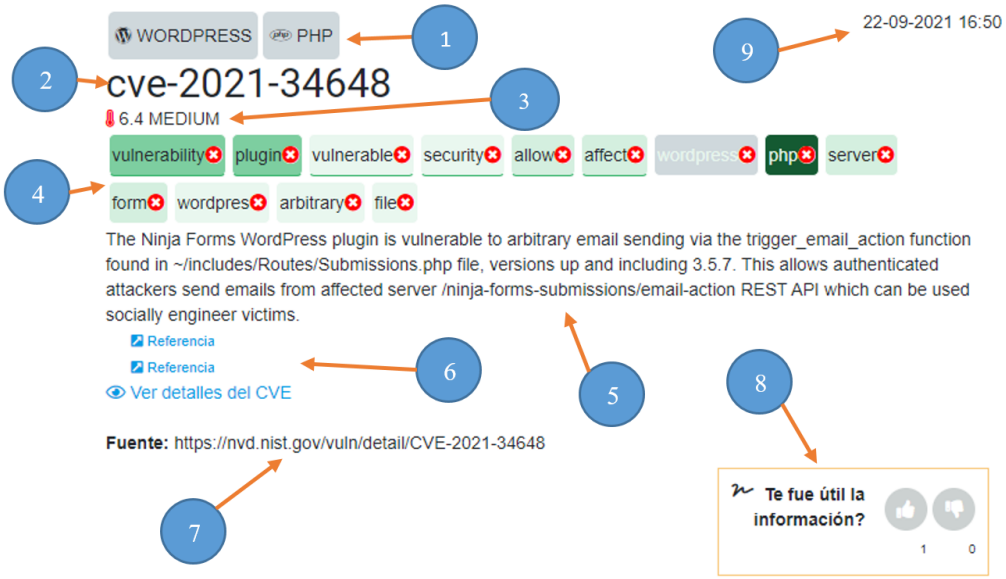

Figura 4.6: Estructura de cada alerta

A continuación explicaremos los componentes principales de cada alerta en su versión web, esta alerta es la visualización ampliada desde la bandeja de entrada, de esta manera se pretende explicar el como visualizan los usuarios las alertas de interés. Para ello presentamos un gráfico el 4.6 y seguidamente con la explicación de cada punto: 1) Las etiquetas tecnológicas del tipo vendor que corresponden con las definidas por el usuario ; 2) el código CVE para el caso que fuera documentada, y estaría sin valor alguno si fuera las no documentadas; 3) el nivel de gravedad del fallo, exlusivo para las alertas documentadas; 4) el *tagging* generado automáticamente e incluso ampliado con términos relacionados; 5) correponde al texto completo de la alerta, en el caso de las no documentadas el texto es de apenas 140 caracteres, pues es la máxima cantidad de información extraíble de *Twitter*; el 6) corresponde a las referencias que se extráen del texto explicativo; 7) la fuente de la alerta; 8) el bloque de valoración de la alerta y por último el punto 9) que sería la fecha de la alerta; en el caso de un Zero Day, y que se haya detectado un hashtag correcto, se coloca una nueva etiqueta para resaltar la misma, ligeramente por debajo del punto 9, con la leyenda Zero Day (ver figura 4.8).

Complementariamente en cada alerta agregamos lo siguiente, ver imagen 4.7: 1) bloque para marcar a solucionado o no una alerta; 2) opción para archivar la alerta para seguimiento; 3) opción para eliminar una alerta, esta opción eliminaría definitivamente de la bandeja de entrada del usuario; 4) historial de tags eliminados, para un

Figura 4.7: Estructura de cada alerta sección 2

seguimiento y eventualmente poder recuperar un tag eliminado erroneamente, y por último destacamos el apartado 5) que es un bloque en el que se muestran alertas similares y que son coincidentes en algunos tags importantes de la alerta principal; esto podría servir para ver la importancia del fallo al estar relacionado con otros similares compartidas en otras fuentes.

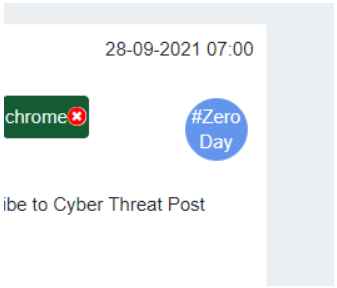

Figura 4.8: Resaltador Zero Day

Uno de los principales ajustes realizados en el marco de curación de los filtros para el descubrimiento de las vulnerabilidades del *software* no documentadas, comúnmente conocidas como *Zero Day*, se refiere a lo siguiente, en nuestros seguimientos de los avisos compartidos en *Twitter*, hemos detectado en un porcentaje importante el uso adecuado de hashtags [1] para etiquetar los avisos del tipo zero day, y descubrimos algu-

[1] https://rockcontent.com/es/blog/hashtags

nos términos hashtag para referirse a ellos: *zero-day, 0Day*, etc.. Entonces, por sobre la detección de cualquier alerta que posea el texto zero day (ver imagen referencial, alerta sin hashtag 4.14), catalogamos explícitamente como *Zero Day* solo aquellas que efectivamente tienen algún término referido a *Zero Day* como hashtag (ver imagen referencial, alerta con hashtag 4.15). En este aspecto, percibimos una mejor categorización automática de los avisos *Zero Day*, pues inicialmente nuestros resultados, arrojaban muchos falsos positivos, que con esta y otras técnicas de mejoramiento, se redujeron notablemente.

En este contexto, es importante comentar que fuímos aprendiendo e indentificando a algunos usuarios que realizaban importantes contribuciones en nuestra área de interés en *Twitter*, inicialmente habíamos seleccionado al azar a algunos, pero a medida avanzamos con nuestras pruebas y conocimientos, fuimos incorporando algunas cuentas primando aquellas con más seguidores y con más tecnicímo en el ámbito de la ciberseguridad, en la Tabla 4.4 se puede ver nuestra lista de usuarios de los que extraemos automáticamente sus publicaciones en tiempo real, en ella se destaca un usuario con solo 15 seguidores, el cual fue uno de nuestros primeros usuarios incorporados.

Para esta segunda etapa de validación, el sistema se empezó a usar masivamente (por 5 usuarios) a partir del 17/08/2021 hasta el 17/09/2021. Los participantes, quienes fueron invitados por e-mail y aceptaron sin remuneración alguna su participación en el estudio, fueron seleccionados por su perfil técnico avanzado y se destaca el uso del sistema dentro la operativa diaria del MITIC (*"Ministerio de Tecnologías de la Información y Comunicación"*), de dicha institución se sumaron 3 usuarios altamente técnicos con énfasis en al área de ciberseguridad, para la utilización del sistema disponibilizaron un monitor de 50 pulgadas empotrado a la pared para su visualización en tiempo real de las alertas, ver fotografías 4.12 y 4.13.

Cada usuario definió su propia realidad tecnológica del *software* mediante la definición de *tags* de preferencias. Las alertas se generaron y descubrieron a partir de dichos *tags* tecnológicos. En ese lapso de tiempo, y enmarcado bajo esos parámetros, se descubrieron más de 870 vulnerabilidades, 96 fueron del tipo documentadas (CVE) y 42 de ellas de gravedad crítica.

Como protocolo de validación de los resultados, los usuarios disponían de dos mecanismos de validación (i) valorar la información como: *útil* y *no útil*. Para esta tarea se disponibilizó un bloque en cada detalle de la vulnerabilidad, ver imagen 4.9 que permitía, mediante el pulgar arriba 4.10, definir la información como relevante y mediante el pulgar abajo 4.11, definir la información como no relevante o no útil. Además, como segunda opción de validación (ii) cada usuario podía eliminar las etiquetas que no correspondían a las vulnerabilidades descubiertas, las mismas que eran generadas automáticamente y por aprendizaje automatizado. Cabe mencionar, que como amenaza y debilidad de la validación se pudieron notar 3 situaciones: (i) que no todos los usuarios habían caracterizado correctamente su realidad tecnológica, algunos se limitaron solo a definir tags muy generales a nivel vendors, por lo tanto alertas generadas muy gené-

Cuenta	Seguidores
Norton	+150mil
MsftSecIntel	+133mil
TheHackersNews	+730mil
TrendMicroRSRCH	+49mil
inj3ct0r	+46mil
cybersec_feeds	+18mil
CyberSecDN	+15mil
TechData_IBM	+3400
cybsecbot	+1900
trendmicro_mea	+1450
threatmeter	+1300
JinibaBD	+800
UITSECGlobal	15

Tabla 4.4: Cuentas seguidas en Twitter

ricas; (ii) no todos marcaban las vulnerabilidades como útil o no útil, y (iii) no todos eliminaron tags incorrectos en las etiquetas generadas automáticamente.

En este contexto, y tomando todo el cúmulo de información ya existente en el sistema desde el 21-02-2021, el sistema arrojó las siguientes estadísticas. ver Tabla 4.5.

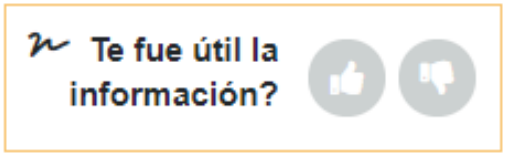

Figura 4.9: Bloque de valoración de la información

Figura 4.10: Pulgar arriba: información útil

Como parte final de la validación se realizaron encuestas que contemplaron 13 temas (ver tabla 4.6) y que produjo el siguiente resumen (ver tabla 4.7).

En esta etapa se responde una de nuestras preguntas de investigación la P1. La

Valores	Tipo	Descripción
444	Alertas	del tipo CVE descubiertas
227	Alertas	de gravedad crítica
4948	Alertas	generadas directamente desde las Redes Sociales
+2030	Alertas	marcadas con el hastash zero day
573	Tags	fueron eliminados por el usuario
31	Alertas	archivadas para seguimiento
30	Alertas	marcadas como solucionadas
+2000	Alertas	enviadas por e-mail
114	Tags	etiquetas de vendor configuradas por usuarios
146	Tags	etiquetas de productos y características adicionales
220	Alertas	marcadas cómo útil
85	Alertas	marcadas cómo no útil
+13000	Word2vec	textos aprendidos (incorporados)
+330000	Word2vec	relaciones aprendidas para incrustación

Tabla 4.5: Estadísticas en el sistema desde el 21-02-2021 al 20-09-2021

Item	Pregunta
1	¿Con qué dificultades se enfrentan día a día en sus operativas respecto a los avisos de vulnerabilidades?
2	Respecto a los avisos/boletines sobre vulnerabilidades y parches de seguridad en las industrias, empresas, y/o personal involucrado/afectado. ¿Qué tanta conciencia de atención perciben?
3	¿Qué les pareció acceder a la información sobre vulnerabilidades del software utilizando éste sistema en relación a otras formas que utilizan normalmente para buscar información sobre vulnerabilidades?
4	¿Qué les pareció el modo de listar las vulnerabilidades encontradas?
5	¿Qué les pareció la funcionalidad de configurar sus preferencias para recibir alertas sobre vulnerabilidades?
6	¿Qué les pareció el etiquetado de vulnerabilidades en base a las palabras claves?
7	¿Qué otras opciones más consideran necesarias para un correcto seguimiento de las alertas?
8	¿En qué les ayuda o ayudaría la herramienta en sus operativas?
9	¿Con qué periodicidad utilizan o utilizarían la herramienta?
10	¿Considera Ud. que se generaron correctamente los avisos y en tiempos correctos? ¿Por qué?
11	¿Qué valor genera la herramienta en sus operativas diarias?
12	En base a una alerta generada por el sistema, ¿Qué acciones realizaron en consecuencia?
13	¿Qué futuro le ven a la herramienta y a quienes podría ayudar?

Tabla 4.6: Cuestionario de validación

4.2 Evaluación de validación de resultados

Item	Pregunta
1	Diariamente se deben revisar y clasificar la información disponible en Twitter y otros medios sociales. Filtrar la información manualmente consume mucho tiempo, esfuerzo e implica gran conocimiento del área.
2	En la generalidad, en las empresas afectadas por algún incidente, se percibe muy poca conciencia respecto a las alertas de seguridad y por consiguiente baja prioridad en la solución o el tratamiento de las vulnerabilidades
3	La herramienta en sí fue muy valorada para un uso en producción
4	El listado de vulnerabilidades, tipo timeline, fue muy bien aceptado por todos los usuarios del sistema
5	En general se considera una de las fortalezas de la herramienta, pues comparando con otros sistemas, realiza un filtrado muy acertado respecto a la realidad tecnológica del software del usuario, evitando atosirgarlo con tanta información y evitando, por lo tanto, la fatiga de alertas. Además se enfatizó la capacidad de la herramienta de descubrir alertas referentes al zero day.
6	Se validó como una opción interesante, por que a simple vista se puede conocer prácticamente de qué se trata la amenaza
7	Se enfatiza la necesidad de generar más reportes estadísticos en paneles tipo dashboard, y además se sugirió mejorar la funcionalidad de los filtros.
8	En la utilización que se le dio en las pruebas, las alertas descubiertas servían como disparador en la generación de sus propios boletines de alertas, al menos si ameritaba pues habían además falsos positivos
9	Se constató 2 maneras de uso, primero la disponibilidad en un monitor exclusivo para visualizar las alertas, y por otro el menos utilizado las alertas por correo, este último de menos valor en las pruebas pues habían agregado, a sus preferencias, casi todos los vendors habilitados, y por lo tanto, al límite de las fatigas de alertas. Pero si se hubiera limitado a menos tecnologías, ambas formas de recibir las alertas hubieran sido optimas.
10	En general todas las alertas y avisos de las redes sociales se generaron en tiempos correctos, solo que al tener demasiadas tecnologías como preferencias, la cantidad de información diaria fue bastante y reclamaron que se perdía en la lista y sugieren una especie de listado, más que por orden de descubrimiento, por orden de relevancia (aunque se aclara que ninguna lógica para el marcado de relevancia fue analizada en la tesis)
11	En general se mencionó, que por la diversidad de fuentes de consultas, y por la calidad de información resultante, ayuda a conocer rápidamente las tendencias en fallos de ciberseguridad en el mundo
12	Se constató específicamente una situación, se generó una alerta de vulnerabilidad grave en un producto, el mismo era uno de los utilizados con mayor asiduidad por el usuario y gracias a esa alerta, y siguiendo los pasos descriptos en el CVE, desconectaron de internet el software realizaron el parchado del mismo y luego lo volvieron a poner en línea. Además otra utilidad verificada, descubierta una alerta y que era relevante para la infraestructura nacional, verificaron mas información al respecto y/o documentaciones oficiales y luego procedieron a generar noticias y boletines al respecto.
13	Recomendarían su uso en CSIRTS (Computer Security Incident Response Team, Equipo de Respuesta ante Incidencias de Seguridad Informáticas) de todos los países.

Tabla 4.7: Resumen de la validación

Figura 4.11: Pulgar abajo: información no útil

Figura 4.12: Fotografía en el MITIC

respuesta es el resultado de experiencias de usuarios expertos en ambientes reales y se resume en lo siguiente:

"Diariamente se deben revisar y clasificar la información disponible en Twitter y otros medios sociales (como por ejemplo Telegram). Filtrar la información manualmente consume mucho tiempo, esfuerzo e implica gran conocimiento del área."

4.2 Evaluación de validación de resultados

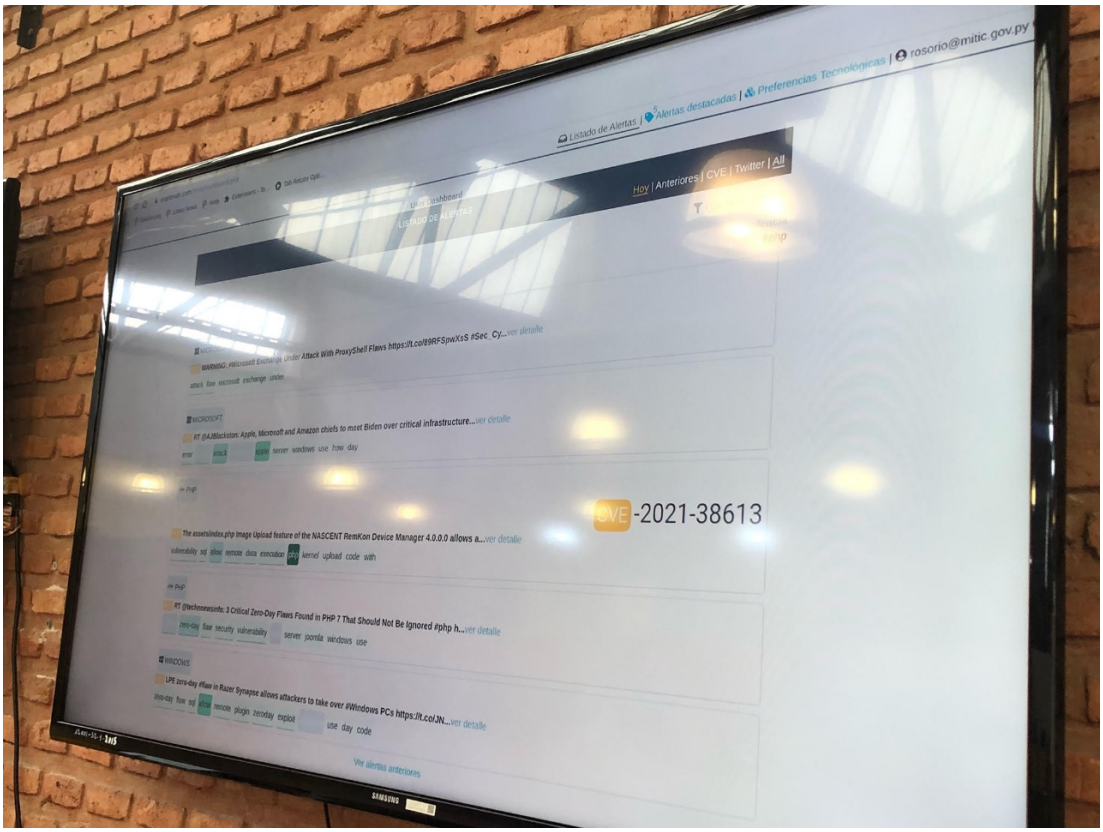

Figura 4.13: Fotografía del monitor en el MITIC

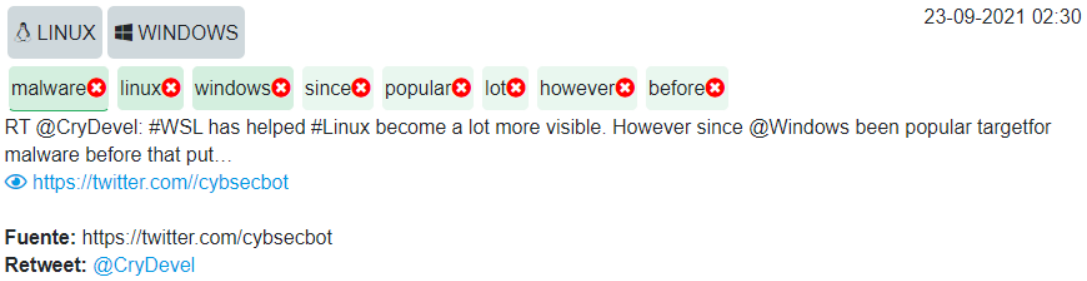

Figura 4.14: Alerta proveniente de Twitter sin hashtag del Zero Day

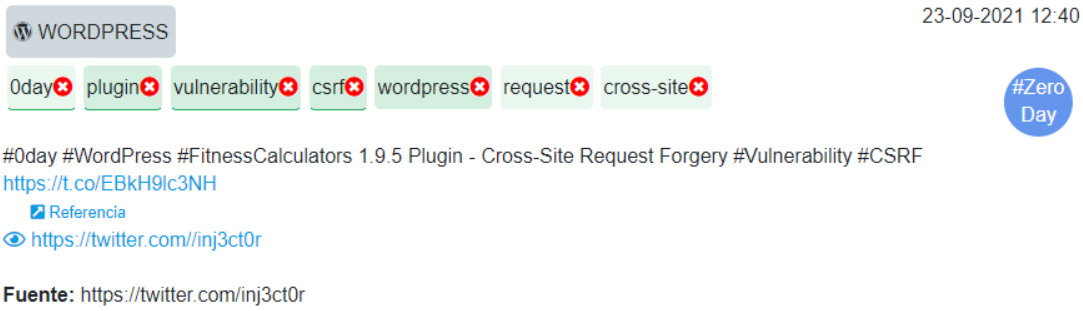

Figura 4.15: Alerta proveniente de Twitter con hashtag del Zero Day

Capítulo 5

Conclusiones y trabajos futuros

5.1. Conclusión

Este artículo presenta una propuesta de alertas tempranas sobre vulnerabilidades formalmente documentadas en repositorios oficiales, así como también sobre potenciales vulnerabilidades del día cero detectadas a partir de las redes sociales. La propuesta apunta a abordar las necesidades creciente y urgente de las organizaciones de mantenerse al tanto de las debilidades del software, fenómeno siempre creciente particularmente en los últimos años. Nuestra propuesta aborda esta problemática mediante una solución que combina técnicas de recuperación de la información [17], expansión de consultas (*query expansion*) [13], categorización y etiquetado inteligente [27] de vulnerabilidades mediante técnicas de *word embeddings* [18], y presentación de la información mediante mecanismos basados en *timelines* [3] y *push notifications* [8].

Los estudios realizados con usuarios representativos en forma de entrevistas contextuales [11] y pruebas de usabilidad [15] han revelado la viabilidad, utilidad y potencialidad de la solución propuesta. Los resultados arrojados por el estudio indican que éstos usuarios representativos encuentran en la solución propuesta una herramienta útil que emplearían en el día a día de sus operaciones diarias. Aspectos altamente positivos, en la perspectiva de los participantes, incluyen la posibilidad de establecer preferencias tecnológicas sobre las cuales recibir información sobre vulnerabilidades (permitiendo lidiar con la sobrecarga de información en este dominio), la precisión de la categorización y etiquetado de las vulnerabilidades, la posibilidad de filtrar información en base a criterios como vulnerabilidades del día cero y vulnerabilidades documentadas, criticidad de las vulnerabilidades, etc. En cuanto a las limitaciones, se ha visto con preocupación la posibilidad de que cualquier usuario pueda modificar las etiquetas (p.ej., elimándola). En consecuencia, se ha propuesto acceder a dicha funcionalidad basada en privilegios ganados por el usuario en base a su reputación.

En la segunda etapa de validación se puso a prueba la herramienta en ambientes reales, y por lo tanto, se hizo el seguimiento de la calidad, cantidad y efectividad de las alertas en tiempo real generadas y descubiertas por el sistema. En esta etapa los usuarios realizaron la clasificación de sus tecnologías de acuerdo a su interés y su realidad tecnológica, por lo tanto, la información resultante, para cada uno, fue distinta y sin sobrecargas de información. En general la herramienta cumplió con el objetivo de proveer la información en tiempo y forma, y cada usuario, en contrapartida, debió actuar en concencuencia. Se evidenció una manera mucho más relajada y enfocada de trabajar en lo que ocurre en su ambiente tecnológico, al no estar el usuario tan pendiente, ni de realizar tan exhaustivamente tareas de búsqueda y descubrimiento de vulnerabilidades, sino se enfocaba en ahondar solo aquellas alertas ya descubiertas y disponibilizadas por la herramienta.

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.

En la siguiente conclusión repondemos la pregunta P2 de investigación, y concluímos 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.

5.2. Trabajos futuros

Como trabajo futuro, proponemos incorporar las mejoras sugeridas por los participantes del estudio y desplegar la solución en un entorno organizacional real para su utilización en operaciones diarias, con propósitos de llevar a cabo un caso de estudio longitudinal sobre los beneficios y limitaciones de la propuesta en la creación de conciencia sobre las vulnerabilidades del software, la influencia de la propuesta en la mitigación de las mismas y el impacto en la mejora de la ciberseguridad de la organización.

En nuestras pruebas en campo, se evidenció un consumo importante de procesador respecto a calculos y procesamiento, especialmente a nivel Base de Datos, por lo tanto, se debe tener muy en cuenta la infraestructura en la que se va a desplegar el sistema; este 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 texo 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 cibertecnologías 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 sistema se enfocó en el área de ciberseguridad pero se evidencia que de la misma forma, utilizando en otro dominio: el aprendizaje, las semillas base, la recopilación y clasificación automatizada; podría perfectamente extrapolarse a otras áreas del conocimiento, por ejemplo en un ambiente financiero.

Por otro lado, como se explicó en capítulos anteriores, la implementación del sistema se hizo en PHP copiando la lógica del Word2Vec implementado en Python; como siguiente trabajo se debería realizar una prueba exhaustiva y comparar los resultados en ambas implementaciones para sacar una conclusión correcta, para saber si realmente se justifica la lógica implementada, pues ciertamente es un modelo similar al Word2vec original, tiene ciertas diferencias programática y lógicas, e intenta solucionar un problema detectado en la lógica original, el cual se refiere a la incorporación de nuevos textos a su base de aprendizaje.

Referencias

[1] D. E. Alexander, "Social media in disaster risk reduction and crisis management," *Science and engineering ethics*, vol. 20, no. 3, pp. 717–733, 2014.

[2] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality control in crowdsourcing systems: Issues and directions," *IEEE Internet Computing*, vol. 17, no. 2, pp. 76–81, 2013.

[3] O. Alonso, R. Baeza-Yates, and M. Gertz, "Exploratory search using timelines," in *Proceedings of the ACM SIGCHI 2007 Workshop on Exploratory Search and HCI*, 2007, pp. 23–26.

[4] S. S. Alqahtani, E. E. Eghan, and J. Rilling, "Tracing known security vulnerabilities in software repositories–a semantic web enabled modeling approach," *Science of Computer Programming*, vol. 121, pp. 153–175, 2016.

[5] L. Atymtayeva, K. Kozhakhmet, and G. Bortsova, "Building a knowledge base for expert system in information security," in *Soft computing in artificial intelligence*. Springer, 2014, pp. 57–76.

[6] S. Beattie, S. Arnold, C. Cowan, P. Wagle, C. Wright, and A. Shostack, "Timing the application of security patches for optimal uptime." in *LISA*, vol. 2, 2002, pp. 233–242.

[7] K. Darabal, "Vulnerability exploration and understanding services," in *Security Vulnerability Information Service with Natural Language Query Support, Advanced Information Systems Engineering*. Springer, 2019, pp. 497–512.

[8] J. D. Falcão, J. Krebs, S. Kumar, and H. Erdogmus, "Openalerts: A software system to evaluate smart emergency alerts and notifications," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium*

on Pervasive and Ubiquitous Computing and Wearable Computers, 2018, pp. 1250–1255.

[9] E. Ferrara, P. De Meo, G. Fiumara, and R. Baumgartner, "Web data extraction, applications and techniques: A survey," *Knowledge-based systems*, vol. 70, pp. 301–323, 2014.

[10] S. Furnell and J. N. Shah, "Home working and cyber security–an outbreak of unpreparedness?" *Computer Fraud & Security*, vol. 2020, no. 8, pp. 6–12, 2020.

[11] K. Holtzblatt and S. Jones, "Conducting and analyzing a contextual interview (excerpt)," in *Readings in Human–Computer Interaction*. Elsevier, 1995, pp. 241–253.

[12] A. Joshi, R. Lal, T. Finin, and A. Joshi, "Extracting cybersecurity related linked data from text," in *2013 IEEE Seventh International Conference on Semantic Computing*. IEEE, 2013, pp. 252–259.

[13] R. P. Khandpur, T. Ji, S. Jan, G. Wang, C.-T. Lu, and N. Ramakrishnan, "Crowdsourcing cybersecurity: Cyber attack detection using social media," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1049–1057.

[14] A. Khazaei, M. Ghasemzadeh, and V. Derhami, "An automatic method for cvss score prediction using vulnerabilities description," *Journal of Intelligent & Fuzzy Systems*, vol. 30, no. 1, pp. 89–96, 2016.

[15] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Cambridge university press, 2008.

[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[19] S. A. Mokhov, J. Paquet, and M. Debbabi, "The use of nlp techniques in static code analysis to detect weaknesses and vulnerabilities," in *Canadian Conference on Artificial Intelligence*. Springer, 2014, pp. 326–332.

[20] S. Mumtaz, C. Rodriguez, and B. Benatallah, "Expert2vec: Experts representation in community question answering for question routing," in *International Conference on Advanced Information Systems Engineering*. Springer, 2019, pp. 213–229.

[21] S. Mumtaz, C. Rodriguez, B. Benatallah, M. Al-Banna, and S. Zamanirad, "Learning word representation for the cyber security vulnerability domain," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[22] C. Rodriguez, S. Zamanirad, R. Nouri, K. Darabal, B. Benatallah, and M. Al-Banna, "Security vulnerability information service with natural language query support," in *International Conference on Advanced Information Systems Engineering*. Springer, 2019, pp. 497–512.

[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[24] C. Sabottke, O. Suciu, and T. Dumitraș, "Vulnerability disclosure in the age of social media: Exploiting twitter for predicting real-world exploits," in *24th {USENIX} Security Symposium ({USENIX} Security 15)*, 2015, pp. 1041–1056.

[25] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 851–860.

[26] R. Sumi, T. Yasseri, *et al.*, "Edit wars in wikipedia," in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 2011, pp. 724–727.

[27] J. Vig, S. Sen, and J. Riedl, "The tag genome: Encoding community knowledge to support novel interaction," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 3, pp. 1–44, 2012.