

Seguridad aplicada en las Redes Sociales mediante extracción de datos

E. Onofre Ruiz^{1*}, B. Y. Ramírez Herrerías², J. Martínez Santiago³, R. Castro Valdivia,⁴ A. Escobar García⁵
¹Ingeniería en Tecnologías de la Información, Universidad Tecnológica del Centro de Veracruz,
Av. Universidad N° 350, Carretera Federal Cuitláhuac - La Tinaja, Localidad Dos Caminos, Cuitláhuac, Ver.
[*eric.onofre@utcv.edu.mx](mailto:eric.onofre@utcv.edu.mx)

Área de participación: Sistemas Computacionales

Resumen

De acuerdo con estudios realizados, actualmente en el mundo hay más de trescientos millones de usuarios que se conectan a diario a las redes sociales, es decir, un 42% de la población mundial. Hoy en día, las redes sociales no solo son exclusivas para jóvenes y adolescentes, también son utilizadas por otras generaciones adultas; sin embargo, los usuarios deben de ser conscientes y saber utilizarlas con responsabilidad, cuidando el contenido que publican y la información personal que comparten, porque este tipo de información puede ser valiosa para ciberdelincuentes. Estos se sirven de este tipo de información para cometer ciberdelitos en contra de los usuarios. Es por ello por lo que SARA (Seguridad a las Redes Sociales), tiene el objetivo de reducir los niveles de riesgo en los usuarios en las redes sociales, para brindarles recomendaciones preventivas que les ayuden a disminuir el riesgo de exposición dentro de ellas.

Palabras clave: análisis, publicaciones, redes sociales, seguridad.

Abstract

According to studies, there are currently more than three hundred million of users in the world who are connected to social media every day, that is, 42% of the world population. Nowadays, social media is not the only one used by young people and teenagers, they are also used by other adult generations; however, they must be aware of the use of the social media, and to know how to use them responsibly by being careful about the content they post, and the personal information they share because this type of information can be valuable for cybercriminals. They use this information to commit cybercrimes against users. That is why SARA (Seguridad Aplicada a las Redes Sociales), aims to reduce the levels of risk in users on social media in order to provide preventive recommendations that help them to reduce the risk of exposure within them.

Key words: analysis, publications, security, social media.

Introducción

En la actualidad las redes sociales se han vuelto la herramienta de comunicación más común entre las personas, a tal grado que gran parte de la vida social de los individuos se desarrolla a través de estos medios. Se ha cambiado la interacción personal y los canales de comunicación habituales por un entorno artificial de comunicación a distancia. Sin duda, las redes sociales hacen la comunicación más fácil, permiten que las personas se comuniquen de una forma más abierta porque el contacto que se tiene con la otra persona no es directo. (Ortega, 2015). Por esta razón, es más fácil expresar opiniones o puntos de vista de determinados temas o externar emociones e ideas de una forma más espontánea, pero no siempre publicar todo lo que se piensa o hace en estos entornos es prudente. Junto con todas las bondades que ofrecen las redes sociales para comunicarse con las personas vienen implícitos factores potenciales que pueden poner en riesgo a los usuarios que hagan un mal uso de ellas. De acuerdo con (Villanueva, 2019) indica que, las malas prácticas, tales como la publicación de información personal, ubicación actual del usuario o estados que expresen alguna vulnerabilidad emocional, exponen a las personas con ciberdelincuentes que se sirven de esta información para cometer una infinidad de ciberdelitos en contra del usuario incauto. Estos delitos van desde el cyberbullying, ciberacoso, robo, extorsión y en casos más graves hasta el sexting, grooming o secuestro.

En un comunicado publicado por la Secretaría de Seguridad Ciudadana (SSC, 2019) de la Ciudad de México, las cifras que se dan a conocer acerca de los ciberdelitos cometidos a través de las redes sociales son preocupantes. De enero a diciembre del 2019, se han levantado diversos reportes en los que se destacan mil novecientos casos de acosos y amenazas, mil trescientos por fraude, cuatrocientos por delitos contra menores y trescientos por suplantación de identidad. Además, se han atendido ocho mil ochocientas llamadas telefónicas, ocho mil correos electrónicos, así como se han dado de baja ciento noventa páginas *web* por diferentes motivos, mismos que van desde el fraude hasta la publicación de contenido poco apropiado.

Como una solución viable a esta problemática se desarrolló una Aplicación *Web* llamada “SARA” (Seguridad Aplicada a las Redes Sociales), que se encargará de prevenir los riesgos a los que se exponen los usuarios de las redes sociales a través del análisis del texto de sus publicaciones. Después, generará una serie de recomendaciones que ayudarán a que los usuarios identifiquen y corrijan sus malas prácticas al realizar una publicación en las redes sociales. El objetivo de la aplicación *web* SARA es prevenir que los usuarios expongan información que los pueda poner en riesgo, y de esta forma poder proteger su integridad emocional, moral, económica y física; además de crear un grado de conciencia en sus usuarios en términos de los riesgos a los que se pueden exponer por realizar malas prácticas en el uso de las redes sociales. Todo esto para que gradualmente se reduzcan las amenazas latentes y puedan tener una experiencia segura al usarlas o publicar en estas.

En el presente artículo, se describe el desarrollo de la aplicación *web*, el uso de herramientas y algoritmos para la extracción de datos, tales como Graph API Explorer bajo el protocolo OAuth; la metodología que se implementará describiendo el desarrollo de sus fases, así como los resultados y conclusiones obtenidas al desarrollar esta aplicación *web*.

Metodología

El Proceso Unificado Rational (RUP) es un proceso de ingeniería de software que suministra un acercamiento disciplinado en la asignación de tareas y responsabilidades en una organización de desarrollo. Este es utilizado como un marco genérico que se especializa en una gran variedad de tipos de sistemas, diferentes áreas de aplicación, tipos de organizaciones, niveles de aptitud y diferentes tamaños de proyectos. (Martínez, 2002).

Según el autor (Martínez, 2002), el principal objetivo de esta metodología es asegurar la producción de software de alta calidad, llegando a las necesidades del cliente final. Se caracteriza por ser incremental, centrado en la arquitectura y guiado por los casos de uso. El desarrollo del proyecto es realizado por iteraciones, cada una de ellas contiene actividades en varios flujos de trabajo; asimismo, las iteraciones se organizan en cuatro fases: inicio, elaboración, construcción y transición.

A continuación, se describe cada una de las fases de RUP especificando en cada una de ellas lo realizado en el desarrollo de la aplicación *web* SARA.

Fase de inicio

Actualmente, en la sociedad cualquier persona cuenta con una red social, sin embargo, se ignora el tipo de riesgo en el que pueden estar expuestos, de acuerdo con este tipo de situaciones, surge la necesidad de recabar información acerca de los peligros expuestos en las redes sociales.

El objetivo de la aplicación *web* SARA es prevenir que los usuarios expongan información que los ponga en riesgo a los usuarios, y de esta forma proteger su integridad emocional, moral, económica y física.

A continuación, se describen los artefactos generados en esta etapa:

Modelado de negocios

El modelado de negocios lo conforman dos tipos de modelos UML, tales como el diagrama de casos de usos (en este modelo describe la funcionalidad propuesta de la aplicación), y modelo de objetos (especifica las clases y atributos por los cuales está compuesto el sistema, así como la terminología de cada uno de ellos).

Al realizar este modelo de negocios los clientes, usuarios finales y desarrolladores conocen con más detalle cuál es el objetivo del desarrollo del proyecto, así como derivan los requisitos necesarios para el apoyo en la organización del mismo (Martínez, 2002).

Levantamiento de requisitos

Es el segundo flujo de trabajo del proceso de la fase inicial y uno de los artefactos fundamentales en esta etapa. Estos son el comienzo del proyecto, y son los que enriquecen el funcionamiento de la aplicación. En este levantamiento de requisitos se utilizaron técnicas de recolección de datos, tales como la entrevista y la encuesta. Estas fueron aplicadas a jóvenes de entre 15 a 18 años y adultos entre 25 a 40 años.

Fase de elaboración

En esta fase se elaboraron los requisitos a nivel diseño, por tanto, ayuda a conocer si el proyecto es técnicamente viable, así como ayuda a seleccionar las tecnologías emergentes que se utilizarán para el desarrollo del proyecto. Con los requisitos recabados en la primera fase, se diseña la primera solución a la problemática planteada, también con la ayuda de los casos de uso, se define la arquitectura base de la aplicación y se desarrolla el primer análisis del dominio del problema.

Una vez definida y validada la arquitectura, se crea un plan fiable para dar inicio a la fase de construcción. Este plan está conformado por iteraciones, las cuales evolucionan de manera sucesiva. Esta muestra que la arquitectura propuesta soportará la visión de un coste razonable, así como el tiempo adecuado para el desarrollo de la aplicación.

Después de realizar lo anteriormente descrito, se diseña cada uno de los casos de uso, especificando en cada uno de ellos las acciones que sean necesarias para entender completamente los requisitos, dando una estructura al modelo de casos de uso, la cual sea fácil de entender.

En esta fase se generan artefactos, tales como el diagrama de casos de uso, con el cual los desarrolladores y clientes están en acuerdo con los requisitos planteados, especificando el actor o actores que interactuarán con el sistema. En la figura 1, se muestra el DCU (Diagrama de Casos de Uso) de la aplicación *web SARA*, en la que se visualizan a terceros fuera de la aplicación que colaboran en esta, y sirven para identificar el entorno externo a la aplicación mediante casos de uso, así como muestra la forma en la que los actores usan la aplicación y el flujo del suceso. Este diagrama es utilizado en cada uno de los casos, en el cual se muestra una descripción de la secuencia de las acciones del mismo, descripción de la arquitectura en la que se incluyen los casos de uso que describen alguna funcionalidad importante que se presente.

Cabe hacer mención que, el glosario es importante porque define los términos que se utilizan con frecuencia para describir la aplicación; estos son utilizados por el analista y el desarrollador. El prototipo de interfaz de usuario es realizado para la comprensión y especificación de las interacciones entre los actores y la aplicación durante la captura de los requisitos.

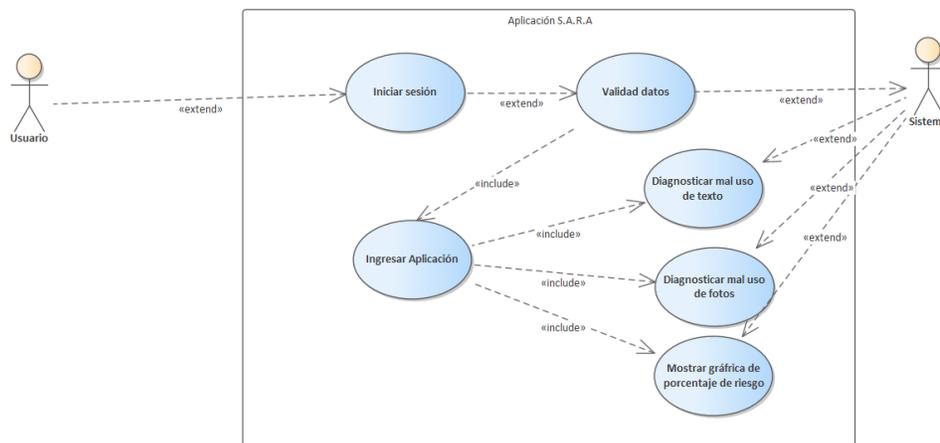


Figura 1. Diagrama casos de uso

Fase de construcción

La finalidad principal de esta fase es alcanzar la capacidad operacional del producto de forma incremental a través de las iteraciones, realizando un enfoque detallado en el diseño, implementación y pruebas para la obtención del resultado. (Martínez, 2002). De acuerdo con lo descrito anteriormente, se inicia con la reducción de costos de desarrollo por medio de la optimización de recursos, evitando trabajos innecesarios.

El desarrollo se realizó a través de iteraciones. En este caso se asignan las tareas específicas a cada uno de los programadores, para que ellos puedan interpretar fácilmente los requisitos de la aplicación con ayuda de los artefactos adquiridos en fases anteriores, tales como los diagramas de casos de uso, clases, el diseño e implementación. En esta etapa se determinó el uso de APIs propias de las redes sociales en conjunto con arreglos en PHP para la extracción de datos, además de explorar algoritmos para el análisis de las palabras y su contexto.

Para la extracción de datos, se utilizó la Graph API Explorer FDK bajo el protocolo OAuth. Primero, se creó una cuenta de Facebook con un perfil sin publicaciones, así como con ninguna configuración de seguridad. Después, la cuenta se registró en Facebook Developers para crear un nuevo proyecto y solicitar permisos en la Graph API (API de Facebook). Posteriormente, se creó un nuevo proyecto con el nombre Graph API. Cabe hacer mención que, cuando se registra un nuevo proyecto como desarrollador en Facebook Developers el proyecto (aplicación) recibe un identificador de la app; es importante conservar este dato porque se utilizará durante la extracción de datos. Luego, se creó una cuenta de hosting gratuito en el sitio 000webhost para subir el proyecto al servidor y se registró a un usuario. Dentro de la cuenta de hosting gratuito en la parte de files se creó una carpeta en la cual se subió el proyecto y después Dentro de la carpeta de public_html, se creó la carpeta Jorge, esta nueva carpeta servirá de contenedor del proyecto.

Luego entonces, dentro de la carpeta anteriormente descrita, se coloca el archivo SDK que servirá de puente con la plataforma de Facebook desde la aplicación PHP. También dentro de esta carpeta se crea un archivo de nombre index.php que es el que contiene el bloque de código en PHP. En este bloque de código se cambia la parte del identificador de la app por el que se le asignó al proyecto en Facebook Developers Una vez realizado lo anterior, se mantiene la cuenta abierta en Facebook Developers, así como se abre el explorador de la Graph API. Finalmente, en la figura 2 se muestra que en la parte inferior derecha se encuentra la opción de agregar un permiso en donde se agregaron algunos de los que se desglosan en la persiana para realizar la extracción de datos.

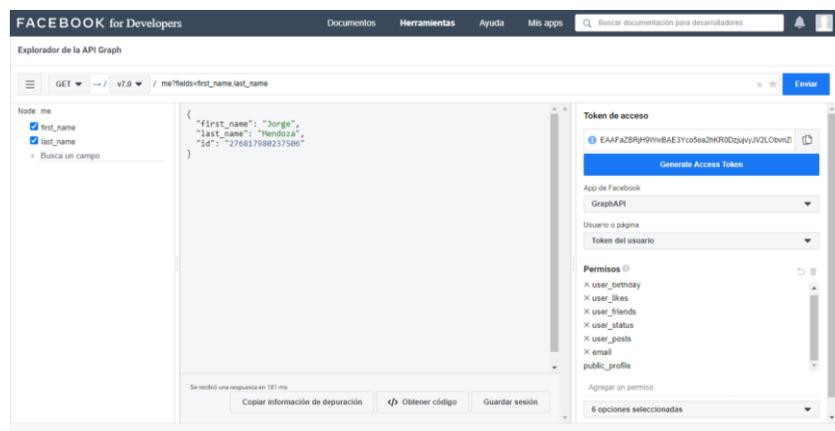


Figura 2. Verificación de datos para la extracción

Además, en esta fase unos de los productos que se generó es la arquitectura integral. En la figura 3 se muestra esta arquitectura, la cual describe el funcionamiento de la aplicación utilizando un modelo en capas. En la capa de presentación, se muestran las tecnologías emergentes utilizadas del lado del cliente. En la capa de integración, se muestran las acciones que realizará el usuario en las redes sociales, tales como el registro y la autenticación. En la siguiente capa, se muestran los servicios proporcionados por la aplicación web, tales como el análisis de las publicaciones, las recomendaciones a los usuarios después de que sus publicaciones hayan sido analizadas, así como las gráficas. En la capa de acceso a los datos, se muestran las tecnologías utilizadas del lado del servidor, así como el uso de API Graph Explorer para realizar la extracción de datos de las publicaciones de los

usuarios en sus perfiles de redes sociales, mismas que serán analizadas a través del algoritmo SARA y el diccionario de palabras a través de arreglos en HTML y PHP para conocer el riesgo de las publicaciones. Finalmente, en la última capa se muestra el almacenamiento en las bases de datos.

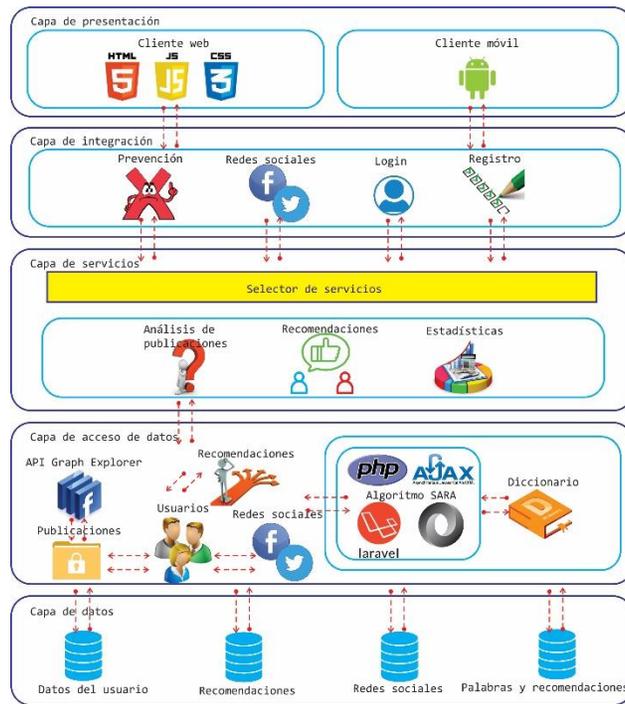


Figura 3. Arquitectura de software utilizando un modelo en capas

Fase de transición

En esta última fase, el producto generado va dirigido hacia los usuarios finales. Se realizaron algunas actividades que incluyen la retroalimentación del usuario de acuerdo con las pruebas realizadas en la etapa de construcción, se realizaron correcciones a la aplicación y se complementó la documentación pendiente basada en la última versión elaborada. En la figura 4, se muestra el formulario corregido con las validaciones correspondientes tomando en cuenta los errores encontrados y la retroalimentación realizada por parte de algunos usuarios.

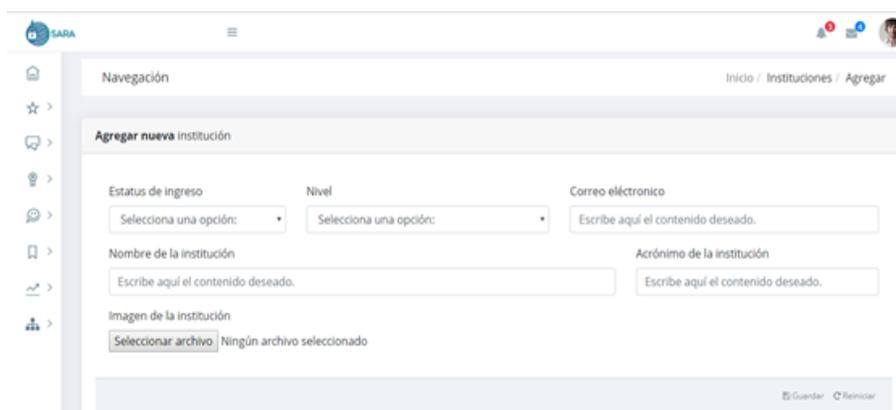


Figura 4. Pruebas de validación

Una de las tareas importantes es capacitar al usuario para que conozca cada uno de los apartados de la aplicación. Esta tarea depende de cada uno de los objetivos planteados al inicio de este proyecto porque conlleva en la configuración, instalación y usabilidad de la aplicación. En la figura 5, se muestra el acceso a aplicación que se puede efectuar a través de un usuario o contraseña o bien utilizando alguna red social (Facebook o Twitter).

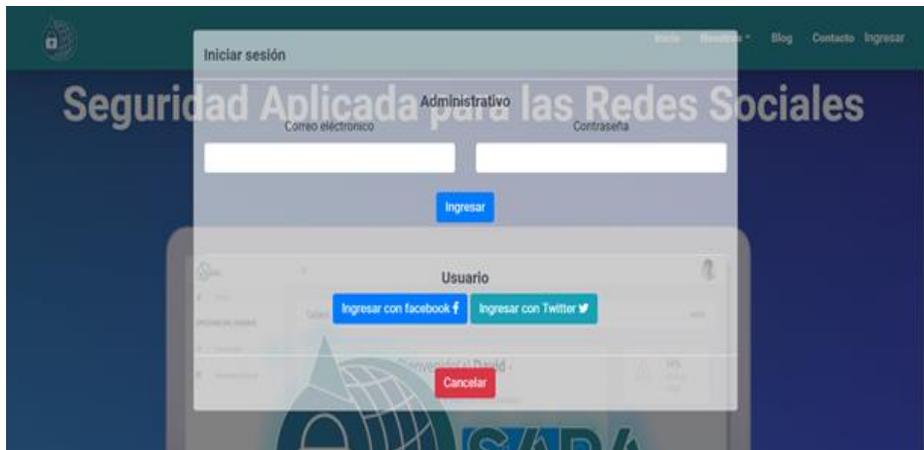


Figura 5. Acceso a la aplicación

Resultados y discusión

Después de haber realizado pruebas en algunos perfiles de usuario Facebook, se comprobó que la extracción de datos se realizó de manera correcta utilizando la API Graph y los permisos del token correspondientes. En la figura 6, se muestra una publicación realizada en Facebook con un número de caracteres mayor, para que posteriormente los datos se extrajeran y analizaran utilizando un diccionario de palabras a través de arreglos HTML y PHP.



Figura 6. Publicación en Facebook con un mayor número de caracteres

Además, para implementar un proyecto de este tipo es importante realizar una efectiva labor de convencimiento en los usuarios participantes basada en el impacto real que tiene el proyecto y los beneficios que se obtendrán. Lo anteriormente descrito se realizó porque la mayoría de las personas no logran dimensionar los riesgos a los que se exponen al publicar cualquier cosa en las redes sociales; en consecuencia, esto ocasiona que se vuelvan vulnerables o se conviertan en víctimas de personas cuyas intenciones no son buenas y pretenden molestar, extorsionar, difamar o atacar al usuario utilizando este medio de comunicación.

Finalmente, como se observa en la figura 7 se extrajeron datos no solo de publicaciones cortas, sino de publicaciones más extensas utilizando Graph API y los permisos del token. Estos datos fueron analizados mediante un diccionario de palabras a través de arreglos en HTML y PHP. Las palabras se almacenan en Arrays y se codifican en formato JSON.

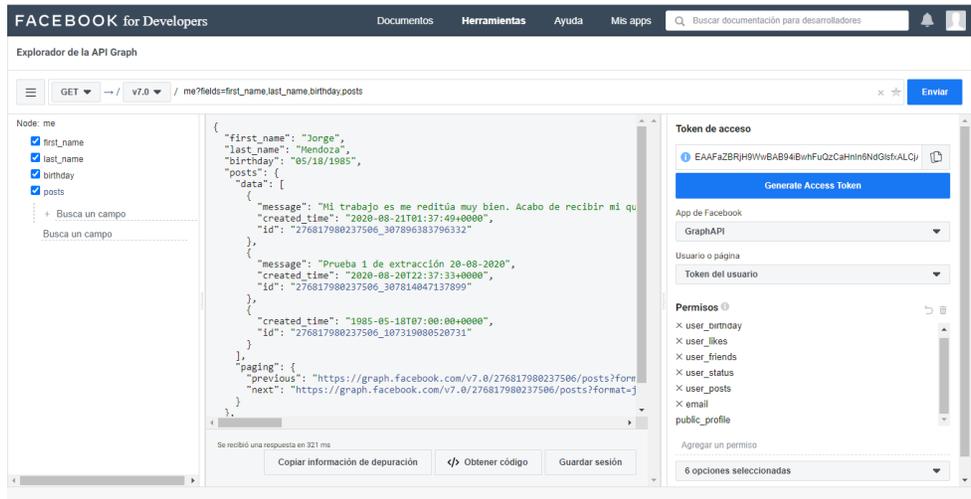


Figura 7. Extracción de datos de una publicación más extensa

Trabajo a futuro

En versiones posteriores de la aplicación, se realizarán mejoras que optimicen el funcionamiento de la misma y aumenten la seguridad de las publicaciones hechas por los usuarios, esto último realizado desde el sistema. Las mejoras contempladas se describen a continuación:

Durante el desarrollo de este proyecto solo se contemplaron las redes sociales Facebook y Twitter para la extracción de datos y el análisis de perfiles. Posteriormente, se integrarán otras redes sociales que permitan el alcance de diferentes personas, pero que de igual forma necesitan seguridad en sus perfiles al publicar.

También se logrará que el usuario pueda hacer publicaciones desde la aplicación de forma más segura, esto a través de filtros de análisis que trabajarán antes de que los comentarios de los usuarios sean publicados, brindando recomendaciones, en los que se indiquen si las palabras utilizadas generan un contexto que los pudiera exponer a posibles riesgos o ataques.

Otra mejora es que las palabras nuevas que no se encuentren en los diccionarios de la aplicación sean aprendidas y agregadas a los diccionarios al momento de encontrarlas en las publicaciones, con el fin de tener un diccionario más amplio que brinde en conjunto con el algoritmo de comparación y filtrado un análisis de las publicaciones más efectivo.

Conclusiones

El impacto actual de una publicación tiene más repercusión de lo que se cree, el contenido que se comparte en ocasiones puede brindar información sobre el estado social y económico de las personas, y esto es aprovechado por personas malintencionadas que buscan obtener beneficios personales, sin importar el daño que puedan ocasionar (Galán, 2015). El uso de la aplicación SARA trae como beneficio a los usuarios un perfil más seguro a través de recomendaciones arrojadas por el análisis de sus publicaciones.

Durante el desarrollo del proyecto se recurrió a la selección de una herramienta para realizar la extracción de datos, tal es el caso de Graph API Explorer bajo el protocolo OAuth. Esta fue de gran ayuda porque se pudieron extraer publicaciones textuales de Facebook tanto cortas como más elaboradas, para que posteriormente estas fueran analizadas mediante un diccionario de palabras a través de arreglos en HTML y PHP. Las palabras se almacenan en Arrays y se codifican en formato JSON.

Además, de acuerdo el análisis realizado y con base en los resultados obtenidos al analizar diferentes perfiles de redes sociales como es el caso de Facebook, se puede concluir que la mayoría de usuarios no es consciente de lo que publica en sus redes sociales, así como no configura la privacidad de su perfil. Por lo anteriormente descrito, se considera necesario el desarrollo de una herramienta tecnológica como la aplicación *web* SARA, para reducir los niveles de riesgo en las publicaciones que realizan los usuarios, así como para brindarles recomendaciones preventivas que ayuden a no convertirlos en víctimas de un ciberdelincuente.

Finalmente, un logro que se alcanzó de forma razonada y reflexiva fue la forma de realizar el análisis de las frases porque las palabras suelen tener un significado de acuerdo al contexto utilizado. De ahí surgió otro aprendizaje en el uso de los filtros para lograr un análisis adecuado. Además, hasta el momento esto se logró realizando un diccionario de palabras a través de arreglos en HTML y PHP que ayudaron a realizar un análisis del riesgo de la publicación que hace una persona en la red social. Esto servirá para que más adelante, se desarrolle un algoritmo propio que sea capaz de realizar comparaciones de acuerdo con las publicaciones extraídas de las diferentes redes sociales.

Referencias

- [1] Bootstrap Documentation, «Bootstrap,» 2019. [En línea]. Available: <https://getbootstrap.com/docs/4.3/getting-started/introduction/>.
- [2] Facebook, «Facebook for Developers,» 2019. [En línea]. Available: <https://developers.facebook.com/docs/graph-api/>.
- [3] P. Galán García, «Supervised machine learning for the detection of troll profiles in twitter social network: application to a real case of cyberbullying,» vol. XXIV, nº 1, pp. 4-9, 2015.
- [4] R. Hernández Sampieri, Metodología de la Investigación, Ciudad de México: Mc Graw-Hill, 2014.
- [5] A. Martínez y R. Martínez, «Guía a Rational Unified Process,» Universidad de Castilla la Mancha, Ciudad Real, 2010.
- [6] V. J. Villanueva Blasco, «Patrón de uso de internet y control parentar de redes sociales como predictor de sexting en adolescentes: una perspectiva de género,» vol. XIV, nº 1, pp. 16-26, 2019.
- [7] R. Ortega Ruiz, «Redes Sociales y Cyberbullying. El proyecto ConRed Convives-Convivencia en la escuela,» vol. III, nº 1, pp. 3-8, 2015.
- [8] Secretaría de Seguridad Ciudadana, «1778 El área de Ciberdelincuencia de la SSC previene el delito en internet,» 14 Julio 2019. [En línea]. Available: <https://www.ssc.cdmx.gob.mx/comunicacion/nota/1778-el-area-de-ciberdelincuencia-de-la-ssc-previene-el-delito-en-internet>.
- [9] Twitter Developer, «API reference index,» 2019. [En línea]. Available: <https://developer.twitter.com/en/docs/api-reference-index>.
- [10] L. Rodríguez García y J. R. Magdalena Benedito, «Perspective of Young People about Security and Privacy on Social Networks,» *Icono 14*, vol. XIV, nº 1, pp. 24-49, 2016.
- [11] F. Minera, PHP Avanzado, Primera ed., Buenos Aires, Fox Andina: Manual Users, 2011.