ISSN 1870-4069

Diseño de un módulo para la detección de ciberbullying en la red social Twitter utilizando lenguaje natural en una aplicación móvil Android implementado en un entorno universitario

Lissete Rosete Rosas[1], Luis Ángel Reyes Hernández[1], Beatriz Alejandra Olivares Zepahua[1], Ignacio López Martínez[1], Laura Angélica Décaro Santiago[2]

[1] Instituto Tecnológico de Orizaba, División de Investigación y Estudios de Posgrado, México

[2] Universidad Autónoma del Estado de México, Centro Universitario UAEM, México

{m16011208, luis.rh, beatriz.oz, ignacio.lm}@orizaba.tecnm.mx, ladecaros@uaemex.mx

Resumen. El ciberacoso es el tipo de acoso que se presenta a través de dispositivos digitales y es un riesgo permanente al que están expuestos los jóvenes debido a la cercanía que tienen con las redes sociales y el Internet. En respuesta al aumento y prevalencia del ciberacoso se ha apostado por implementar técnicas que prevengan este tipo de incidencias, sin embargo, pocas soluciones se encargan de ofrecer una detección temprana de indicios de este fenómeno entre los jóvenes. En este artículo, se propone una arquitectura para el desarrollo de un módulo que detecte situaciones de ciberacoso en la red social Twitter empleando procesamiento de lenguaje natural, dicho modulo será implementado en una aplicación móvil base desarrollada para Android, que tiene como objetivo detectar ocurrencias de bullying tradicional. Además, se describe la obtención de la bolsa de palabras comprendidas en el lenguaje verbal violento de las redes sociales empleado por jóvenes universitarios; así como los diseños de interfaces para el desarrollo del módulo de una aplicación de monitorización de casos de ciberacoso para contribuir en la reducción de ocurrencias de este fenómeno.

Palabras clave: Aplicación móvil, bolsa de palabras, ciberacoso, procesamiento de lenguaje natural.

Design of a Module for the Detection of Cyberbullying in the Social Network Twitter Using Natural Language in an Android Mobile Application Implemented in a University Environment

Abstract. Cyberbullying is the type of bullying that occurs through digital devices and is a permanent risk to which young people are exposed due to the

proximity they have to social networks and the Internet. In response to the increase and prevalence of cyberbullying, there has been a commitment to implement techniques to prevent this type of incident, however, few solutions are responsible for providing early detection of signs of this phenomenon among young people. In this article, we propose an architecture for the development of a module that detects situations of cyberbullying in the social network Twitter using natural language processing, this module will be implemented in a mobile application developed for Android, which aims to detect occurrences of traditional bullying. In addition, it is described the obtaining of the bag of words included in the violent verbal language of social networks used by young university students; as well as the interface designs for the development of the module of an application for monitoring cases of cyberbullying to contribute to the reduction of occurrences of this phenomenon.

Keywords: Mobile app, bag of words, cyberbullying, natural language processing.

1. Introducción

En la actualidad la sociedad se encuentra sumergida en una constante evolución tecnológica, donde el Internet y las comunicaciones son los protagonistas. Estos cambios en la sociedad impactan en el contexto social donde los jóvenes o adolescentes se desenvuelven e interactúan; así como existen numerosas ventajas que trae consigo la pequeña brecha entre la tecnología y los adolescentes, también existe la preocupación de que el entorno se vuelva dañino, pues pueden ocurrir situaciones de riesgo en las que el estado emocional, físico o mental de los adolescentes se vea afectado. Por otro lado, la violencia y agresión en el sector educativo es un problema persistente alrededor del mundo; existe evidencia de esfuerzos para prevenir el bullying y ciberbullying, además de acciones para corregir esta grave situación.

En primer lugar, el cyberbullying ha tomado diversas definiciones a lo largo de su prevalencia y como es mencionado en el trabajo de Chun [1] aún existen debates sobre la definición exacta o final del término de ciberbullying; una definición de ciberbullying desarrollada a partir de la definición de bullying tradicional, es la propuesta por Hinduja y Patchin [2] que conceptualizan al fenómeno como "un daño intencional y repetido infligido mediante el uso de computadoras, teléfonos celulares y otros dispositivos electrónicos". Por lo que, sin duda, el objetivo principal de este fenómeno consiste en el uso de medios digitales para acosar, intimidad, amenazar, amedrentar o molestar a una persona o un grupo de personas mediante ataques personales o divulgación de información personal privada o falsa.

De acuerdo con el módulo sobre ciberacoso 2021 del Instituto Nacional de Estadística y Geografía (INEGI) [3] el 21.7% de la población usuaria de Internet fue víctima de ciberacoso. Además, se estima que el ciberacoso más frecuente se encuentra relacionado con el aspecto físico, a la forma de vestir y al estilo de vida.

El desarrollo de aplicaciones móviles ha ido en constante aumento, pues cada vez existen más aplicaciones que tratan de resolver una problemática en la sociedad y a su vez, son herramientas con una gran accesibilidad, debido a su uso en dispositivos informáticos inalámbricos pequeños, como teléfonos o *tablets* [4]. Por consiguiente, debido al incremento y perseverancia del ciberacoso es necesario crear herramientas

que prevengan o detecten este tipo de situaciones, con la finalidad de disminuir la cantidad de personas afectadas. Las redes sociales son el lugar más común donde las personas son intimidadas, acosadas o ridiculizadas, lo que puede tener graves consecuencias para la salud mental y emocional de las víctimas. Twitter es de las redes sociales más violentas, pues en su plataforma existen miles de casos de ciberacoso; además, la poca o nula censura en los tweets pueden indicar un alto potencial de incidentes de ciberacoso [5].

Por lo que, se pretende desarrollar un módulo que detecte situaciones de ciberacoso en la red social Twitter, este módulo será implementado en una aplicación base desarrollada para Android en el tema "Aplicación móvil de monitorización de bullying en salones de clase empleando DSM-V y el algoritmo FOAF" culminado por el alumno egresado de la maestría en sistemas computacionales, René Navarrete Tenco. La aplicación base se enfoca como medio de recolección de datos para realizar la detección de eventos de bullying dentro de un salón de clases, por medio de la observación del docente, para posteriormente hacer la monitorización del caso hasta llegar a la posible solución del evento.

En la literatura analizada existen trabajos que implementan soluciones que tienen como objetivo prevenir situaciones de ciberbullying, pero tienen la limitante de solo ser informativas. De igual forma, las aplicaciones desarrolladas para que identifiquen ciberbullying [6–11], se encuentran destinadas a usarse en países diferentes a México, por lo que el idioma es un limitante, ya que el vocabulario cambia y también cambia la manera en la que las personas se expresan de forma maliciosa.

La estructura de este artículo es la siguiente: la sección 2 abarca el estado del arte englobando los trabajos más relacionados con la detección y puntos clave del ciberacoso, en la sección 3 especifica la descripción de la arquitectura propuesta para el desarrollo del módulo sobre ciberbullying, en la sección 4 se presenta la construcción del instrumento y los resultados obtenidos para formar la bolsa de palabras que contemple las palabras o frases empleadas por los adolescentes para agredir u ofender a alguien más, a través de redes sociales; los mockups realizados para ilustrar una aproximación a las interfaces del módulo de la aplicación se exponen en la sección 5, en la sección 6, se presenta la discusión de los resultados obtenidos y finalmente en la sección 7, las conclusiones y el trabajo a futuro.

2. Trabajos relacionados

En esta sección se analiza la literatura identificada más relacionada con el ciberacoso, los factores involucrados y los trabajos de previas propuestas para la prevención y detección de este fenómeno.

En primer lugar, en [12] se obtuvieron perfiles psicológicos de adolescentes donde se enfatizaba que los jóvenes con un alto índice en conducta antisocial, tenían mayor uso de las estrategias agresivas para la resolución de conflictos y un mayor nivel en las posibilidades de estar implicados en situaciones de bullying/cyberbullying; en cualquiera de los roles posibles, víctimas, agresores y observadores. Por otro lado, en [13] se propuso un léxico de ciberacoso compuesto de 13 palabras clave (tomadas de Enchanted Learning) acompañadas de su definición y su categoría. Se propuso que el léxico presentado, tiene la capacidad de ser utilizado como diccionario en las redes

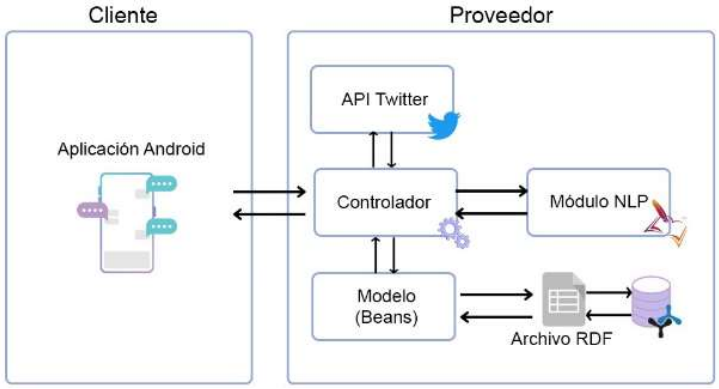

Fig. 1. Patrón arquitectónico MVC propuesto para el desarrollo del módulo de la aplicación.

sociales para considerar palabras o frases que insinúen una situación de posible ciberacoso. Noviantho *et al*. [14] construyó un modelo de clasificación para identificar conversaciones de ciberacosadores, empleando el método de minería de texto Naive Bayes y Máquinas de Vector Soporte.

El método de clasificación obtuvo 4 clases con una precisión promedio del 92.81% para Naive Bayes y un 97.22% para SVM; se concluyó que el modelo más óptimo para separar la muestra en diferentes clases fue SVM. Por su parte, en Upadhyay, *et al*. [6] se diseñó un sistema que emplea las publicaciones y los comentarios de los usuarios en Facebook, para clasificar el contenido en alguna de las categorías que se plantearon. El análisis se llevó a cabo mediante el algoritmo de Procesamiento del Lenguaje Natural permitiendo detectar los sentimientos de los usuarios y bloquear de inmediato el contenido que sea dañino para los usuarios.

De acuerdo con las investigaciones de Farag *et al*. [15] se determinó que los métodos distintos a los supervisados empleados para la detección del ciberacoso eran: codificadores automáticos, aprendizaje profundo, semisupervisado, modelado de series temporales y clustering (agrupamiento), pues se emplearon en diferentes plataformas, tales como MySpace, Twitter, YouTube, Formspring, ask.fm e Instagram para detectar situaciones de acoso.

El resultado de la investigación, demostró que el método para detectar situaciones de cyberbullying depende de la plataforma que se esté analizando. En [7] se desarrolló un modelo automatizado para identificar y medir el grado de cyberbullying en las redes sociales, además del diseño de una aplicación de Facebook que notificaría a los padres en caso de que los adolescentes sean víctimas de ciberacoso.

Como resultado del proyecto, se obtuvo la aplicación *BullyBlocker*, que identifica si un adolescente se encuentra implicado en una situación de cyberbullying y brinda información a los padres del mismo para tomar las medidas adecuadas. De forma similar, Salawu *et al*. [8] propuso una aplicación móvil diseñada para detectar y prevenir el ciberacoso en las redes sociales.

Dicha aplicación se integró por una arquitectura en la que se utiliza un modelo de aprendizaje profundo generalizado, empleado para identificar casos de ciberacoso a partir de la información básica del usuario; el modelo se entrenó para predecir etiquetas de ciberacoso y cuando lleguen nuevos mensajes o comentarios estos se clasifican de

acuerdo al modelo y si son indicios de algún tipo de ciberacoso, se eliminan o se bloquea al remitente temporalmente.

En [9] se diseñó y desarrolló una aplicación móvil (nombrada #StopBully) enfocada en mejorar la comprensión de los conceptos de bullying y cyberbullying entre los estudiantes a través del entretenimiento educativo. La aplicación se diseñó en Android y cuenta con un botón de emergencia, el cual comunica al usuario con alguna autoridad capacitada para denunciar algún caso de bullying/cyberbullying.

Foong y Oussalah [10] propusieron un sistema online que permitió la detección y el seguimiento de casos de ciberacoso en foros y comunidades en línea; este sistema se conformó por componentes básicos del lenguaje natural, contemplando insultos, amenazas y oraciones escritas en segunda persona. Se usó un sistema de aprendizaje automático y ontologías para clasificar la aparición de ese tipo de vocabulario, con la finalidad de que se emita un mensaje a la seguridad del sitio y este tome medidas necesarias.

El objetivo de [11] fue examinar los métodos existentes de detección del ciberacoso para integrarlos en una aplicación móvil que alerta a los padres en caso de que su hijo sea una potencial víctima o autor de ciberacoso. La aplicación se conformó por dos componentes, el de la aplicación móvil y el componente de aprendizaje automático. El modelo que se empleó para incorporar al aprendizaje automático fue una recolección de datos extraídos de Twitter. Teniendo como resultado buenas métricas incluidas la precisión, exhaustividad y exactitud que mostraron una eficacia notable del modelo propuesto.

3. Arquitectura

En esta sección se presenta la arquitectura propuesta para el desarrollo del módulo. La aplicación base está diseñada empleando el patrón de diseño MVC (*Model-View-Controller*, Modelo-Vista-Controlador); por consiguiente, el módulo continúa empleando esta arquitectura. MVC es un patrón de diseño ampliamente utilizado para crear aplicaciones web y es usado en casi todos los marcos de desarrollo web [16]. La arquitectura se muestra en la figura 1.

Para efectos de este proyecto se empleará la red social Twitter debido a las ventajas que proporciona, como sus pocas restricciones de privacidad ya que la gran mayoría de usuarios tiene público su perfil. La API de Twitter se puede utilizar para recuperar y analizar datos de Twitter de forma programática [17], además de que proporciona facilidades para ejecutar sin restricciones un gran número de acciones.

En MVC, el modelo consta de los objetos de dominio que modelan los problemas del mundo real. Las vistas contienen el código que permitirá la visualización de las interfaces con las que el usuario interactúa y los controladores contienen el código que permite responder a las acciones que se solicitan a la aplicación y que, de este modo, se vean reflejadas en las vistas [18].

La aplicación se encuentra desarrollada en dos partes, el lado del cliente y el lado del proveedor, lo que permite una escalabilidad factible para implementar nuevas necesidades que se requieran.

3.1. Cliente

Se le conoce de esta forma al dispositivo que realiza peticiones y consume los servicios generados por el proveedor a través de la aplicación. El cliente dispone la vista de la aplicación que está conformada por interfaces compuestas por formularios, listas, imágenes, tablas, por mencionar algunos; contiene procedimientos relacionados con la interfaz de usuario [19].

La aplicación móvil base se encuentra desarrollada para el sistema operativo Android por medio del Entorno de Desarrollo Integrado (IDE) Android Studio, que integra archivos XML (*Extensible Markup Language* – Lenguaje de Marcado Extensible); además se plantea utilizar Java como lenguaje de programación para el consumo de servicios.

3.2. Proveedor

La aplicación base se encuentra gestionada por servicios web basados en REST. Existen diferentes alternativas de tecnologías para desarrollar este tipo de servicios, para el desarrollo del módulo se optará por el marco de servicios *Jersey RESTful Web Service in Java* por medio del lenguaje de programación Java.

- Modelo: Representa la lógica de la aplicación en distintas clases encargadas de estructurar e interactuar con la información de los usuarios y eventos de posible ciberacoso.
- Controlador: Este componente se encarga de gestionar, atender y procesar las solicitudes recibidas del cliente. Con esto, el modelo y la vista se comunican para solicitar, procesar y pasar los datos necesarios a la vista para que pueda mostrarlos. Al ser un proveedor de servicios web basados en REST, tanto el proveedor como el cliente, devolverán y recibirán los datos en formato ligero para el intercambio de datos Notación de Objetos JavaScript (JSON). En el controlador se contempla la implementación de la conexión con la API de Twitter, en donde a través del registro del username de cada alumno se obtendrán los tweets emitidos respectivamente; por lo cual, para realizar solicitudes HTTP se necesita de una conexión segura utilizando la autenticación OAuth 2.0 por medio del Baerer Token, el cual es un tipo de token de seguridad compuesto de una serie de caracteres alfanuméricos que se utiliza para identificar al usuario y otorgar acceso a recursos protegidos en la plataforma de Twitter.
- Vista: El cliente descrito en la sección 3.1 será responsable de disponer la interfaz de la aplicación visible para el usuario. La vista (archivo de diseño XML) muestra los datos del modelo al usuario.

Por otro lado, para el módulo de NLP (*Natural Language Processing*), se seleccionó la herramienta Apache OpenNLP, debido a la gran cantidad de documentación y herramientas que dispone, además de la buena combinación con el entorno de desarrollo NetBeans que provee Apache. Esta tecnología, cuenta con la detección de idiomas, por lo que es adecuada para el desarrollo de este proyecto debido a que se enfocará en el idioma español.

Los tweets recuperados de los alumnos, serán preprocesados para limpiar el texto de ruido y estandarizar dicho texto para que sea fácil de comparar con respecto a la bolsa

Violencia verbal entre jóvenes

Objetivo: El objetivo del presente cuestionario, es identificar frases o palabras ofensivas dentro del entorno virtual-digital como parte del fenómeno del ciberbullying.

Los resultados serán utilizados únicamente con fines académicos y de investigación. Su contribución apoyará al desarrollo de estrategias para disminuir este fenómeno entre jóvenes, por lo que su participación es muy importante. Cabe mencionar, que los datos obtenidos serán utilizados de forma anónima.

Fig. 2. Formulario creado en Google Forms.

1. ¿Qué frases o palabras has identificado, con el propósito de discriminar a una persona?	2. ¿Qué chistes has identificado cuyo objetivo sea ridiculizar a una persona?	3. Si has observado alguna ironía que tenga por objetivo humillar a una persona, descríbela.	4. ¿Con que frases o palabras has identificado que una persona amenaza a otra?
negro, negra, azteca, deberías s	esquizofrénico, así hacía mi primo	Cuando dicen "Quiero tu autoestim	Cuando dicen que los van a dox
Negro, gorda, anorexica, delgada	Eres un enano	Cuando el chic@ es pobre y se en	Te asesinare
Naco, pendejo, estupido	Burlas hacia el físico de las perso	Cuando hablan sobre una mujer	Tu que
Negrata, jodido, indio, puto, mari	Conozco 5 gordos y tú eres 4 de e	Prófugo del ácido folico.	Desearás que tu madre te haya
Negro, mexicano, retrasado	Eres una mierda	Decirle que no tiene cerebro o que	Te voy a matar
Las mujeres no pueden porque s	Ninguno	Evidenciarlo delante de gente	Aquí no te hago nada porque ha
Negro,Gei,pobre,feo	Machistas,clasistas,humor negro e	No sé que es ironía	No quiero que hables con nadie
Comentarios machistas	Chistes sobre el color de piel de la	No la he notado	No he visto amenazas
Ese es bastante feo	Jajajajaja mira este chocolate se p	No ninguna	Vete con cuidado conmigo
Negro	Te ves muy demacrado	Cuando te tiran una indirecta de có	Te voy a agarrar a golpes

Fig. 2. Previsualización de las respuestas obtenidas.

de palabras. Cabe señalar que en la etapa actual del desarrollo del proyecto se encuentra realizándose el análisis para la selección de la técnica NLP más adecuada a las necesidades del proyecto.

4. Bolsa de palabras del lenguaje verbal violento

En esta sección se detalla la creación del instrumento para obtener las palabras o frases con las que los jóvenes se expresan violentamente a través de medios digitales, así como el resultado obtenido, contemplado en una lista o bolsa de palabras que se empleará para alimentar el algoritmo que permita identificar eventos de ciberacoso.

Ya que el objetivo de este proyecto es la detección de ciberbullying a través de lenguaje natural, será analizada la parte textual de los medios digitales y es necesario obtener la bolsa de palabras empleadas por los jóvenes para agredir, humillar, desacreditar u ofender a alguien más.

En México no se contemplan diccionarios o bolsas de palabras que permitan identificar las agresiones verbales a través de medios digitales por parte de los jóvenes o adolescentes. Para lo cual, como primer punto se clasificaron las dimensiones a abordar en relación con las manifestaciones del ciberbullying; de esta manera, se contempló la dimensión de burla, amenaza e insulto.

La tabla 1 expone cada una de las dimensiones a abordar y para que dichas dimensiones se puedan operacionalizar se estableció su significado dirigido a este proyecto, así como los medios a través de los cuales se manifiesta y los resultados o consecuencias del acto.

Tabla 1. Dimensiones de las manifestaciones de ciberbullying.

Dimensión	**Significado**	**Medio**	**Resultado**
Burla	Es el acto o conducta de provocar la vergüenza de una persona por diversión para ridiculizarlo; puede ser divertido u ofensivo según la ambigüedad de la situación, debiéndose a las interacciones personales de los humanos.	– Chistes – Bromas – Ironía	– Herir – Humillar – Ridiculizar – Desacreditar – Discriminar
Amenaza	Se refiere a la acción de expresar o hacer algo que sugiere la posibilidad de causar daño o peligro a alguien o algo. El objetivo de la amenaza es infundir miedo o intimidación en la otra persona para lograr algún resultado deseado.	– Frases que infunden miedo. – Ataque a la vulnerabilidad de la persona.	– Intimidar – Controlar
Insulto	Los insultos son utilizados de manera intencional para causar daño emocional a otra persona. Puede ser una palabra, frase, comentario o gesto que se utiliza para atacar a alguien de manera ofensiva e hiriente.	– Apodos – Palabras anti sonantes – Palabras despectivas	– Ofender – Menospreciar – Herir – Humillar – Discriminar – Excluir

Tabla 2. Ítems clasificados por dimensión.

Dimensión	Ítems
Burla	En las redes sociales, por medio de los chats, en comentarios, publicaciones… – ¿Qué frases o palabras has identificado, con el propósito de discriminar a una persona? – ¿Qué chistes has identificado cuyo objetivo sea ridiculizar a una persona? – Si has observado alguna ironía que tenga por objetivo humillar a una persona, descríbela.
Amenaza	En las redes sociales, por medio de los chats, en comentarios, publicaciones… – ¿Con que frases o palabras has identificado que una persona amenaza a otra? – ¿Qué frase considerarías una amenaza? – ¿Qué frase considerarías una amenaza con la finalidad de controlar a otra persona?
Insulto	En las redes sociales, por medio de los chats, en comentarios, publicaciones… – ¿Cuál es el insulto más común que hayas visto? – ¿Cuál es el insulto más ofensivo que hayas visto? – ¿Qué groserías has identificado? – ¿Qué insultos, a través de abreviaturas has visto? Y ¿Cuál es su significado? – ¿Qué apodos ofensivos has visto?

Ya delimitadas las manifestaciones de ciberacoso a abordar, se formularon una serie de preguntas enfocadas a cada una de las dimensiones, teniendo un total de 11 ítems (tabla 2) creadas con el propósito de formar un cuestionario.

Objetivo del cuestionario: Recolectar palabras y frases con la finalidad de insultar, burlarse o amenazar a través de las redes sociales por parte de jóvenes universitarios.

Palabra	Total	Largo	1. Dimensión burla	%	2. Dimensión amenaza	%	3. Dimensión insulto	%	Total %
pendejo	179	7	8	0.21%	3	0.07%	168	4.16%	1.96%
puta	130	4	0	0.00%	3	0.07%	127	3.15%	1.42%
madre	127	5	0	0.00%	19	0.47%	108	2.68%	1.39%
negro	110	5	64	1.66%	1	0.03%	45	1.11%	1.20%
verga	99	5	1	0.03%	5	0.12%	93	2.30%	1.08%
hijo	82	4	2	0.05%	2	0.05%	78	1.93%	0.90%
puto	80	4	10	0.26%	0	0.00%	70	1.73%	0.88%
gordo	65	5	29	0.75%	0	0.00%	36	0.89%	0.71%
idiota	61	6	6	0.16%	1	0.03%	54	1.34%	0.67%
persona	54	7	30	0.78%	11	0.27%	13	0.32%	0.59%
chinga	53	6	0	0.00%	1	0.03%	52	1.29%	0.58%
color	51	5	45	1.17%	0	0.00%	6	0.15%	0.56%
ctm	45	3	0	0.00%	0	0.00%	45	1.11%	0.49%
físico	36	6	19	0.49%	1	0.03%	16	0.40%	0.39%
mierda	36	6	1	0.03%	1	0.03%	34	0.84%	0.39%
más	35	3	8	0.21%	14	0.34%	13	0.32%	0.38%
ninguno	35	7	8	0.21%	0	0.00%	27	0.67%	0.38%
gorda	33	5	15	0.39%	0	0.00%	18	0.45%	0.36%

Fig. 3. Previsualización de la bolsa de palabras.

Posteriormente se añadieron las preguntas a un cuestionario en Google Forms (figura 2) con la finalidad de que su distribución fuera más accesible y se distribuyó el enlace con jóvenes universitarios de la licenciatura en Administración de la Universidad Autónoma del Estado de México y de estudiantes de las ingenierías de Sistemas Computacionales, Gestión Empresarial y Química del Instituto Tecnológico de Orizaba. Enlace al cuestionario: https://forms.gle/t8sXH1pHbP4wHkDX9.

La recopilación de la información fue autoadministrada y se llevó a cabo en un periodo de aproximadamente dos semanas; gracias a la herramienta de Google Forms, las respuestas se almacenaron en hojas de cálculo de Excel y se obtuvo una muestra total de 213 jóvenes universitarios que respondieron a cada uno de los cuestionamientos planteados (figura 3). Con el propósito de incrementar la bolsa de palabras, se pretende obtener actualizaciones contemplando un incremento de participantes; por consiguiente, se agregarán al análisis las nuevas entradas de respuestas al cuestionario.

Las respuestas obtenidas fueron tratadas a través del software Atlas.ti [20] realizando un análisis cualitativo para determinar la repetitividad de las palabras y contemplar las palabras más comunes en el lenguaje verbal violento empleado en el grupo poblacional comprendido por jóvenes universitarios.

De igual forma se analizó la repetitividad de las palabras en cada una de las dimensiones comprendidas y posteriormente se realizó la depuración de aquellas palabras que no se relacionaban directamente con el propósito inicial como, por ejemplo, conectores y artículos. Con la finalidad de formar la bolsa de palabras a utilizar para la detección de ciberacoso (figura 4).

De igual forma se realizó el análisis por cada dimensión abordada, obteniendo una lista final por cada una de estas.

– Dimensión Burla: Se obtuvo una mayor cantidad de palabras, ya que en esta categoría se obtuvieron frases más largas por parte de los encuestados; finalmente la lista de palabras contemplo un total de 931 palabras.

Fig. 4. Vista de carga, inicio de sesión para la aplicación y menú principal.

Fig. 5. Menú principal del módulo, interfaz de grupo y envío de solicitud.

– Dimensión Amenaza: Filtrando y depurando conectores y artículos, se obtuvo un total de 522 palabras.
– Dimensión Insultos: En esta categoría se contemplan 519 palabras.

Cabe señalar que en el conteo se contemplaron palabras en singular, plural y sustantivos o adjetivos tanto en masculino como en femenino, ya que son las diferentes maneras en las que se pueden presentar en un contexto determinado.

5. Mockups

A continuación, se presentan los mockups utilizados para representar cómo se verá el diseño final del módulo de la aplicación en su contexto real. Cabe señalar que se continuará trabajando con el diseño de la aplicación base, por lo que se utiliza la misma paleta de colores.

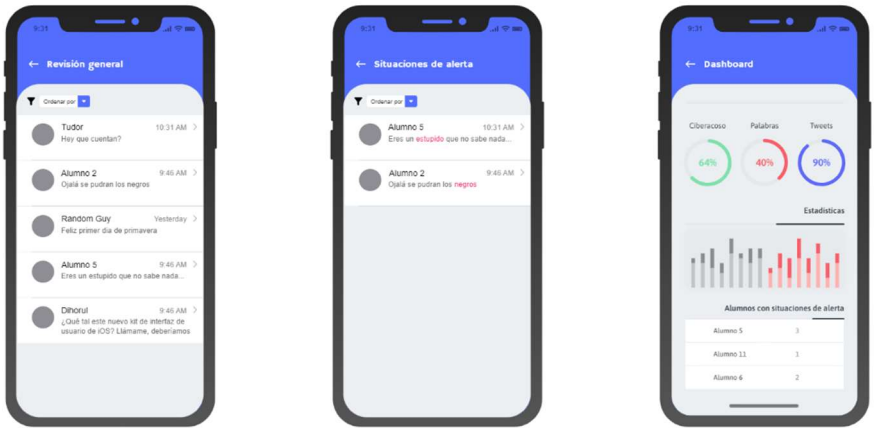

Fig. 6. Interfaz de la revisión general, situaciones de alerta y *dashboard*.

Interfaz de carga de aplicación (*Splash Screen*) que muestra la leyenda identificativa del proyecto "Monitorización de bullying y ciberbullying". Por otro lado, la interfaz de inicio de sesión muestra la solicitud de los campos necesarios para ingresar a la aplicación (correo electrónico y contraseña). Una vez que se autentifique el usuario, se muestra la pantalla principal de la aplicación con las herramientas utilizadas para la sección comprendida por el bullying.

En la parte inferior se muestran cuatro opciones, contemplando la opción de ciberacoso, cuya navegación se dirige al módulo del proyecto (figura 5). En la interfaz del menú principal del módulo sobre ciberacoso, se muestran las herramientas principales acompañadas de la bienvenida al usuario; estas herramientas, son: Grupo, Revisión general, Situaciones de alerta y *Dashboard*.

En la sección de grupo la interfaz presentada, mostrará la lista de los alumnos del grupo, con la opción de añadir su perfil de la red social (*Twitter*), la cual mostrará una ventana flotante con el campo de entrada de texto para colocar el *username* del alumno y el botón de enviar solicitud para que el alumno acepte los permisos necesarios (figura 6).

En la sección de "Revisión general" se mostrarán todas las publicaciones o comentarios que han realizado los alumnos, con la posibilidad de ordenarlas por más recientes o más antiguas comprendidas en un lapso de tiempo. En la interfaz de "Situaciones de alerta", se mostrarán aquellas publicaciones que detecten eventos de violencia verbal y que permita llevar un seguimiento para determinar un posible caso de ciberacoso. Por último, se presenta el *Dashboard* con algunos datos significativos (figura 7).

Es importante hacer mención de que pueden existir nuevas interfaces de acuerdo a las necesidades que se presenten a lo largo del desarrollo del proyecto; además se contempla añadir reportes generados por la aplicación. Las interfaces que si fueron mostradas se consideran las más importantes a incluir para el desarrollo del módulo de la aplicación.

6. Discusión

La arquitectura propuesta para el desarrollo del módulo sobre detección de ciberacoso describe cómo los diferentes componentes se relacionan entre sí y cómo se comunican entre sí para lograr los objetivos del sistema, la arquitectura de software es importante porque ayuda a planificar y organizar el desarrollo del software de manera efectiva y eficiente.

Por otro lado, la clasificación de las manifestaciones comprendidas por el ciberacoso, permitió identificar la delimitación de este proyecto, contemplando las manifestaciones de violencia verbal a través de textos; además, permitió la construcción del instrumento de recolección de las palabras o frases empleadas por los adolescentes para expresarse de forma agresiva u ofensiva en las redes sociales o medios digitales.

Los resultados obtenidos del análisis cualitativo aplicado a las respuestas correspondientes a la muestra a través de Atlas.ti, ayudó en la determinación de la bolsa o lista de palabras a emplear para identificar o detectar posibles situaciones de ciberacoso por medio del procesamiento del lenguaje natural. De igual forma, los mockups presentados brindan una apreciación previa de la distribución de elementos, diseños e interfaces con los que contará el módulo de la aplicación, respetando el diseño de la aplicación móvil base.

7. Conclusiones y trabajo a futuro

A pesar que el fenómeno de cyberbullying no es algo nuevo, tiene un aumento considerable en los últimos años, ya que la pandemia de COVID-19 iniciada en 2020 provoco que los adolescentes se encontraran más presentes en las redes sociales y a su vez más expuestos a sufrir algún tipo de acoso.

En el análisis presentado se hace notar que actualmente existen aplicaciones que brindan información sobre la prevención del ciberacoso, dando consejos o recomendaciones que tienen como objetivo prevenir este tipo de acoso entre los jóvenes; asimismo, existen aplicaciones que detecten casos de ciberacoso pero dichas aplicaciones se encuentran destinadas a usarse en otros países, lo que conlleva que el lenguaje sea diferente al español, cambiando también las palabras o términos que se suelen usar para ofender o humillar a una persona.

Debido a esto y con el objetivo de disminuir el ciberacoso, en este artículo se presentó la arquitectura para desarrollar un módulo implementado en una aplicación móvil base, que mediante el procesamiento de lenguaje natural identifique ocurrencias de ciberacoso; de igual forma, se abordó la obtención de la bolsa de palabras que se empleará para el algoritmo que permita detectar posibles situaciones de este fenómeno ya abordado, así como los mockups principales que se involucran en las vistas del sistema abarcado por el módulo a desarrollar.

Como trabajo a futuro se contempla realizar la conexión con la API de la red social Twitter a través del protocolo oAuth con las credenciales correspondientes, además de integrar algoritmos que sean capaces de detectar posibles casos de ciberbullying, lo que consecuentemente dirige al desarrollo el módulo de la aplicación.

Agradecimientos. Se agradece al Tecnológico Nacional de México por el apoyo otorgado, mencionando al Instituto Tecnológico de Orizaba por ser el anfitrión del desarrollo de este proyecto. Este proyecto cuenta con el apoyo del Consejo Nacional de Ciencia y Tecnología (CONACyT).

Referencias

1. Chun, J. S., Lee, J., Kim, J., Lee, S.: An international systematic review of cyberbullying measurements. Computers in Human Behavior, vol. 113 (2020) doi: 10.1016/j.chb.20 20.106485
2. Hinduja, S., Patchin, J. W.: Cyberbullying: An exploratory analysis of factors related to offending and victimization. Deviant Behavior, vol. 29, no. 2, pp. 129–156 (2008) doi: 10.1080/01639620701457816
3. INEGI: Módulo sobre ciberacoso. Comunicado de prensa, no. 364 (2022)
4. Weichbroth, P.: Usability of mobile applications: A systematic literature study. IEEE Access, vol. 8, pp. 55563–55577 (2020) doi: 10.1109/ACCESS.2020.2981892
5. Ho, S. M., Kao, D., Chiu-Huang, M. J., Li, W., Lai, C. J.: Detecting cyberbullying "Hotspots" on Twitter: A predictive analytics approach. Forensic Science International: Digital Investigation, vol. 32 (2020) doi: 10.1016/j.fsidi.2020.300906
6. Upadhyay, A., Chaudhari, A., Arunesh, Ghale, S., Pawar, S.: Detection and prevention measures for cyberbullying and online grooming. In: Proceedings of the International Conference on Inventive Systems and Control, ICISC, pp. 1–4 (2017) doi: 10.1109/ICISC. 2017.8068605
7. Silva, Y. N., Hall, D. L., Rich, C.: BullyBlocker: Toward an interdisciplinary approach to identify cyberbullying. Social Network Analysis and Mining, vol. 8, no. 18, pp. 1–15 (2018) doi: 10.1007/s13278-018-0496-z
8. Salawu, S., He, Y., Lumsden, J.: BullStop: A mobile app for cyberbullying prevention. In: Proceedings of the 28th International Conference on Computational Linguistics: System Demonstrations, 70–74 (2021). doi: 10.18653/v1/2020.coling-demos.13
9. Neo, H. F., Teo, C. C., Han-Boon, J. L.: Mobile edutainment learning approach: #stopbully. In: ICDTE: Proceeding of ten 2nd International Conference on Digital Technology in Education, pp. 6–10 (2018) doi: 10.1145/3284497.3284500
10. Foong, Y. J., Oussalah, M.: Cyberbullying system detection and analysis. In: Proceedings - 2017 European Intelligence and Security Informatics Conference, EISIC, pp. 40–46 (2017) doi: 10.1109/EISIC.2017.43
11. Thun, L. J., Teh, P. L., Cheng, C. Bin: CyberAid: Are your children safe from cyberbullying? Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 7, pp. 4099–4108 (2021) doi: 10.1016/j.jksuci.2021.03.001
12. Garaigordobil, M.: Conducta antisocial: Conexión con bullying/cyberbullying y estrategias de resolución de conflictos. Psychosocial Intervention, vol. 26, no. 1, pp. 47–54 (2017) doi: 10.1016/j.psi.2015.12.002
13. Hang, O. C., Dahlan, H. M.: Cyberbullying lexicon for social media. In: International Conference on Research and Innovation in Information Systems, ICRIIS. pp. 1–6 (2019) doi: 10.1109/ICRIIS48246.2019.9073679
14. Noviantho., Isa, S. M., Ashianti, L.: Cyberbullying classification using text mining. In: Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS, pp. pp. 241–245 (2017)
15. Farag, N., McKee, G., El-Seoud, S. A., Hassan, G.: Bullying hurts: A survey on non-supervised techniques for cyber-bullying detection. ACM International Conference Proceeding Series, pp. 85–90 (2019) doi: 10.1145/3328833.3328869

16. Microsoft: Información general sobre ASP.NET MVC https://docs.microsoft.com/es-es/aspnet/mvc/overview/older-versions-1/overview/asp-net-mvc-overview
17. Twitter Inc.: Developer Platform, https://developer.twitter.com/en/docs/twitter-api
18. Sharan, K.: Model-view-controller pattern. Learn JavaFX, vol. 8, pp. 419–434 (2015)
19. Bertocco, M., Ferraris, F., Offelli, C., Parvis, M.: A client–server architecture for distributed measurement systems. IEEE Transaction on Instrumentation and Measurement, vol. 47, no. 5, pp. 1143–1148 (1998)
20. ATLAS.ti: ATLAS.ti, https://atlasti.com/es