ISSN 1870-4069

Generación de un corpus en español con expresiones agresivas

César Jesús Núñez-Prado[1], Liliana Chanona-Hernández[2], Grigori Sidorov[1]

[1] Instituto Politécnico Nacional,
Centro de Investigación en Computación,
México
[2] Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica,
México

{cesar.jnprado, lchanona}@gmail.com,
sidorov@cic.ipn.mx

Resumen. El bullying es un concepto que se refiere al acoso físico, verbal o psicológico entre una o más personas hacia un individuo o un grupo de individuos. Este tipo de agresión evolucionó al mismo paso que los avances tecnológicos y es conocido como cyberbullying. Ahora los agresores hacen uso de las redes sociales para continuar con tales ataques hacia sus víctimas. Por otra parte, las redes sociales no cuentan con filtros efectivos que sean capaces de identificar este tipo de conductas entre sus usuarios y ello complica la detección a tiempo y las consecuencias psicológicas que sufren las víctimas. Se busca crear un corpus en español que contenga las expresiones más utilizadas entre un grupo específico de personas para identificar la violencia en los mensajes publicados en Twitter.

Palabras clave: Bullying, cyberbullying, corpus, redes sociales, Twitter.

Generation of a Corpus in Spanish with Aggressive Expressions

Abstract. Bullying is a concept that refers to physical, verbal or psychological harassment between one or more people towards an individual or a group of individuals. This type of aggression evolved at the same time as technological advances and is known as cyberbullying. Now the attackers use social networks to continue such attacks on their victims. On the other hand, social networks do not have effective filters that are capable of identifying this type of behavior among their users and this complicates early detection and the psychological consequences suffered by victims. The aim is to create a corpus in Spanish that contains the most widely used expressions among a specific group of people to identify violence in messages published on Twitter.

Keywords: Bullying, corpus, social networks, twitter.

1. Introducción

La violencia no es algo nuevo, ha existido desde siempre, y en el presente las formas de materialización han ido evolucionando. Se han ideado nuevas maneras de llevarla a cabo e inclusive se han servido de los avances tecnológicos para ello. La violencia que se vive día a día en nuestro país y en el mundo entero va en incremento, ya que en todos los medios de transmisión (radio, televisión, periódicos, internet, entre otros) se ve reflejada como tema de morbosidad para atraer a un número mayor de audiencia y son los niños la parte de la población más afectada y vulnerable ya que buscan repetir las acciones que les rodean o que observan, sin detenerse a pensar si sus acciones pueden repercutir en alguien.

Con el creciente uso y desarrollo de redes sociales, una nueva modalidad de esa violencia ha aparecido y es conocida como *cyberbullying* o ciber-acoso. En ésta nueva forma, las víctimas son exhibidas de muchas maneras, tales como intimidación pública, publicación de información falsa, insultos hacia su persona, insultos hacia sus preferencias, publicación de fotos privadas, etc.

De acuerdo con [1] "El *cyberbullying* es un daño intencional y repetido a través del uso de computadoras, teléfonos celulares y otros dispositivos electrónicos" y posee ciertas características tales como: el comportamiento debe ser deliberado, la intimidación refleja un patrón de comportamiento, no solo un incidente aislado y que el objetivo debe percibir que el daño fue infligido.

Tabla 1. Diferentes tipos de acoso cibernético.

Tipos de acoso	Método
Grooming	Engatusamiento de menores por parte de pedófilos
Cebo	Uso de artimañas para atraer a niños a encuentros fuera de la red
Sexting	Enviar y pedir fotografías sexualmente explícitas
Ciber-acecho	Seguimiento de los pasos *online* de la víctima
Cyberbullying	Ofender, burlarse o agredir utilizando la lengua escrita o por medio de imágenes

Las principales consecuencias mostradas en las víctimas de este tipo de violencia son: baja autoestima, actitudes pasivas, trastornos emocionales, depresión, ansiedad, pensamientos suicidas, deseos de dejar de asistir a clases, aislamiento y en el peor de los casos el intento de suicidio u homicidio.

2. Trabajos relacionados

La aplicación en [2] se desarrolló por una joven norteamericana motivada por la noticia del suicidio de una niña menor a ella, la cual había sido acosada por sus compañeros de clase. La idea principal en este trabajo es detectar y detener el acoso en la fuente, es decir; antes de que ocurra el acoso y el daño sea irreversible. La aplicación

Tabla 2. Resultados de encuestas a estudiantes.

Porcentaje de encuestados	Tipo de interacción
99 %	Utilizan redes sociales
41%	No tienen agregados a sus padres
57 %	Sus padres no usan ningún tipo de red social
30 %	Ha sido víctima de violencia cibernética
15 %	Ha realizado *bullying* o *cyberbullying* por lo menos en una ocasión

está diseñada para el idioma inglés y al ejecutarse muestra una pantalla emergente cada vez que detecta palabras que puedan ser ofensivas y le pide al emisor que se tome un momento para pensar si realmente quiere enviar un mensaje el cual pueda herir al receptor.

En [3] se analiza el contenido de los mensajes publicados en *Twitter* y se determina si el mensaje contiene acoso o no, mediante el uso de algoritmos de aprendizaje automático, n-gramas, errores gramaticales y está diseñado para el idioma español.

3. Sujeto de estudio

Gracias a las autorizaciones por parte del Coordinador de Área de la Secretaría de Educación Pública (SEP) y del director de una secundaria técnica en la Ciudad de México, se pudo trabajar con una población mixta de 900 alumnos. El objetivo principal de involucrar a alumnos en ese rango de edad (entre 12 y 15 años) es la de conocer las interacciones de los estudiantes con las redes sociales, la manera en que se expresan y las redes sociales que utilizan.

El 41 % que afirmó, que en sus cuentas en redes sociales no tienen agregados a sus padres ni a ningún familiar, lo hacen debido a que se sienten hostigados y vigilados. Esta encuesta revela que un gran porcentaje de los padres de familia no tienen conocimiento de las interacciones de sus hijos dentro de las redes sociales, lo cual conlleva como consecuencia, al desconocimiento de si su hijo es posible víctima o victimario de acoso cibernético.

Dentro de ésta población, un 30 % aceptó haber sido víctima de *cyberbullying* por lo menos en una ocasión y un 15 % confirmó haber realizado por lo menos una vez *cyberbullying*.

Al ser menores de edad los alumnos de la secundaria con los cuales se trabajó, se pidió autorización a los padres de familia. En estas pláticas los padres comentaron que se han presentado algunos casos de acoso dentro de la comunidad estudiantil pero que no han podido terminar con el hostigamiento hacia las víctimas.

4. Generación de diccionario con malas palabras

A partir de las encuestas realizadas a los estudiantes, se pudo crear un diccionario con un total de 115 malas palabras consideradas a partir de las expresiones que utilizan

Tabla 3. Muestra del diccionario.

alv	estupido	putito
culo	fuck	homosexual
gorda	trans	pendejo
puta	culero	huevos
joto	mamon	wey
verga	naco	guey

con mayor frecuencia en sus interacciones en las redes sociales. Las palabras que se incluyeron aparecieron con al menos un 50 % en las encuestas y se agregaron inclusive cuando las palabras están mal escritas (con faltas de ortografía), ya que los encuestados refirieron que escribir correctamente les quita mucho tiempo.

Estas palabras se agregaron al diccionario tanto en femenino como en masculino, en singular y en plural, con abreviaturas e incluso con faltas ortográficas ya que los encuestados refirieron que escribir correctamente les quita demasiado tiempo.

5. Sistema de extracción de *tweets*

Para la descarga de mensajes de la plataforma *Twitter* se aperturó una aplicación de desarrollo interna (disponible para programadores), la cual proporciona los permisos y las claves necesarias para realizar una correcta recuperación de información dentro de la misma red social. El sistema se desarrolló en el lenguaje de programación *Python* y para realizar la conexión con *Twitter* se utilizó la librería *Tweepy*. Para la búsqueda y descarga de los *tweets* se utilizaron las palabras del diccionario creado a partir de las encuestas realizadas. Se extrajeron cerca de 2000 mensajes a los cuales se les realizó un filtrado para eliminar los mensajes repetidos y los mensajes escritos en un idioma diferente al español. Una vez que se contó con mensajes únicos (1300 mensajes no repetidos y en español) se realizó una fase de procesamiento de lenguaje natural utilizando librerías de *Natural Language ToolKit* (NLTK) y librerías internas de *Python*. Esta fase incluyó:

- Conversión del texto a minúsculas,
- Eliminación de enlaces en internet,
- *Tokenización*,
- Eliminación de caracteres especiales y signos de puntuación,
- Eliminación de *stickers* y
- Eliminación de palabras de parada (*stop words*).

6. Evaluación y generación del corpus

El archivo con los mensajes descargados de *Twitter* y ya procesados se envió a un jurado conformado por 5 personas. El jurado es mixto, es decir; con hombres y mujeres

con un rango de edad entre los 20 y los 45 años y que tienen contacto diario con redes sociales. El jurado se encargó de calificar los mensajes con solo dos opciones: "con acoso" o "sin acoso". La medida que se utilizó para considerar si el mensaje se agregaba al corpus, fue que al menos 3 miembros del jurado calificaran de acoso al mismo mensaje. El corpus final quedó compuesto por 600 mensajes en español con acoso cibernético.

7. Conclusiones y trabajo a futuro

El manejo de las redes sociales en la actualidad es casi inevitable y no importa si el usuario es adulto mayor, estudiante de primaria, alto funcionario, hombre o mujer. En algunas ocasiones, las vacantes para empleos requieren que se dominen las redes sociales y el manejo de dichas redes puede traer consigo ventajas y desventajas.

Fue a partir de las desventajas, que se pensó en contribuir con la creación de un corpus que sirva como herramienta para la detección de este tipo de violencia cibernética y con ello ayudar a las personas más vulnerables.

De las encuestas realizadas a los alumnos se concluye que uno de los principales factores que propician el ser víctimas de acoso cibernético, radica en que no se cuenta con la información necesaria para realizar una denuncia y con ello limitar en cierta proporción el ciber-acoso sufrido en las redes sociales.

El corpus creado refleja de manera real, el uso de la lengua empleado por estudiantes de nivel secundaria.

Como trabajo a futuro se busca que el corpus creado pueda crecer en dimensión para que pueda contener más expresiones agresivas utilizadas y el error al detectar el ciber-acoso se pueda reducir.

También se piensa en realizar un sistema que haga uso de inteligencia artificial para lanzar una alerta al usuario que sea una posible víctima de violencia en redes sociales.

Agradecimientos. Agradecemos a CONACYT, SNI, IPN (SIP, COFAA), apoyo de proyectos SIP 20200859 y 20200797 y Conacyt A1-S-47854.

Referencias

1. Ruiz-Palacios, F.: Corea capacita a SSP contra ciberdelitos. El Universal, http://www.eluniversal.com.mx/articulo/metropoli/cdmx/2016/11/2/corea-capacita-ssp-contra-ciberdelitos (2016)
2. Rethink Words: ReThink before the damage is done. Rethinkwords, http://www.rethinkwords.com (2016)
3. Ramos-Márquez, J..: Detección de Acoso en mensajes de Twitter. Centro de Investigación en Computación, Tesis de Maestría (2017)
4. Sidorov, G.: Construcción no lineal de n-gramas en la lingüística computacional, n-gramas sintácticos, filtrados y generalizados (2013)
5. Sanz, E.: ¿En qué consiste el "sexting"? Muy interesante. http://www.muyinteresante.es/curiosidades/preguntas-respuestas/ien-que-consiste-el-sexting (2016).
6. Russell, M.A.: Mining the Social Web, Data Mining Facebook, Twitter, Linkedin, Google+, Github, and More. http://www.webpages.uidaho.edu/~stevel/504/Mining-the-Social-Web-2nd-Edition.pdf (2013)

7. Bird, S., Klein, E., Loper, E.: Natural language processing with python, analyzing text with the natural language toolkit. O'Relly Media (2010.