

Detección de cyberbullying a través de un sistema de reconocimiento óptico de caracteres en imágenes de páginas web

MARTÍNEZ-BARRAGÁN, Rosalía†, SABINO-MOXO, Beatriz Adriana, MARQUÉZ-DOMÍNGUEZ, José Alberto y SÁNCHEZ-ACEVEDO, Miguel Ángel

Universidad de la Cañada

Recibido Octubre 3, 2017; Aceptado Diciembre 14, 2017

Resumen

En este documento se presenta un software para detectar texto ofensivo en imágenes alojadas en páginas web, el objetivo principal es proveer una herramienta informática que coadyuve en la detección de cyberbullying. La metodología empleada considera las siguientes etapas: extracción de imágenes de páginas web a través de una URL usando funciones de PHP; enseguida se realiza la segmentación que consiste en aplicar una serie de filtros de procesamiento digital de imágenes, posteriormente la imagen resultante con el texto es convertida a matriz binaria para dividir la palabra en letras; el entrenamiento y reconocimiento se efectuó adaptando la librería *phpOCR*; finalmente se busca el texto en una base de datos que contiene las palabras ofensivas, éstas fueron recopiladas a través de un cuestionario aplicado a adolescentes, el software muestra una alerta cuando encuentra alguna coincidencia con la información almacenada. La contribución de este trabajo es proveer a la sociedad una herramienta de apoyo para la detección de cyberbullying en páginas web con imágenes que incluyen texto ofensivo, además se brinda una alternativa para adaptar la librería *phpOCR* en la detección de texto a color, debido a que ésta actualmente solo reconoce caracteres en imágenes monocromáticas.

PhpOCR, segmentación, entrenamiento

Abstract

This document presents software to detect offensive text in images hosted in web pages, the main objective is to provide a computer tool that helps in the detection of cyberbullying. The methodology used has the following stages: extraction of images in web pages through its URL and PHP functions; Segmentation is performed with digital image processing filters, then the resulting image with text is converted to binary matrix to divide the word into letters; Training and recognition was done by adapting the *phpOCR* library; Finally the text is searched in a database of offensive words, these were collected through a questionnaire applied to young people, the software shows an alert when it finds some coincidence with the information stored. The contribution of this work is to provide society with a support tool for the detection of cyberbullying on websites with offensive content, in addition, it provides an alternative to adapt the *phpOCR* library in color text detection, because it currently only recognizes characters in monochrome images.

PhpOCR, segmentation, training

Citación: MARTÍNEZ-BARRAGÁN, Rosalía, SABINO-MOXO, Beatriz Adriana, MARQUÉZ-DOMÍNGUEZ, José Alberto y SÁNCHEZ-ACEVEDO, Miguel Ángel. Detección de cyberbullying a través de un sistema de reconocimiento óptico de caracteres en imágenes de páginas web. Revista de Tecnología Informática 2017, 1-3: 8-15

† Investigador contribuyendo como primer autor.

Introducción

El Bullying se define como “una situación social en la que uno o varios toman como objetivo de su actuación injustificadamente agresiva a otro compañero y lo someten, por tiempo prolongado, a agresiones físicas, burlas, hostigamiento, amenazas, aislamiento social o exclusión social aprovechándose de su inseguridad, miedo o dificultad personal para pedir ayuda o defenderse” (Ortega, 1994). Otra definición la proporciona Dan Olweus quién fue el primer investigador sobre el maltrato en el ámbito educativo desde 1999 y lo interpreta de la siguiente manera: Un estudiante está siendo intimidado cuando otro estudiante o grupo de estudiantes le dice cosas mezquinas o desagradables, se ríe de él o ella o le llama por nombres molestos o hirientes. Le ignora completamente, le excluye de su grupo de amigos o le retira de actividades a propósito Garaigordobil (2011).

En los últimos años se han observado diferentes modalidades de acoso, una de éstas es el ciberbullying, este fenómeno es más reciente debido a los avances tecnológicos, según la opinión Kowalski *et al.*, (2010) el acoso electrónico o cibernético o crueldad social en línea se define como el acoso a través de la difusión maliciosa de información en la red, en mensajes de textos, redes sociales, correo electrónico, páginas web, blogs, salas de chat, etc. Puede ser información en texto, fotografías, o imágenes modificadas o editadas, todo a través de una computadora o teléfono móvil. Otra opinión acertada es de López (2012) este considera que el acoso escolar se define como la conducta agresiva y deliberada que implica un desequilibrio del poder o fuerza.

El ciberbullying causa daños psicológicos como depresión, baja autoestima, ideas suicidas e incluso, el propio suicidio de la víctima.

La Fundación Vodafone (2015) efectuó un estudio a 5,000 jóvenes de entre 13 y 18 años provenientes de 11 países, obteniendo que más de la mitad de los adolescentes piensan que el ciberbullying es peor que la intimidación cara a cara, y el 43% cree que es un problema mayor que el abuso de drogas. En promedio general uno de cada cinco (18%) afirman haber sufrido ciberbullying en algún momento de su vida, y de ellos, una quinta parte ha tenido incluso pensamientos suicidas.

Una forma de hacer ciberbullying se lleva a cabo por medio de ‘memes’. Algunos pensadores hablan de los ‘memes’ como un nuevo lenguaje utilizado por los nativos digitales del siglo XXI, quienes cuentan con una gramática propia que despliega nuevos sentidos a través de la web y que, como se observa, puede llegar a influir en los pensamientos y/o acciones de los usuarios Castagno (2013).

La palabra “meme” tiene origen en el libro “El gen egoísta” del biólogo y escritor Richard Dawkins. Este autor propone la idea de que los genes de los seres vivos tienen como único fin, el replicarse y perpetuarse por medio de la reproducción de su código genético. De ese modo, postula que el egoísmo inherente de los genes se da debido a su naturaleza replicadora. (Chamorro, 2014). No obstante, Arnao *et al.*, (2015) describen el término de memes como imágenes graciosas y ofensivas sobre alguna persona o tema en particular, su fin es sarcástico y busca hacer quedar mal a quien lo recibe. Generalmente cada imagen tiene unos mensajes ofensivos o sarcásticos, pero su fin es dañar.

A partir de esto surge la idea de desarrollar una herramienta de apoyo a la sociedad para la detección de ciberbullying en páginas web que contienen imágenes denominadas ‘memes’ con texto ofensivo, debido a que en la actualidad el ciberbullying es un problema que provoca daños psicológicos principalmente.

La hipótesis central de este proyecto es la siguiente “A partir de la URL de un sitio web se puede extraer texto ofensivo de sus imágenes usando reconocimiento óptico de caracteres”.

La funcionalidad de la propuesta de software es analizar el contenido de páginas web públicas para obtener información que indiquen casos de ciberbullying. El software extrae imágenes de un sitio web, posteriormente las analiza y verifica si ésta contiene palabras ofensivas. Lo anterior contribuye en la detección del ciberbullying a través de las imágenes denominada ‘memes’.

Este documento se compone de los siguientes apartados: Software de apoyo para el bullying y Ciberbullying, en esta sección se describe un panorama acerca de los sistemas o aplicaciones de apoyo para la prevención, detección y difusión de información sobre el bullying y ciberbullying; posteriormente se muestra la metodología que se llevó a cabo para el desarrollo de esta aplicación que abarca los siguientes pasos: Extracción de páginas web, segmentación, pre-procesamiento, entrenamiento y reconocimiento con phpOCR y búsqueda de información ofensiva de la BD. Después, se muestran las pruebas y resultados obtenidos, trabajo a futuro, conclusiones y las referencias utilizadas para el desarrollo del tema.

A continuación, se describe algunas aplicaciones que apoyan a la problemática del ciberbullying.

Software de apoyo para el bullying y ciberbullying

En la actualidad existen diferentes aplicaciones que apoyan en la problemática del bullying y ciberbullying, en la Tabla 1 se muestran una comparativa de éstos.

Aplicaciones	Botón de alerta	Envío de videos e imágenes	Control ciber acoso
Stop it	SI	SI	NO
Bully Semáforo	NO	NO	NO
Basta de bullying	NO	NO	SI
Bully Block	SI	SI	SI
Bully button	SI	SI	SI
Know Bullying	NO	NO	NO
Break UP	SI	SI	NO
My Mobile Watch Dog	NO	NO	SI
NetNanny	NO	NO	SI
Bully Tag	SI	SI	NO
Anonymus alert	SI	SI	SI
Bullying es acoso escolar	NO	NO	NO
No al acoso escolar	SI	SI	NO
Bullying No More Free	NO	NO	NO
Basta de bullying (juegos)	NO	NO	NO
CiberBully Hostline	SI	SI	SI

Tabla 1 Aplicaciones contra el acoso escolar

Fuente: Tomado de Peinado et al., (2016)

La funcionalidad de las aplicaciones móviles es la protección de los menores en su entorno educativo, pero no es lo suficiente para reducir el acoso escolar, la mayoría de ellas se basa en acciones informativas, almacenamiento de imágenes o videos con envío y no hay una plataforma de control real de servidores Peinado et al., (2016).

Por otra parte, *rethink* es un software que mide la voluntad de los adolescentes al publicar palabras dañinas en redes sociales, por otra parte, si el adolescente decide mandar un mensaje dañino, le pedirá que reconsidere su decisión de publicar el mensaje mostrándole un mensaje de alerta.

Los resultados demostraron que los adolescentes estaban 93,43% menos dispuestos a publicar mensajes dañinos usando el sistema (Prabhu, 2014).

Pujol (2016) presenta una propuesta de un prototipo de un sistema de reconocimiento facial en imágenes digitales mediante la identificación del género y la franja de edad de un sujeto. Lo anterior se empleará con perfiles públicos en redes sociales, como Facebook o Instagram, en los casos en los que se produzca un reconocimiento de la víctima potencial de ciberbullying, pero los datos de los perfiles no se correspondan con el del usuario analizado, se procederá a activar una alarma sobre posible suplantación de identidad. A continuación, se presenta la metodología que se usó para el desarrollo del software.

Metodología

El proceso para llevar a cabo la realización de esta aplicación se muestra en la Figura 1.

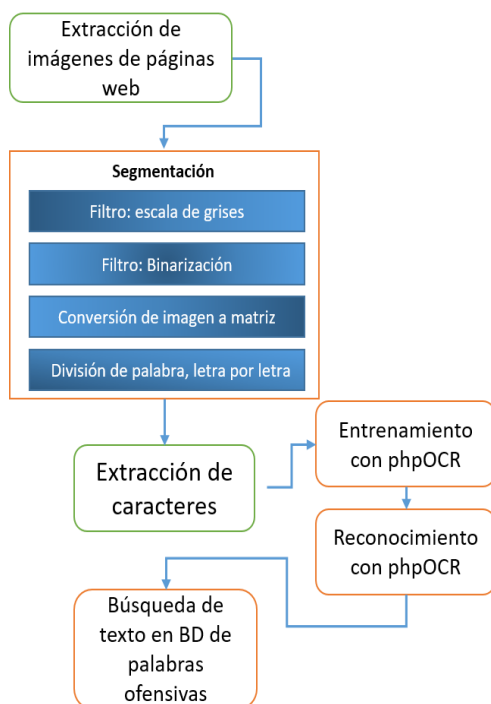


Figura 1 Etapas del proceso para el desarrollo de la aplicación

Fuente: Elaboración propia.

Extracción de imágenes de páginas web

Para la extracción de imágenes de una página web, se utilizó un formulario en el cual el usuario coloca la url del sitio web que desee analizar, obteniendo las imágenes (Figura 2), para esto se utilizó funciones de *PHP*, como *imagesHTML* estas funciones obtienen las imágenes de la página web y se guardan en una ruta específica. A continuación, el siguiente paso es la segmentación.



Figura 2 Interfaz de la aplicación

Fuente: Elaboración propia

Segmentación

La segmentación es un proceso que consiste en dividir la imagen en sus partes constitutivas con el fin de extraer regiones, texturas y contornos presentes, es una etapa esencial de procesamiento ya que permite identificar contenido en la imagen y sus resultados permiten analizar la información. Comas *et al.*, (2014).

La segmentación puede ser de utilidad en diferentes áreas de investigación tales como Medicina, Industria, entre otras (Babu *et al.*, 2010; Tehsin *et al.*, 2014; Rajab *et al.*, 2014).

En la segmentación inicialmente se prepara la imagen con el fin de mejorar su contraste, aplicando filtros de PDI (Ortiz, 2013), la función seleccionada se aplica a un punto o pixel de la imagen en la posición (i, j) y como resultado se generan un nuevo valor que se asigna a la imagen transformando en coordenada (i, j).

Los filtros que se aplicaron en esta etapa son: escala de grises y binarización; el primero es la representación de una imagen en la que cada pixel se dibuja usando un valor numérico individual que representa su luminancia, en una escala que se extiende entre el blanco y negro. Posteriormente, se aplica el segundo, utilizándolo para separar objetos o regiones de interés en la imagen. Es decir, que se puede separar el fondo de la imagen y los objetos que se desean analizar.

Ésta se obtiene por medio de un barrido en la matriz en la imagen digital, por medio de bucles o recursividad, con el fin de que el proceso produzca la reducción de escala de grises a dos únicos valores. Negro (0) y blanco (255) o un sistema binario de ausencia y presencia de color 0-1. La comparación de cada píxel de la imagen viene determinada por un umbral. Por ejemplo, los valores que sean mayores que el umbral toman un valor 255 y los menores 0 (Magro, 2013) como se muestra en la Figura 3.



Figura 3 Imágenes con filtros

Fuente: *Elaboración propia*

En el siguiente apartado se describe el pre-procesamiento que se llevó a cabo para la segmentación.

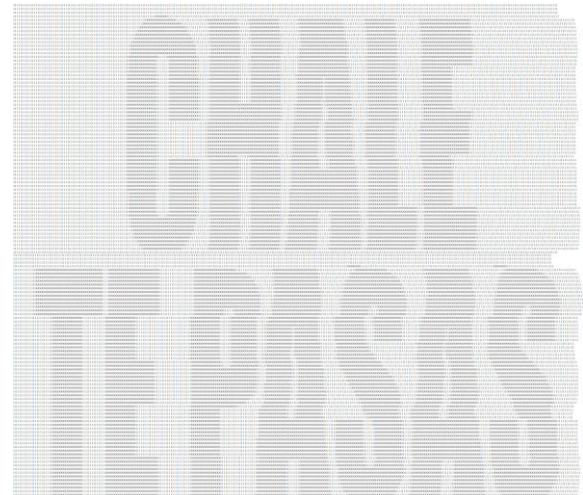


Figura 4 Matriz binaria de la imagen procesada

Fuente: *Elaboración propia.*

Pre-procesamiento

Se creó una función en la cual convertirá la imagen resultante a una matriz binaria para poder dividir la palabra en letras (Figura 4).

Obteniendo la matriz binaria se realizó la segmentación, y así se filtró solamente la información que corresponde a las letras (Figura 5).



Figura 5 Segmentación de las letras a partir de la matriz binaria

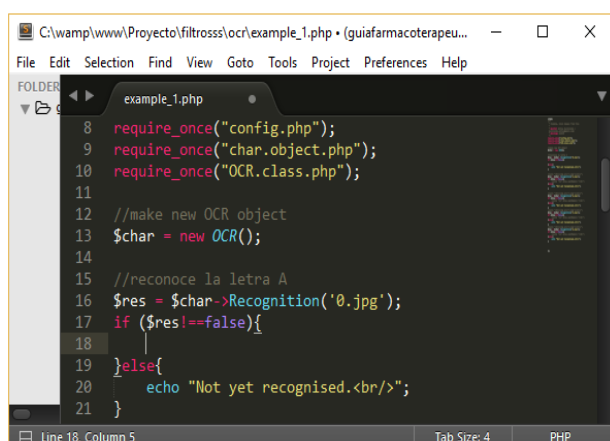
Fuente: *Elaboración propia.*

Entrenamiento con *phpOCR*

El entrenamiento con la librería *phpOCR* permite clasificar una letra escrita en un tipo de fuente con un carácter que la representa, para esta etapa se clasificó con respecto al tipo de fuente *Arial Black*, debido a que es la comúnmente empleada en las imágenes de análisis, para ello se generó un archivo con las 27 letras del alfabeto.

Reconocimiento con *phpOCR*

En esta etapa se genera un archivo de texto que contiene todas las letras que conforman las palabras o frases que tiene la imagen de entrada. En la Figura 6 se muestra la función que se utilizó para el reconocimiento.



```

8  require_once("config.php");
9  require_once("char.object.php");
10 require_once("OCR.class.php");
11
12 //make new OCR object
13 $char = new OCR();
14
15 //reconoce la letra A
16 $res = $char->Recognition('0.jpg');
17 if ($res!==false){
18
19 }else{
20     echo "Not yet recognised.<br/>";
21 }
  
```

Figura 6 Función para el reconocimiento

Fuente: Elaboración propia

Búsqueda de texto de la BD

Las palabras obtenidas en el paso anterior se buscan en una base de datos de palabras ofensivas implementada localmente y se muestra un mensaje de alerta si hay una coincidencia entre éstas.

Pruebas y resultados

En este apartado se muestran las pruebas realizadas del software iniciando con la extracción de imágenes en páginas web, después la segmentación de imágenes, posteriormente con el entrenamiento y reconocimiento de la librería *phpOCR*, y por último la detección de ciberbullying.

Con respecto a la extracción de imágenes se analizaron 15 sitios web, éstos contenían en promedio 25 imágenes cada uno, de los cuales se extrajeron el 100%, en total se obtuvieron 375, de las cuales el 70.1% contenía texto y el 30% restante no incluía información adicional.

Ahora bien, la segmentación de las 263 imágenes se realizó correctamente con el 60% de éstas, el 40% restante no arrojó los resultados esperados debido a que éstas tenían ruido y además estaban escritas con otro tipo de fuente.

Por lo que se refiere al reconocimiento, de las 158 imágenes segmentadas se logró reconocer todas ya que como anteriormente se ha mencionado solo contenían el mismo tipo de fuente del que fue entrenada la librería.

Por último, de esas 158 imágenes se obtuvieron con un promedio de 1,300 palabras aproximadamente, éstas se compararon con las palabras ofensivas que se tenían localmente, encontrando un 60% de coincidencia, mostrando el mensaje de alerta.

En la Tabla 2 se muestra tres tipos de imágenes a los que se les aplicó el proceso de segmentación, con la imagen del inciso (a) la segmentación y reconocimiento se realizó correctamente.

Sin embargo, el caso de presentado en el inciso (b) la segmentación no se realizó adecuadamente porque la imagen tenía información con ruido, en el último caso (inciso c) la segmentación se realizó correctamente, sin embargo, no se reconoció el tipo de fuente, debido a que el usado en el entrenamiento es diferente al de la imagen.

Imagen de entrada	Salida
 <p>a)</p>	
 <p>b)</p>	<p>¡imagen invalida!</p>
 <p>c)</p>	 <p>Reconocimiento de letra errónea.</p>

Tabla 2 Tres tipos de imagen para el proceso de segmentación

Fuente: *Elaboración propia.*

Trabajo a futuro

Se considera importante considerar las siguientes actividades para mejorar la propuesta:

- Agregar imágenes con diferente tipo de fuente para el entrenamiento y reconocimiento.
- Incluir otros tipos de filtros que permitan la eliminación completa del ruido y obtener las letras con mayor fidelidad.

Conclusiones

Cabe mencionar que durante la investigación sobre el ciberbullying no se encontró aplicaciones que realicen el análisis de texto ofensivo de imágenes, ya que otros trabajos de investigación se basan principalmente en el análisis de texto, mientras que otros se dedican a difundir buenas practicas contra el ciberbullying, por este motivo es una buena alternativa de desarrollo.

Para finalizar es importante tomar en cuenta que solo se puede realizar la extracción de imágenes de blogs y páginas web que no sean privados ya que de lo contrario no se puede obtener.

Podemos concluir, que las pruebas que se obtuvieron la mayoría de páginas contenían por lo menos 5 o 6 imágenes con palabras ofensivas, sin embargo, más adelante se puede mejorar el software para que segmente y reconozca el 100% de las imágenes con texto que se extraen de los sitios web.

Referencias

Arnao Marciani, J., y Surpachín Miranda, M. (2015). Cyberbullying, violencia relacional en la sociedad de la información.

- Babu, G., Srimaiyee, P., y Srikrishna, A. (2010). Text extraction from heterogeneous images using mathematical morphology. *Journal of Theoretical y Aplicada Tecnología de la Información*, 16.ISO 690.
- Castagno, L. (2013). ¿Memes sólo diversión?: el fenómeno de los memes de internet más allá de la humorada (Bachelor's thesis, Facultad de Ciencia Política y Relaciones Internacionales).
- Chamorro Bernal, J. A. (2014). Viralización de contenidos y memes en internet (memoria para optar al título de periodista). Universidad de Chile instituto de la comunicación e imagen. Santiago de Chile.
- Comas, D. S., & Meschino, G. J. (2014). Segmentación de Imágenes mediante Reconocimiento de Patrones.
- Fundación Vodafone, 2015. Vodafone global survey on cyberbullying. Recuperado de <https://markets.ft.com/data/announce/full?dockey=1323-12509304-1Q8SVKQBQINRPVIK2IJ53QU6KG4&mhq5j=e3>
- Garaigordobil, M. (2011). Bullying y cyberbullying: conceptualización, prevalencia y evaluación. FOCAD Formación Continuada a Distancia. Duodécima edición enero, 1-22.
- Kowalski, J. R. y Limber, S. y Agatston, P. W. (2010). Cyberbullying, el acoso escolar en la er@ digit@l. España: Ed. Desclée Ce Brower.
- López, E. M. (2012). Acoso cibernético o cyberbullying: Acoso con la tecnología electrónica. *Pediatría de México*, 14(3), 133-146.
- Magro R, (2013). Binarización de imágenes digitales y su algoritmia como herramienta aplicada a la ilustración entomológica. *Boletín de la SEA*, España, 443-464.
- Ortega Ruiz, R. (1994). Violencia interpersonal en los centros educativos de enseñanza secundaria. Un estudio sobre maltrato e intimidación entre compañeros. *Revista de Educación*. Madrid, (304), 253-280.
- Ortiz, M. (2013). Procesamiento digital de imágenes. Benemérita universidad Autónoma de Puebla <http://www.cs.buap.mx/mmartin/pdi>.
- Peinado Miguel, F., & Mateos Abarca, J. P. (2016). Aplicaciones móviles contra el acoso escolar. *Revista del centro de investigación y estudios gerenciales*, 296-314.
- Prabhu. (2014). ReThink: An Effective Way to Prevent Cyberbullying in Adolescents. Recuperado de https://getinspired.mit.edu/sites/default/files/documents/ST307_Report.pdf
- Pujol, F. A., Jimeno-Morenilla, A., Pertegal Felices, M. L., & Gimeno Nieves, E. (2016). Detección automática de cyberbullying a través del procesamiento digital de imágenes.
- Rajab, L., Mohammad, H., Karajeh, H., & Al Khatib, T. (2014). Una técnica de extracción de texto mejorado basado en la transformación lineal. *Ciencias de la Vida Diario*, 11(7).
- Reconocimiento texto en imágenes con php Recuperado de <http://www.tufuncion.com/phpocr>
- Tehsin, S., Masood, A., Kausar, S., y Arif, F. (2014). Segmentación Fuzzy-base para la Variable-Font Sized extracción de texto de las imágenes / vídeos. *Problemas Matemáticos en Ingeniería*, 2014.