



LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - E-Revistas - Google Scholar
DOI - REBID - Mendeley - DIALNET - ROAD - ORCID

Title: Detección de ciberbullying a través de un sistema de reconocimiento óptico de caracteres en imágenes de páginas web

Author: Rosalia MARTÍNEZ BARRAGÁN

Editorial label ECORFAN: 607-8324
BCIERMIMI Control Number: 2017-02
BCIERMIMI Classification (2017): 270917-0201

Pages: 28

Mail: li_rosaliamtzb@unca.edu.mx

RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.

244 – 2 Itzopan Street

La Florida, Ecatepec Municipality

Mexico State, 55120 Zipcode

Phone: +52 1 55 6159 2296

Skype: ecorfan-mexico.s.c.

E-mail: contacto@ecorfan.org

Facebook: ECORFAN-México S. C.

Twitter: @EcorfanC

www.ecorfan.org

Holdings

Bolivia	Honduras	China	Nicaragua
Cameroon	Guatemala	France	Republic of the Congo
El Salvador	Colombia	Ecuador	Dominica
Peru	Spain	Cuba	Haití
Argentina	Paraguay	Costa Rica	Venezuela
Czech Republic			

Contenido

- Introducción
- Software de apoyo para el bullying y cyberbullying
- Desarrollo
 - Extracción de imágenes de páginas web
 - Segmentación
 - Entrenamiento con *phpOCR*
 - Reconocimiento con *phpOCR*
 - Búsqueda de texto de la BD
- Pruebas y resultados
- Trabajo a futuro
- Conclusiones
- Referencia

INTRODUCCIÓN



**Congreso Interdisciplinario de Energías Renovables,
Mantenimiento Industrial, Mecatrónica e Informática**

2017



Bullying

Dan Olwues 1999



Ciberbullying



Ciberbullying por medio de memes

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

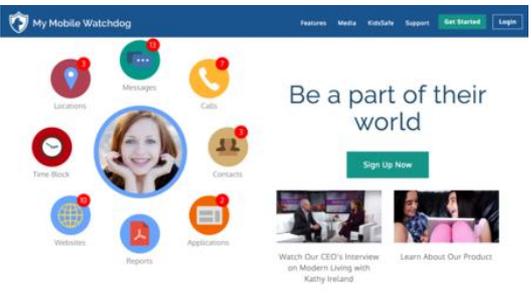
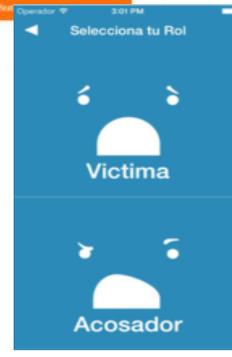
A partir de esto surge la idea de desarrollar una herramienta de apoyo a la sociedad para la detección de ciberbullying en páginas web que contienen imágenes denominadas ‘memes’ con texto ofensivo.

Objetivo principal

Es proveer una herramienta informática que coadyuve en la detección de ciberbullying.

Hipótesis central

“A partir de la URL de un sitio web se puede extraer texto ofensivo de sus imágenes usando reconocimiento óptico de caracteres”.



SOFTWARE DE APOYO PARA EL BULLYING Y CIBERBULLYING



Congreso Interdisciplinario de Energías Renovables, Mantenimiento Industrial, Mecatrónica e Informática

2017



Difusión de información

- Aplicaciones móviles es la protección de los menores en su entorno educativo como: Stop it, Basta de bullying, Bully Block, Bully button, Know Bullying, Break UP, My Mobile Watch Dog, etc.

Rethink

- Si el adolescente decide mandar un mensaje dañino, le pedirá que reconsidere su decisión de publicar el mensaje mostrándole un mensaje de alerta.

Propuesta de un prototipo de un sistema de reconocimiento facial en imágenes digitales

- Procederá a activar una alarma sobre posible suplantación de identidad en los casos en los que se produzca un reconocimiento de la víctima potencial de ciberbullying



**Congreso Interdisciplinario de Energías Renovables,
Mantenimiento Industrial, Mecatrónica e Informática**

2017

METODOLOGÍA

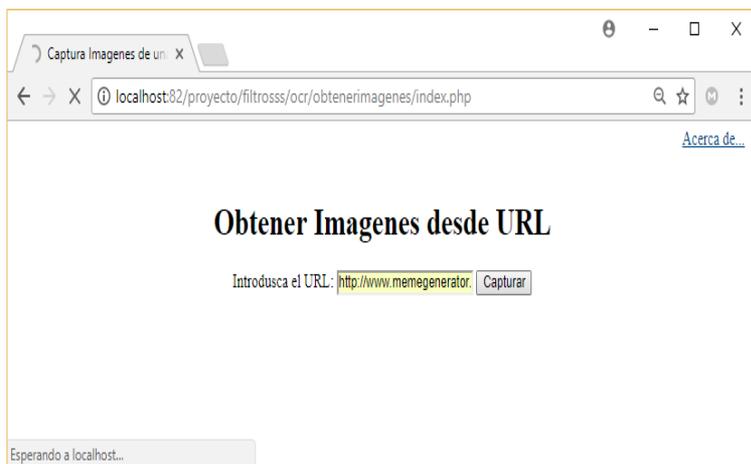


Figura 1 Etapas del proceso para el desarrollo de la aplicación.

Fuente: *Elaboración propia.*

Extracción de imágenes de páginas web

Figura 2 Interfaz de la aplicación.



Fuente: *Elaboración propia.*

Se utilizaron funciones de *PHP*, como *imagesHTML* estas funciones obtienen las imágenes de la página web y se guardan en una ruta específica.

Segmentación

Los filtros que se aplicaron en esta etapa son: escala de grises que se extiende entre el blanco y negro; binarización que nos permitirá separar objetos o regiones de interés en la imagen.



Original



Escala de grises



Binarización

Segmentación

Se creó una función en la cual convertirá la imagen resultante a una matriz binaria para poder dividir la palabra en letras (Figura 3).

Figura 3 Matriz binaria de la imagen procesada.



Fuente: *Elaboración propia.*

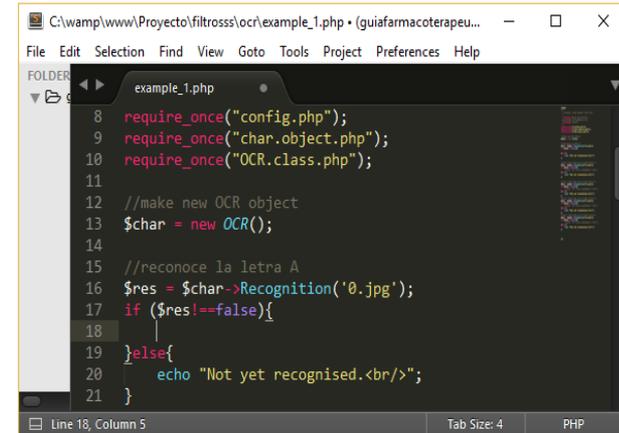


Figura 5 Segmentación de las letras a partir de la matriz binaria.

Entrenamiento con *phpOCR*

Se clasifican las letra escrita en un tipo de fuente con caracteres que la representa, en este caso se utilizo un 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.

Figura 4 Función 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 }
```

Fuente: *Elaboración propia.*

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.

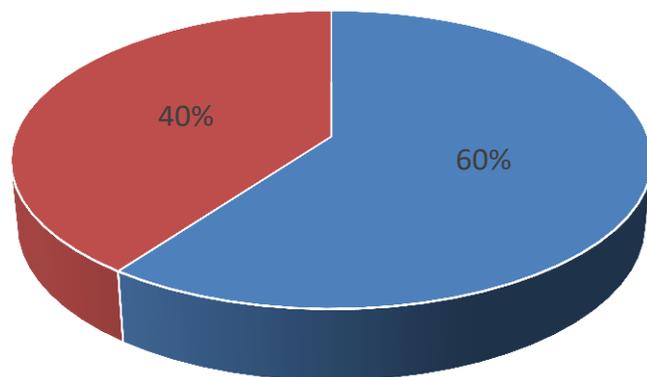
Búsqueda de texto de la BD

- Se analiza las palabras obtenidas y muestra un mensaje de alerta si hay una coincidencia entre éstas.

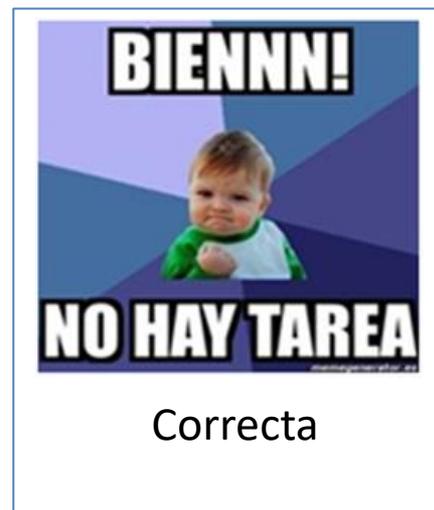
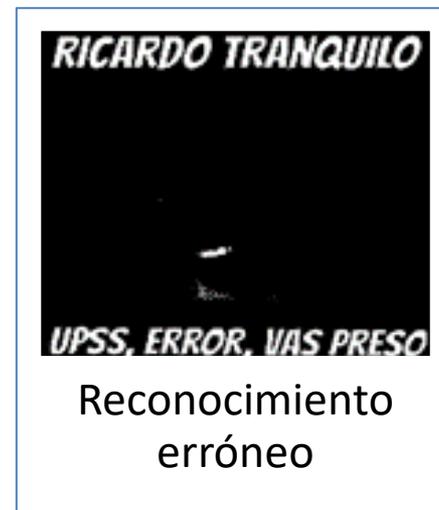
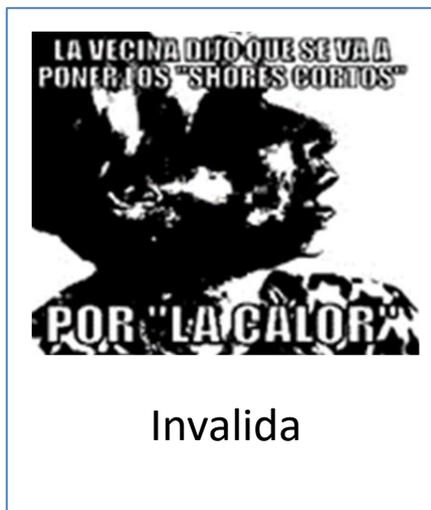


PRUEBAS Y RESULTADOS

SEGMENTACIÓN



■ Correctas ■ Incorrectas



- En promedio se obtuvieron aproximadamente 1,300 palabras ofensivas que se tenían localmente, encontrando un 60% de coincidencia, mostrando el mensaje de alerta.

TRABAJO A FUTURO



**Congreso Interdisciplinario de Energías Renovables,
Mantenimiento Industrial, Mecatrónica e Informática**

2017

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 fiabilidad.

CONCLUSIÓN

Cabe mencionar que durante la investigación sobre el ciberbullying no se encontró aplicaciones que realicen el análisis de texto ofensivo de imágenes, solo en el:

- Análisis de texto como mensajes.
- Difundir buenas practicas contra el ciberbullying.

- 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.

- 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). Cyberbullyng, violencia relacional en la sociedad de la información.
- Babu, G., Srimaiyee, P., y Srikrishna, A. (2010). Text extraction from hetrogenous 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=1323-12509304-1Q8SVKBQINRPVIK2IJ53QU6KG4&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>.



Universidad de la Cañada

- 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 ciberbullying 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.



**Congreso Interdisciplinario de Energías Renovables,
Mantenimiento Industrial, Mecatrónica e Informática**

2017



ECORFAN®

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BCIERMIMI is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- (www.ecorfan.org/ booklets)