



UNIVERSIDAD DEL BÍO-BÍO, CHILE

FACULTAD DE CIENCIAS EMPRESARIALES

Departamento de Sistemas de Información

APLICACIÓN WEB PARA ANÁLISIS DE  
AFECTOS EN TEXTOS EN ESPAÑOL  
PARA DETECCIÓN DE CYBERBULLIYNG  
EN REDES SOCIALES

TESIS PRESENTADA POR OMAR VEGA BUSTOS  
PARA OBTENER EL GRADO DE INGENIERO CIVIL INFORMÁTICO  
DIRIGIDA POR CHRISTIAN VIDAL CASTRO

Agosto, 2018

# Agradecimientos

Primero que todo agradecer a Dios por todo lo que me ha entregado.

Agradecer sin duda a mi familia, mis padres Sergio, Nancy y a mi hermano Sergio, por su apoyo incondicional en todo momento, por sus enseñanzas y todo el esfuerzo que han puesto para ayudarme a salir adelante. Sin duda dar las gracias a mi pareja Marcela por su apoyo, paciencia y comprensión, en toda ésta etapa ha sido un pilar fundamental para llevar a cabo este proceso.

Mencionar ciertamente a mi profesor guía Christian Vidal Castro, de quien siempre tuve mucho apoyo y orientación, además siempre estuvo comprometido con la realización de este proyecto.

## Resumen

Éste trabajo se enfoca en realizar análisis de subjetividad en textos en español de redes sociales para detección automática de cyberbullying mediante una aplicación web diseñada para este propósito. El análisis se efectúa mediante el corpus ingresado al sitio web, en donde los datos son enviados al servidor para aplicar un pre-procesamiento de los datos haciendo una limpieza del corpus. Luego se identifican las palabras que contienen las frases de la conversación y se incluyen en alguna categoría de afecto, donde se volverán a procesar con otro lexicón de malas palabras, lo que nos da una indicación si existe o no agresión. Como resultado se obtiene el porcentaje de cyberbullying contenido en la conversación, al posible agresor y a la posible víctima.

El modelo toma como referencia 6 afectos básicos definidos por Paul Ekman, que son: **Alegría**, **Asco**, **Enojo**, **Miedo**, **Sorpresa** y **Tristeza**, además de un lexicón con bad words (malas palabras) más utilizadas en insultos y cyberbullyig.

## Abstract

This work focuses on the analysis of subjectivity in Spanish texts of social networks for automatic detection of bullying through a web application designed for this purpose. The analysis is carried out through the registry entered into the website, where the data was sent to the server to perform a pre-processing of the data by cleaning the corpus. Then the words that contain the phrases of the conversation are identified and included in the category of affection, where they can be read again with another lexicon of bad words, which gives us an indication if there is aggression or not. As a result, the percentage of cyberbullying contained in the conversation, possible aggressor and possible victim is obtained.

The model takes as reference 6 basic affects defined by Paul Ekman, which are: Joy, Disgust, Anger, Fear, Surprise and Sadness, as well as a lexicon with bad words most used in insults and cyberbullyig.

# Índice general

<b>1. Introducción</b>	<b>9</b>
1.1. Motivación . . . . .	9
<b>2. Objetivos e Hipótesis de Investigación</b>	<b>12</b>
2.1. Objetivos . . . . .	12
2.1.1. Objetivo General . . . . .	12
2.1.2. Objetivos Específicos . . . . .	12
2.2. Alcance de la Investigación . . . . .	13
<b>3. Marco Teórico</b>	<b>14</b>
3.1. Revisión Sistemática de Literatura . . . . .	14
3.2. Conceptualización . . . . .	15
3.2.1. Bullying - Acoso Escolar . . . . .	15
3.2.2. Cyberbullying . . . . .	15
3.2.3. Agresor . . . . .	16
3.2.4. Víctima . . . . .	17
3.2.5. Consecuencias . . . . .	18
3.3. Emociones . . . . .	19
3.4. Sentimientos - Polaridad . . . . .	19
3.5. Lexicón . . . . .	20
<b>4. Caso de Estudio</b>	<b>24</b>
4.1. Diagrama . . . . .	24
4.1.1. Servidor . . . . .	25
4.1.2. StopWords . . . . .	25

<i>Índice general</i>	<i>6</i>
<hr/>	
4.1.3. Stemming . . . . .	26
4.1.4. Lexicón de afectos . . . . .	26
4.1.5. Lexicón de Bad Words [Malas Palabras] . . . . .	32
4.1.6. Principales pantallas aplicación y ejemplo . . . . .	33
<b>5. Resultados</b>	<b>36</b>
5.1. Detalle de resultados . . . . .	36
5.1.1. StopWords . . . . .	36
5.1.2. Stemming . . . . .	39
5.1.3. Obtención de porcentajes . . . . .	41
<b>6. Conclusión de pruebas y Trabajos futuros</b>	<b>43</b>
<b>Bibliografía</b>	<b>45</b>

## Índice de figuras

3.1. Extracto de la adaptación de ANEW al español.[1]	21
3.2. Extracto del lexicón creado por Pérez-Rosas.[1]	22
3.3. Extracto de Wordnet-Affect.[1]	23
4.1. Diagrama del procesamiento para el análisis de afecto.	24
4.2. Pantalla principal aplicación web.	33
4.3. Captura de pantalla con el texto insertado listo para realizar el análisis de la conversación.	34
4.4. Resultados obtenidos luego del procesamiento de la conversación.	34
4.5. Captura de pantalla la respuesta en la consola del servidor.	35

## Índice de cuadros

3.1. Ejemplo de clasificación de polaridad en una frase. . . . .	19
4.1. Ejemplos de stopwords contenidos en el documento. . . . .	26
4.2. Ejemplos de palabras de alegría. . . . .	28
4.3. Ejemplos de palabras de enojo. . . . .	28
4.4. Ejemplos de palabras de miedo. . . . .	29
4.5. Ejemplos de palabras de asco. . . . .	30
4.6. Ejemplos de palabras de sorpresa. . . . .	31
4.7. Ejemplos de palabras de tristeza. . . . .	32
4.8. Ejemplos de Bad Words. . . . .	33
5.1. Tiempo de ejecución de procesamiento del corpus con stopwords. . . .	37
5.2. Porcentaje de palabras encontradas dentro de una categoría con stop- words. . . . .	37
5.3. Tiempo de ejecución de procesamiento del corpus sin stopwords. . . .	38
5.4. Porcentaje de palabras encontradas dentro de una categoría sin stop- words. . . . .	38
5.5. Porcentaje de palabras encontradas dentro de una categoría con stem- ming. . . . .	40
5.6. Porcentaje de palabras encontradas dentro de una categoría sin stem- ming. . . . .	40

# Capítulo 1

## Introducción

Primeramente es importante destacar que el acoso escolar, bullying o cyberbullying ha sido y sigue siendo en la actualidad un gran problema. A través de éste proyecto se pretende identificar un método el cual permita realmente disminuir las cifras actuales o aportar en una parte en ello. Abordaremos los proyectos que existen actualmente con respecto a éste tema, enfocándonos también un aplicación web realizada con el propósito de identificar si existe cyberbullying en conversaciones de redes sociales analizando la subjetividad de 6 afectos de Ekman, principalmente en la aplicación móvil Whatsapp, en la cual según ANAR (Ayuda a Niños y Adolescentes en Riesgo) contiene el 81 % de los afectados.

Antes de la realización de éste proyecto no se habían encontrado trabajos similares para el idioma español, ni basados en lexicón ni machine learning, por lo tanto era un motivación adicional el poder adquirir resultados y mostrar el enfoque propuesto, yendo de la mano con el poder aportar en la disminución de casos de cyberbullying en países de habla hispana.

### 1.1. Motivación

La necesidad de resolver éste problema viene de la mano con la disminución en la edad de acceso a internet y el aumento del acoso cibernético que se ve reflejado en estudios realizados obtenidos desde el sitio web [ciberbullying.com](http://ciberbullying.com)[15] que nos

indica que el ciberbullying aumentó en los EE. UU. más del 80% en 10 años. Los investigadores apreciaron un auge desde el 6% que había en 2000 al 9% en 2005 y al 11% en 2010. [CCRC, Universidad de New Hampshire, 2010 EE.UU], en España Según el estudio “Juventud y Violencia”, de la Fundación Pfizer, el 11,6% de los adolescentes españoles de entre 12 y 18 años ha sufrido maltrato psicológico a través de la Red y un 8,1% lo ha sufrido a través del móvil. [Fundación Pfizer, 2010 España]. En Chile no son distintas las cosas donde el 87,8% de los estudiantes de 7° básico (12 años) a 4° medio (17 años) ha sido víctima de ciberbullying por Internet o vía celular. [UAB y Paz Ciudadana, 2011 Chile][15]. En 2015 se registraron aumentos en las denuncias por ciberbullying en un 33% con respecto al año anterior, y en 2016 se registró un aumento del 27% con respecto a 2015, lo que nos indica bajo todos los puntos de vista que a medida que pasa el tiempo todas estas cifras van en alza.

Es lamentable nombrar y recordar un gran número de casos en los cuales las víctimas del cyberbullying han terminado en suicidios. Casos dramáticos como el de Brandy Vela, una joven estadounidense de 18 años en ese entonces, a quien le hacían bullying por su peso terminó disparándose en el pecho frente a su familia. Sin embargo, aún después de su muerte siguió recibiendo mensajes odiosos y burlescos en sus redes sociales. Existe también el caso de Allem Halkic, quien tras todo el ciberbullying que recibió a través de las redes sociales terminó con su vida lanzándose desde un puente. Y así mismo también existen muchos casos en Chile, que según la OCDE es el segundo país con la mayor tasa de suicidios adolescente después de Corea del Norte.

Mediante éste proyecto se pretende detectar tempranamente y de forma automática los comentarios en español de redes sociales que contengan cyberbullying o acoso en cualquiera de sus formas (intimidación, burlas, amenazas, rumores, etc), analizando afectos en textos utilizando algún modelo de detección. Con esto se está atacando una de las bases del problema, previniendo a su vez todo lo que desencadena el ciberacoso, tales como la depresión, estrés, baja de autoestima, posibles secuelas a largo plazo, incluso el suicidio si se detecta y reacciona a tiempo. La investigación y posterior desarrollo toma sentido ya que en español no se encuentra disponible ninguna aplicación por lo que conlleva un estudio completo del método a utilizar.

Es importante mencionar que se tomarán en cuenta dos proyectos de título an-

teriores, donde se estudia el análisis automático de sentimientos sobre opiniones y/o comentarios de novelas en español[3] y detección automática de ironía en textos en español[2].

## Capítulo 2

# Objetivos e Hipótesis de Investigación

### 2.1. Objetivos

#### 2.1.1. Objetivo General

Se propone un modelo de detección de cyberbullying en comentarios de redes sociales en español, analizando sentimientos y subjetividad en textos. Para ello se estudiarán diversos métodos para el análisis de los comentarios en redes sociales.

#### 2.1.2. Objetivos Específicos

- Estudiar propuestas existentes para la detección de cyberbullying que utilicen técnicas de análisis de sentimientos.
- Proponer un módulo de detección de cyberbullying en redes sociales a partir de los métodos estudiados.
- Construir una aplicación que detecte el cyberbullying en comentarios de redes sociales en español.
- Evaluar los resultados obtenidos mediante el estudio de los métodos y la aplicación realizada.

## 2.2. Alcance de la Investigación

Con el gran crecimiento que ha tenido el internet, su accesibilidad y masificación, se ha logrado aumentar la conectividad a lo largo del planeta. Sin duda muchos beneficios son los que trae, como en el último tiempo las redes sociales, las cuales son utilizadas de manera muy ventajosa, como el compartir socialmente a través de la web. Sin embargo también trae consigo cosas negativas, así como un delicado tema que proponemos en éste proyecto como es el **cyberbullying**, **ciberbullying** o **ciberacoso** (*se utilizarán indistintamente los 3 términos a lo largo del documento*).

Hasta la fecha no se han encontrado trabajos que hagan referencia a la detección de cyberbullying en español, por lo tanto se propondrá un método en el cual se analizarán los comentarios con un corpus validado por psicólogos

## Capítulo 3

### Marco Teórico

En este capítulo se presentará una explicación de lo que es el cyberbullying, sus consecuencias y la conceptualización para interiorizarse en el tema.

#### 3.1. Revisión Sistemática de Literatura

Se realizó una revisión sistemática en la literatura con la finalidad de identificar los trabajos realizados anteriormente con respecto a la temática de éste proyecto. Se hizo una revisión de distintos textos y tesis con una similitud en el enfoque. Sin embargo, la mayor parte de la investigación se desarrolló en la red buscando trabajos previos en idioma español y métodos de análisis.

De tal modo se pudo observar que en el idioma español no se encuentran trabajos realizados con éste enfoque. Si hay artículos de los que se puede rescatar información en inglés para hacer adaptaciones a nuestro idioma como Detecting Offensive Language in Social Media to Protect Adolescent Online Safety de Ying Chen, Yilu Zhou y Sencun Zhu, además Automatic Detection of Cyberbullying in Social Media Text de Cynthia Van Hee , Gilles Jacobs , Chris Emmery , Bart Desmet , Els Lefever , Ben Verhoeven , Guy De Pauw , Walter Daelemans y Véronique Hoste, entre otros.

## 3.2. Conceptualización

La detección de qué está sintiendo alguien cuando escribe, qué afecto tiene un texto cuando es escrito, es realmente un desafío debido a la subjetividad que este contiene, por lo tanto a continuación se presentan algunos conceptos que ayudarán a interiorizarse en el tema.

### 3.2.1. Bullying - Acoso Escolar

El acoso escolar (*también conocido como hostigamiento escolar, matonaje escolar, maltrato escolar o en inglés bullying*) es cualquier forma de maltrato psicológico, verbal o físico producido entre estudiantes de forma reiterada a lo largo de un tiempo determinado tanto en el aula, como a través de las redes sociales, con el nombre específico de ciberacoso. Estadísticamente, el tipo de violencia dominante es el emocional y se da mayoritariamente en la clase y en los patios escolares. Los protagonistas de los casos de acoso escolar suelen ser niños y niñas en proceso de entrada en la adolescencia. Siendo ligeramente mayor el porcentaje de niñas en el perfil de víctimas.

### 3.2.2. Cyberbullying

El ciberacoso (*derivado del término en inglés cyberbullying*) también denominado acoso virtual o acoso cibernético, es el uso de medios de comunicación digitales para acosar a una persona o grupo de personas, mediante ataques personales, divulgación de información confidencial o falsa entre otros medios. Es decir, se considera ciberacoso, o ciberagresión a todo aquello que se realice a través de los dispositivos electrónicos de comunicación con el fin intencionado de dañar o agredir a una persona o a un grupo. Además, estos actos de ciberagresión poseen unas características concretas que son el anonimato del agresor, su inmediatez y su alcance[18].

Las víctimas de acoso hoy enfrentan amenazas más allá de la violencia física o los encuentros cara a cara. En éstos días el bullying ha evolucionado y se ha extendido a todos los rincones del mundo digital.

Formalmente se define como **“daño intencional y repetido infligido a través del uso de computadoras, teléfonos celulares y otros dispositivos electrónicos”** [6].

Se desarrolla ésta definición que es simple pero a su vez implica los elementos más importante como:

- **Intencionalidad:** es una acción que debe ser deliberada, no accidental.
- **Repetido:** no se trata de un incidente puntual, sino que un patrón repetitivo a través del tiempo.
- **Daño:** quien es acosado debe percibir el daño que es infligido.
- **Computadoras, celulares y otros dispositivos electrónicos:** esto diferencia el ciber acoso del bullying tradicional.

Varias son las formas en que puede ser llevado el cyberbullying, una de ellas son las actualmente populares redes sociales, donde la gran mayoría de jóvenes se encuentra activamente participando. Según el grupo de investigación EU Kids Online de la Universidad del País Vasco **la edad de inicio de acceso a internet ha disminuido considerablemente en los últimos años** pasando desde los 10-11 años a los 7 años en promedio aumentado las posibilidades que se produzca el **acoso, comentarios hirientes, amenazas, rumores**, etc, en menores de edad.

### 3.2.3. Agresor

Podríamos definir el perfil genérico del acosador como el de una persona fría, con poco o ningún respeto por los demás. El acosador disfruta y muestra su poder persiguiendo y dañando moralmente a esa persona.[10]

Se describen cuatro tipos[9]:

- El “Angel Vengador”: se ve a sí mismo como vigilante, y usa este medio para proteger amigos que podrían estar sufriendo de bullying.
- El “Hambriento de Poder”: quiere imponer su autoridad y control sobre otros a través del miedo, siendo por lo general pequeño físicamente, y etiquetado por los demás como “nerd”, o poco atractivo socialmente.
- Las “Chicas Malas”: actúan en grupo, por lo común se hallan aburridos y buscan entretenerse de esta forma.

- El “Inadvertido”: normalmente no pretende realizar el daño, pero responde a la situación (por ejemplo, reenviando o comentando) sin pensar seriamente en las consecuencias de sus acciones.

### 3.2.4. Víctima

Aunque algunos estudios identifican una serie de características psicológicas o de actitud y comportamiento comunes en las víctimas de bullying, lo cierto es que constituyen un grupo tan variado y heterogéneo que por desgracia cualquier alumno puede convertirse en blanco de las burlas de un grupo de agresores, muchas veces sin otro motivo que el puro azar. No obstante, existen algunos factores que multiplican las posibilidades de sufrir acoso:

- Sufrir alguna discapacidad física o psíquica.
- Tener algún tipo de trastorno del aprendizaje, especialmente los problemas relacionados con el lenguaje oral porque son muy evidentes y fácil objeto de burlas.
- Haber padecido acoso escolar con anterioridad.
- Pertenecer a un grupo étnico, religioso o cultural.

Un vez ha comenzado el acoso, tener una personalidad introvertida o reservada o una mala comunicación con padres o profesores dificulta que los chicos denuncien las situaciones a tiempo, con el consiguiente peligro de que el acoso vaya más, agravándose sus consecuencias y complicando su resolución[17].

Según ANAR el 95 % de las víctimas sufre algún tipo de secuelas. En cuanto a emociones, la principal consecuencia que se provoca es la ansiedad con un 79,5 %.

La **tristeza** es la segunda secuela que se más se repite en las víctimas con un **66,1 %**, seguida por el **miedo** con un **64,6 %**. La misma entidad declara que **Whatsapp** es la principal herramienta que se utiliza por parte de los acosadores, al menos así es en el **81 %** de los casos.

### 3.2.5. Consecuencias

#### A corto y mediano plazo

El ciberacoso está presente las 24 horas del día. Siempre está en línea. Incluso si se apaga el ordenador la víctima sabe qué página web está accesible, o qué personas están propagando ese rumor sobre ti. La dureza de esto es psicológicamente devastadora.[14] Los efectos secundarios de la violencia sistemática (ya sea psicológica, física o sexual) de páginas difamatorias, suelen incluir, en la mayoría de los usuarios agredidos, estrés, humillación, ansiedad, ira, impotencia y fatiga; y, aunque en pocos casos se han presentado enfermedades físicas, en gran parte de estas situaciones el individuo acosado tiene una enorme pérdida de confianza en sí mismo. Para la creación de este ambiente nocivo, los acosadores recurren a las diversas técnicas ya descritas con el fin de generar una especie de delirio de persecución en sus víctimas, de herir y asustar a la víctima.[8] Buscarán crear un estado de paranoia en la cual la víctima dude de sí misma, intentando así deshacer su vida en todos los aspectos: profesional, social, amoroso y familiar.

#### Secuelas del acoso prolongado

Los supervivientes de acosos prolongados desarrollan cambios de personalidad característicos del desorden de estrés postraumático, incluyendo deformaciones en la capacidad de relacionarse.[11]

Como los síntomas postraumáticos son tan persistentes y tan variados, pueden ser confundidos con rasgos de la personalidad de la víctima. Tiempo después de que haya acabado el acoso, muchos afectados, siguen sintiendo que ha muerto una parte de ellos, y algunas víctimas, las más afligidas, desearían estar muertas. La amenaza de aniquilación que define el acoso puede atormentar al superviviente mucho después de que haya pasado el peligro. Las personas traumatizadas se sienten absolutamente abandonadas, y solas, exiliadas del sistema humano de cuidado y protección.

El acontecimiento traumático que significa ser víctima del hostigamiento del entorno ha destruido la creencia de que el individuo puede ser “uno mismo” en relación con los demás.

Así por definición, los acontecimientos traumáticos frustran la iniciativa y des-

trozan la competencia individual. Con independencia de la actitud de la víctima, el acoso ha proseguido, es decir que no importa lo valiente y lo llena de recursos que esté la víctima; sus acciones fueron insuficientes para evitar el desastre. Después de los acontecimientos violentos del acoso las víctimas revisan y juzgan su propia conducta, y los sentimientos de culpa y de inferioridad son prácticamente universales.

Se ha comentado anteriormente que los acontecimientos traumáticos destrozan los sistemas de protección normales que dan a las personas una sensación de control, de conexión y de significado. Así las reacciones traumáticas tienen lugar cuando la acción no sirve para nada, es decir, cuando no es posible ni resistirse ni escapar, entonces el sistema de autodefensa humano se siente sobrepasado y desorganizado. Es en estas situaciones cuando los acontecimientos traumáticos producen profundos y duraderos cambios en la respuesta fisiológica, las emociones, lo cognitivo y la memoria.[12]

### 3.3. Emociones

Las emociones no son un fenómeno simple, en donde muchos factores influyen sobre ello. Izard declaró que para tener una definición completa de las emociones se debe tener en cuenta el sentimiento consciente de la emoción, los procesos que ocurren en el sistema nervioso, en el cerebro y los modelos expresivos observables de emoción. En si, un conjunto de reacciones neuroquímicas y hormonales que nos predisponen a reaccionar de cierta manera ante un estímulo externo.

### 3.4. Sentimientos - Polaridad

Hay una forma de clasificar los sentimientos, ya sean positivos, negativos o neutros, consiguiendo la polaridad de una palabra, frase, oración, etc.

Polaridad	Frase
Positiva	“La inteligencia artificial es una buena rama de la informática”
Negativa	“Él no está contento con su nuevo trabajo”
Neutra	“Mi amigo usa redes sociales”

Cuadro 3.1: Ejemplo de clasificación de polaridad en una frase.

### 3.5. Lexicón

**Lexicón:** Se entiende por lexicón a un diccionario de palabras predefinido para algún propósito a realizar.

Existen lexicones para:

**Sentiment:** Las palabras contienen una etiqueta indicando si la palabra es positiva, negativa o neutra. Puede contener más información como porcentajes, promedios, valores, etc. Como por ejemplo:

#### **Lexicón propuesto SANEW (Redondo et. al, 2007)**

Es una adaptación al español realizada por el grupo de Jaime Redondo et al. , basándose en el Affective Norms for English Words, (ANEW) compuesta de 1034 palabras en inglés, que posteriormente fueron traducidas. Para crear este lexicón se reunió a 720 personas, quienes evaluaron estas palabras, y las clasificaron en tres dimensiones que se definen a continuación. La dimensión valencia (que oscila entre lo agradable y desagradable), la dimensión excitación (que oscila desde la calma a la agitación) y finalmente la tercera dimensión dominio o control (que oscila desde el control a la salida de control). Un ejemplo sería, la palabra funeral, que tiene una valencia de 1.48 y una excitación de 5.06, es decir, representa una palabra agradable para el usuario y que produce una excitación intermedia. Otro ejemplo, es la palabra madre, con una valencia 8.19, y excitación 5.19, que representa una palabra agradable con una excitación neutra.[3]

#### **Lexicón propuesto por Perez - Rosas et. al, 2012**

Este lexicón fue realizado por Verónica Pérez-Rosas et al., quienes elaboraron dos lexicones ambos en español, el primero *fullStrengthLexicon* contiene un léxico proveniente del *OpinionFinder* lexicón (Janyce Wiebe, 2005), conformado por 1347 palabras que fueron clasificadas de manera manual, consultando las distintas ponderaciones que estaban asociadas a una palabra en particular, y tomando su valor más alto. El segundo lexicón, *mediumStrengthLexicon*, conformado por 2496 palabras, que fue generado de manera automática, basándose en el *SentiWordNet* (Sebastiani,

2006), seleccionando el synset que contiene una puntuación superior a 0.5. Ambos lexicones fueron etiquetados como positivo (pos) y negativo (neg).[3]

Number	E-Word	S-Word	Val-Mn-All	Aro-Mn-All	Dom-Mn-All	Val-Mn-Fem	Aro-
722	derelict	abandonado	1,68	5,88	3,11	1,61	
583	bees	abejas	3,18	6,59	3,42	2,99	
622	abortion	aborto	2,80	6,07	4,17	2,64	
975	scorching	abrasador	2,46	6,00	4,00	2,41	
218	hug	abrazo	8,13	5,70	6,12	8,23	
300	overwhelmed	abrumado	2,90	5,36	3,84	2,97	
623	absurd	absurdo	3,80	4,90	4,44	4,01	
624	abundance	abundancia	6,80	5,40	5,68	7,10	
48	bored	aburrido	2,33	2,90	4,78	2,16	
2	accident	accidente	1,32	7,58	2,62	1,19	
989	sour	ácido	3,41	5,57	4,29	3,43	
88	cozy	acogedor	7,64	4,37	6,08	7,80	
740	event	acontecimiento	5,99	5,81	5,01	6,09	
948	pungent	acre	4,23	4,19	4,50	4,34	
4	activate	activar	6,00	6,50	5,82	5,87	
631	agreement	acuerdo	7,24	4,83	6,22	7,20	
404	snuggle	acurrucarse	6,98	4,28	6,01	6,96	
581	addict	adicto	2,41	5,77	3,28	2,47	
488	wealthy	adinerado	6,21	5,58	5,69	6,07	
5	admired	admirado	7,33	6,53	5,48	7,37	
6	adorable	adorable	7,48	5,24	5,76	7,67	
546	adult	adulto	5,68	4,90	5,16	5,63	
808	impair	afectar	3,48	5,61	3,77	3,26	
7	affection	afecto	8,10	6,19	5,83	8,20	
381	secure	afianzar	5,93	5,46	5,66	5,99	
127	distressed	afligido	1,96	4,41	3,93	1,99	

Figura 3.1: Extracto de la adaptación de ANEW al español.[1]

### Enfoque de lexicón para Affect

Otro enfoque que se le puede dar al lexicón es para affect analysis, donde las palabras son clasificadas en categorías según las emociones que representan, tales como **Developing affective lexical resources**.

bueno	1166413	pos	pos	
burla	1222859	neg	neg	
calvario	7287730	neg	neg	
cansancio	14016361	neg	neg	
censura	6713512	neg	neg	
confianza	3833898	pos	pos	
confusion	7507912	neg	neg	
contento	908929	pos	pos	
conveniente	51373	pos	pos	
correccion	4912732	pos	pos	
corrupcion	4850996	neg	neg	
cortes	1947741	pos	pos	
criterio	5614175	pos	pos	
crueldad	3797439	neg	neg	
curioso	665156	neg	pos	
desconocido	1122595	neg	neg	neg
destacable	2586957	pos	pos	
dictadura	8440630	neg	neg	
diestro	2226162	pos	pos	
difícil	746047	neg	neg	
dificultad	5686955	neg	neg	
digno	2586206	pos	pos	
educado	1947741	pos	pos	
entusiasmo	4634540	pos	pos	
erupcion	7435273	neg	neg	
estima	6206800	pos	pos	
estimulante	1642245	pos	pos	pos
estricto	711059	neg	neg	
facultades	5624042	pos	pos	
feroz	1263013	neg	neg	

Figura 3.2: Extracto del lexicón creado por Pérez-Rosas.[1]

**Wordnet-Affect (Valitutti et. al, 2004)**

Este léxico fue realizado por Alessadro Valitutti et al., quienes elaboraron el WordNetAffect a partir del WordNet, por medio de la selección y etiquetado del syntec (relaciones semánticamente equivalentes), representando afectivamente el concepto. El syntec nos proporciona una correlación entre un concepto y las palabras correspondientes. Paralelamente se le añade una etiqueta de dominio, que representa el concepto afectivo que personifica un estado emocional, entre los que se encuentran: alegría, sorpresa, ira, tristeza, disgusto y miedo. (A Valitutti, 2004) Por ejemplo, podemos tener un conocimiento afirmativo de que la gente le tiene miedo a los terremotos. Basándonos en algunas declaraciones y analizándolas pueden entregar relaciones semánticas que las relacionan entre sí. Así que, cada vez que la palabra terremoto

aparezca, el sistema podría saber que produce miedo, y por lo tanto clasificar el texto con esta dimensión adicional.

<b>Anger</b>	<pre> n#05588321 wrath n#05588822 umbrage offense n#05589637 pique temper irritation n#05588618 lividity n#05614716 irascibility short_temper spleen quick_temper n#05588725 infuriation enragement n#05588960 indignation outrage n#05589074 huffiness n#05590065 harassment torment n#05588413 fury rage madness n#05589778 frustration n#10337658 fit tantrum scene n#05588172 dudgeon high_dudgeon n#05605389 displeasure n#05589169 dander hackles n#05589301 bad_temper ill_temper n#05589430 annoyance chafe vexation n#05587878 anger cholera ire n#05589957 aggravation exasperation a#00117684 aggravated provoked a#00117308 angrv </pre>
<b>Fear</b>	<pre> n#05593296 unassertiveness n#05592071 trepidation n#05592739 timidity timidness timorousness n#05592642 suspense n#05591681 stage_fright n#05592914 shyness n#05592428 shadow n#05591591 scare panic_attack n#05592546 presage n#05591377 panic terror n#05593389 intimidation n#05591290 hysteria n#05591212 horror n#05593183 hesitance hesitancy n#05579669 heartlessness coldheartedness hardheartedness n#05592156 foreboding premonition presentiment boding n#05590260 fear fearfulness fright n#05593002 diffidence self-doubt self-distrust n#05579830 cruelty mercilessness pitilessness ruthless n#05590900 creeps n#05603115 chill pall </pre>

Figura 3.3: Extracto de Wordnet-Affect.[1]

# Capítulo 4

## Caso de Estudio

### 4.1. Diagrama

El siguiente diagrama representa el camino que toma el procesamiento de la información para este proyecto. Sin embargo, el detalle de éste se realizará a lo largo del capítulo.

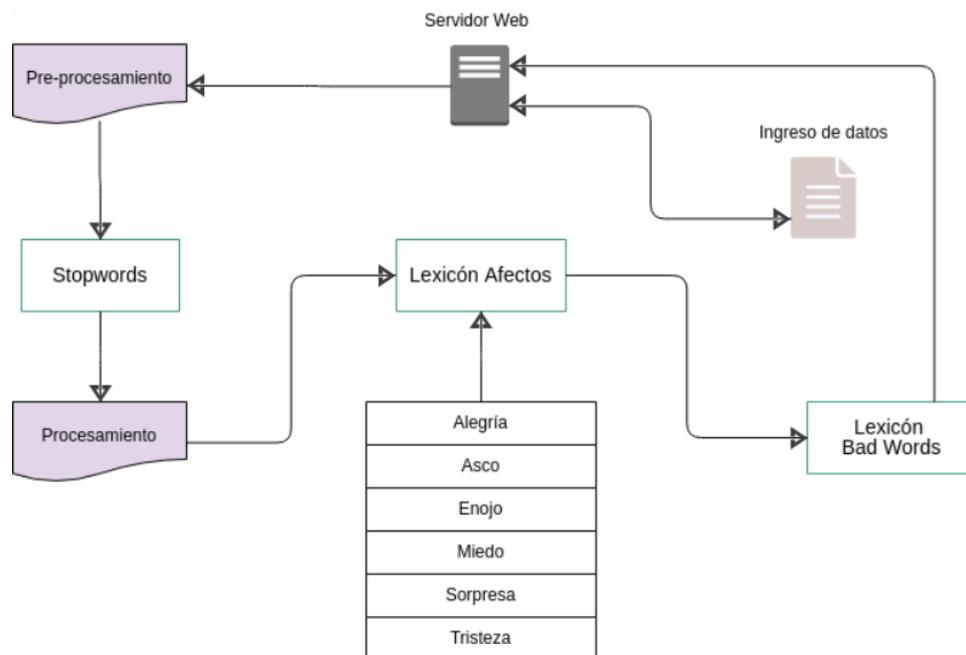


Figura 4.1: Diagrama del procesamiento para el análisis de afecto.

En la figura anterior se indica el proceso que se sigue para realizar el análisis de la información.

Comienza con el ingreso de datos desde el usuario a través del sitio web hacia el servidor, donde lo primero que se realiza es un pre-procesamiento de los datos eliminando palabras innecesarias dentro del análisis (stopwords). Luego se procesa la información verificando a qué categoría de afecto corresponde. Dependiendo de los porcentajes obtenidos en cada categoría anterior se realiza un nuevo análisis con otro diccionario que contiene *bad words* (“malas palabras”). Nuevamente bajo ciertos porcentajes se obtienen resultados que son entregados desde el servidor hacia el cliente mediante el sitio web.

#### 4.1.1. Servidor

El servidor recibe la información de la conversación ingresada por el cliente desde el sitio web, verificando que el campo no esté vacío para el posterior procesamiento de la información.

#### 4.1.2. StopWords

Cuando hablamos de stop words o palabras vacías nos referimos a todas aquellas palabras que carecen de un significado por sí solas. Las palabras vacías suelen ser artículos, preposiciones, conjunciones, pronombres, etc.

Para entenderlo mejor podemos analizar la frase “optimizar el SEO”, en ella las palabras “optimizar” y “SEO” tienen significado por sí solas y nos dan información acerca del contenido del texto mientras la palabra “el” por sí sola no nos dice nada, carece de significado.[16]

Es importante recalcar que no se debe asociar los stopwords de una forma negativa como “palabras malas o perjudiciales”, ya que son palabras necesarias que forman parte de nuestro idioma.

Para la realización de éste proyecto se generó un documento con una recopilación la mayor cantidad de stopwords posibles obtenidos a través de la red [345 elementos], incluyendo artículos en masculino y femenino, preposiciones, conjunciones, pronom-

bres, algunas abreviaciones, etc.

A continuación se presenta una tabla con algunas palabras contenidas en el documento de stopwords del proyecto.

A	Ajenas	Algo	Alguno
Acá	Ajeno	Algún	Algunos
Ahí	Ajenos	Alguna	Allá
Ajena	Al	Algunas	Allí
Ambos	Ante	Antes	Aquel
Aquella	Aquellas	Aquello	Aquellos
Aquí	Aqui	Arriba	Así
Atrás	Aún	Aunque	Bajo
Bien	Cabe	Cada	Casi
Cierta	Ciertas	Cierto	Ciertos
Como	Cómo	Con	Conmigo
Conseguimos	Conseguir	Consigo	Consigue

Cuadro 4.1: Ejemplos de stopwords contenidos en el documento.

### 4.1.3. Stemming

Stemming es un método para reducir una palabra a su raíz o (en inglés) a un stem. Hay algunos algoritmos de stemming que ayudan en sistemas de recuperación de información. Stemming aumenta el recall que es una medida sobre el número de documentos que se pueden encontrar con una consulta. Por ejemplo una consulta sobre "bibliotecas" también encuentra documentos en los que solo aparezca "bibliotecario" porque el stem de las dos palabras es el mismo ("bibliotec"). El algoritmo más común para stemming es el algoritmo de Porter. [13]

### 4.1.4. Lexicón de afectos

**Lexicón:** Se entiende por lexicón a un diccionario de palabras predefinido para algún propósito a realizar.

En este caso se pretende utilizar un lexicón para cada categoría de afecto propuesto por Ekman, psicólogo pionero en el estudio de las emociones y expresiones faciales, considerado dentro de los mejores psicólogos del siglo XXI.

Según Ekman existen 6 emociones básicas; en palabras del autor, esto significa que es posible identificar ciertas señales fisiológicas específicas y acontecimientos que anteceden a su aparición, y que son compartidas por todas las culturas. Estas emociones básicas son el **miedo**, la **tristeza**, la **alegría**, la **ira**, el **asco** y la **sorpresa**. [4, 5]

### Alegría

La alegría es un sentimiento grato y vivo que suele manifestarse con signos exteriores. Palabras, gestos o actos con que se expresa el júbilo. La alegría es una de las emociones que experimenta el ser humano en la vida.

Es un estado interior, fresco y luminoso, generador de bienestar general, altos niveles de energía, y una poderosa disposición. La alegría es un sentimiento o valor, la acción constructiva, que puede ser percibida en toda persona, siendo así que quien la experimenta, la revela en su apariencia, lenguaje, decisiones y actos. La tristeza es el sentimiento o emoción contraria.

También se puede definir como el estado de ánimo más confortable por el cual se puede pasar. La tristeza juega aquí un rol, ya que sin ella no se podría sentir alegría, y según estudios médicos, la alegría ayuda al mejoramiento de la salud.

### Enojo

La ira, rabia, enojo o furia es una emoción que se expresa a través del resentimiento o de la irritabilidad. Los efectos físicos de la ira incluyen aumento del ritmo cardíaco, de la presión sanguínea y de los niveles de adrenalina y noradrenalina. Algunos ven la ira como parte de la respuesta cerebral de atacar o huir de una amenaza o daño percibidos. La ira se vuelve el sentimiento predominante en el comportamiento, cognitivamente, y fisiológicamente cuando una persona hace la decisión consciente de tomar acción para detener inmediatamente el comportamiento amenazante de otra fuerza externa.

Admirable	Afectivo	Agradable	Alegre
Alegremente	Alegría	Amable	Amigable
Amor	Amar	Amoroso	Animar
Apreciado	Bueno	Cariño	Cercano
Compasivo	Confortable	Disfrutar	Empatía
Empático	Encantar	Encanto	Entusiasmo
Estima	Estimulante	Euforia	Fascinar
Felicidad	Felicitar	Feliz	Fervoroso
Festivo	Fraterno	Gozo	Graciosamente
Gracioso	Grandioso	Gratificante	Gustar
Hermanable	Idolatrable	Idolatrar	Impresionar
Impulsar	Inspirar	Jubilo	Lealtad

Cuadro 4.2: Ejemplos de palabras de alegría.

Abominación	Aborrecer	Acoso	Afectar
Agraviar	Agresividad	Alterarse	Amargura
Ambicioso	Antagonismo	Antipatía	Atribulado
Avaro	Avaricioso	Berrinche	Celos
Celópata	Codicioso	Criminalmente	Crucificar
Desagradable	Desagrado	Desanimado	Descaro
Desconcertante	Despecho	Despreciar	Desprecio
Detestar	Disgusto	Dolido	Enajenado
Endiablar	Enfurecer	Enojado	Enojo
Enrabiado	Envidia	Envidiable	Exacerbar
Exasperar	Fastidiado	Fastidioso	Frustración
Furia	Hostil	Indignación	Ira

Cuadro 4.3: Ejemplos de palabras de enojo.

## Miedo

El miedo se puede definir como una emoción que aparece en un gran número de especies animales cuando percibimos una amenaza. Se distingue de la ansiedad, un término muy manejado por la psicología, en el hecho de que el miedo se asocia a un peligro inminente, sea éste real o potencial, mientras que la ansiedad tiene un carácter anticipatorio y con frecuencia más abstracto.

Esta percepción de amenaza desencadena un cambio en la conducta del organismo; en este sentido cabe hacer mención al concepto de la respuesta de lucha-huida, que se debe a la liberación de neurotransmisores que hacen que el nivel de activación del sistema nervioso se intensifique, especialmente la adrenalina y la noradrenalina. La paralización es otro patrón comportamental asociado al miedo.

Cuando sentimos miedo aumentan nuestras frecuencias cardíaca y respiratoria (lo cual lleva a la hiperventilación), así como la tensión muscular y la intensidad de la sudoración. En cuanto a la expresión facial, el miedo se asocia con la elevación de las cejas y de los párpados superiores, con la retracción de los labios y con la tensión de los párpados inferiores.

Acobardar	Acobardarse	Agredido	Alarmado
Alerta	Ansioso	Aprensión	Asustar
Aterrado	Atroz	Crueldad	Espantar
Feo	Frio	Histeria	Horrendo
Horrible	Horroroso	Inquietamente	Inseguridad
Inseguro	Intimidación	Inmisericorde	Nerviosa
Nerviosismo	Nervioso	Miedoso	Miedo
Monstruo	Monstruosamente	Pánico	Paralizado
Pesimista	Presagiar	Presentimiento	Retraidamente
Retroceder	Sombra	Sumiso	Suspense
Temeroso	Temor	Terrible	Terror
Timidez	Vacilación	Vergonzoso	Verguenza

Cuadro 4.4: Ejemplos de palabras de miedo.

## Asco

El asco es la emoción que expresa el rechazo a estímulos determinados que resultan desagradables para alguno de los sentidos; por ejemplo, puede aparecer en respuesta a objetos que huelen mal pero también ante personas que provocan en el individuo una especie de idea de sufrir algún tipo de contaminación, sea esta consciente o no.

Según Jack y colaboradores, la emoción de asco no se distingue de la de ira, sino que fisiológicamente se trataría de la misma. Así, por ejemplo, arrugar la nariz o levantar el labio superior son indicadores tanto de lo que denominamos “asco” como de la ira; en este mismo sentido, cuando sentimos miedo o sorpresa una de nuestras reacciones naturales es alzar las cejas.[7]

Abominablemente	Abominable	Aborrecible	asco
Asqueado	Desagradable	Desagradar	Despreciable
Detestable	Enfermante	Enfermo	Fétido
Hastiado	Inmoral	Malvado	Nauseabundo
Obsceno	Odioso	Odiosa	Odiosamente
Ofensivo	Ofensivamente	Ofensiva	Ofender
Repugnante	Repugnar	Repelente	Repulsión
Ruidoso	Desagrado	Disgusto	Hastío
Desgana	Empalago	Empaco	Hartura
Escrúpulo	Aprensión	Recelo	Fastidio
Fastidiar	Rechazo	Asquerosidad	Inmundicia
Arcada	Náusea	Vómito	Porquería

Cuadro 4.5: Ejemplos de palabras de asco.

## Sorpresa

A diferencia del resto de emociones básicas, la sorpresa no es considerada positiva ni negativa (es decir, agradable o desagradable) sino que puede incluir componentes fisiológicos propios tanto de la alegría como del miedo.

Siguiendo con las propuestas de Paul Ekman, la expresión facial de la sorpresa consistiría en la apertura de la boca y de los ojos junto con la elevación de la musculatura asociada a las cejas.

Admiración	Anonadado	Apabullar	Asombrado
Admirable	Asombrada	Asombro	Asombroso
Asombrosa	Atonito	Atontado	Atontada
Aturdido	Aturdir	Atontar	Confundir
Confundido	Confundida	Desconcertar	Desconcertado
Desconcertada	Desconcertante	Despistar	Despistado
Despistada	Engatuzar	Engatuzado	Engatuzada
Enredar	Enredado	Espectacular	Esplendido
Estúpido	Excelente	Fantástico	Fabuloso
Increíble	Genial	Intrigar	Maravilloso
Perplejo	Rimbombante	Sensasional	Sobrecogido
Sorprendente	Sorprendido	Superiormente	Terrible

Cuadro 4.6: Ejemplos de palabras de sorpresa.

## Tristeza

Cuando estamos tristes tenemos menos energía, y esto facilita que nos tomemos un tiempo para estar con nosotros mismos y pensemos más profundamente en el suceso que nos provoca esta emoción.

Por otro lado, la tristeza también tiene un papel importante en la comunicación con los demás. Las expresiones faciales asociadas a la tristeza expresarán de manera no verbal que necesitamos ayuda. Por lo tanto, esto facilitará que la gente de nuestro alrededor se preocupe y nos preste la atención que necesitamos, sin ni siquiera tener que pedirlo. Dada la importancia que tiene la tristeza tanto a nivel de comunicativo como para ayudarnos a gestionar los cambios y llevarnos al crecimiento personal, no es de extrañar que se trate de otra de las emociones que han perdurado hasta nuestros días.

Para efectos del proyecto se analizan las palabras de una frase en las conversaciones ingresadas por el usuario con cada una de las categorías indicadas anteriormente, para luego obtener los porcentajes de cada afecto dentro de la frase y en el texto completo.

Abandonado	Abatido	Aburrido	Acongojado
Acongojar	Acosar	Acoso	Acosado
Aflijido	Angustiante	Apagado	Apagar
Apenado	Apesadumbrado	Aproblemado	Arrepentido
Arrepentimiento	Arruinado	Avergonzado	Bajoneado
Cansado	Compungido	Compungimiento	Consternación
Culpa	Deplorable	Depresión	Depresivo
Deprimente	Deprimido	Deprimir	Desafortunado
Desalentador	Desalentar	Desaliento	Desamparado
Desamparo	Desanimado	Desanimar	Desanimo
Desconsuelo	Desmoralizado	Desolado	Desventurado
Difícil	Dolor	Infeliz	Lacrimoso

Cuadro 4.7: Ejemplos de palabras de tristeza.

#### 4.1.5. Lexicón de Bad Words [Malas Palabras]

Este lexicón es una segunda etapa para el análisis de las frases contenidas dentro de la conversación proporcionada desde el cliente. Se hace una comparación de la palabra con el lexicón de *bad words* para obtener si corresponde e indica una tendencia hacia el bullying dependiendo algunos porcentajes.

Usada en el contexto de identificar las palabras más usadas en el cyberbullying, ya sean insultos, agresiones, ataques, garabatos, etc, utilizados con el fin de denostar y agredir a la víctima.

Por lo tanto se hizo una recolección de todas éstas palabras, incluyendo todas sus formas escritas, reconociendo todas sus modificaciones para obtener un mejor desempeño y una mejor precisión a la hora de obtener resultados. Con la recopilación se logró generar un diccionario de 917 palabras.

A continuación se presenta una tabla con algunas de la palabras que contiene este lexicón:

Es de recalcar que en la tabla anterior no se colocó garabatos ni insultos de mayor envergadura para que el proyecto pueda ser leído por cualquier persona sin restricción. Además está escrito de la manera más entendible, ya que las palabras en el diccionario se encuentran estipuladas con todas sus variantes.

Abobada	Abobado	Aborto	Abuso
Acusete	Adicto	Adoquín	Afeminao
Afeminado	Afilar	Aforrar	Africano
Agarron	Agila	Agilao	Allegado
Amarrete	Amateur	Amermelao	Amorfo
Analfabeto	Apitutao	Apretao	Asopao
Barsa	Bastardo	Bellaco	Bisexual
Blancucho	Bruto	Cabezón	Cabron
Caca	Cafiche	Cahuinero	Califa
Callampa	Canuto	Cartucha	Chancho
Copuchento	Cuatica	Cuico	Cuma
Descerebrado	Enano	Enfermo	Estúpido

Cuadro 4.8: Ejemplos de Bad Words.

Luego de haber obtenido los porcentajes de malas palabras dentro de las frases se analizan los resultados y son enviados al servidor, donde se reenvía una respuesta al cliente redireccionándolo en el sitio web a una pestaña la cual contiene información del porcentaje de posible cyberbullying dentro de la conversación. Además se logra identificar al posible agresor y a la posible víctima.

#### 4.1.6. Principales pantallas aplicación y ejemplo



Figura 4.2: Pantalla principal aplicación web.

Para mostrar el proceso de uso de la aplicación se muestra en la siguiente figura

como se pega la conversación que se quiere analizar. En la foto se muestra una extracto de la conversación utilizada para propósitos de análisis dentro del proyecto.

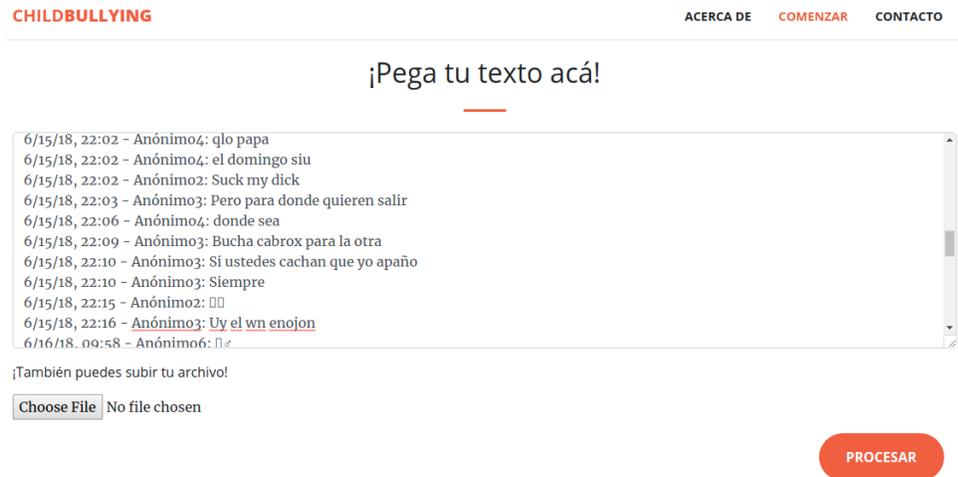


Figura 4.3: Captura de pantalla con el texto insertado listo para realizar el análisis de la conversación.

Luego de darle al botón "procesar" de la pantalla anterior, el servidor comienza a procesar, valga la redundancia, la conversación realizando el análisis de las frases y palabras contenidas en estas, agregando éstas palabras a una categoría específica de afecto descritos anteriormente.

La siguiente figura nos muestra la pestaña a la cual nos redirecciona la aplicación con los resultados obtenidos luego del procesamiento desde el servidor.



Figura 4.4: Resultados obtenidos luego del procesamiento de la conversación.

La Figura 4.4 nos muestra los resultados obtenidos luego del procesamiento. Nos entrega como información el porcentaje de posible cyberbullying en la conversación, para éste caso un 10.59 %. Además nos entrega el posible agresor y posible víctima, que son Anónimo2 y Anónimo3 respectivamente. No está demás recordar que los nombres de los participantes han sido cambiados para mantenerlos en el anonimato.

Desde el lado del servidor obtenemos la siguiente respuesta:

```
count_frases: 3200
afecto: [551, 50, 205, 102, 69, 173]
Posible agresor: anónimo2:
Posible víctima: anónimo3
cantidad frases: 3200
cantidad de frases con prob de bullying: 339
porcentaje de bullying: 10.59375
EL tiempo de ejecución es: 6.035345077514648
127.0.0.1 - - [16/Aug/2018 22:28:49] "POST / HTTP/1.1" 200 -
■
```

Figura 4.5: Captura de pantalla la respuesta en la consola del servidor .

# Capítulo 5

## Resultados

### 5.1. Detalle de resultados

Los análisis descritos en los siguientes párrafos fueron realizados desde una conversación real de whatsapp, con un corpus de 3200 frases donde hay 6 participantes, los cuales fueron identificados cambiando sus nombres por Anónimo1, Anónimo2, Anónimo3, Anónimo4, Anónimo5 y Anónimo6, ocultando las identidades evitando cualquier tipo de conflicto.

#### 5.1.1. StopWords

Primero que todo vemos un análisis en tiempos de ejecución, haciendo una comparativa del software con y sin stopwords.

#### **Con stopwords**

El Cuadro 6.1 fue realizado con el análisis de 5 pruebas del corpus con stopwords, del cual se rescata el tiempo de ejecución de cada una.

Además se obtiene la cantidad de palabras que encontró dentro de las categorías de afectos y los porcentajes.

Num. de Prueba	Tiempo de ejecución en segundos
1	8.384556770324707
2	8.606800317764282
3	8.387563705444336
4	8.425436019897461
5	8.3577299118042
Promedio	<b>8.43241735</b>

Cuadro 5.1: Tiempo de ejecución de procesamiento del corpus con stopwords.

Afecto	Cantidad de palabras	Porcentaje
Alegría	8172	24.38 %
Asco	2893	8.63 %
Enojo	4560	13.60 %
Miedo	7226	21.56 %
Sorpresa	3257	9.71 %
Tristeza	7403	22.09 %

Cuadro 5.2: Porcentaje de palabras encontradas dentro de una categoría con stopwords.

Con un total de 33511 palabras encontradas dentro de alguna categoría, podemos decir que hay una tendencia hacia algunos afectos, aún así la alta cantidad de palabras encontradas nos dice que el mantener los stopwords nos genera también confusión dentro del análisis. [Ir a la comparativa con stopwords] 50.34375 %

### Sin stopwords

El Cuadro 6.3 fue realizado con el análisis de 5 pruebas del corpus sin stopwords, del cual se rescata el tiempo de ejecución de cada una.

Num. de Prueba	Tiempo de ejecución en segundos
1	5.830798864364624
2	5.75787091255188
3	5.619127511978149
4	5.659636735916138
5	5.705440044403076
Promedio	<b>5.71457481</b>

Cuadro 5.3: Tiempo de ejecución de procesamiento del corpus sin stopwords.

Afecto	Cantidad de palabras	Porcentaje
Alegría	551	47.91 %
Asco	50	4.34 %
Enojo	205	17.82 %
Miedo	102	8.87 %
Sorpresa	69	6.00 %
Tristeza	173	15.04 %

Cuadro 5.4: Porcentaje de palabras encontradas dentro de una categoría sin stopwords.

Con un total de 1150 palabras encontradas dentro de alguna categoría, podemos decir que hay una tendencia hacia algunos afectos, y obtenemos datos más objetivos de lo que realmente contienen las frases sin confundir con algunas palabras, sucediendo algo parecido con el stemming en la sección siguiente. Además, podemos apreciar una clara diferencia en la optimización en los tiempos de ejecución, pasando de 8.43

segundos a 5.71 segundos, una diferencia de 2.72 segundos solamente eliminando los stopwords. A medida que sea una mayor cantidad de frases esa diferencia se hará aún mayor.

### 5.1.2. Stemming

En una primera instancia se realizaron pruebas con **stemming** (def. 5.1.3) para adaptar de una mejor manera las palabras y así obtener mejores resultados en las pruebas.

Al menos para éste proyecto, el aplicar stemming a las palabras de cada frase no provoca buenos resultados debido a que se generan conflictos con algunas palabras de contiene el lexicón de afecto, como por ejemplo si tenemos la frase:

*“Que lindo el sol”*

Eliminando stopwords nos quedan las siguientes palabras:

*“lindo sol”*

Luego aplicando stemming a las palabras obtenemos:

*“lind sol”*

Como podemos ver la palabra *“lindo”* vuelve a su raíz, mientras tanto la palabra *“sol”* queda tal cual debido a que se encuentra en su raíz. Aún así al comparar éstas dos palabras con el lexicón obtenemos que *lindo* correspondía a la categoría de *alegría*. Hasta el momento todo correcto, en tanto si comparamos la palabra *sol* podemos encontrarnos en que se puede ubicar en la categoría *tristeza* donde no tiene relación alguna, debido a que la tomaría como raíz de la palabra *soledad* o *solo*, ubicándola en una categoría la cual no corresponde tergiversando las palabras contenidas en las frases y modificando y alterando erróneamente los resultados.

Para tener una mejor idea del párrafo anterior se pueden realizar tablas con la comparativa de ambas ejecuciones para luego sacar conclusiones.

Afecto	Cantidad de palabras	Porcentaje
Alegría	639	36.08 %
Asco	90	5.08 %
Enojo	385	21.73 %
Miedo	183	8.87 %
Sorpresa	125	7.06 %
Tristeza	349	19.70 %

Cuadro 5.5: Porcentaje de palabras encontradas dentro de una categoría con stemming.

Se encontraron 1771 palabras en alguna categoría, donde el porcentaje de posible cyberbullying es un 14.153 % dentro de toda la conversación. Además arrojó los siguientes resultados para identificar al agresor y victima:

Agresor: Anónimo3

Victima: Anónimo3

Como se puede apreciar lo que describe al comienzo de la sección 6.1.2 éstos resultados no son totalmente objetivos en cuanto a cantidad de palabras pertenecientes a un afecto determinado por los conflictos expuestos anteriormente.

Afecto	Cantidad de palabras	Porcentaje
Alegría	551	47.91 %
Asco	50	4.34 %
Enojo	205	17.82 %
Miedo	102	8.87 %
Sorpresa	69	6.00 %
Tristeza	173	15.04 %

Cuadro 5.6: Porcentaje de palabras encontradas dentro de una categoría sin stemming.

Se encontraron 1150 palabras en alguna categoría, donde el porcentaje de posible cyberbullying es un 10.59375 % dentro de toda la conversación. Además arrojó los

siguientes resultados para identificar al agresor y victima:

Agresor: Anónimo2

Victima: Anónimo3

En los resultados obtenidos de ésta última prueba revisada en la tabla 6.6 podemos encontrarnos con valores representativos de las palabras asignadas a los diferentes afectos definidos.

### 5.1.3. Obtención de porcentajes

Para poder entregar los resultados, decir si una frase contiene cyberbullying o no y así determinar si hay un agresor y una victima es necesario determinar algunos valores.

En el corpus nombrado anteriormente se determinó cierta cantidad de frases que contienen cyberbullying, y en ello se fueron ajustando los valores de los porcentajes hasta lograr acercarse lo más posible a estos, obteniendo así los siguientes resultados:

- Si el porcentaje de enojo dentro de una frase es igual o mayor al 25 % es considerado cyberbullying.
- Si el porcentaje de bad words dentro de una frase es igual o mayor al 20 % es considerado cyberbullying.
- Si la suma del porcentaje de enojo y el porcentaje de bad words dentro de una frase es mayor o igual al 20 % es considerado cyberbullying.
- Si el porcentaje de tristeza dentro de una frase es igual o mayor al 25 % es considerada una posible victima

De la lista anterior podemos concluir:

$$cyberbullying = \left( \frac{(0,25 * enojo)}{total} \geq 25 \right) \vee \left( \frac{(0,2 * badwords)}{total} \geq 20 \right) \quad (5.1)$$

$$cyberbullying = \frac{(badwords + enojo) * 0,2}{total} \geq 20 \quad (5.2)$$

$$victima = \left( \frac{(0,25 * tristeza)}{total} \geq 25 \right) \quad (5.3)$$

Cuando una frase es considerada contenedora de cyberbullying el autor es guardado para luego hacer una comparación y retornar quien haya realizado la mayor cantidad de frases de éste tipo. De la misma forma es almacenado una posible victima, cuando el software determina que la frase contiene las características correspondientes.

## Capítulo 6

# Conclusión de pruebas y Trabajos futuros

A continuación se presenta la discusión de los resultados de las pruebas realizadas.

En el presente proyecto se obtuvo un corpus con datos reales pero manteniendo en el anonimato a los participantes de la conversación para asegurarse que no haya ningún tipo de problema a futuro. La dificultad fue identificar aquellas frases que tenían indicios de cyberbullying para poder realizar los análisis posteriormente, lo cual se hizo manualmente para así hacer una proyección de los datos y obtener las métricas que determinan si existe una probabilidad real de cyberbullying.

En términos generales el modelo propuesto es capaz de identificar si en una conversación existe cyberbullying, el porcentaje que éste se encuentra, además del posible agresor y posible víctima.

Es interesante el análisis que se puede realizar después haber desarrollado éste proyecto teniendo en cuenta por ejemplo que fue complicado encontrar la respuesta a la inexactitud de los resultados al aplicar stemming en las frases, donde uno pensaría que los efectos que produciría serían positivos; para éste caso al menos no lo fue. Aparte de los cambios que se generan en los tiempos de ejecución son sumamente notorios cuando se eliminan los stopwords, disminuyendo en casi 3 segundos la respuesta, pensando en una cantidad de datos mayor.

Por otro lado, es importante recalcar que para utilizar éste tipo de métodos es necesario tener lexicones robustos, abarcando todas las posibilidades del lenguaje

para aumentar la precisión en los resultados, de lo contrario se harán estimaciones no reales de lo que se quiere comprobar.

Además podemos concluir que para éste tipo de trabajos éste método no es la manera más óptima de hacer un análisis. Si bien funciona cuando el proyecto es pequeño, cuando éste mismo es de mayor envergadura el cambio es muy notorio y no posee una base tan sólida para obtener buenos resultados. Lo mejor para un caso así sería aplicar Machine Learning y no con Lexicón, debido al constante cambio que nuestro lenguaje está sometido en cortos plazos de tiempo en donde un método basado en lexicón es una de sus peores desventajas, esto no sucedería con un enfoque basado en aprendizaje.

Finalmente podemos agregar que para el idioma español no se habían realizado trabajos con la temática del cyberbullying utilizando ésta metodología, por lo tanto servirá para poder realizar comparativas y sacar mejores conclusiones.

## Bibliografía

- [1] Chistian Vidal C. *Análisis de la subjetividad en textos*. U. Bío-Bío.
- [2] Marcelo Pinto C. *Modelo de Detección Automática de Ironía en Textos en Español*. U. Bío-Bío, 2017.
- [3] Carol Oyarzún D. *Análisis automático de sentimientos sobre opiniones y/o comentarios de novelas en español*. U. Bío-Bío, 2014.
- [4] Paul Ekman. *An argument for basic emotions*. *Cognition and Emotion*. 1992.
- [5] Paul Ekman. *Facial expression and emotion*. *American Psychologist*. 1993.
- [6] Sameer Hinduja y Justin W. Patchin. *Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying, 2nd Edition*. Sage-Publications, 2015.
- [7] R. E. Jack, O. G. Garrod, y P. G Schyns. *Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time*. 2014.
- [8] Vania Tovilla Quesada Jéssica Dorantes Segura y Patricia Trujano Ruiz. Cyberbullying, acoso online. 2011. URL <https://web.archive.org/web/20110709013540/http://www.depsicoterapias.com/articulo.asp?IdArticulo=436>.
- [9] Kowalski, Limber, y Agatston. *Cyber bullying: El acoso escolar en la era digital*. 2010.
- [10] Antonio Chacón Medina. Una nueva cara del internet: El acoso. *Universidad de granada*, pág. 4, 2003.

- 
- [11] Marina Parés Soliva. Peritación social del mobbing. 2007. URL <https://web.archive.org/web/20161203211652/http://www.acosomoral.org/pdf/peritajesocial/peritajesocial1.PDF>.
- [12] VVAA. Ciberacoso, consecuencias del ciberacoso. . URL [https://es.wikipedia.org/wiki/Ciberacoso#Consecuencias\\_del\\_ciberacoso](https://es.wikipedia.org/wiki/Ciberacoso#Consecuencias_del_ciberacoso).
- [13] VVAA. Stemming. . URL <https://es.wikipedia.org/wiki/Stemming>.
- [14] VVAA. El ciberacoso: un problema en auge. 2013. URL <http://web.archive.org/web/20130405071823/http://www.laflecha.net/canales/blackhats/noticias/el-ciberacoso-un-problema-en-auge>.
- [15] VVAA. Incidencia/prevalencia del ciberbullying. 2013. URL <http://www.ciberbullying.com/cyberbullying/incidencia-del-ciberbullying/comment-page-6/>.
- [16] VVAA. ¿qué son las stop words o palabras vacías? 2014. URL <http://www.vozidea.com/que-son-las-stop-words-o-palabras-vacias>.
- [17] VVAA. Uno de cada cuatro casos de acoso escolar es por ciberbullying. 2016. URL <http://www.anar.org/estudio-ciberbullying/>.
- [18] David Álvarez García, Alejandra Barreiro-Collazo, y José-Carlos Núñez. *Cyberaggression among Adolescents: Prevalence and Gender Differences*. 2017.

# Herramientas Utilizadas y Recursos

## Servidor

### Características del Ordenador

Todo el trabajo realizado se desarrolló bajo un ordenador Lenovo Thinkpad con un procesador Intel Core i7-3740QM 2.7Ghz, 8GB de memoria RAM, GPU integrada Intel HD4000, bajo la distribución GNU/Linux ManjaroLinux 17.1.11.1 64-bit.

## Plataforma

Como base del sistema se realiza en **Python** en su versión 3.6 debido a que es un lenguaje de programación muy reconocido, por lo tanto posee una cantidad de librerías extraordinarias para todo lo que es procesamiento del lenguaje natural.

## Framework

Ya que se basa en una plataforma web, el desarrollo se realiza mediante el micro framework de python **Flask**, que permite realizar páginas web de forma fácil y rápida, minimizando las líneas de código. Está basado en la especificación WSGI de Werkzeug y el motor de templates Jinja2 y tiene una licencia BSD.

## **Librerías adicionales**

Como librería adicional se agregó NLTK para el procesamiento del lenguaje natural y poder aplicar stemming en las frases de la conversación