



UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES
CENTRO DE CIENCIAS BÁSICAS
DEPARTAMENTO DE SISTEMAS DE INFORMACIÓN

TESIS

DETECCIÓN DE CIBER-VIOLENCIA EN LOS SITIOS DE NOTICIAS DE
AGUASCALIENTES, MEDIANTE MINERÍA DE CONTENIDO WEB

PRESENTA

Jorge Luis García Rodríguez

PARA OBTENER EL GRADO DE MAESTRÍA EN INFORMÁTICA Y
TECNOLOGÍAS COMPUTACIONALES

TUTOR

Dr. Jaime Muñoz Arteaga.

COMITÉ TUTORAL

Dra. Loecelia Ruvalcaba Sánchez.

Dr. Francisco Javier Álvarez Rodríguez.


Aguascalientes, Ags; 01 de Junio de 2012.

Por este conducto autorizamos a:

L.I. Jorge Luis García Rodríguez


La impresión de su documento final de tesis, ya que cumple con los requisitos de contenido y forma exigidos por la Universidad Autónoma de Aguascalientes

Director de Tesis




Dr. Jaime Muñoz Arteaga

Sinodales



Dra. Loecelia Ruvalcaba Sánchez



Dr. Francisco Javier Álvarez Rodríguez



UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES

JORGE LUIS GARCÍA RODRÍGUEZ
ALUMNO DE LA MAESTRÍA EN INFORMÁTICA Y TECNOLOGÍAS
COMPUTACIONALES
P R E S E N T E

Estimado Sr. García:

Por medio de este conducto me permito comunicarle a usted, que el trabajo tesis o caso práctico titulado "Detección de situaciones de ciber-violencia en los sitios de noticias de Aguascalientes, mediante minería de contenido web", está autorizado y será bajo la dirección del Dr. Jaime Muñoz Arteaga; y como revisores el Dr. Francisco Javier Álvarez Rodríguez y la Dra. Loeceña Ruvalcaba Sánchez, para que pueda obtener así el grado de Maestría en Informática y Tecnologías Computacionales.

Sin otro particular me permito saludarle (a) muy afectuosamente.

ATENTAMENTE

Aguascalientes, Ags., 1 de junio del 2012

"SE LUMEN PROFERRE"

LA DECANO

MTRA. MARTHA CRISTINA GONZÁLEZ DÍAZ



VoBo M en C JORGE EDUARDO MACÍAS LUEVANO
SECRETARIO TÉCNICO-MAESTRÍA EN INFORMÁTICA
Y TECNOLOGÍAS COMPUTACIONALES

- c.c.p. Dr. Alejandro Padilla Díaz- Secretario de Investigación y Posgrado.
- c. c. p. M en C Jorge Eduardo Macías Luevano.- Secretario Técnico de la Maestría en Informática y Tecnologías Computacionales.
- c. c. p. Dr. Jaime Muñoz Arteaga
- c. c. p. Dr. Francisco Javier Álvarez Rodríguez
- c. c. p. Dra. Loeceña Ruvalcaba Sánchez
- c. c. p. Archivo.

AGRADECIMIENTOS.

Quiero expresar mi agradecimiento a:

- Mi asesor, el Dr. Jaime Muñoz Arteaga, por su apoyo y paciencia.
- Los profesores, Dra. Loecelia Ruvalcaba Sánchez y Dr. Javier Álvarez Rodríguez, por haber aceptado ser parte del jurado de esta tesis.
- Mis padres, por brindarme todo su apoyo, ya que sin él no hubiera podido hacer posible este trabajo.
- Mis hermanos, por su apoyo incondicional, paciencia y por estar conmigo en toda esta etapa.
- A todos los profesores, que a lo largo de este tiempo, me han enseñado e inspirado para continuar aprendiendo.
- A la **Universidad Autónoma de Aguascalientes**, al **Centro de Ciencias Básicas** y en especial al **Departamento de Sistemas de Información** por permitirme ser parte de una generación de triunfadores y líderes en informática.

DEDICATORIAS.

A mis padres porque siempre creyeron en mí, porque siempre me apoyaron, y porque en gran parte gracias a ellos, hoy puedo ver alcanzada mi meta.

A mis hermanas y hermanos por todo ese apoyo incondicional que siempre me brindaron, por su tiempo, por su comprensión, por todas las recomendaciones y consejos que me dieron.

Gracias por haber fomentado en mí ese deseo de superación y de ganas de triunfar.

Jorge Luis García Rodríguez.

Índice

RESUMEN.....5

ABSTRACT.....6

CAPITULO 1.....7

 1. Introducción.....7

 1.1. Descripción del contexto de investigación.....7

 1.2. Relevancia y justificación de la investigación.....8

 1.3. Planteamiento del problema.....10

 1.4. Descripción general de la contribución de la investigación.....12

 1.5. Descripción de los capítulos de la tesis.....12

 1.6. Objeto de estudio.....13

 1.7. Preguntas/Objetivos de la investigación.....13

 1.8. Metodología.....14

 1.8.1. Material de investigación.....15

 1.8.2. Procedimiento.....16

 1.8.3. Captura y procesamiento de la información.....16

 1.9. Hipótesis.....17

CAPITULO 2.....18

 2. Marco Teórico.....18

 2.1. Historia de internet.....18

 2.2. Concepción de la violencia a través de la historia.....19

 2.3. Violencia.....20

 2.4. Tipos de Violencia.....21

 2.4.1. Violencia social.....21

 2.4.2. Violencia económica.....22

 2.4.3. Violencia de género.....22

 2.4.4. Otras definiciones de violencia.....22

 2.5. Causas de la violencia.....23

 2.6. Mecanismos de defensa psicológico.....24

 2.7. La violencia en la actualidad.....29

 2.8. Las redes sociales.....30

 2.9. Las aplicaciones sociales.....33

 2.10. Ciber-violencia en las redes sociales.....34

2.11.	Las redes sociales y la edición colaborativa.....	36
2.12.	La edición online.	37
2.13.	Periodismo en México.....	37
2.14.	Periodismo online.....	39
2.15.	Los blogs como herramienta periodística.....	39
2.16.	La Minería web.....	40
2.16.1.	Minería de contenido web.....	41
2.16.2.	Minería de estructura de la web.....	41
2.16.3.	Minería de los registros de navegación de la web.....	42
2.16.4.	Recuperación de información (IR).	42
CAPITULO 3.....		44
3.	Estado del Arte.....	44
CAPITULO 4.....		56
4.	Modelo de la ciber-violencia en los sitios web.....	56
4.1.	Modelo conceptual de ciber-violencia en los sitios de noticias.....	56
4.1.1.	Descripción del modelo de ciber-violencia.....	57
4.1.2.	Roles involucrados en el modelo de ciber-violencia.....	58
4.2.	Caso de estudio. Violencia textual en el sitio de noticias Aguascalientes.com.....	59
CAPITULO 5.....		62
5.	Sistema de Detección de Situaciones de Ciber-Violencia en los sitios de noticias y blogs.....	62
5.1.	Modelo arquitectónico de la aplicación.....	62
5.2.	Componentes principales del sistema.....	64
5.3.	Tecnologías elegidas para el desarrollo del sistema.....	65
5.4.	Análisis de contenidos.....	67
5.4.1.	Fases del análisis.....	67
5.4.2.	Selección de grupos de periódicos online.....	68
5.4.3.	Extracción de información.....	69
5.4.4.	Minería de contenido.....	70
5.4.5.	Comparación mediante el Algoritmo de la Distancia de Levenshtein.....	70
5.4.6.	Evaluación.....	71
5.4.7.	Difusión.....	71
CAPITULO 6.....		72

6.	Validación y resultados del análisis en la detección de situaciones de ciber-violencia.	72
6.1.	Desarrollo sustancial del análisis de detección de situaciones ciber-violencia.....	72
6.1.1.	Selección de grupo de noticias online.	72
6.1.2.	Extracción del texto mediante minería de contenido utilizando el Sistema de Detección de Situaciones de Ciber-Violencia.....	76
6.1.3.	Comparación.....	78
6.1.4.	Evaluación.	78
6.2.	Resultados obtenidos.....	79
6.2.1.	Resultados derivados de la teoría.	79
6.2.2.	Resultados derivados de la aplicación del instrumento.....	79
6.2.3.	Resultados derivados del Sistema de Detección de Situaciones de Ciber-Violencia.	80
CAPITULO 7.....		86
7.	Conclusiones.....	86
7.1.	Aportación de éste trabajo a investigaciones futuras.....	89
GLOSARIO.....		90
BIBLIOGRAFIA.....		91
ANEXOS.....		95

Índice de Tablas.

Tabla 1. Roles involucrados en el modelo de ciber-violencia.58

Tabla 2. Selección y evaluación68

Tabla 3. Sitios de noticias online en aguascalientes.73

Tabla 4. Selección de sitios de noticias.73

Tabla 5. Establecimiento del orden de la población.74

Tabla 6. Resultados de la muestra.81

Tabla 7. Resultados porcentuales por sitio de noticias online.82

Índice de Figuras.

Figura 1. Representación de las redes sociales31

Figura 2. Concepto de las aplicaciones web enriquecidas.....31

Figura 3. Redes sociales colaborativas.37

Figura 4. Taxonomía de la minería web.43

Figura 5. Modelo de ciber-violencia en los comentarios.....57

Figura 6. Diagrama de bloques.63

Figura 7. Arquitectura de la aplicación.....64

Figura 8. Componentes del sistema.65

Figura 9. Fases del análisis de contenidos.67

Figura 10. Contenido noticioso y comentarios.69

Figura 11. Algoritmo de la distancia de levensthein.71

Figura 12. Interface del sistema de detección de ciber-violencia.77

Figura 13. Extracción de texto de la noticia.77

RESUMEN.

En la presente tesis se encuentra el estudio realizado a sitios de noticias y blogs de Aguascalientes, donde la finalidad fue determinar la existencia de ciber-violencia en los mismos.

El análisis fue aplicado a una muestra tomada de una población seleccionada de sitios de noticias online y consistió en encontrar palabras consideradas agresivas en los comentarios textuales hechos por los usuarios sobre las noticias publicadas.

El análisis fue realizado a través de una metodología diseñada para la selección, procesamiento, y detección de texto considerado como agresivo mediante minería de contenido y comparación de textos, utilizando para el análisis, un sistema basado en web diseñado específicamente para realizar la extracción, búsqueda y comparación de textos.

ABSTRACT.

This thesis contains a study made on web sites of online news and blogs from Aguascalientes, that it is intended to determine the existence of ciber-violence in its content.

The analysis was applied to a sample taken from online news sites where it could be determined the aggressive text on the comments made by the users about the news published online.

This analysis was performed through a designed methodology to select process and detect the text defined as aggressive by using Web mining content and text comparison, and using to analyze the text, a web-base system designed specifically to extract, find and text comparison.

TESIS TESIS TESIS TESIS TESIS

CAPITULO 1.

1. Introducción.

1.1. Descripción del contexto de investigación.

Las tecnologías de la información y la comunicación (TICS) han contribuido en los avances tecnológicos, de tal manera que se han convertido en herramientas indispensables para el desarrollo de la vida humana, las cuales han hecho posible crear nuevas formas de interacción en el ciberespacio. Estas herramientas han evolucionado la forma de interacción entre las personas, pues actualmente representan un medio de comunicación inmediata, cómoda y de fácil acceso.

En el presente proyecto se hace patente el uso de la tecnología, la cual, como ya se mencionó, ha evolucionado la forma de vida del hombre, pero también se aborda la existencia de un fenómeno social que en aras de la evolución humana, ha penetrado en el área virtual, para crear un concepto relativamente nuevo, denominado ciber-violencia.

Para entrar en materia y comprender el tema, es necesario definir algunos conceptos que serán utilizados a lo largo de la presente investigación. En este sentido tenemos que:

- **Ciber:** Es el elemento prefijal que entra en la formación de palabras con el significado de cibernético o informático¹, por ende, es todo aquello relativo a la informática.
- **Violencia:** Es el uso intencional de la fuerza física o psicológica para generar una amenaza, contra uno mismo, otra persona, o contra un grupo de personas, que cause o tenga muchas probabilidades de causar lesiones, muerte, daños psicológicos, trastornos del desarrollo, o privaciones².
- **Red social:** es un conjunto de estructuras sociales formadas por varias personas, las cuales están conectadas por uno o varios tipos de relaciones como amistad, parentesco, o intereses comunes.

¹ <http://buscon.rae.es/draeI/Srvlt/ObtenerHtml?origen=RAE&IDLEMA=87478&NEDIC=Si> Sitio de la Real Academia Española, consultado el miércoles veintitrés de mayo, del dos mil doce, a las 23:20 horas.

² Alexander Butchart, «Violence and public health: an integrated, evidence-based approach to preventing domestic violence and child abuse.». *130th international training course visiting experts' papers*

Por lo tanto, la **ciber-violencia**, es la interacción humana realizada a través de las herramientas de tecnologías de la información y comunicación, para generar una amenaza o intimidación, mediante el uso de la fuerza psicológica, dirigida a otra persona o a un grupo de personas.

Ahora, por naturaleza, las redes sociales en internet tienen como objetivo compartir un perfil, formar parte de una red de amigos, publicar contenidos digitales, o simplemente compartir aficiones o material digital, tales como fotos y videos. Entre los usos más recurrentes que se les da a las redes sociales están el de comunicarse con amigos o familiares, dar seguimiento a temas, realizar contribuciones a temáticas publicadas en línea, crear nuevos contenidos, así como conocer los acontecimientos diarios que se publican en internet sobre temas, productos o servicios, los cuales son de su interés.

Pero estos no han sido los únicos usos que han tomado las redes sociales, ya que la tendencia que están desarrollando las personas en el uso de estos sitios, está dando un cambio drástico, debido al contenido que los mismos usuarios publican sobre los temas. Los problemas sociales, políticos y económicos, tales como la pobreza, el desempleo y la inseguridad, han ocasionado que las personas desahoguen su enojo usando las herramientas que las mismas redes sociales les proveen, utilizándolas para manifestar su inconformidad o insatisfacción sobre algún tema de actualidad, publicando comentarios ofensivos, o agrediendo directamente a los demás usuarios en las publicaciones de la red.

1.2. Relevancia y justificación de la investigación.

Por un lado, el crecimiento de la red de internet y los avances tecnológicos en términos de redes sociales y, por el otro, los problemas sociales, políticos y de inseguridad que acechan a nuestro país, nos han llevado a buscar alguna forma de manifestar nuestra inconformidad ante tales situaciones.

Una de estas formas es mediante el uso de las herramientas que las redes sociales proveen para manifestar las diferencias de opiniones que puede originar una noticia o tema en particular, por medio de la publicación de comentarios textuales, por parte de las personas, o usuarios de la red, en los sitios de noticias o blogs. Así, dependiendo del impacto que

tengan sobre la sociedad, la noticia publicada en el sitio web, provocará que los usuarios realicen comentarios positivos o negativos, a través de su participación mediante las herramientas que integra el sitio.

Esto está creando un fenómeno en los sitios web de noticias en México, específicamente en los sitios de noticias en línea de Aguascalientes, ya que es común el que se observen situaciones de ciber-violencia, la cual es ocasionada por la diferencia de opiniones y puntos de vista respecto a los temas.

Por otro lado, el aumento de la ciber-violencia en los sitios web está creciendo considerablemente debido a la problemática social, a la situación de inseguridad y al entorno político por el que está atravesando nuestro país. Estas situaciones ya se han estado observando en Aguascalientes, donde los sitios de noticias y blogs están siendo utilizados para realizar esta práctica.

Aguascalientes cuenta con una población de 1, 884,996 habitantes³, de los cuales el 81% se encuentra en la zona urbana y el 19% en la zona rural. Ahora bien, respecto a las personas que tienen acceso a internet, Aguascalientes se ubica en el lugar número 11 con respecto al total nacional, con el 24.05% de los hogares conectados a la red⁴.

El derecho de expresión forma parte del catálogo de derechos fundamentales del hombre, sin embargo, frente a todo derecho siempre existe un deber, pues derechos y obligaciones deben cohabitar e interrelacionarse a efecto de garantizar una convivencia armónica y fraternal entre los hombres. En este sentido, los instrumentos internacionales de los que México forma parte, consagran el derecho de expresión como un derecho inherente a la naturaleza del hombre, así la Declaración Universal de Derechos Humanos, en su artículo 19⁵ contempla la libertad de expresión como garantía de hombre para no ser molestado por

³ FUENTE: INEGI. Censo de Población y Vivienda 2010.

Link:

<http://cuentame.inegi.org.mx/monografias/informacion/ags/poblacion/distribucion.aspx?tema=me&e=01> consultado el 20 abril de 2012, a las 20:48 horas.

⁴ <http://www.desdelared.com.mx/2010/universidad/0517-internet.html> consultado el 9 de mayo de 2012 a las 20:52 horas.

⁵ Declaración Universal de Derechos Humanos, consultada en el sitio oficial de las Naciones Unidas: <http://www.un.org/es/documents/udhr/> el 19 de mayo del 2012 a las 09:12 horas.

causa de sus opiniones; por su parte, la Declaración Americana de Derechos y Deberes del Hombre⁶, instrumento internacional que postula principios básicos de derechos humanos, el cual México ha ratificado, es decir, se ha obligado a su observancia y cumplimiento, contempla en su artículo IV, el derecho de libertad de investigación, opinión, expresión y difusión, mientras que en su artículo XXIV, habla respecto al deber de toda persona frente a la sociedad, en donde debe existir una convivencia tal, de manera que pueda asegurarse el desarrollo integral de la personalidad de cada individuo.

Pese a que los instrumentos internacionales mencionados, obligan a México para que reconozca y aplique los principios fundamentales de derechos humanos que en ellos se consagran, en nuestro país, no existe todavía alguna legislación que regule el problema de la ciber-violencia, en la que se plasmen objetivamente los derechos y obligaciones de los usuarios de internet para que exista una convivencia virtual armónica entre la población cibernauta, donde se respete el derecho a la libre expresión, pero con la limitante del respeto frente al derecho de expresión del otro usuario; por ello, la ciber-violencia continuará mientras no exista una normatividad que la regule y persista la problemática social en nuestro país.

La ciber-violencia se puede llevar acabo de diferentes formas y en diferentes niveles, y puede ser percibida por los usuarios como violencia en línea, por lo que actualmente es necesario identificar las situaciones de ciber-violencia que se presentan en los sitios web de noticias y blogs.

1.3. Planteamiento del problema.

El problema que actualmente enfrenta la sociedad Aguascalentense, es la publicación de comentarios ofensivos en los sitios web de noticias, tal es el caso de sitios informativos como diarios y blogs.

⁶ Declaración Americana de Derechos y Deberes del Hombre, consultada en el sitio oficial de la Comisión Interamericana de Derechos Humanos: <http://www.cidh.org/basicos/Basicos1.htm>, el diecinueve de mayo del dos mil doce, a las nueve horas con treinta minutos.

En la actualidad, con la Web 2.0 los sitios web permiten que los usuarios puedan contribuir en los temas publicados en línea, con la finalidad de que el usuario ejerza el derecho de expresión y mediante un nombre de usuario, como identificador, pueda contribuir con su opinión acerca de un tema en específico. Esto se realiza a través de los llamados wikis⁷, ya que esta tecnología coloca a los usuarios al mismo nivel que los colaboradores del sitio.

Las redes sociales han entrado en la vida de millones de personas sin importar su edad, condición social, sexo, religión, etc., y cada vez más están siendo adoptadas, con fines comerciales, por grupos empresariales, pues mediante ellas publican los productos y servicios que ofrecen.

Muchos sitios presentan aportaciones bastante enriquecedoras, pero muchos otros presentan aportaciones ofensivas y amarillistas, donde pareciera que las reglas en la participación son el insulto y la descalificación. Esto ocasiona una cadena de insultos y agresiones entre los usuarios. Cabe mencionar que actualmente los sitios de noticias y blogs utilizan moderadores, con la finalidad de bloquear los comentarios ofensivos realizados por los usuarios, aunque esta práctica aun no se realiza en todos los sitios.

Sobre la participación en los periódicos de noticias online, un estudio publicado por la AMIPCI⁸, sobre los principales hábitos de los usuarios en internet, reveló que el 22 % de los usuarios mexicanos acostumbra leer diarios/periódicos en línea, el 36% acostumbra únicamente leer las noticias locales y el resto se divide entre escuchar radio, leer diarios, revistas, etc.

En otro estudio sobre el acceso a redes sociales en México, realizado también por la AMIPCI⁹, se observó en base a una muestra, que un 53% de los usuarios utilizan las redes sociales para el seguimiento de opiniones sobre cultura, entretenimiento y deportes, y el 46% para dar seguimiento y opinión a las últimas noticias acontecidas.

⁷ Una wiki es un sitio web cuyas páginas pueden ser editadas por múltiples voluntarios a través del navegador web. Los usuarios pueden crear, modificar o borrar un mismo texto que comparten.

⁸ <http://estudios.amipci.org.mx:8080/mashboard/main.jsp>, consultado el día 15 de febrero de 2012, a las 20:48 horas.

⁹ <http://www.slideshare.net/gpbolde/estudio-amipci-de-redes-sociales-2011>, consultado el 15 de febrero de 2012, a las 21:43 horas. Este estudio fue sobre el acceso a redes sociales en México, el cual se realizó tomando como base un total de 705 entrevistados.

De acuerdo a lo anterior, la problemática que se presenta se muestra a continuación:

Existen situaciones de ciber-violencia en los sitios de noticias y blogs, ya que estos permiten que los usuarios aporten comentarios sobre un tema abierto o un tema en específico, dichos comentarios pueden ser aportaciones enriquecedoras, pero también pueden ser aportaciones denigrantes y ofensivas.

Cabe mencionar que, en la actualidad no todos los sitios web permiten la participación de los usuarios en los temas publicados.

Para fines académicos y específicamente para el desarrollo de esta tesis, se analizarán algunos sitios sociales de periodismo y blogs de Aguascalientes.

1.4. Descripción general de la contribución de la investigación.

La contribución lograda en esta investigación es la siguiente:

- EL uso de la minería web, específicamente de la minería de contenido, para detectar situaciones de violencia que se presentan en los sitios de noticias y blogs en el Estado de Aguascalientes.

1.5. Descripción de los capítulos de la tesis.

Los capítulos se encuentran organizados de la siguiente manera.

- 1- **Introducción.** Describe el contexto de la investigación, la relevancia, justificación, planteamiento del problema, contribución. Describe de manera general el enfoque de la investigación.
- 2- **Marco Teórico.** Describe de manera general los conceptos teóricos tomados como base para el desarrollo de la tesis.
- 3- **Estado del Arte.** Describe las investigaciones que fueron realizadas utilizando técnicas de minería web, específicamente de minería de contenido, las cuales fueron aplicadas en diferentes áreas. Así como también se describen las investigaciones realizadas para la detección de ciber acoso en internet.

- 4- **Modelo de la Ciber-Violencia en los sitios web.** Este capítulo describe el modelo conceptual de la ciber-violencia en los sitios de noticias, así como los roles involucrados. También se hace mención de un caso de estudio.
- 5- **Sistema de Detección de Situaciones de Ciber-Violencia en los Sitios de Noticias y Blogs.** En este capítulo se describe la arquitectura del sistema utilizado para la búsqueda de ciber-violencia en los comentarios de los sitios de noticias y blogs. Además se describe las fases para el desarrollo del análisis.
- 6- **Validación y Resultados del Análisis en la Detección de la Ciber-Violencia.** En este capítulo se presenta el desarrollo del análisis así como los resultados obtenidos.
- 7- **Conclusiones.** Este apartado muestra las conclusiones sobre los resultados obtenidos, sobre el método de investigación empleado, sobre el aprendizaje personal así como las recomendaciones finales.

1.6. Objeto de estudio.

El objeto de estudio de esta tesis se centra en las situaciones de ciber-violencia que se presentan en las redes sociales, tales como los sitios de noticias en línea y blogs en el estado de Aguascalientes.

1.7. Preguntas/Objetivos de la investigación.

El objetivo general de la presente investigación es detectar situaciones de violencia que se presentan en los comentarios de los sitios de noticias y blogs de Aguascalientes.

Además de:

- Conocer los antecedentes de la violencia.
- Identificar los principales tipos de violencia.
- Determinar la existencia de agresividad en el contenido de los comentarios realizados en los sitios de noticias o blogs.

Las preguntas que guiaron la presente investigación son las siguientes:

P1. ¿Es posible identificar cuáles son las palabras más utilizadas en los sitios de noticias y blogs en Aguascalientes para ejercer ciber-violencia?

P2. ¿Existe ciber-violencia en los sitios de noticias y blogs en Aguascalientes?

1.8. Metodología.

Desde el punto de vista de la metodología, la investigación que se realizará será del tipo correlativa, debido a que se refiere a una comprobación sin prueba, basado en la relación entre variables que forman dichas hipótesis a comprobar. Es descriptiva, pues se indagó sobre cómo se manifiesta el fenómeno de la ciber-violencia en los sitios de noticias y blogs en Aguascalientes.

Al mencionar las bases de la investigación, es necesario señalar que se utilizarán los siguientes tipos de observación: indirecta, pues debido a que es descriptiva necesita de la fundamentación bibliográfica y de la información de datos muertos para sustentar los antecedentes de la presente investigación, por lo tanto los datos que se van a obtener por medio de ésta serán de tipo cualitativo; directa, por la necesidad de la investigación, pues es necesario complementar con éste tipo de observación, ya que la observación de los grupos se hará en base a la aplicación de los instrumentos de análisis que nos permitirán rescatar datos de tipo cualitativo y cuantitativo.

El instrumento que se aplicará, es un instrumento semi-estructurado, ya que está formulado con preguntas cerradas y abiertas, de las cuales se obtendrán las estadísticas necesarias para realizar la investigación.

La población que se eligió para la observación directa son alumnos en un rango de edad de 23 a 27 años, quienes pertenecen a la población estudiantil de la Universidad La Concordia, con un total de 1,495 alumnos correspondientes al turno matutino, que con la aplicación de la fórmula de muestreo finito arrojó un total de 305 encuestas como muestra para aplicar el cuestionario.

La fórmula para determinar el tamaño de la muestra es la siguiente:

$$n = \frac{Z^2(p)(q)(N)}{(N)(E)^2 + (Z)^2(p)(q)} \quad n = \frac{(1.96)^2(0.5)(0.5)(1,495)}{(1,495)(0.05)^2 + (1.96)^2(0.5)(0.5)}$$

Donde:

n:	Tamaño de muestra	p:	Variabilidad positiva
N:	Tamaño de la población	q:	Variabilidad negativa
Z:	Nivel de confianza	E:	Precisión de error

De las cuales:

n=	Tamaño de muestra	p=	0.5
N=	1,495	q=	0.5
Z=	1.96	E=	0.05

Por tanto, n= 305

1.8.1. Material de investigación.

Para realizar la investigación sobre la ciber-violencia en los sitios de noticias y blogs en el Estado, fue necesario iniciar con un instrumento que ayudara a recabar la información elemental para crear un catálogo en el cual se generara un inventario de texto, entendido éste como palabras y expresiones escritas con significado agresivo, por lo cual se elaboró un instrumento de evaluación en el que se plantearon preguntas en torno a la ciber-violencia. El instrumento creado contiene un total de 8 preguntas, abiertas, cerradas y de selección múltiple.

Las preguntas del instrumento para detectar la ciber-violencia son: 1.¿Visitas sitios de noticias o Blogs en Internet?, 2.¿Qué secciones del sitio visitas?, 3.¿Con qué frecuencia visitas estos sitios?, 4.¿Utilizas las herramientas de comentarios o wikis que contienen los sitios de noticias o blogs para participar en los comentarios de los contenidos publicados?, 5.¿Alguna vez te ha agredido algún usuario de internet cuando publicas algún comentario en alguna noticia utilizando la herramienta de comentarios o wikis, las cuales te permiten participar en los temas publicados en los sitios de noticias o blogs?, 6.¿Alguna vez has agredido a algún usuario de internet cuando publica algún comentario en alguna noticia utilizando las herramientas de cometarios o wikis que te permiten participar en los temas

publicados en los sitios de noticias o blogs?, 7.Describe lo que sientes y que haces cuando te agreden en internet, y finalmente 8.Escribe una lista de 20 palabras, frases o ideas agresivas que has leído en los comentarios realizados por los usuarios de los sitios de noticias o blogs

1.8.2. Procedimiento.

Primeramente se solicitó el apoyo a profesores de la Universidad Autónoma de Aguascalientes para realizar el pilotaje de la encuesta, aplicado a los grupos de Tecnologías de Información, con la finalidad de detectar posibles errores de redacción, los cuales fueron corregidos en las observaciones realizadas. Una vez corregidos los errores de redacción de la encuesta, se seleccionó a los alumnos de la Universidad la Concordia Campus Paraíso, para la aplicación de la misma. La aplicación se realizó en los salones de clases donde previo a la aplicación, se informó a los alumnos el motivo de la encuesta, señalándoles que sería anónimo y se les solicitó que contestaran con sinceridad. La aplicación duró en promedio 10 minutos, debido a que la encuesta no fue extensa. Una vez terminada, los alumnos entregaron las encuestas al profesor.

1.8.3. Captura y procesamiento de la información

La información fue capturada en una hoja de Excel de la paquetería Microsoft Office 2007, ordenando el contenido verticalmente con la finalidad de separar las preguntas y poder generar las respuestas de una forma clara.

Cabe mencionar que el contenido de la pregunta numero 8 se procesó por separado, debido a la preparación que se tuvo que realizar para poder generar un catalogo de agresión. Dicho catálogo está conformado por el inventario de palabras y frases catalogadas como agresivas por la población seleccionada a la que se le aplicó el cuestionario.

1.9. Hipótesis.

A continuación se plantea la siguiente hipótesis:

Hipótesis de investigación.

H_i: Los sitios web de noticias y blogs en Aguascalientes contienen ciber-violencia.

Hipótesis correlativa de causalidad.

H_{cc}: A mayor participación en los comentarios de los sitios de noticias y blogs, es mayor la tendencia a generar ciber- violencia.

Para efectos de la presente investigación se considerará que los sitios de noticias y blogs en Aguascalientes contienen ciber-violencia, siempre y cuando en los resultados se encuentre la presencia de 1 o más texto considerado agresivo, tomando como base un porcentaje de agresión aceptable cuando la evidencia sea menor o igual al 10% en relación al total de palabras que integran los comentarios. Este criterio de comparación fue aplicado debido a que no se encontró material bibliográfico que sustentara dicho dato.

Variables Independientes.

- Texto encontrado en sitios web de noticias y blogs en Aguascalientes.

Variable Dependiente.

- Texto agresivo.

CAPITULO 2.

2. Marco Teórico.

2.1. Historia de internet.

Internet empezó en 1957, cuando lo que antes era la Unión Soviética envió al espacio el primer satélite llamado Sputnik. Esa superioridad tecnológica despertó el temor de los norteamericanos. La RAND Corporation, fábrica de ideas numero uno durante la Guerra Fría, tuvo que vérselas ante un insólito problema estratégico: ¿cómo podría el gobierno de los Estados Unidos llevar a cabo con éxito las comunicaciones en caso de una guerra nuclear, sin un centro informático?, pues con los ordenadores de alta tecnología de esa época, no estaba lo bastante bien protegido y las líneas podrían ser destruidas en caso de un ataque nuclear.

Atendiendo a éste problema, los trabajadores de la RAND estuvieron trabajando en una serie de soluciones hasta encontrar la clave: no debería haber una autoridad central en las redes, es decir, no debería existir un centro informático que moviera todos los hilos. El nuevo tipo de red, según la RAND, debería ser operativo aun en un caso de destrucción, por lo que crearon un principio muy sencillo: cada ordenador de la red repartida por los Estados Unidos debería de tener el mismo estatus a la hora de enviar, recibir o reproducir un mensaje.

En la década de los 60's la red de RAND Corporation, basada en el envío de paquetes a prueba de bomba, fue presentada en diferentes instituciones estatales, que quedaron más que entusiasmadas por las posibilidades que ofrecía. Poco después, en Estados Unidos se fundó la Advance Research Project Agency, o Agencia para proyectos de investigación avanzada ARPA (por sus siglas en inglés), subordinada al Ministerio de Defensa norteamericano Department of Defense. En otoño de 1969 se instalo el primer pre ordenador en la Universidad de Los Ángeles, con la intención de formar una red siguiendo las especificaciones establecidas por RAND Corporation.

En diciembre de 1969 había un total de 4 nodos formando una red repartida por Estados Unidos y con un patrocinador publico que no estaba en el pentágono. Éste fue el

TESIS TESIS TESIS TESIS TESIS

nacimiento de ARPANET. Gracias a esta agencia, los científicos e investigadores podían utilizar a distancia las instalaciones informáticas de otros centros académicos y en consecuencia, se promovió un activo intercambio académico entre universidades.

Desde entonces se fue avanzando paso a paso. En 1971 ARPANET ya tenía 15 nodos. En 1972 ya eran 37 y así fue extendiéndose cada vez más. En el año de 1983 ARPANET fue creciendo, apuntando hacia la formación de una red internacional, a la que cada vez se añadían mas universidades y centros de estudios superiores. En ese año los patrocinadores fuera del pentágono excedieron sus atribuciones, por lo que se corrió el peligro de perder el control. Eso fue lo que generó la división de ARPANET en MILNET, destinada a la comunicación militar y la ARPANET, destinada a los centros de investigación.

Después de esto, ingenieros e investigadores desarrollaron redes a escala mundial que debían conectarse a ARPANET. También había empresas privadas las cuales estaban interesadas en acceder a ARPANET. Gracias a un estándar único para la transferencia de paquetes, cada vez más redes de todo el mundo se incorporaban a ARPANET. Con este acontecimiento había nacido internet.

2.2. Concepción de la violencia a través de la historia.

La violencia es un comportamiento humano que ha existido desde siempre y se ha practicado de diferentes formas, por ejemplo: la violencia para sobrevivir, la violencia para controlar el poder, la violencia física, entre otras.

Este comportamiento humano ha sido un factor que siempre ha estado en las distintas formas de la sociedad, desde la violencia territorial y la violencia de caza, dada por los primeros hombres que caminaron por la tierra, hasta los asesinos en serie que han recorrido nuestra historia.

Un ejemplo de la violencia nos lo proporcionan los etólogos, en investigaciones sobre el comportamiento innato de los animales, quienes han llegado a la conclusión de que el instinto agresivo tiene un carácter de supervivencia, ya que la agresión existente entre los animales no es negativa para la especie, si no que es un instinto necesario para su

existencia[1]. Lo mismo sucede con los seres humanos, ya que la violencia es el resultado que se genera a partir del instinto humano para sobrevivir ante su entorno.

El hombre, desde la primera vez que levantó una piedra y la lanzó contra su adversario, utilizó un arma de defensa y sobrevivencia, muchísimo antes de que el primer trozo de sílex hubiese sido convertido en lanza. Desde la más remota antigüedad, los hombres se han enfrentado entre sí por diversos motivos. En los últimos 5,000 años de la historia, la humanidad ha experimentado miles de guerras y en todas ellas se han usado armas más poderosas que la fuerza humana. La historia de la humanidad es una historia de guerras y conquistas, donde el más fuerte se impone al más débil y si de los textos de historia quitásemos las guerras, se convertirían en un puñado de páginas en blanco[1].

Cuenta una leyenda del siglo XX antes de Cristo, que existió un soldado romano llamado Cognomen, quien era tan frío de corazón que fue capaz de matar a toda su familia con su propia espada para pagar una apuesta. Esta historia relata que el soldado Cognomen estaba de campaña cerca de Italia y en medio de una noche de juegos entre los soldados, el coronel del pelotón, Marcos ex Antonio, mencionó que no había un soldado suficientemente valiente como para derramar la sangre de sus seres más queridos. Cognomen, quien no carecía de arrogancia, cuando volvió a su casa tomó su espada y apuñaló a su esposa, degolló a sus tres hijos y a su hija mayor. Para demostrar su honor, llevo las cabezas de su familia ante su coronel. Cuando éste vio las cabezas, sintió tanto repudio hacia Cognomen por la atrocidad de su acto que él mismo lo mató con su espada cuando aquél se acercó a mostrarle las cabezas de su familia degolladas.

2.3. Violencia.

La violencia es un término que se utiliza para describir situaciones diversas, razón por la cual se generan muchas confusiones y controversias. Existen diferentes maneras de clasificar la violencia, cada una sirve por lo general a un propósito determinado y la bondad de la clasificación está estrechamente relacionada con la utilidad de la misma.

El autor Alexander Butchart[2] según un informe mundial sobre violencia y salud, define la violencia como: El uso intencional de la fuerza física o el poder, amenaza o efectivo, contra

uno mismo, otra persona o contra un grupo o comunidad, que cause o tenga muchas probabilidades de causar lesiones, muerte, daños psicológicos, trastornos del desarrollo o privaciones.

2.4. Tipos de Violencia.

Existen diferentes tipos de violencia. A continuación se describe cada una de ellas:

2.4.1. Violencia social.

La violencia social se divide en:

- **Violencia física:** es el atentado contra la integridad física que se realiza de una persona físicamente superior, hacia otra más débil. Ésta se caracteriza por lastimar cualquier parte del cuerpo del atacado, utilizando en ocasiones el propio cuerpo del atacante, como por ejemplo usando las manos, los pies, la dentadura o mediante el empleo de objetos físicos.
- **Violencia emocional:** este tipo de violencia no se percibe tan fácilmente como la física, pero también lastima. Consiste principalmente en manifestar actitudes de rechazo, humillación, intimidación o vergüenza, con el objeto de hacer sentir mal a una persona, deteriorando su imagen y su propio valor, dañando de esta manera su estado de ánimo y afectando en consecuencia su capacidad para tomar decisiones. Para este tipo de violencia existen 2 variantes:
 - **La violencia verbal:** este tipo de violencia tiene lugar cuando una persona, con el uso de la palabra, ridiculiza, insulta, humilla e intimida a otra.
 - **La violencia no verbal:** se manifiesta en actitudes corporales de agresión, miradas de desprecio, muestras de rechazo, indiferencia, silencio y gestos insultantes para descalificar a una persona.

También existe una variante de la violencia que comparte características de la violencia emocional y de la violencia física, la cual es:

- **La violencia sexual:** este tipo de violencia ocurre cuando se le obliga a una persona a tener cualquier contacto sexual contra su voluntad y daña física y emocionalmente a la persona.

A su vez este tipo de violencia se presenta de diversas maneras:

- **Acoso:** Es la persecución, mediante agresiones o molestias, de una persona que generalmente tiene un nivel jerárquico superior, hacia otra de nivel inferior, que puede llegar a involucrar actividad sexual. Generalmente se da en ambientes de trabajo.
- **Abuso:** Consiste en tocar o acariciar el cuerpo de otra persona en contra de su voluntad.
- **Violación:** Es el acto de extrema violencia contra una persona que incluye la penetración.
- **Incesto:** Es el contacto sexual entre familiares o que tienen un tipo de parentesco.

2.4.2. Violencia económica.

Es toda acción u omisión del agresor que afecta directamente la supervivencia económica de la víctima. Este tipo de violencia se manifiesta a través de limitaciones enfocadas a controlar el ingreso de sus percepciones económicas, así como la percepción de un salario menor por igual trabajo, dentro de un mismo centro laboral. En pocas palabras, el agresor hace lo posible por controlar el acceso de la víctima al dinero.

2.4.3. Violencia de género.

El género se refiere al sistema de características asignadas a los hombres y las mujeres en una sociedad[3]. Se entiende por violencia de género aquella que sufren las mujeres por razones sexistas o basadas en su género[4]. En este tipo de violencia las mujeres suelen ser las víctimas principales tanto en el ámbito público como en el privado, debido a la situación de desigualdad y discriminación en la que viven.

2.4.4. Otras definiciones de violencia.

El autor Jamil Salmi[5] clasifica la violencia en cuatro tipos:

- **La violencia directa,** se refiere a los actos físicos que resultan del daño deliberado a la integridad humana. Esta categoría incluye todos los tipos de homicidio (genocidios, crímenes de guerra, masacres de civiles, asesinatos), así como todo tipo de acciones brutales que implican un sufrimiento físico o psicológico (encarcelamiento, secuestro, toma de rehenes, trabajo forzado, torturas, violaciones, malos tratos).

- **La violencia indirecta**, es una categoría destinada a cubrir situaciones perjudiciales, a veces mortales, que aunque debido a la intervención humana, no implican una intervención directa entre las víctimas y las instituciones, grupos de poblaciones o personas responsables de esa situación. Existen dos sub categorías para este tipo de violencia: la violencia por omisión y la violencia mediada.
- **La violencia represiva**, se refiere a las formas más comunes de violencia de derechos humanos, documentados y monitoreados por las ONG's internacionales como Amnistía Internacional¹⁰ o Human Rigths Watch¹¹. Las violaciones de los derechos civiles se producen cuando a las personas se les niegan la libertad de pensamiento, de religión, y de movimiento, o cuando no hay igualdad ante la ley, incluido el derecho a un juicio justo.
- **La violencia alienante**, se refiere a la privación mayor de derechos de una persona, tales como los derechos psicológicos, emocionales, culturales o intelectuales. Se basa en las suposiciones de que el bienestar de una persona no solo se satisface de necesidades materiales.

2.5. Causas de la violencia.

Para Miquel Domènech y Lupicinio Íñiguez Rueda, en su artículo La construcción social de la violencia[6], la familia es uno de los contextos donde la violencia se produce de una forma más espectacular, alarmante, inesperada y cruel. Todas las combinaciones posibles, y de ello dan testimonio los medios de comunicación. Padres y madres contra hijos, integrantes de la pareja, unos contra otros, hermanos contra hermanas e hijos contra padres.

Existen otros factores que, según los socio-psicólogos, pueden contribuir a la probabilidad de la violencia individual o colectiva:

¹⁰ La Amnistía Internacional es una organización no gubernamental (ONG) internacional de carácter humanitario que trabaja para promover los derechos humanos en el marco de la Declaración Universal de los Derechos Humanos aprobada en 1948 y otros tratados internacionales como los Pactos Internacionales de Derechos Humanos que constituyen la denominada carta Internacional de Derechos Humanos.

¹¹ Human Rigths Watch es una de las organizaciones no gubernamentales (ONG) líderes en el mundo dedicada a la investigación, defensa y promoción de los derechos humanos.

- **Edad y sexo.** Aunque hay excepciones, en general la violencia es más frecuente entre varones. Puede haber factores hormonales que determinen una mayor agresión en el sexo masculino, pero las presiones culturales y sub-culturales desempeñan evidentemente un papel importante.
- **Clase social.** Se ha mencionado este factor en relación con los tumultos en los ghettos¹², pero en términos generales puede decirse que cuanto más baja es la clase socioeconómica, mayor es la frecuencia de la violencia. Se ha apuntado, sin embargo, que en los grupos que padecen mayor privación, la apatía y la impotencia pueden reducir la violencia.
- **Raza y grupo étnico.** Es difícil poner en duda que los negros en los Estados Unidos tienen tasas de homicidio superiores a las de los blancos, aun admitiendo que, en igualdad de condiciones, los negros tienen más probabilidades de ser detenidos y condenados, aunque esta situación está mejorando. No hay pruebas, sin embargo, de que ello se deba a factores genéticos.

2.6. Mecanismos de defensa psicológico.

Los mecanismos de defensa (proyección, racionalización, sublimación, represión, etc.) son funciones psíquicas reguladoras que se asemejan a la manera de trabajar. Su misión dentro del psiquismo, es la de regular la carga de energía (disminuyendo la tensión psíquica) para proteger el equilibrio y evitar toda clase de trastornos o perturbaciones producidas por exceso de excitación emocional[7].

Los mecanismos de defensa según Augusto Vels.[7] son:

- **La represión.**

Bajo el punto de vista psicoanalítico, la represión es el aprisionamiento en el subconsciente de recuerdos, ideas, emociones, etc., cuya exteriorización, a través de la

¹² Según la Real Academia de la Lengua Española, la palabra ghetto o gueto, significa a) Judería marginada dentro de una ciudad; b) Barrio o suburbio donde viven personas marginadas por el resto de la sociedad; c) Situación o condición marginar en que vive un pueblo, una clase social o un grupo de personas. Información consultada en el sitio web http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=ghetto el día 4 de abril del dos mil doce, a las 20:39 horas.

conciencia, está impedida por las barreras psíquicas de la censura. Por ejemplo, la atracción sexual, o el odio hacia el padre o la madre, o hacia un hermano o hermana. Otro ejemplo sería el niño que hace algo reprobable y sus padres le castigan por ello. Esto genera en el niño cierta ansiedad asociada al temor a perder el cariño de sus padres. Para evitarlo, el niño reprime el impulso a hacer de nuevo aquello que fue motivo de castigo, con lo cual evita también la ansiedad correspondiente.

Cuando en el subconsciente se almacenan una serie de emociones rechazadas (fuerzas, impulsos, deseos, etc.) que acumulan cargas tensionales, más o menos irritativas, dentro del proceso químico celular correspondiente, es probable que esto desencadene una neurosis. Este estado tensional anormal puede liberarse de varias maneras, como por ejemplo a través del sueño y con disfraces simbólicos o a través de manifestaciones de duda, temor, ansiedad y aprensión poco normales. En algunos casos, los estados tensionales excesivos producidos por la represión se manifiestan también a través de actos o de hábitos simbólicos o simulando algún padecimiento físico como expresión de lo reprimido.

- **La regresión.**

La regresión es lo que en el A.T. (Análisis Transaccional) sería un retorno al yo infantil a consecuencia de un enturbiamiento del yo adulto. Ese enturbiamiento podemos considerarlo como una especie de fracaso del yo frente al objeto. El sujeto, o bien no ha recogido informaciones para actuar eficazmente sobre su realidad circundante, o no es lo suficientemente activo y emprendedor para informarse y asegurar la acción adecuada para dominar los hechos, circunstancias o acontecimientos.

En cualquiera de los casos (impotencia, fracaso, fatiga, pereza, deterioro mental, etc.), el sujeto opta, como solución a su problema, por volver a un estadio anterior de desarrollo de su vida afectiva y mental en donde se sintió más cómodo, seguro y protegido. De este modo, se convierte en un niño interior reactivo, ya sea adoptando la fase de docilidad (dependencia-sumisión), o bien, adoptando un rol de rebeldía, negativismo, oposición o tozudez. Toda regresión supone una vuelta atrás del proceso de maduración psicológico, tanto afectivo como mental. Por tanto, es un claro síntoma de neurosis, de desadaptación. El ello domina sobre el yo, lo que tiende a crear un disturbo de la personalidad, pues así

como la progresión es la tendencia a mejorar, a superarse para alcanzar una mayor adaptabilidad a los cambios y evoluciones de la vida, el sujeto con una libido en regresión huye de situaciones insostenibles refugiándose en pensamientos, sentimientos y modos de conducta primitivos, aunque cree que con su modo de pensar, sentir y actuar no elude las dificultades y problemas que tiene delante.

- **La sublimación.**

Dentro de los mecanismos de defensa más positivos para la descarga de las tensiones podemos considerar la sublimación. La sublimación viene a ser como un medio de alcanzar satisfacción, en forma sustitutiva o imaginaria, a las dos tendencias básicas de nuestros instintos: la sexualidad y la agresividad. Se trata de un tipo de comportamiento o conducta en el que tendencias, impulsos instintivos, deseos, etc. que son moral y culturalmente rechazables por la conciencia y por la convivencia social se descargan canalizando su energía en torno a comportamientos socialmente aceptables. Según Freud, todas las actividades científicas, artísticas, intelectuales, religiosas y culturales, en general, son consecuencia de la sublimación. Así por ejemplo, ciertas tendencias sádicas se descargan en actividades como las de cirujano, dentista, escultor, grabador, matarife, luchador, cazador, etc., profesiones donde la agresividad sádica está justificada por su beneficio a la colectividad.

- **La proyección.**

La proyección es un mecanismo de defensa mediante el cual escupimos sobre el mundo o sobre los demás aquellas emociones, vivencias o rasgos de carácter, que deseamos desalojar de nosotros mismos, por inaceptables. Todo lo que resulta placentero es aceptado como algo perteneciente al propio yo. Todo cuanto resulta desagradable, molesto o doloroso es sentido como ajeno al yo.

La frustración desaparece cuando creemos que nuestros defectos, nuestras deficiencias y nuestros fracasos también los tienen los demás, o son los demás los causantes de esos fracasos, de esas deficiencias o de algo que no deseamos tener. Por ejemplo, las personas que piensan mal creen que los demás también piensan del mismo modo. El pequeño

comerciante que roba 50 gramos en cada kilo, cuando compra en otra tienda será el cliente más desconfiado.

La proyección, bien canalizada, puede generar creaciones artísticas o científicas más o menos originales, pero en el caso del científico la proyección debe estar controlada por una inteligencia analítico-sintética que revise y sistematice las propias ideas con un serio cotejo de la realidad.

- **La introyección.**

Este es un mecanismo de defensa que consiste en introyectar, mediante la absorción, identificación o imitación, ciertas cualidades que tienen los objetos externos. El sujeto actúa imitando al padre o a la madre, al profesor, al actor de moda o a aquellas personas con las que se identifica. Cuando el Rey Sol decía Francia soy yo, o cuando utilizamos frases como mi empresa, mi partido, mi equipo, etc., estamos indicando un proceso de identificación introyectiva en el que asociamos cosas externas a nuestro yo, como si fuesen parte de nosotros mismos. Cuando la introyección es exagerada, se produce lo que se ha venido a llamar el canibalismo psíquico. Así, el celoso desarrolla tanto el sentimiento de posesión del ser amado como parte de sí mismo, que no puede tolerar ni una simple sonrisa de su pareja hacia alguien del otro sexo, o el sujeto que no puede tolerar que otro compañero se siente en su mesa o utilice su teléfono o sus utensilios de trabajo.

- **La conversión.**

La conversión o somatización es un fenómeno que se da principalmente entre los histéricos. Consiste en convertir en trastorno o enfermedad física las frustraciones o contrariedades sufridas. Mediante este mecanismo de defensa, ciertos histéricos, transformando en enfermedad las contrariedades sufridas, eligen este medio para dominar, castigar o retener a las personas de su círculo íntimo. Por ejemplo, habiéndose negado el esposo a llevar a su mujer a un viaje de negocios que tiene que hacer a Italia, ésta cae enferma de cualquier cosa de apariencia grave e impide el viaje. ¿Cómo podría el marido abandonar a su esposa en semejante estado? Coincide además, que la conversión o somatización se produce siempre que es contrariado un deseo, sea o no razonable.

• **La compensación.**

Cuando fracasamos en algo o nos sentimos menos dotados de lo normal en algún aspecto, en muchos casos los mecanismos de defensa estimulan a triunfar en la misma dirección o en otra esfera sustitutiva. Por ejemplo, Demóstenes, tartamudo de nacimiento, logra, mediante ejercicios en la playa y en solitario, convertirse en uno de los oradores más famosos de la antigüedad. Beethoven, sordo como una tapia desde los 28 años, logra sus mejores composiciones a partir de su sordera. Pio Baroja no logra su triunfo como médico, pero en cambio logra ser uno de los mejores escritores de la generación del 98.

• **La racionalización.**

Es una forma de negación en la que, para evitar el conflicto o la frustración, se dan razones o se expresan argumentos que ocultan, justifican, o encubren los fallos o contrariedades. Mediante este mecanismo, el sujeto se defiende del efecto frustrante y trata de convencerse que, en el fondo, no deseaba aquello que no ha conseguido. Un ejemplo es el de la zorra que después de varios saltos no logra alcanzar las uvas de la parra y exclama: ¡Vah, están verdes!. Otro ejemplo es el del joven que se siente frustrado por no conseguir el amor de una muchacha de la que está enamorado y dice para sí mismo: Es estúpida, no sabe lo que se pierde, ¡con la gran cantidad de chicas que hay, incluso mejores que ella! O el del empleado que aspira a un cargo que la empresa ha dado a otro compañero de su departamento. Cuando el empleado ascendido le dice: siento que me hayan elegido a mí pues, por antigüedad, te tocaba a ti ascender, el aludido contesta: No te preocupes, mis aspiraciones son más altas.

• **La fijación.**

Es una intensa adhesión a algo o a alguien, como la fijación de la madre en el complejo de Edipo, o al padre en el de Electra. La fijación se refiere, mayormente, a adherencias desarrolladas en la infancia que persisten de una manera inmadura o neurótica en el adulto. Su consecuencia es la ineptitud para desarrollar otras adhesiones normales desplazando la libido hacia otras personas u objetos. En estos casos, el sujeto conserva las mismas ideas y la misma manera de hacer las cosas. La fijación conduce a la rigidez

mental. El apego exagerado a personas u objetos puede volverse ambivalente, es decir, el sujeto odia y ama a la vez el objeto de su fijación. La fijación a la madre, por ejemplo, puede llevar a un apego al estado oral en aquellos sujetos que la madre les prolongó el pecho demasiado, o se mostró excesivamente protectora en la infancia, lo que entraña más tarde en el adulto un comportamiento infantil tiránico y absorbente. En algunos casos, la fijación puede provocar una transformación de la libido en perversiones.

Según Freud, las fijaciones en las etapas oral, anal y genital primarias pueden provocar neurosis, psicosis y criminalidad. La fijación no sólo es un estado de estancamiento o regresión a cualquiera de las fases de evolución infantil (oral, anal, genital, etc.), sino que puede referirse también a ciertos contenidos representativos (experiencias, imágenes, fantasías, etc.) que persisten de manera fija e inalterada en el inconsciente y a las cuales se ligan cargas energéticas (pulsiones, impulsos) que tienden a alcanzar un fin.

- **La formación reactiva.**

La formación reactiva es una actitud o hábito de reacción opuesto al deseo reprimido. El sujeto lucha directamente contra toda representación penosa, frustrante o dolorosa, sustituyéndola por un síntoma primario de defensa o contra síntoma, consistente en adoptar una conducta o reacción que excluye de la conciencia a los elementos que intervienen en el conflicto (la representación sexual y el reproche que ésta suscita, por ejemplo), en favor de virtudes morales llevadas al extremo, como pueden ser el pudor, la escrupulosidad, la persecución de las conductas inmorales, etc.

2.7. La violencia en la actualidad.

En la actualidad en todos los ámbitos somos testigos de la violencia y se ha convertido en parte de la vida diaria de las personas. Cada día vemos, por ejemplo, violencia en las calles, violencia en el hogar, violencia en la escuela, violencia en el trabajo, violencia contra la mujer, violencia contra niños y ancianos, etc. La violencia es una mezcla irracional de emociones que se expresan a través de enojo, celos, ira, entre otras cosas.

Expertos en el área de estudio, sugieren que la violencia debe entenderse como la interacción entre factores que tienen que ver con el desarrollo psicosocial de los individuos,

TESIS TESIS TESIS TESIS TESIS

sus referencias neurológicas y hormonales, y los procesos sociales que se dan a su alrededor[8]. Para la sociología, la violencia se ubica en el marco de las relaciones macro sociales, donde violencia y poder parecen ser conceptos inseparables. La sociología identifica dos formas de materialización de la violencia:

- a) **La manifiesta**, que afecta con la integridad física de los individuos o grupos, de manera que sus manifestaciones son cuantificables. En ella se inscriben actos como el homicidio, los golpes, la violación, etc.
- b) **La estructural**, cuya casualidad se encuentra en las condiciones estructurales de la sociedad y cuya consecuencia no se pueden atribuir a sujetos específicos.

La violencia constituye una de las tres fuentes principales de poder humano, las otras dos son el conocimiento y el dinero[9]. Estas tres fuerzas afectan nuestra vida desde que nacemos hasta que morimos.

En la última década la violencia se ha manifestado de diferente manera dependiendo del contexto en el que se ejerza. El internet es una herramienta de comunicación que ha sido explotada por niños, jóvenes y adultos. Una herramienta para que investiguen, busquen información y se comuniquen con familiares y amigos. Sin embargo los riesgos de violencia son grandes, ya que al navegar libremente se está expuesto a una infinidad de material e información, que en muchas ocasiones puede conllevar a la denigración del ser humano. Pensemos, por ejemplo, en el material pornográfico que abunda en internet, o la facilidad que existe en internet para propiciar la corrupción de menores. Los factores que mayor influencia tienen en estos casos, son la facilidad para acceder a páginas con contenidos inapropiados y el anonimato.

2.8. Las redes sociales.

Los sitios de redes sociales son estructuras compuestas de grupos de personas, las cuales están conectadas por una o varios tipos de relaciones, tales como amistad, parentesco, intereses comunes, o alguna otra.

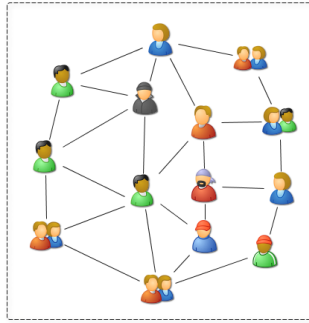


FIGURA 1. REPRESENTACIÓN DE LAS REDES SOCIALES

Desde 1992 la web ha proporcionado muchas formas de interacción entre las personas[10], las redes sociales son una de ellas.

Estos sitios son, sin duda, una de las aplicaciones más populares que han afectado al mundo del Internet recientemente[11], siendo la base de estos, la web 2.0, la cual permite crear aplicaciones web enriquecidas que se caracterizan básicamente por una variedad de controles de operación interactiva e interface de usuario. Las aplicaciones web enriquecidas (Rich Internet application en ingles) es la intersección entre la funcionalidad de las aplicaciones web, con la apariencia de las aplicaciones de escritorio, y a esto agregándole las tecnologías de comunicaciones.

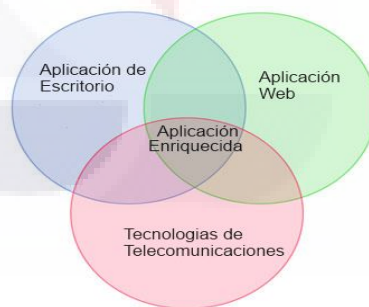


FIGURA 2. CONCEPTO DE LAS APLICACIONES WEB ENRIQUECIDAS

En las redes sociales, gente de todas las edades comparte información con amigos, amigos de amigos y algunas veces con personas que no tienen un vínculo directo. Por ejemplo

Facebook¹³ que es la red social más popular con más de 665 millones de usuarios hasta abril del 2011¹⁴, Twitter¹⁵, Friendster¹⁶, LinkedIn¹⁷, MySpace¹⁸, por nombrar algunas, permiten a los usuarios contactar a cualquier persona que tenga un perfil en línea. Estos sitios han ganado popularidad por la oportunidad que le dan a los usuarios de poder contactar a otros, además de permitir intercambiar y compartir información de manera sencilla[12].

En los sitios de redes sociales destaca el grado de centralidad[13] que dispone cada una de las personas. Esto se refiere al número de contactos directos que se mantienen con las otras personas de la red social, de este modo los usuarios que se mantengan con un mayor número de contactos se postulan como más centrales, por lo que un elevado grado de centralidad les proporciona privilegios.

Las redes sociales son estructuras que mapean las relaciones semánticas de sus miembros. En estas redes, la conexión por lo general sigue el modelo no estructurado, es decir, los miembros están conectados a otros con características similares[14]. La revolución de internet está cambiando la manera en que usamos los sitios de redes sociales. Cada uno de los usuarios tiene un enfoque como usuario final a cada uno de estos sitios, por ejemplo, los estudiantes utilizan las redes sociales para intercambiar fotos, mensajes, correo electrónico, planificación de eventos, escribir en el muro de otros, entre otras. Las empresas las están usando para atraer más clientes, los sitios de noticias y blogs las están integrando a sus contenidos para permitirles a los usuarios una mayor interacción, etc.

Como ya se mencionó, los sitios de redes sociales están atrayendo a empresas para hacer negocios en línea, las cuales buscan acercarse a sus consumidores, accediendo a un mundo de información de los usuarios y generando nuevas oportunidades en el mercado[15].

¹³ <http://www.facebook.com/>

¹⁴ Información recabada en el sitio: https://www.facebook.com/note.php?note_id=10150164303572961 consultada el 21 de abril de 2012, a las 13:28 horas.

¹⁵ <http://www.twitter.com/>

¹⁶ <http://www.friendster.com/>

¹⁷ <http://www.linkedin.com/>

¹⁸ <http://www.myspace.com/>

2.9. Las aplicaciones sociales.

La historia de las aplicaciones sociales tiene su comienzo en el año 2003. En ese momento nacen las primeras redes sociales en la historia de internet: Tribe.net, LinkedIn y Friendster. El objetivo principal de estas redes sociales era trasladar a la red una teoría sociológica conocida como los seis grados de separación en virtud de contactar a cualquier persona en el mundo, con máximo, a través de otras cinco personas.

A medida que estas redes se iban conociendo mundialmente, se empezaron a crear otras como eConosco en diciembre del 2003 y Orkut a principios del 2004, la cual estaba orientada a administrar contactos de tipo personal y que además solo admitía miembros que hayan recibido una invitación de otros. Ambos sitios son de origen hispano. También se crearon en el 2003 otras redes sociales, tal es el caso de Myspace, la cual estaba orientada a la música y, Facebook creada en el 2004, la cual estaba enfocada a ser punto de encuentro de estudiantes universitarios.

Las aplicaciones sociales se clasifican de acuerdo a su función. A continuación se listan las principales.

- **Blogs:** este tipo de aplicación permite que un único usuario lo alimente de información. Esta era la idea original con la que fueron creados, sin embargo, actualmente muchos sitios de este tipo permiten que otros usuarios publiquen sus propios informes o comenten los de los demás.
- **Foros:** es un tipo de aplicación de las más antiguas, ya que ofrecen muchas características a los usuarios y los agrupa de acuerdo a sus gustos.
- **Wikis:** son tipos de herramientas o aplicaciones que permiten crear y hacer páginas web fáciles, sin necesidad de conocimientos informáticos.
- **Sistemas de etiquetado:** este tipo de aplicaciones permiten la búsqueda de afinidades en la red utilizando etiquetas o tags.
- **Redes sociales:** son sistemas que agrupan usuarios bajo diferentes criterios y permiten que los usuarios se conozcan y establezcan un contacto frecuente.

Según los autores Juan José Flores Cueto, Jorge Joseph Morán Corzo y Juan José Rodríguez Vila [24] una aplicación online podrá considerarse como parte de la web 2.0, es decir, como aplicación social cuando permita:

- Procesos de interactividad de contenidos contributiva; es decir, cuando el usuario pueda añadir y compartir información con otros usuarios como en Flickr¹⁹.
- Procesos de interacción de contenidos combinatoria; es decir, cuando posibilite la interrelación de contenidos de diferentes bases de datos como en los denominados mashups²⁰. Para ello, es importante indicar que se utilizan API (Application Program Interface), las cuales son una serie de instrucciones conocidas como rutinas, que permiten intercomunicar o pasar los datos desde un entorno o programa a otro.
- Procesos de interacción de interface, que pueden estar relacionados con las preferencias estéticas con funciones o con los procesos automáticos de generación de interface (cuando el sistema, a partir del análisis del modo de operar del usuario con la interface, decide por el usuario cómo o qué datos presentar).

2.10. Ciber-violencia en las redes sociales.

La participación en los sitios de redes sociales ha aumentado dramáticamente en la actualidad, haciendo a un lado algunas reglas como la validación de la semántica, permitiendo a sus usuarios independientemente de su género o edad, enviar o postear, es decir, publicar mensajes a sus conocidos sin que previamente se valide el contenido; este es el caso de Facebook o Twitter, donde el usuario tiene hasta 140 caracteres para expresar sus ideas. Estos mensajes de texto pueden ser ofensivos, denigrantes, amenazadores o ridiculizantes, pero también pueden ser fotografías y videos ofensivos.[16].

En septiembre del 2007, la red social más popular Facebook, introdujo el News feed [17] el cual es una herramienta que realiza un seguimiento a las actividades actuales de los amigos o contactos de un usuario, es decir, este servicio permite que un usuario pueda conocer

¹⁹ www.flickr.com

²⁰ Un mashup es un término relativamente nuevo y se refiere a una aplicación web que permite la integración de datos a partir de otras fuentes. Es decir, se tiene la posibilidad, por ejemplo, de utilizar datos de diferentes sitios web que resulten importantes para una determinada persona o grupo.

cuándo uno de sus contactos actualiza su estatus, publica fotografías, adquiere nuevos amigos a través de la red social, publica mensajes en su muro o en el de los demás, etc.

Esta característica hace vulnerable la entrada de información que se divulga en la red social por parte de los usuarios, siendo completamente pública e indignando a los usuarios por sentirse expuestos y privados respecto al control de su información.

Las redes sociales carecen de mecanismos de seguridad para la validación de información, lo cual hace un problema en la participación de los usuarios para la generación de contenidos electrónicos[18].

Los sitios de redes sociales permiten que cualquier persona pueda hacerse usuario y pueda crear un perfil en línea sin ninguna restricción, únicamente debe contar con una cuenta de correo electrónico válida, aprobando así al usuario para el acceso a la red social, por lo que es muy sencillo postear información sin validar su significado.

El aumento en la participación de usuarios de los sitios de redes sociales, se ha visto acompañado de una progresiva diversificación y sofisticación, tanto de propósitos, como de patrones de uso, a través de una multitud de sitios diferentes[19], produciendo en consecuencia un problema en la validación semántica.

Los sitios de redes sociales se han visto envueltos en una serie de contrariedades, siendo utilizados no solamente para crear vínculos sociales entre personas, sino que también como herramienta para crear intimidación y violencia cibernética, también llamada ciber-violencia o bullying cibernético. Este problema está ocasionando que los usuarios entren en un juego de agresión convirtiendo dicha conducta, de cierta manera, en un algo que podemos catalogar como normal en los sitios de internet.

En muchos países no hay una legislación que regule la ciber-violencia. Tal es el caso de México. Esto representa un problema, porque puede relacionarse con conductas delictivas como el acoso sexual y la pornografía infantil.

Sobre la seguridad en la validación semántica en los sitios de redes sociales en México, aun no se realiza esta práctica, y sólo se basan en las quejas y denuncias por parte de los

usuarios, quienes tienen que justificar mediante un patrón de quejas, cual es el problema encontrado en el contenido publicado en la red social. La magnitud de este tipo de problemas se desconoce, ya que la policía cibernética en México rechaza dar información sobre este tipo de denuncias.

En México, el Instituto Nacional de Estadística y Geografía (INEGI) reporta que hasta Mayo del 2010 el 34.6% de la población mexicana son usuarios de computadora, y de ellos, el 84.4% son usuarios de internet[20].

Estudios aun no publicados de los autores Gerardo García, Víctor Manuel Joffre, Gerardo Jesús Martínez y Arturo Llanes[20], realizados al sur del estado de Tamaulipas a finales del 2010, en los que participaron 603 alumnos de secundaria, se encontró que durante los 6 meses previos al estudio, el 3.5% de los encuestados había sido ciber intimidado, el 2.8% había ciber agredido y el 1.3% había mostrado un rol mixto; en cuanto al género, el 63% de los involucrados fueron varones y el 37% fueron mujeres.

2.11. Las redes sociales y la edición colaborativa.

Cuando hablamos de edición colaborativa, no pasamos por alto la transición de la web 1.0 a la web 2.0, ya que la web 1.0 se componía de HTML²¹ estático el cual únicamente permitía desplegar texto plano entre etiquetas, las cuales pre formateaban el texto dándole un estilo amigable para el lector. Después vendría la web 1.5 también llamada la web del punto-com²² la cual permitía generar un texto más dinámico con un toque un poco más interactivo, donde la estética y los hits²³ eran lo más importante.

Poco tiempo después aparecería la web 2.0 también llamada la web para las personas, la cual permitiría tener más interacción con la información y los contenidos y, de aquí, el nacimiento de las redes sociales. Con la web 2.0 vendría el poder de la inteligencia colectiva, el cual permitiría la aportación de contenidos por parte de los usuarios, lo que convertiría a la web en un gran repositorio de información distribuida[21].

²¹ Hyper Text Mark of Language, por sus siglas en inglés. Es el lenguaje predominante para la elaboración de páginas web.

²² Forma parte del sistema de nombres de dominio que utilizan las empresas para sus portales en internet.

²³ También llamada page hit, es la forma de acceder a otros sitios, como páginas o gráficos de un servidor web.



FIGURA 3. REDES SOCIALES COLABORATIVAS.

2.12. La edición online.

El término Wiki es una aplicación de informática que reside en un servidor web a la que se accede con un navegador²⁴ y se caracteriza por que permite a los usuarios añadir contenidos y editar los existentes[21][22][23]. Este término, comienza en 1995 cuando, el padre de la idea, Ward Cunningham buscaba el nombre para una aplicación que había diseñado, la cual permitía que la comunidad dispersa de usuarios editara páginas web rápidamente, usando su navegador y formularios web.

Las características de los wikis son las siguientes:

- Cualquiera puede cambiar cualquier cosa sin necesidad de saber HTML.
- Los wikis son flexibles, ya que es sencillo al no tener estructuras predefinidas, lo que permite crear nuevas páginas o vincular contenidos con páginas ya existentes.

2.13. Periodismo en México.

Según el Diccionario de la Real Academia de la Lengua Española, el periodismo es la captación y tratamiento, escrito, oral, visual o gráfico, de la información en cualquiera de

²⁴ También llamado navegador web, es un programa que permite ver la información que contienen las páginas web.

TESIS TESIS TESIS TESIS TESIS

sus formas²⁵. El periodismo es una actividad encaminada a la recolección, sintetización, jerarquización y publicación de información relativa a los hechos de vida actual. Las bases del periodismo en nuestro país se remontan a la época colonial, en el siglo XVI, cuando circularon por las calles de la capital de la Nueva España los famosos pregoneros, que a pulmón abierto gritaban en plazas públicas o sitios de gran concurrencia como los mercados, las noticias de actualidad. Obviamente que esta labor era realizada bajo la supervisión de las autoridades coloniales, que a través del cabildo otorgaban permiso expreso a estos ingeniosos hombres.

La llegada de la imprenta a México ocurre en 1539, cuando a instancias del arzobispo Fray Juan de Zumárraga llega a radicar en la capital de la Nueva España el impresor italiano Juan Pablos. De manera paulatina se fueron instalando más talleres de impresión, con lo que inicia la circulación de hojas volantes. La primera de ellas que fue impresa en nuestro país es el Mercurio Volante, editado en 1693 por el intelectual Carlos de Sigüenza y Góngora, con noticias de carácter histórico y científico.

Una de las hojas volantes más antiguas que se conocen y que ha servido como fuente de datos para varios historiadores es la relación del terremoto de Guatemala, evento que ocurrió en noviembre de 1541, aunque el volante circuló en México hasta 1542.

En enero de 1722, el clérigo católico y doctor Juan Ignacio de Castorena y Ursúa fundó la Gaceta de México y noticias de Nueva España, considerado como el primer periódico de México, donde publicó informaciones religiosas, oficiales, comerciales, sociales, mineras y marítimas. Este órgano mensual duró de enero a junio de ese año.

Desde entonces, los periódicos y las hojas volantes en México y en el resto de Hispanoamérica tuvieron una utilización práctica. Jorge Calvimontes, en su libro: El Periódico[25] establece que en una gaceta de la época colonial (justo del año de 1760) era posible encontrar una columna de acontecimientos o la historia de la época, un ensayo de origen local o importado sobre cualquier tema, desde astronomía, hasta el cultivo de las zanahorias.

²⁵ http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=periodismo, consultado el quince de abril del dos mil doce, a las 21:20 horas.

2.14. Periodismo online.

A principios de 1995 existían alrededor de 100 periódicos online, pero la popularidad de este sistema de información se incrementó a grado tal que, a inicios de 1996 la cifra ascendía a 750. En mayo del mismo año, la cifra de periódicos online era de 1,115 y en diciembre de ese año fue de 1,587. Esto representó un crecimiento importante en el periodismo, aunque cabe mencionar que los periódicos norteamericanos representaron más del 60% de los periódicos online; otros países de habla anglosajona obtuvieron el 11.7%, mientras que los países de la unión europea llegaron al 17.4% y finalmente el resto de mundo con únicamente el 11%, de entre los cuales se incluye nuestro país[26].

El periodismo en línea en México forma parte de las nuevas tecnologías de la información y comunicación, y representa una herramienta útil para informarse y comunicarse. Con la aparición de los diarios online en México comenzó a hablarse del concepto de periodismo digital, como una nueva forma de hacer periodismo.

2.15. Los blogs como herramienta periodística.

El Blog, también llamado Weblog[21] en un término moderno, es una página web sencilla, consistente en párrafos de información y opiniones, ordenada cronológicamente desde el más reciente, al primero que fue publicado.

Todo blog se compone de una estructura, tal como se describe a continuación:

- **El post de entrada**, es la parte principal en la que el autor publica o escribe según la finalidad del blog (diario personal, opiniones, noticias, etc.)
- **Los comentarios**, de los visitantes del blog. La mayoría de los blogs permiten a sus visitantes realizar publicaciones para cada entrada del post²⁶. La interacción entrada (post) – comentarios (comments) representa todo un sistema de comunicación con los lectores que dejas sus opiniones y enriquecen al post.
- **Su temporalidad o periodo de publicación**, mes, día, año y hora para cada post publicado.

²⁶ Es una entrada de texto permitida por un sitio web.

- **Clasificación de las entradas**, temporal y temática, dependiendo de la vigencia de la noticia.

Los orígenes del blog han causado cierto debate. Los autores Jeremy B. Williams y Joanne Jacobs [27], explican que la frase web log fue usada por primera vez por Barger en 1997 y la versión abreviada por Merholz en 1999. Los autores comentan que el fenómeno de los blogs empezó a aumentar de manera constante, donde después hubo una explosión en el número de blogs cuando salió la versión gratuita para el público. Las herramientas para los blogs están disponibles desde mediados de 1999, y el más notable es el sitio Blogger²⁷, el cual es uno de los más conocidos y utilizados por los blogueros, ya que son muy fáciles de configurar y utilizar.

Cabe mencionar que una parte importante en los blogs son los enlaces, ya que éstos permiten la referencia hacia otras noticias o publicaciones relacionadas sobre el mismo tema. Existen varios tipos de enlaces los cuales pueden existir en un blog, por ejemplo:

- **Enlaces permanentes:** son URL²⁸ fijas, los cuales son generados por el sistema y se aplica para cada blog.
- **Trackback:** permite a un blogger notificar a otro blogger que ha sido comentado alguno de los post o entradas.
- **Blogroll:** es una lista de enlaces de otros blogs, que un blogger encuentra interesante.
- **Sindicación:** es el software del blog que permite la sindicación de titulares o entradas vía RSS²⁹.

2.16. La Minería web.

Existen varias definiciones atinadas sobre la minería web. En la definición más completa, según los autores Bamshad Mobasher, Namit Jain, Eui-Hong, y Jaideep

²⁷ <http://www.blogger.com>

²⁸ Acrónimo de Uniform Resource Locator, por sus siglas en inglés, el cual hace referencia a un localizador uniforme de recursos, que es una secuencia de caracteres de acuerdo a un formato melódico y estándar, usado para nombrar recursos en internet.

²⁹ Really Simple Syndication, por sus siglas en inglés. Es un formato XML para syndicar o compartir contenidos en la web.

Srivastava, la minería web se define como: el análisis de información útil en la world wide web[28].

La minería web es la integración de información obtenida mediante métodos tradicionales de la minería de datos, con información recogida sobre la web. Esta se divide en tres tipos, los cuales utilizan diferente método de extracción y análisis de información, pero con la misma finalidad.

La minería web se clasifica en función de la parte de la web que se analiza, por lo que existen tres aéreas principales:

2.16.1. Minería de contenido web.

La minería de contenido de la web recoge datos e identifica patrones relativos al contenido de la web y a las búsquedas que se realizan sobre los mismos. También extrae patrones existentes en los contenidos de las páginas web, utilizando como entrada:

- Texto libre.
- Código HTML.
- Código XML.
- Recursos multimedia.
- Cualquier otro contenido que se encuentre en la web.
- Minería de resultados de búsquedas.
- En los patrones de resultados de los motores de búsqueda.

El principal problema con el que se enfrenta la minería de contenido es el cambio de los datos estructurados, no-estructurados o semi-estructurados, ya que por la misma naturaleza de la web esta tiende a evolucionar constantemente.

2.16.2. Minería de estructura de la web.

La minería de estructura de la web pretende revelar la organización real de un sitio web a través de la recolección de datos referente a su estructura y principalmente a su conectividad. Este tipo de minería puede usarse para categorizar páginas web y es útil para

generar información como la similitud y relación entre diferentes sitios Web. En el artículo Excavando la web, El profesional de la información[29], el autor Ricardo Baeza-Yates comenta: “la estructura puede ser usada por los buscadores para jerarquizar los resultados en base a las paginas mas referenciadas” utilizando heurísticas como pagerank[30], usando los buscadores de sitios web.

2.16.3. Minería de los registros de navegación de la web

La minería de registros de navegación de la web es el proceso de descubrimiento automático de patrones de acceso o uso de servicios de la web, centrándose en el comportamiento de los usuarios cuando interactúan en internet.

Estos registros se hacen en el log³⁰ del servidor. El log del servidor es el que se emplea para el análisis de tráfico y uso. Este archivo se almacena en texto plano (ASCII) y puede estar en diferentes formatos. Otra manera de encontrar esta información puede ser directamente en las bases de datos, ya que la configuración del almacenaje depende del administrador del servidor. El contenido del log del servidor depende de su configuración. Generalmente la información que se almacena es: fecha de la petición, IP³¹ del cliente que hizo la solicitud, el método (POST-GET)³², el archivo acceso, etc.

2.16.4. Recuperación de información (IR).

Los objetivos originales de la recuperación de información eran realizar una indexación para facilitar la búsqueda de documentos útiles en una colección[31]. Ahora se define como el proceso de extracción de los contenidos web, es decir, tomar la información del sitio web y almacenarla de manera local para su procesamiento, lo que incluye la organización, representación y acceso a la información. Para esta técnica, existen herramientas que permiten detectar patrones de texto de acuerdo a los criterios fijados por el usuario y en relación a la necesidad.

³⁰ Registro o bitácora que se almacena en el servidor de la aplicación.

³¹ Internet Protocol, por sus siglas en ingles. Es una etiqueta numérica que identifica cada dispositivo en la red.

³² Se refiere al envío y recepción de parámetros, que se envían a través de HTTP.

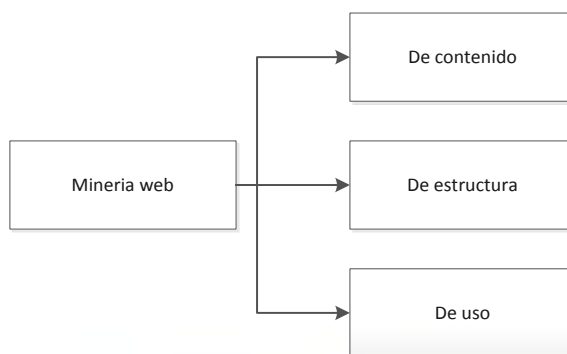


FIGURA 4. TAXONOMIA DE LA MINERIA WEB.



CAPITULO 3.

3. Estado del Arte.

En esta sección se presenta el estado del arte en la minería web, así como las técnicas utilizadas para la extracción de información de la WWW³³.

Actualmente existen numerosas metodologías para la lectura de información en la web, tal es el caso de la minería web, la cual utiliza herramientas y técnicas de minería de datos para extraer información de las páginas web. Cabe mencionar que la minería web es un campo de investigación bastante amplio, ya que involucra técnicas como: extracción y recuperación de información, entre otras.

La aplicación de estas metodologías de extracción de conocimiento de la web, se ha realizado en diferentes áreas. A continuación, se muestran algunos trabajos relacionados con las disciplinas en las que se está aplicando actualmente y se exponen los trabajos más significativos.

Yury Yineth Niño Roa[32], realizó un reporte sobre la minería de textos web de bibliografías médicas, con un enfoque hacia la clasificación y categorización de documentos. La implementación de la minería web optimizó el proceso de recolección, clasificación y análisis de datos en internet, con el fin de obtener un conocimiento que sea útil en la prevención y tratamiento de enfermedades.

Un inconveniente encontrado por el autor, fue que cuando intentó hacer minería de contenido sobre textos médicos, mucha de esta información científica se encontraba en artículos digitalizados de publicaciones medicas, así que para atacar ese problema, fue necesario recurrir a otro campo de investigación relacionado con la minería de texto, el cual ofreció conceptos y técnicas más apropiados para el tratamiento de información no estructurada.

³³ World Web Wide, por sus siglas en inglés. Es un sistema de distribución de información, basada en hipertexto o hipermedios enlazados y accesibles a través de internet.

A continuación, se muestran algunas herramientas nombradas en el artículo en comento, las cuales fueron utilizadas para la minería de texto en el campo médico.

Año	Herramienta
2003	MedScan: Motor NPL. Software orientado al dominio biomédico, que procesa frases de los abstract ³⁴ de MEDLINE y produce un conjunto de estructuras lógicas regularizadas que presentan el significado de cada frase.
2004	TXTGate: Software que combina índices de literatura de fuentes biológicas disponibles en el sistema de minería de texto para el análisis de grupos de genes. El clustering ³⁵ se hace sobre información contenida en múltiples bases de datos bilógicas en línea, entre ellas MEDLINE.
2003	PubMatrix: Herramienta basada en web que permite hacer minería de texto sobre literatura de CNBI1. Específicamente presenta comparaciones en dos dimensiones de los nombres de los genes y términos funcionales, basados en la combinación de resultados de múltiples queries ³⁶ , hechos sobre PUBMed.
2004	Borat: herramienta para extracción de información, específicamente biomédica. La herramienta es capaz de acceder y analizar el abstract y el documento completo. Combina un diseño de plantilla con un web spider ³⁷ que ubica y recupera artículos de journals.
1990	CONTRUE: Sistema desarrollado por el grupo Carnegie Mellon University. Su característica especial consiste en que se usa un dominio específico desarrollado manualmente. De acuerdo con los autores, adaptarlo con otros dominios de aplicación podría ser costoso y representaría una labor intensiva debido a que tiene reglas de aplicación específica

³⁴ Resumen que cubre los principales puntos de un escrito.

³⁵ Conjunto o conglomerados de computadoras.

³⁶ Consulta realizada a través de una base de datos, utilizada para obtener datos, modificarlos o borrarlos.

³⁷ Robot de buscadores que accede a las páginas web y las indexa.

En el trabajo de Niño Roa, se puede ver el uso de la minería en las herramientas listadas, las cuales tienen una aplicación en diferentes áreas, pero con un mismo objetivo, que es la minería de texto.

Por otro lado, en el artículo: Tratamiento automático de noticias empresariales en la Web, mediante minería textual, I. Esquivel[33] buscó aplicar técnicas de minería textual para realizar la búsqueda, filtrado, procesamiento y descubrimiento de noticias, que permitieran determinar el tipo y frecuencia de las actividades realizadas por los participantes del área de noticias.

El sistema que se propuso en ese trabajo, es un sistema que navega través de los sitios web previamente identificados de las empresas de interés y extrae los informes noticiosos y los procesa para descubrir nuevo conocimiento, e intenta responder las siguientes preguntas:

- ¿Qué están realizando los competidores del sector?
- ¿Qué está haciendo específicamente el competidor x?
- ¿Cuáles actividades se dan en momentos específicos del año?
- ¿A qué tipo de actividades parece darles más importancia cada empresa?
- ¿Existen tendencias de estabilidad o cambio, con respecto a ciertas actividades?

De igual manera, se listan los principales trabajos que sirvieron de base para el proyecto.

Proyecto	Descripción
Clasificación automática de desastres naturales en México	Sistema de clasificación automática de noticias de desastres naturales en México. Los resultados obtenidos permiten clasificar una página web dentro de las categorías de huracán, inundación, sequia y no relevantes, con exactitud aproximada de un 97%.
Sistema para la clasificación de notas periodísticas.	Sistema para la clasificación de noticias. Este sistema funciona en base a criterios, donde clasifica las noticias de acuerdo a notas negativas y positivas para una entidad específica.
Nominalización en	Sistema para detección de información a partir de los títulos de

<p>títulos: una manera de extraer detalles de documentos.</p>	<p>los documentos e identificación del nivel de contenido específico.</p>
--	---

Las noticias que las empresas publican en sus sitios web generalmente corresponden a eventos que fueron, están, o serán realizados y se presentan a manera de avisos, dirigidos hacia la comunidad que sirven. En algunos casos los formatos son los siguientes:

- a) Una lista de titulares de los eventos.
- b) Un listado de títulos de los eventos con enlace hacia páginas que informan a detalle.
- c) Una relación de encabezados, seguidos de la descripción completa del evento.
- d) Una lista de noticias con encabezado, una descripción parcial y un enlace a otra página complementaria.

Para el procesamiento automático de los titulares de eventos, el sistema propuesto en ese trabajo realiza las actividades de configuración, pre procesamiento y descubrimiento, las cuales se describen a continuación:

- **Configuración:** Esta actividad consistió en la adición y/o modificación de la empresa competidora de interés, así como su nombre y dirección web en donde publican avisos.
- **Pre procesamiento:** En esta actividad se realizaron las operaciones de lectura y escritura, en una base periódica, como se describe a continuación:

Lectura:

- Se obtuvo del registro de empresas, la dirección electrónica donde se publican las noticias.
- Se obtuvo una copia del archivo HTML.
- Se convirtió el archivo en texto sin etiquetas.
- Se procedió a transformar el archivo resultante, eliminando de cada noticia, el texto fuera del título.

- Se compararon los títulos obtenidos con una versión anterior y en caso de coincidir cada uno de ellos, se eliminó el archivo de trabajo; en caso contrario se mantiene para su posterior tratamiento.
- El archivo que resultó se duplicó como versión anterior para su posterior comparación.
- El nombre del archivo se agregó al listado de empresas por trabajar.

Escritura:

- Se obtuvo el nombre del archivo de entrada.
- Se leyó el archivo correspondiente cada uno de los títulos que lo conforman.
- Se eliminaron del texto aquellas palabras no significativas.
- Cada una de las palabras resultantes se comparó con la lista de tópicos previamente configurada.
- Al coincidir, se incrementaron los contadores empresa, categoría y tópico correspondientes.
- En caso contrario, se verificó la existencia del término en la lista de palabras no referenciadas.

Una vez terminado el procesamiento de todos los archivos, se registraron los datos relativos a los titulares de las noticias: fecha, empresa, tópico, categoría y frecuencia. Además de esto, se generó un archivo nombrado de acuerdo a la fecha de corrida, con los términos no referenciados más frecuentes para su futura consulta y posible actualización de datos.

Descubrimiento:

El autor, basándose en el trabajo: Mining the News: Trends, Associations, and Deviations [34] utilizó representaciones estadísticas (frecuencias y distribuciones de probabilidad) de los reportes noticiosos y medias estadísticas (mediana, desviación estándar, coeficiente de cambios y otros) para los hallazgos.

Por su parte, en el trabajo denominado: Estado del Arte del Web Mining, José E. Medina Pagola[35] presentó una estructura y descripción de los tipos de minería web, y sus principales aplicaciones. Además, exhibió algunas técnicas, métodos y análisis de diversas tecnologías asociadas con ese tipo de minería.

En su artículo, el autor comenta que en la tipificación de aplicaciones y problemática de la minería web se han encontrado diversas clasificaciones. Comenta que algunos autores han considerado 2 tipos de minería, siendo éstas, minería de contenido de la web y minería de uso de la web. Pero la mayoría de los especialistas en esta área, consideran 3 tipos de minería web, atendida por el tipo de información a minar, las cuales son:

- Minería de estructura de la web.
- Minería de uso de la web.
- Minería de contenido de la web.

La recuperación de información (IR) es un término que se aplica a la búsqueda y recuperación de información, sobre un tema en específico o una colección de documentos o textos. Estas técnicas han sido aplicadas en los últimos 20 años en la indexación de documentos, pasando por diferentes aplicaciones desde la multimedia e hipertextos, hasta las que se realizan sobre la web.

Las principales actividades de IR son:

- Modelación de conocimiento a indexar;
- Clasificación y categorización de los documentos;
- Filtraje de la información; y
- Perfeccionamiento de las interfaces de usuarios.

De estas actividades, según el autor, la clasificación y categorización presentan una evidente relación con las técnicas de minería de contenido, permitiendo una mejor indexación.

De acuerdo al autor, existen 2 tipos principales de búsqueda en la web: los directorios como Yahoo, Netscape, entre otros y los motores de búsqueda tipo Lycos, Google, etc. Comenta

que los sistemas tipo directorio están orientados a búsquedas mas interactivas, por lo que en general, aseguran una eficiente y rápida organización y clasificación de la información que muestran y recomiendan.

La extracción de información es una forma de aprendizaje o descubrimiento de patrones textuales, a partir de información no estructurada o semi-estructurada.

Menciona también que los sistemas de agentes son sistemas en los que varios agentes, usualmente inteligentes, interactúan entre sí para lograr determinados objetivos o realizar ciertas tareas. La relación con la minería web, se observa en que estos sistemas actúan en un ambiente distribuido como son las redes locales e internet. Un tipo especial de los agentes distribuidos, es el conocido como agencities³⁸, el cual es un conjunto de sistemas de software conectado a redes de internet publica, donde cada una de esas plataformas posee sistemas de agentes capaces de comunicarse con el mundo exterior usando mecanismos estándares de comunicación.

Las plataformas suelen estructurarse con diferentes tipos de agentes. Entre los tipos relevantes en la minería web se tienen los siguientes:

- Agentes de interfaces de usuarios: que son aquellos que incluyen técnicas de aprendizaje de los intereses de los usuarios (agentes de recuperación de información, agentes recomendadores, agentes personales, etc.)
- Agentes distribuidos: que conforman grupos de agentes que actúan de forma cooperada para descubrir determinado conocimiento (agentes de filtraje cooperativo entre otros).

Los sistemas multi-agentes están siendo aplicados en muy diversas situaciones y utilizan muchas de las técnicas antes expuestas; ejemplo de estos son:

- El proyecto @LIS TechNet, el cual pretende crear un ambiente de enseñanza y experimentación entre varios países de Europa y América latina³⁹.
- El proyecto de tesis de maestría de Alfredo Rojo, el cual propone diferentes tipos de agentes para recomendar recursos digitales considerando la métrica TF-IDF y el

³⁸ <http://www.agencities.org>

³⁹ <http://alis.cs.bath.ac.uk/alis/>

algoritmo KEA, el cual fue basado en el proyecto de bibliotecas digitales de Nueva Zelanda⁴⁰.

Por último, el autor en su trabajo, lista algunos especialistas vinculados con la minería web, donde los ordena alfabéticamente por su primer apellido y junto a cada especialista, señala sus direcciones de correo y páginas personales.

En ese trabajo, el autor presenta de una forma muy sintética algunos de los principales tipos de aplicaciones de la minería web, los cuales fueron expuestos a partir de una estructuración de los tipos de minería de este medio, tratando de considerar el criterio de la mayoría de especialistas de esta temática.

En cuanto a los autores Richard S. Segall y Qingyu Zhang[36], en su artículo: Web Mining of Hotel Customer Survey Data, plantean técnicas de minería web que se aplica en las encuestas realizadas a los clientes de un hotel, utilizando el software Megaputer PolyAnalyst®. Las técnicas de minería empleadas son clustering, análisis de vínculos, extracción de frases, taxonomías y matrices de dimensión.

En un repaso en la literatura, diferentes autores nombrados en ese artículo hablan de sus investigaciones tales como técnicas para analizar y recuperar patrones de los clientes de bases de datos de sistemas construidos en la WWW[37]. Otra investigación señalada en el artículo en comento, fue la realizada por Murphy, et al[38], la cual consistió en la investigación de mercado mediante el seguimiento de rastros dejados por los visitantes en los sitios de algunas principales marcas hoteleras. Por mencionar solo alguna, una investigación realizada sobre minería de contenido y extracción web reportada en ese artículo, es la del autor Scharlr[39], quien utilizó un método de extracción de contenido web para investigar el comercio electrónico de los viajes y turismo competitivo.

El software utilizado para esa investigación es el Megaputer PolyAnalyst®, el cual es un sistema analítico que tiene una característica única de integración de la minería web, junto con los datos de la minería de texto. El sistema puede cargar datos de entrada de las fuentes de los sitios web, así como de las diferentes bases de datos más populares, generar

⁴⁰ <http://www.sadl.uleth.ca/nz/cgi-bin/library>

estadísticas, y crear datos a partir de hojas de cálculo. También puede cargar documentos tales como hojas HTML, archivos DOC, PDF y archivos de texto, entre otros.

En ese trabajo, los autores introducen una herramienta que es utilizada para analizar los datos de una encuesta aplicada a los clientes de un hotel. La cual también demuestra la utilidad de la minería web en la extracción de palabras claves, análisis de los vínculos de éstas y toda la consulta en la búsqueda de datos de la encuesta de los clientes del hotel. Esto con la finalidad de dar seguimiento a los servicios de éste comercio que necesitan ser mejorados.

En México, el autor Luis Antonio Lucio López[40]. Presenta una investigación realizada a 34 centros educativos del nivel medio superior de la Universidad Autónoma de Nuevo León, donde dio a conocer como es la convivencia de los estudiantes en el ciberespacio.

Esta fue una investigación exploratoria y descriptiva, debido a que no hay antecedentes en México sobre esta clase de estudios. En la metodología utilizada se aplicó un cuestionario en la modalidad de auto informe a una muestra de 1,066 alumnos con un 95% de confianza y con un límite de error no mayor al 0.3 (+- %). El objetivo general de la investigación fue determinar la incidencia de acoso e intimidación en el ciberespacio, entre los alumnos de la preparatoria. Además de que otra meta fue conocer la incidencia de actos deplorables como atentados al honor e intimidación de los alumnos, así como afectar la reputación de los alumnos en sitios como fotolog⁴¹, youtube⁴² y otras redes sociales a través de fotografías, videos y escritos ofensivos.

La pregunta que guió al autor fue: ¿cuál es la incidencia de conductas de acoso e intimidación a través de internet que sufren como víctimas los estudiantes de las escuelas preparatorias de la UANL?

La población de participantes fue de 44 mil 900 alumnos de las escuelas preparatorias durante el semestre Agosto-Diciembre del 2008. El tipo de muestreo probabilístico fue aleatorio, estratificado con asignación proporcional por centro y genero.

⁴¹ <http://www.fotolog-mx.com/>

⁴² <http://www.youtube.com/>

El instrumento utilizado como guía fue generado previamente para realizar otro estudio sobre el mismo tema, únicamente se le realizó algunas modificaciones y adaptaciones en cuanto a conductas de bullying, pero esta vez enfocado al ciberespacio. El estudio piloto que se realizó en este trabajo se aplicó en una preparatoria con una participación de 50 alumnos que contestaron las preguntas.

Una vez realizadas las entrevistas, los datos fueron procesados y codificados en el programa estadísticos SPSS versión 16. Los resultados obtenidos en este estudio se listan a continuación:

- El 17.3% de los estudiantes de las escuelas preparatorias de la UANL ha insultado a alguien en internet.
- El 22.2% ha amenazado a alguien a través de la red.
- El 10% ha acosado sexualmente a alguien, a través de internet.
- El 5.1% ha chantajeado a algún compañero pidiéndole cosas a cambio de no divulgar sus cosas íntimas en la red.
- El 7.8% de los alumnos ha enviado correos con mensajes desagradables.
- El 11.5% ha afectado la reputación de compañeros en la red con comentarios que atentan contra el honor.
- El 10.7% de los alumnos dijo haber difundido malos rumores de alguien en fotolog, páginas personales, foros, youtube y otros portales.
- El 8.1% de los estudiantes señaló haber subido a youtube u otras páginas similares, vídeos desagradables para ofender a alguien.
- El 7.9% dijo haber subido fotografías a la red con la intención de ridiculizar o desprestigiar a alguien.
- El 9.3% del alumnado confesó haber robado contraseñas para impedir a compañeros entrar a sus fotolog.
- El 11.1% señaló haber incursionado ilegalmente en equipos de sus compañeros con técnicas hacker.

En ese trabajo, el autor pudo demostrar que el cyberbullying es una realidad en las escuelas, pues es común que alumnos y alumnas agredan a sus compañeros a través de formas que

expresan en el ciberespacio, tales como la publicación de fotografías, videos, comentarios, correos, chat, etc.

Otros trabajos relacionados.

Artículo	Prevalencia y consecuencias del cyberbullying: una revisión.
Autores	Maite Garaigordobil.
Datos observados	Estudios sobre la prevalencia del cyberbullying en España, países europeos, América Latina, Oceanía y Asia.
Objeto de estudio	Estudiantes de escuelas preparatorias.
Instrumento utilizado	Mediante la recolección de información.
Artículo	Cyberbullying Offending and Victimization.
Autores	Hinduja, S. & Patchin, J. W.
Datos observados	<ul style="list-style-type: none"> • El promedio de los entrevistados pasa por semana 18 horas en línea. • El 33 % de los hombres y el 36 % de las mujeres reportaron haber sido víctimas alguna vez, siendo los chats, la forma más común. • El 18% de los hombres y casi el 16 % de las mujeres informaron de acoso cibernético. • Los investigadores también midieron los comportamientos que se asocian típicamente con bullying tradicional. • Los jóvenes que intimidaron a otros, o fueron intimidados en la vida real en los últimos seis meses, tuvieron 2.5 veces más probabilidades de ser molestados o ciberacosados.
Objeto de estudio	1,378 adolescentes.
Instrumento utilizado	Cuestionario.
Artículo	An Exploratory Study of Cyberbullying with Undergraduate University Students.
Autores	Carol M. Walker, Beth Rajan Sockman, Steven Koehn.
Datos observados	Análisis a 120 estudiantes no graduados quienes señalaron que: <ul style="list-style-type: none"> • 54% conocen a alguien que ha sufrido cyberbullying. • 11% han experimentado cyberbullying. • 29% han experimentado el cyberbullying de 4 a 10 veces. • 14% lo han experimentado más de 10 veces.
Objeto de estudio	Estudiantes universitarios.

Instrumento utilizado	Cuestionario.
Artículo	Cyberbullying. El crudo problema de la victimización en línea.
Autores	Luz María Velázquez Reyes.
Datos observados	Se investigaron 2 poblaciones: <ul style="list-style-type: none"> • 594 alumnos de secundaria (48% hombres y 52% mujeres). • 31 alumnos de licenciatura (61% hombres y 39% mujeres).
Objeto de estudio	Estudiantes de secundaria y licenciatura.
Instrumento utilizado	Investigación por medio de relatos narrados por estudiantes.
Artículo	El tratamiento en Internet de las noticias de violencia de género.
Autores	Javier Borrás Torregrosa y Elvira García de Torres.
Datos observados	Tratamiento de una noticia relacionada a la agresión contra las mujeres y la relación con el soporte y el tema de las noticias que se publican en la red.
Objeto de estudio	Diario Levante-EMV.
Instrumento utilizado	Mediante la selección de unidades informativas que hagan referencia a la agresión, y mediante la minería de contenido.

En base a los estudios expuestos en esta sección, se puede observar las investigaciones desarrolladas por los diferentes autores en el campo de la minería y su aplicación en las diferentes áreas, así como las diferentes herramientas desarrolladas para el análisis de contenidos. También se expusieron algunos trabajos relacionados con la ciber-violencia, así como los instrumentos que cada autor utilizó para realizar sus descubrimientos.

Cabe mencionar que no se encontró bibliografía referente al uso de la minería de contenido web para detectar situaciones de ciber-violencia en un contexto determinado, por lo que con esta investigación se está experimentando en un campo nuevo de investigación.

CAPITULO 4.

4. Modelo de la ciber-violencia en los sitios web.

4.1. Modelo conceptual de ciber-violencia en los sitios de noticias.

Los contenidos digitales disponibles en la WWW, son divididos en temáticas que van dirigidas a diferentes tipos de lectores. Con la llegada de la web 2.0 cada contenido digital puede ser realimentado de manera participativa por los usuarios, esto gracias a las herramientas web colaborativas que hoy integran los sitios web.

Dichas herramientas permiten que los usuarios que se encuentren registrados en el portal y que cuenten con un correo electrónico válido, utilizando seudónimos o, mediante las aplicaciones desarrolladas por terceros, puedan contribuir con las opiniones acerca del tema publicado. Esto da cabida a que los usuarios, ejerciendo libremente su derecho de libre expresión, puedan realizar comentarios utilizando el vocabulario (adecuado o inadecuado), sobre el tema.

En otros países como China, Chile y Estados Unidos, se están desatando controversias entre la población, debido a la publicación de comentarios en algunos sitios de noticias realizados por los usuarios. Por documentar un ejemplo en esta tesis, una noticia publicada en el sitio de noticias CNN MEXICO⁴³ el lunes 4 de abril del 2012, que lleva por título: “China cierra sitios web por difundir rumores sobre un golpe de Estado”, explica lo siguiente: “El mayor sitio de microblogging de China suspendió las secciones de comentarios este fin de semana, después de ser ‘castigado por permitir que se difundieran rumores’ de un intento de golpe de Estado en Beijing, la cual fue una noticia que reportaron los medios oficiales de ese país⁴⁴”.

⁴³<http://mexico.cnn.com/>

⁴⁴Noticia publicada en: <http://mexico.cnn.com/tecnologia/2012/04/02/china-cierra-sitios-web-por-difundir-rumores-sobre-un-golpe-de-estado>

4.1.1. Descripción del modelo de ciber-violencia.

En la Figura 5. Modelo de ciber-violencia en los contenidos, se describe el modelo conceptual de la ciber-violencia en los sitios de noticias en internet, el cual se explica a continuación:

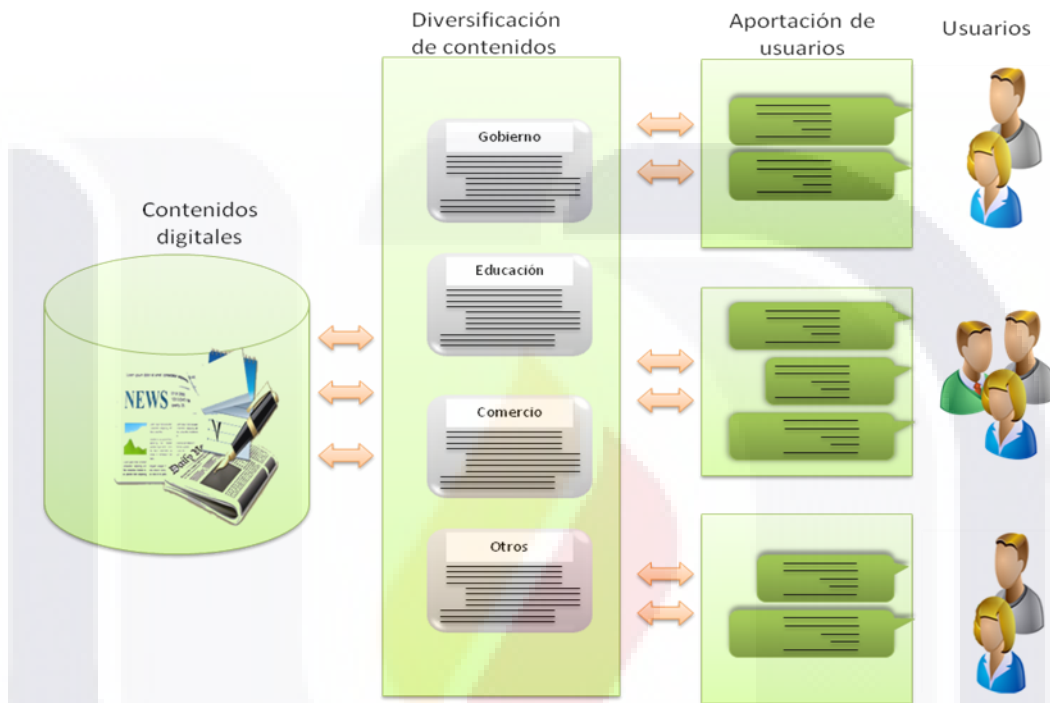


FIGURA 5. MODELO DE CIBER-VIOLENCIA EN LOS CONTENIDOS.

- **Los contenidos digitales.** En primer lugar se tiene una base de datos de contenidos digitales en línea, los cuales provienen de diferentes fuentes informativas como sitios de noticias o blogs, los cuales son alimentados día a día.
- **Diversificación de contenido.** En segundo lugar se muestra la diversificación de contenidos, los cuales se encuentran divididos y/o clasificados, dependiendo de su naturaleza, llámese temas sociales, deportivos, policíacos, entretenimiento, etc. Estos contenidos se generan al día, dependiendo de los acontecimientos y el impacto que tengan sobre la sociedad.
- **Aportación de usuarios.** En la actualidad, los sitios web de noticias y blogs contienen herramientas que permiten a los usuarios participar en los temas publicados. Estas

herramientas están disponibles para los usuarios que deseen contribuir con sus comentarios sobre los temas.

4.1.2. Roles involucrados en el modelo de ciber-violencia.

Rol	Característica principal
Periodista	Es la persona que se dedica profesionalmente al periodismo, en cualquiera de sus formas, ya sea en la prensa escrita, radio, televisión o medios digitales.
Editor	Editor o jefe de redacción es una persona o entidad que edita o publica una obra ajena costeándola y administrándola.
Usuario	Es la persona que utiliza el sitio para consultar información que se publica día a día, además de que es quien puede participar en los comentarios sobre las publicaciones. El usuario puede tomar dos papeles, el de agresor y el de agredido.
Usuario agresor	Es quien agrede a otros usuarios mediante los comentarios en las publicaciones.
Usuario agredido	Es quien recibe las agresiones textuales por parte de un usuario agresor.
Fuente de información	Es toda materia que genere, contenga, suministre o transfiera información.
Contenidos digitales	Es una imagen, un documento digitalizado, un libro, electrónico, una página HTML, etc.

TABLA 1. ROLES INVOLUCRADOS EN EL MODELO DE CIBER-VIOLENCIA.

4.2. Caso de estudio. Violencia textual en el sitio de noticias Aguascalientes.com.

Los objetivos de este caso de estudio son:

- Conocer el diario online Aguascalientes.com.
- Presentar el contexto de agresión textual que se observa en las noticias online.

El diario online Aguas digital.

El Diario Aguas es una publicación que circula de lunes a domingo en Aguascalientes, en su zona conurbana y en los municipios del interior del Estado. Fue creado el 3 de enero del año 2000, por la Empresa Editorial de Aguascalientes, la cual es una organización periodística fundada en 1981. Es formato de este diario es de tipo tabloide.

Así mismo se creó Aguasdigital.com⁴⁵, la cual es la edición para internet, que apareció por primera vez en febrero de 2007 y fue reformada y ampliada a partir de mayo de 2008, como parte de una estrategia para ampliar las posibilidades de dar voz a nuestros lectores.

La redacción está integrada por jóvenes universitarios, egresados de las instituciones educativas del estado, y es una empresa en la que jóvenes profesionistas conforman la mayoría, lo cual brinda la característica de un diario joven y diferente.

Aguasdigital.com tiene un público de todas las edades, aunque su presencia en las universidades del estado de Aguascalientes lo ha convertido en la opción preferida de los jóvenes.

El contenido que se publica en Aguasdigital.com es estrictamente informativo, contiene desde noticias locales, hasta noticias nacionales e internacionales. Cabe mencionar que el sitio de Aguasdigital.com contiene hipervínculos hacia otros periódicos en línea, tal es el

⁴⁵ <http://www.aguasdigital.com>

TESIS TESIS TESIS TESIS TESIS

caso de hidrocálido digital⁴⁶ que también es un periódico de Aguascalientes, entre otros diarios.

El sitio Aguasdigital.com también pone a la disposición de sus lectores servicios como: paginas amigas, callejero, horóscopos, videos, galerías, encuestas, etc. Convirtiendo a este periódico en un diario muy completo.

Las secciones que contiene Aguasdigital.com son:

- **Principal.** La cual es una sección en la que se despliegan los encabezados principales de las noticias locales, nacionales e internacionales.
- **Actualidad.** Que es la sección especial en la que se publican las últimas noticias que acontecen en el estado, en el país, o en el mundo.
- **Aguascalientes.** Es la sección especial para las noticias locales actuales y anteriores.
- **Marcador.** Es la sección especial dedicada a los eventos deportivos.
- **Las plumas.** Esta sección está enfocada a columnistas y críticos.
- **Tendencia.** Sección dedicada a eventos sociales.

Cada una de estas secciones es actualizada constantemente dejándole saber al usuario la fecha de actualización y periodicidad de la información.

Contexto de la agresión.

Internet, como muchos otros avances tecnológicos de la humanidad, fue concebido para lograr fines benéficos, y aún cuando sigue siendo utilizado para ello en su mayoría, también se ha prestado para la realización de conductas inapropiadas, entre ellas se encuentra la lesión a la fama pública y al honor de una persona, mediante la humillación u ofensa, que un usuario practique en perjuicio de otro u otros, mediante la publicación de comentarios ofensivos en algún sitio público, tal es el caso de los blogs, foros o sitios de noticias. Dicha práctica es alentada por el anonimato en el cual pueden ampararse los usuarios.

⁴⁶ <http://hidrocalidodigital.com>

Actualmente el sitio de noticias Aguasdigital.com permite que los usuarios puedan expresar su opinión en relación a un determinado tema de actualidad, que conforma su acervo de noticias. De esta manera, el usuario, mediante las herramientas de participación del diario digital, puede incursionar en la discusión del tema, publicando sus comentarios, los cuales en muchas ocasiones son expresados mediante el uso de lenguaje ofensivo, dirigido a determinado usuario o a un grupo de usuarios, generando o, en ocasiones, involucrando la participación de otros usuarios, quienes se integran en la discusión del tema, expresando su punto de vista, pero siguiendo un mismo patrón de agresividad en el lenguaje.

En relación a lo anterior, si una persona es agredida, humillada u ofendida públicamente a través de los comentarios plasmados en el sitio de noticias Aguasdigital.com, no existe hasta ahora una disposición normativa mediante la cual se le pueda fincar responsabilidad al agresor.

Aguasdigital.com cuenta con mecanismos de seguridad y herramientas que permiten que los administradores moderen los comentarios publicados, pero en algunos casos estas herramientas no son utilizadas correctamente, lo cual propensa la creación de situaciones incómodas y humillantes para los usuarios, sin importar que dichas situaciones fueron creadas o no por ellos mismos.

Por lo tanto se ha encontrado que el sitio Aguasdigital.com, así como otros sitios de noticias, están siendo utilizados por muchos de sus usuarios para expresar comentarios ofensivos en las noticias publicadas, lo que ha ocasionado que el nivel de agresión que se está presentando, vaya en aumento. Además de que los comentarios ofensivos que generan los usuarios, incitan a otros usuarios a crear cadenas de agresión. Lo anterior indica que, en un futuro, el sitio Aguasdigital.com podría convertirse en un periódico de baja reputación, al seguir presentando este tipo de agresiones.

CAPITULO 5.

5. Sistema de Detección de Situaciones de Ciber-Violencia en los sitios de noticias y blogs.

Con base en la investigación realizada en torno a la problemática actual de violencia en los sitios de noticias y blogs, la cual se encuentra plasmada en el cuerpo del presente trabajo, se propone una herramienta tecnológica que mediante la minería de contenido web, identifique la presencia de ciber-violencia y, en consecuencia, determine la existencia de ciber-violencia en los comentarios publicados por los usuarios en los sitios de noticias y blogs en Aguascalientes.

Por lo tanto, la solución que se propone se basa en la combinación de herramientas tecnológicas, como lo es un sistema informático basado en una plataforma web, la cual permitirá realizar análisis de texto de sitios online, mediante minería de contenido web, para detectar las situaciones de ciber-violencia en los comentarios de los sitios de noticias y blogs en internet.

Lo anterior se realizará mediante la implementación de técnicas de Minería de la web, específicamente con el dominio de Minería de contenido, utilizando técnicas de recuperación y extracción de información[31], así como de comparación de cadenas de texto, es decir, esta combinación permitirá explorar el contenido de los sitios web mediante la herramienta, con la finalidad de detectar las situaciones de ciber-violencia.

5.1. Modelo arquitectónico de la aplicación.

La arquitectura utilizada en el sistema de detección de situaciones de ciber-violencia en los contenidos sociales se basa completamente en una aplicación web, por lo que se decidió seguir la arquitectura Model-View-Controller la cual aumenta las capacidades de la arquitectura conocida como Modelo Cliente/Servidor. Al implementar la aplicación siguiendo esta arquitectura, se obtienen todas las características utilizadas en la actualidad para el desarrollo de las aplicaciones web y las ventajas de los sistemas distribuidos.

En la Figura 6. Diagrama de bloques., se muestra el diagrama del sistema, en el que se observa tanto la parte de programación, como la parte de infraestructura que se utilizara para su desarrollo.

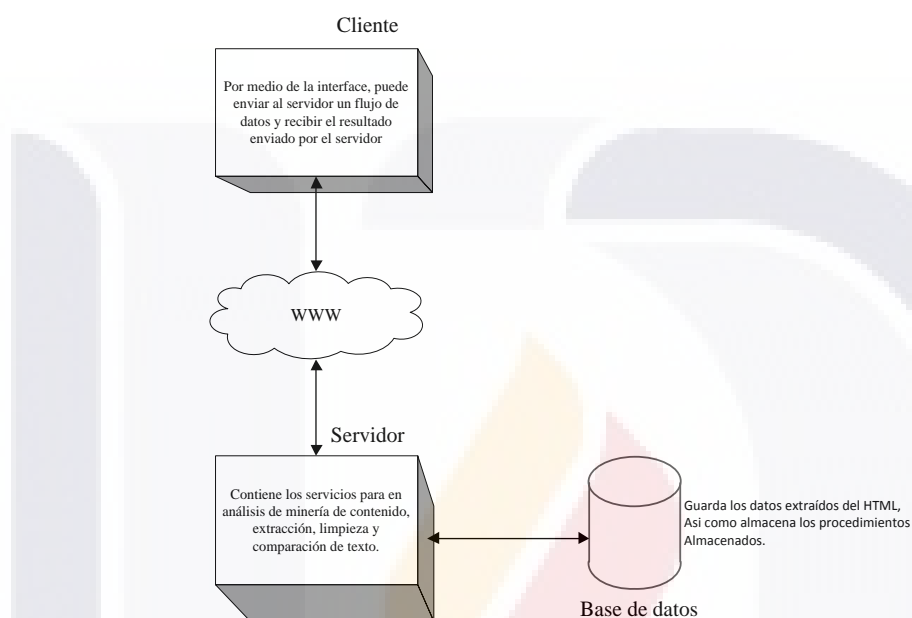


FIGURA 6. DIAGRAMA DE BLOQUES.

La arquitectura cliente/servidor es el modelo más utilizado tanto para aplicaciones web, como para servicios de alto nivel o paginas dinámicas. En este modelo de desarrollo de aplicaciones web, las operaciones del sistema se dividen en procesos independientes que cooperan para proporcionar la respuesta. Esta arquitectura consta de tres componentes: 1) el proceso que solicita el servicio de visualización, extracción, limpieza y comparación (cliente); 2) el proceso que atiende la solicitud del servicio (servidor); y finalmente 3) el componente intermedio que comunica a los clientes con el servidor (redes LAN o WAN).

En una descripción más detallada se muestra la Figura 7. Arquitectura de la aplicación., la cual muestra los componentes que intervienen en dicha arquitectura. Como parte central se muestra el servidor web, que contiene la aplicación con la lógica para procesar, almacenar, limpiar, comparar y exponer el resultado del análisis.

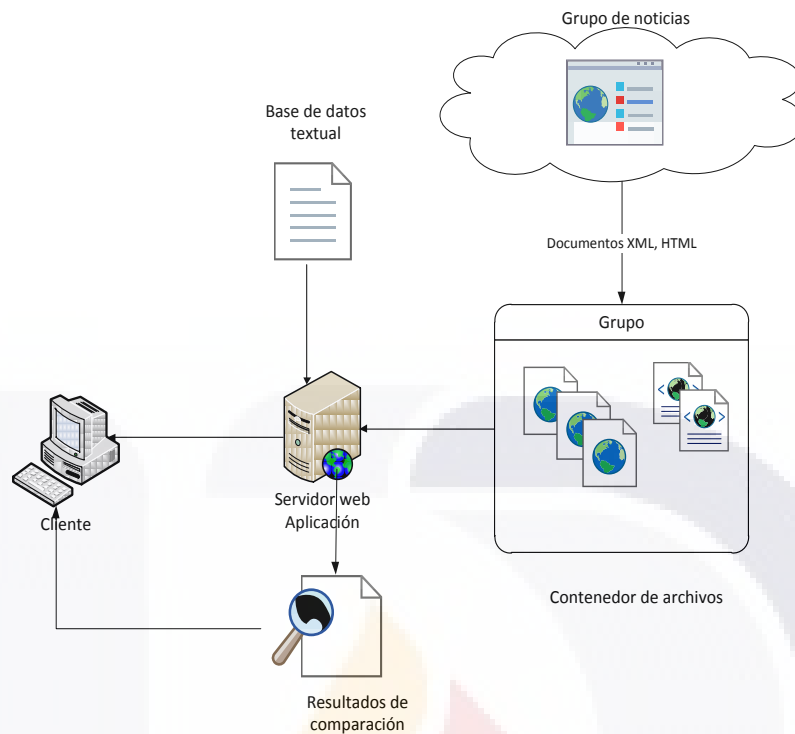


FIGURA 7. ARQUITECTURA DE LA APLICACIÓN.

5.2. Componentes principales del sistema.

Como se muestra en la Figura 8. Componentes del sistema., el sistema se basa en la arquitectura de tres capas, las cuales se describen a continuación:

- **Interface de usuario.** Conformada por una interface principal, en el cual el funcionamiento en la captura y despliegue de información se realiza mediante la utilización de iframes y ventanas modales.
- **Lógica del sistema.** Conformada por las instrucciones recibidas desde la interface del usuario y las operaciones de análisis almacenadas en el sistema.
- **Acceso a datos.** Controla la lectura y escritura de la información y es donde se almacenan los datos encontrados.

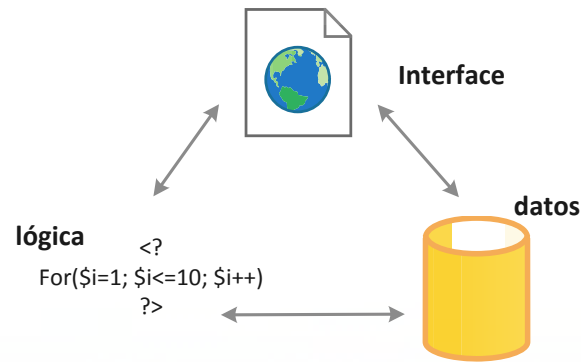


FIGURA 8. COMPONENTES DEL SISTEMA.

5.3. Tecnologías elegidas para el desarrollo del sistema.

Las tecnologías para el desarrollo del sistema fueron elegidas de acuerdo a la funcionalidad y consumo de recursos que demanda la aplicación. A continuación se describen las tecnologías implementadas en el desarrollo del sistema y la razón por la cual se decidió utilizarlas en el desarrollo de cada capa de la aplicación.

- **Interface de usuario:**

Tecnología	Razón
JQuery 1.7.2 ⁴⁷	<ul style="list-style-type: none"> • Permite realizar aplicaciones AJAX ligeras y sencillas. • Permite hacer llamadas asíncronas al servidor. • Compatibilidad con la mayoría de navegadores. • Framework java script. • Manipulación del DOM con menos código. • Componentes de la interface sencillos de programar.
HTML5 ⁴⁸	<ul style="list-style-type: none"> • Permite añadir formas y tablas en la

⁴⁷ <http://jquery.com/>

⁴⁸ <http://dev.w3.org/html5/spec/>

	aplicación. <ul style="list-style-type: none"> • Compatibilidad con la mayoría de los navegadores.
CSS(Cascade Style Sheets) ⁴⁹	<ul style="list-style-type: none"> • Separación del contenido con la presentación. • Optimización de la carga de las páginas. • Unificación del diseño de las páginas. • Compatibilidad con los navegadores.

• **Lógica del sistema:**

Tecnología	Razón
PHP V5.3 ⁵⁰	<ul style="list-style-type: none"> • Velocidad de ejecución en las operaciones de la aplicación. • No requiere demasiados recursos. • Estabilidad de la aplicación. • Seguridad. • Simplicidad para generar nuevo código.

• **Acceso a datos:**

Tecnología	Razón
MySQL 5.5 ⁵¹	<ul style="list-style-type: none"> • Velocidad de ejecución en las rutinas y procedimientos almacenados en la base de datos.

⁴⁹ <http://www.w3.org/standards/webdesign/htmlcss#whatcss>

⁵⁰ <http://www.php.net/>

⁵¹ <http://www.mysql.com/>

5.4. Análisis de contenidos.

En la presente tesis, el análisis de contenidos se refiere a la metodología utilizada para realizar la extracción, limpieza y análisis textual, para la detección de situaciones de ciber-violencia a través de un sistema informático.

En este tema se presenta las fases del ciclo del análisis de contenido, desde la etapa selección, hasta la etapa de presentación de los resultados.

La Figura 9. Fases del análisis de contenidos., se muestra el procesamiento de información a partir del contenido de una página web.

5.4.1. Fases del análisis.

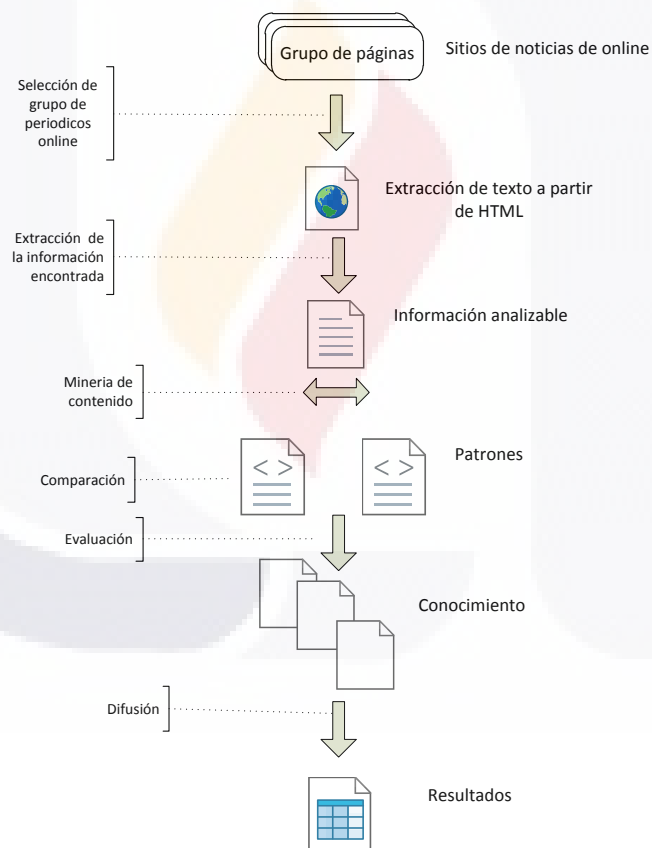


FIGURA 9. FASES DEL ANÁLISIS DE CONTENIDOS.

El análisis de contenidos se compone de las siguientes fases:

- **Fase de selección** de blogs y/o noticias destacadas. Para este caso, la selección será a partir de la existencia de un catalogo de sitios de noticias o blogs online que cumplan con los criterios que previamente sean fijados.
- **Fase de extracción de texto** a analizar mediante un parseo de código HTML, para descargar el texto, limpiarlo y almacenarlo en la base de datos por medio de la aplicación.
- **Minería de contenido** para la búsqueda de situaciones de ciber-violencia sobre un contenido específico, en el cual se ejecutarán los algoritmos de limpieza y comparación entre el texto extraído y el catalogo con el cual se quiera realizar la comparación.
- **Evaluación** de los resultados encontrados en el análisis, el cual tendrá por objetivo detectar las situaciones de ciber-violencia.
- **Difusión** y presentación de los resultados y hallazgos encontrados.

5.4.2. Selección de grupos de periódicos online.

Este proceso consiste en la búsqueda y selección de periódicos online y/o blogs, los cuales serán analizados. Para poder desarrollar el análisis, los periódicos deberán permitir la participación de los usuarios en los temas, esto es, que las noticias una vez publicadas permitan que los usuarios puedan contribuir con su opinión sobre el tema publicado mediante la herramienta de comentarios incluida en el sitio web.

Una vez encontrados los periódicos online, se creará un listado por nombre, sección, noticia y fecha, el cual ayudará a tener un mejor control de los resultados, tal como se muestra en la siguiente tabla:

Nombre del periódico en línea:							
Sección analizada:							
Fecha de análisis							
Título de la noticia	URL	Fecha de análisis	Fecha de publicación	Total de palabras encontradas	Total de comentarios	Total de palabras en comentarios	Total de texto agresivo encontrado

TABLA 2. SELECCIÓN Y EVALUACIÓN

5.4.3. Extracción de información.

La calidad de la extracción de texto, no solo depende del algoritmo con el que se realice la extracción, sino también dependerá de la calidad de los datos. La detección del texto en el contenido HTML nos ayudará a identificar las palabras clave en el tema tratado.

La extracción del contenido será realizada mediante un parseo de código el cual almacenará todo el texto en un contenedor, donde posteriormente se realizará limpieza de texto para eliminar todo caracter innecesario, tales como caracteres especiales y caracteres numéricos. La limpieza de texto ayudará a acelerar el proceso de comparación de cadenas de texto.

En el ejemplo de la Figura 10. Contenido noticioso y comentarios. Se muestra el tipo de elemento a extraer para el análisis de contenido.



FIGURA 10. CONTENIDO NOTICIOSO Y COMENTARIOS.

La estructura actual de las páginas web de noticias y blogs se integra por un titulo, el cuerpo donde se encuentra el contenido de la noticia, un frame que contiene publicidad (imágenes GIF, JPG, FLASH etc.) y, en la parte inferior, un apartado diseñado especialmente para los comentarios, el cual permite que los usuarios lectores puedan

contribuir con sus opiniones. Cada sitio noticioso utiliza un estándar de diseño propio, por lo que los elementos que se contienen en el sitio pueden variar en la posición. Los comentarios están sujetos a una identificación mediante un correo electrónico válido, un nombre, el cual no es forzoso pero se utiliza para la identificación del comentario con otros usuarios, y finalmente el comentario referente al tema publicado.

5.4.4. Minería de contenido.

En cuanto a la minería de contenido, es indispensable identificar todo el texto de la noticia, además de los comentarios generados por los usuarios. Cabe precisar que no todos los sitios de noticias online y/o blogs permiten que los usuarios puedan publicar comentarios debido a políticas administrativas del mismo sitio, pero tampoco todos los sitios de noticias online contienen comentarios, por lo que en este caso la minería de contenido no arrojaría los datos que se necesitan para el análisis.

Una vez extraído el texto de la noticia, se realizará una comparación, entre el texto extraído de los comentarios de la noticia, y el contenido en el catálogo de entrada. Para este caso se utilizará un algoritmo para realizar una comparación del texto extraído de la página HTML, con la finalidad de validar la similitud de las cadenas de texto.

5.4.5. Comparación mediante el Algoritmo de la Distancia de Levenshtein.

La distancia de Levenshtein es una medida de la similitud entre dos cadenas: la cadena fuente y la cadena de destino. La distancia es el número de eliminaciones, inserciones o sustituciones requeridas para transformar una cadena en otra. Cuanto mayor sea la distancia de Levenshtein, más diferentes son los textos. El algoritmo utiliza dos cadenas, la cadena de origen como entrada y la cadena de destino, como punto de comparación, o entradas del diccionario[41].

Una vez seleccionado el texto del HTML proveniente de la noticia o blog, el contenido se depositará en una base de datos para realizar la comparación y búsqueda de similitudes en los textos, frente al catálogo.

Esta comparación deberá arrojar un resultado igual a cero al encontrar similitudes en las palabras o frases, caso en el cual el resultado se incrementará para contabilizar los hallazgos encontrados. Ahora, si dicha comparación arroja un resultado mayor a cero, significa que el comentario no presenta ninguna similitud entre los textos contenidos.

A continuación se presenta el algoritmo:

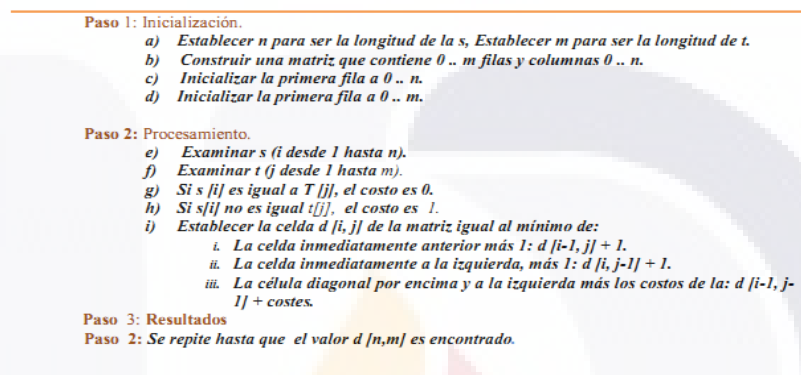


FIGURA 11. ALGORITMO DE LA DISTANCIA DE LEVENSTHEIN.

5.4.6. Evaluación.

Una vez analizado y comparado el texto mediante el algoritmo de la distancia de Levenshtein, se procederá a determinar la agresión que se presenta en los comentarios de la noticia. Esto se realizará tomando como base la tabla de selección y evaluación propuesta en el tema 5.4.2 Selección de grupos de periódicos online, para hacer el registro de los hallazgos encontrados y determinar si presenta ciber-violencia.

5.4.7. Difusión.

Los datos arrojados del análisis se publicarán en la tabla de selección y evaluación, en la cual se mostrarán los totales de acuerdo a los siguientes criterios:

- Total de palabras encontradas;
- Total de palabras por comentario;
- Total de comentarios; y
- Total de palabras agresivas encontradas.

CAPITULO 6.

6. Validación y resultados del análisis en la detección de situaciones de ciber-violencia.

6.1. Desarrollo sustancial del análisis de detección de situaciones ciber-violencia.

Para el desarrollo del análisis en la Detección de Ciber-Violencia, se verificaron las fases de análisis de contenido descritas en el tema 5.4, del capítulo anterior. Ahora se mostrará el seguimiento dado a cada fase, así como los resultados que fueron obtenidos.

6.1.1. Selección de grupo de noticias online.

Para determinar si los sitios de noticias online en Aguascalientes contienen o no ciber-violencia, se llevó a un procedimiento de selección para determinar qué periódicos formarían parte de la población que se analizaría, por lo que se verificaron los siguientes puntos:

- a) **Definición de criterios a considerar para la población.** Los criterios que se consideraron para determinar la población sujeta a análisis fueron los siguientes tres:
 - Que se trate de un sitio de noticias online, es decir, que sea un periódico que cuente con versión en internet.
 - Que sea del Estado de Aguascalientes.
 - Que cuente con una sección de comentarios donde se permita a los lectores o usuarios publicar comentarios.

- b) **Búsqueda e identificación de periódicos online del Estado de Aguascalientes.** Mediante una investigación realizada en el buscador de Google⁵², se determinó la existencia de 16 sitios de noticias online en el Estado de Aguascalientes, listados en Tabla 3. sitios de noticias online en Aguascalientes. Cumpliéndose con ello los dos primeros requisitos.

⁵² <http://www.google.com>

Nombre del sitio	URL
La Jornada Aguascalientes	http://www.lajornadaaguascalientes.com.mx/
Desde la red	http://www.desdelared.com.mx/
Visión urbana	http://www.visionurbananoticias.com/
El Herald de Aguascalientes	http://www.heraldo.mx/
Ultra Noticias	http://www.ultra.com.mx/noticias/aguascalientes.html
Aguas Digital	http://www.aguasdigital.com/
Revista Cambio	http://revistacambio.mx/
Crisol Plural	http://crisolplural.com/
Noticiero el Circo	http://www.noticieroelcirco.com/
El Hidrocálido	http://www.hidrocalidodigital.com/
La Grilla	http://lagrilla.blogspot.mx/
El Sol del Centro	http://www.oem.com.mx/elsoldelcentro/
Blog Aguascalientes	http://blogaguascalientes.blogspot.mx/
Página 24	http://www.pagina24.com.mx/
Picacho	http://picacho.com.mx/inicio/
El Clarinete	http://www.elclarinete.com.mx/

TABLA 3. SITIOS DE NOTICIAS ONLINE EN AGUASCALIENTES.

c) **Selección de los periódicos online de Aguascalientes que contengan una sección de comentarios.** Luego de la búsqueda e identificación de periódicos en línea que fueran editados en Aguascalientes, se realizó un diagnóstico de los mismos para identificar aquellos que permitieran a los lectores o usuarios comentar la noticia, verificándose con ello el tercer y último requisito para identificar a la población sujeta al análisis. En este proceso 9 periódicos fueron los que cumplieron con este criterio, de los 16 que se obtuvieron inicialmente.

Nombre del sitio	URL
Aguas digital	http://aguasdigital.com/index.php
Blog Aguascalientes	http://blogaguascalientes.blogspot.mx/
El Clarinete	http://www.elclarinete.com.mx/
Crisol Plural	http://crisolplural.com/
La Grilla	http://lagrilla.blogspot.mx/
La Jornada Aguascalientes	http://www.lajornadaaguascalientes.com.mx/
Noticiero El Circo	http://www.noticieroelcirco.com/
Picacho	http://picacho.com.mx/inicio/
Revista Cambio	http://revistacambio.mx/

TABLA 4. SELECCIÓN DE SITIOS DE NOTICIAS.

d) **Establecimiento del orden en los periódicos seleccionados.** Para establecer un orden en los periódicos que conformaron la población a analizar, se utilizó el

criterio alfabético, ordenándolos en relación a la primera letra del nombre del periódico; luego de ordenarlos alfabéticamente, fueron numerados del 1 al 9.

#	Nombre del sitio	URL
1	Aguas digital	http://aguasdigital.com/index.php
2	Blog Aguascalientes	http://blogaguascalientes.blogspot.mx/
3	Clarinete El	http://www.elclarinete.com.mx/
4	Crisol Plural	http://crisolplural.com/
5	Grilla La	http://lagrilla.blogspot.mx/
6	Jornada Aguascalientes La	http://www.lajornadaaguascalientes.com.mx/
7	Noticiero el Circo	http://www.noticieroelcirco.com/
8	Picacho	http://picacho.com.mx/inicio/
9	Revista Cambio	http://revistacambio.mx/

TABLA 5. ESTABLECIMIENTO DEL ORDEN DE LA POBLACIÓN.

e) **Método de selección de la muestra.** Después de haber obtenido la población sujeta a análisis, fue necesario extraer una muestra que representara la totalidad de la población. Para ello se empleó el muestreo aleatorio sistemático, en el cual, una vez que los periódicos a analizar fueron ordenados de manera alfabética y numérica, se realizó el siguiente procedimiento:

Uno. Se determinó que el tamaño de la muestra a analizar sería de 3, es decir, se analizarían 3 sitios de noticias online de una población de 9.

Dos. Se determinó el coeficiente de elevación utilizando la fórmula:

$$\frac{n}{m} = \frac{9}{3} = 3$$

Dónde las variables corresponden a los siguientes valores:

n= población;

m= unidades que contiene la muestra;

Por tanto, el coeficiente de elevación corresponde a 3.

Tres. Se determinó el origen de la muestra en base a un número aleatorio, dando como resultado el número 1, del orden numérico que se estableció previamente.

Cuatro. Para seleccionar el segundo sitio de noticias online que conformaría la muestra, se tomó el origen (1), más el coeficiente de elevación (3), de la siguiente manera:

$$1 + 3 = 4$$

Por tanto, el siguiente sitio de noticias online para la muestra es el correspondiente al número 4.

Cinco. Para obtener el último sitio de noticias online que conformaría el tamaño de la muestra, se sumó la unidad resultante anterior (4), más el intervalo constante (3), de la siguiente manera:

$$4 + 3 = 7$$

En consecuencia, el sitio de noticias online resultante es el que correspondió al número 7.

Como resultado del procedimiento anterior, se obtuvo que los sitios de noticias online que formarían parte de la muestra, serían el número 1, 4 y 7, en el orden alfabético-numérico que previamente se les había asignado.

Cabe apuntar que los periódicos que conformaron la población contienen las secciones de: noticias locales, sociales, ciencia y tecnología, cultura, deportes, educación y otras; sin embargo, para efectos de esta investigación, sólo se analizaron los artículos de la sección de noticias locales de los periódicos que forman parte de la muestra y que tienen una antigüedad no mayor a 6 meses a la fecha de publicación del presente trabajo, por tanto sólo se analizaron aquellas que fueron publicadas entre el mes de enero y mayo de dos mil doce. **Ver Anexo A.**

6.1.2. Extracción del texto mediante minería de contenido utilizando el Sistema de Detección de Situaciones de Ciber-Violencia.

Esta fase fue realizada utilizando el Sistema de Detección de Ciber-Violencia, el cual opera mediante seis funciones básicas, las cuales se describen a continuación:

- **Visualizar.** Sirve para presentarle al usuario el contenido de la noticia que se analizará para una revisión visual rápida y esencial.
- **Extraer texto.** Función que realiza un parseo de texto, para extraer todo el contenido dentro de las etiquetas HTML, y al mismo tiempo almacenarlo en una tabla de la base de datos, creada específicamente para almacenar el contenido extraído.
- **Preparar.** Esta función despliega una ventada que contiene a su vez dos sub funciones que se encargan de preparar el texto extraído para su análisis:
 - **Limpiar contenido.** Elimina todos los caracteres que no son necesarios para el análisis.
 - **Analizar contenido.** Mediante éste, el sistema identifica el total de palabras encontradas en el contenido de la noticia, el total de comentarios y el total de palabras encontradas en los comentarios.
- **Analizar.** Determina si existe o no agresión en la publicación analizada, en base a la comparación con el catalogo generado específicamente para esta investigación, utilizando el algoritmo de la distancia de Levensthain.
- **Cerrar frame.** Este botón cierra el menú.

A continuación se muestra la interface principal del Sistema de Detección de Situaciones de Ciber-Violencia.

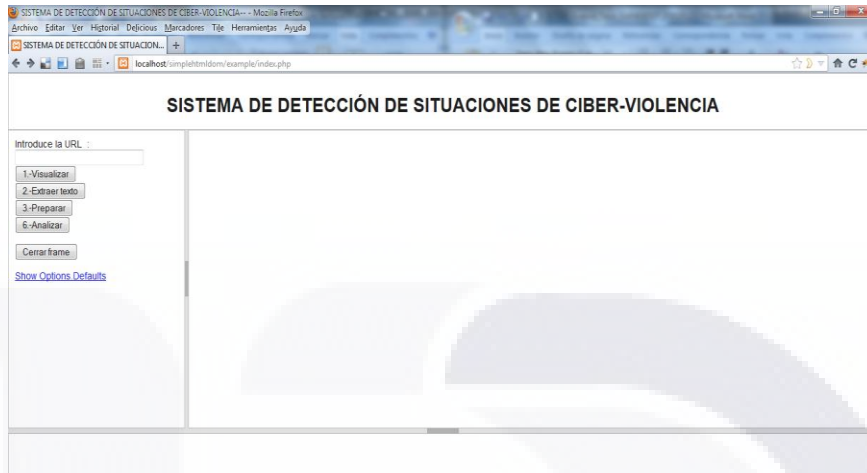


FIGURA 12. INTERFACE DEL SISTEMA DE DETECCIÓN DE CIBER-VIOLENCIA.

Las especificaciones de requerimiento del Sistema de Detección de Ciber-Violencia, se explican en el **Anexo B**.

Una vez determinada la muestra, se hizo un listado con las noticias comprendidas entre enero y mayo de dos mil doce, las cuales están contenidas en el **Anexo C**. Cada una de las noticias, fue procesada a través del Sistema de Detección de Ciber-Violencia, en el cual se extrajo y se almaceno el texto contenido dentro de las etiquetas HTML en una tabla de la base de datos del propio sistema.

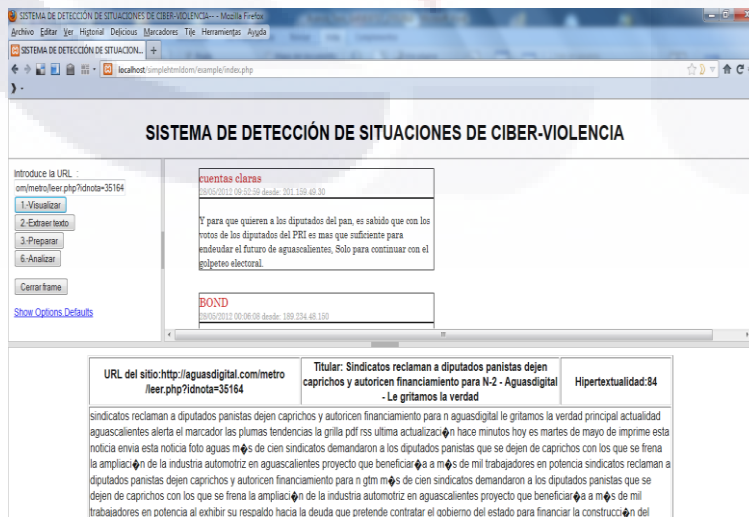


FIGURA 13. EXTRACCIÓN DE TEXTO DE LA NOTICIA.

Es necesario señalar que la publicación de artículos o noticias de la muestra, no tienen una periodicidad constante, por lo que se seleccionaron noticias de fechas diversas, pero comprendidas dentro del periodo de enero a mayo de dos mil doce.

6.1.3. Comparación.

Una vez realizada la extracción, el sistema analizó la similitud del texto extraído del contenido publicado, contra las palabras que conforman el catálogo de agresión, mediante el algoritmo de la distancia de Leveinshtein. El sistema evaluó la similitud de las palabras para detectar coincidencias existentes en el contexto en que se puso a prueba, desplegando todas aquellas palabras, consideradas agresivas, según el catálogo de agresión.

6.1.4. Evaluación.

Luego de haber realizado el análisis y comparación mediante el Sistema de Detección de Situaciones de Ciber-Violencia, para cada una de las noticias que conformaron la muestra, se procedió a vaciar los resultados arrojados por el sistema a la tabla de selección y evaluación en la cual se concentro la información obtenida. **Ver Anexo D.**

6.2. Resultados obtenidos.

En el desarrollo de la presente investigación se utilizó tanto información teórica, esto es, aquella extraída de la literatura, como información práctica o de campo, es decir, información obtenida mediante la aplicación del instrumento de evaluación y los resultados obtenidos en el Sistema de Detección de Situaciones de Ciber-Violencia.

Como consecuencia de lo anterior, los resultados obtenidos en la investigación se dividen en tres vertientes.

6.2.1. Resultados derivados de la teoría.

Toda vez que la hipótesis planteada en la presente investigación se centra en detectar la existencia de ciber-violencia en los sitios online y blogs de Aguascalientes, fue necesario acudir a la literatura y delimitar el concepto de violencia, las causas que la originan y los antecedentes de ésta. Así se encontró que existe violencia no sólo social, económica o de género, sino también que existen determinantes genéticas, psicológicas y contextuales que predisponen al ser humano para desplegar una conducta, hacía sí mismo o hacía otro, la cual puede ser nociva tanto física como psicológicamente.

Además de lo anterior, y partiendo del hecho de que por naturaleza el hombre tiende a adaptarse a las nuevas condiciones de vida, así como de la innegable y creciente evolución tecnológica que se ha desarrollado en los últimos tiempos, especialmente en el área de las comunicaciones, se descubrió que el comportamiento humano ya no se restringe tan sólo a la comunicación física, sino que se ha extendido también a la comunicación virtual y es ahí precisamente en donde se posiciona esta nueva figura creada como consecuencia de la adaptación del hombre a la vida moderna: la ciber-violencia.

6.2.2. Resultados derivados de la aplicación del instrumento.

Por su parte, los resultados prácticos se refieren a aquellos obtenidos del instrumento de evaluación o cuestionario, y los obtenidos del Sistema de Detección de Situaciones de Ciber-Violencia.

En cuanto al instrumento de medición, el resultado principal que éste arrojó fue la aportación de materia prima para la creación de un catálogo de agresión, compuesto por todas aquellas palabras o expresiones consideradas como agresivas por la población a la que se le aplicó, lo cual a su vez, constituyó uno de los elementos esenciales del sistema propuesto, pues mediante él, el Sistema de Detección de Situaciones de Ciber-Violencia detectó contenido agresivo en los sitios de noticias online de Aguascalientes que fueron analizados. **Ver Anexo E.**

Ahora bien, para el desarrollo del catálogo de agresión, el instrumento de evaluación fue aplicado a un grupo de 305 estudiantes del Estado de Aguascalientes, con un rango de edad de 23 a 27 años; el cuestionario se formuló con 8 preguntas, enfocadas todas ellas al tema de ciber-violencia, de las cuales, las seis primeras arrojaron datos cuantitativos y las dos últimas, cualitativos.

Los resultados conseguidos en dicho instrumento fueron capturados en una tabla de Excel de la paquetería Microsoft Office 2007; luego, las respuestas de la pregunta 1 a 6 fueron representadas cada una en una gráfica de barras verticales, cuya finalidad fue mostrar de manera clara los resultados obtenidos en cada pregunta; la respuesta a la pregunta número 7 no fue plasmada en los resultados toda vez que la respuesta obedece a términos subjetivos que sólo fue necesario interpretar, más no cuantificar; en cuanto a la respuesta a la pregunta número 8, ésta fue vaciada en una tabla de datos, ordenando su información por moda y frecuencia. Cabe señalar que ésta fue la base para conformar el inventario de palabras y expresiones textuales del catálogo de agresión. **Ver Anexo F.**

6.2.3. Resultados derivados del Sistema de Detección de Situaciones de Ciber-Violencia.

En cuanto a los resultados obtenidos de la aplicación del Sistema de Detección de Situaciones de Ciber-Violencia, se observó lo siguiente:

El análisis de cada uno de los tres sitios de noticias online del Estado de Aguascalientes que conformaron la muestra representativa de la población que se analizó a través del sistema, mostró de manera cuantitativa la existencia de texto agresivo en los comentarios publicados

en las noticias analizadas. Sin embargo debe considerarse que los resultados obtenidos en dicho análisis son consecuencia de la comparación de la muestra con un vocabulario considerado agresivo en una determinada zona geográfica, respecto de personas con un específico rango de edad y en un mismo nivel educativo, pues el cuestionario del cual se creó el catálogo de agresión, fue aplicado únicamente en Aguascalientes, a estudiantes universitarios de entre 23 y 27 años de edad; por lo tanto, dicho catálogo puede incrementar su inventario si es enriquecido con aportaciones derivadas de la búsqueda de vocabulario agresivo en otros contextos.

A continuación se presentan los resultados del total de texto que fue encontrado por el sistema en los comentarios de los sitios de noticias online que formaron parte de la muestra, así como el total de texto agresivo contenido en dichos comentarios:

Sitios de noticias online	Total de texto encontrado en comentarios	Total de texto agresivo encontrado en comentarios
Aguas digital	2,831	41
El Circo	4,048	89
Crisol Plurar	1,989	1

TABLA 6. RESULTADOS DE LA MUESTRA.

Cabe señalar que una palabra o expresión textual pudo haberse presentado en más de una ocasión en los comentarios, por lo que al haber sido detectada y catalogada como agresiva por el sistema, ésta fue contabilizada tantas veces como ocurría en los comentarios.

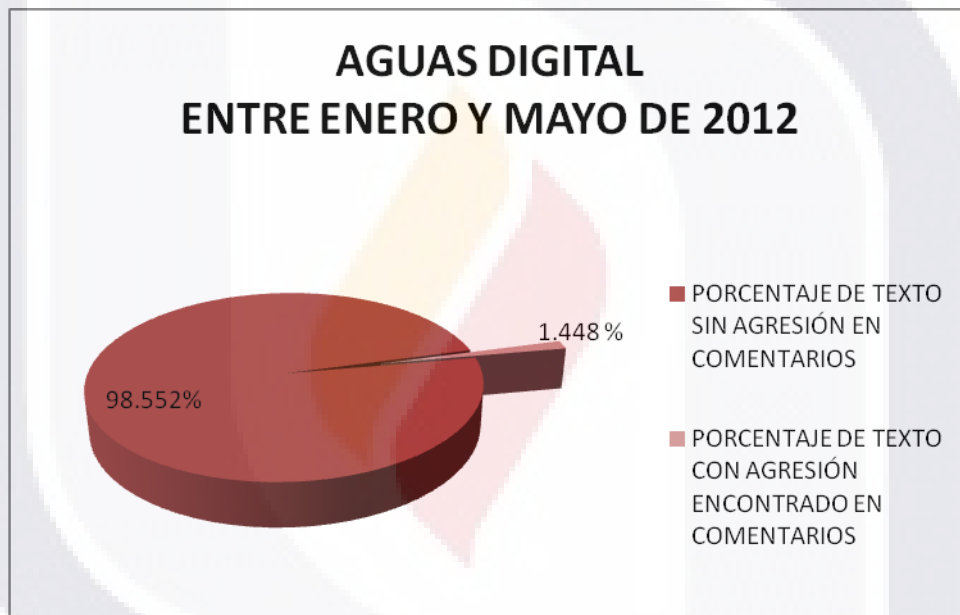
En la siguiente tabla se muestra el porcentaje total de texto agresivo que el Sistema de Detección de Situaciones de Ciber-Violencia encontró en los comentarios, por cada uno de los sitios online, así como el porcentaje que correspondió al texto clasificado como no agresivo del total de los comentarios de cada sitio. Cabe señalar que para obtener el porcentaje del resultado de texto agresivo, se dividió éste entre el total de texto encontrado en los comentarios y finalmente se multiplicó por 100 para agregar el factor porcentual a cada uno de ellos.

Sitios de noticias online	Porcentaje de texto sin agresión	Porcentaje de texto con agresión
Aguas digital	98.552%	1.448%
El Circo	97.801%	2.199%
Crisol Plurar	99.95%	0.050%

TABLA 7. RESULTADOS PORCENTUALES POR SITIO DE NOTICIAS ONLINE.

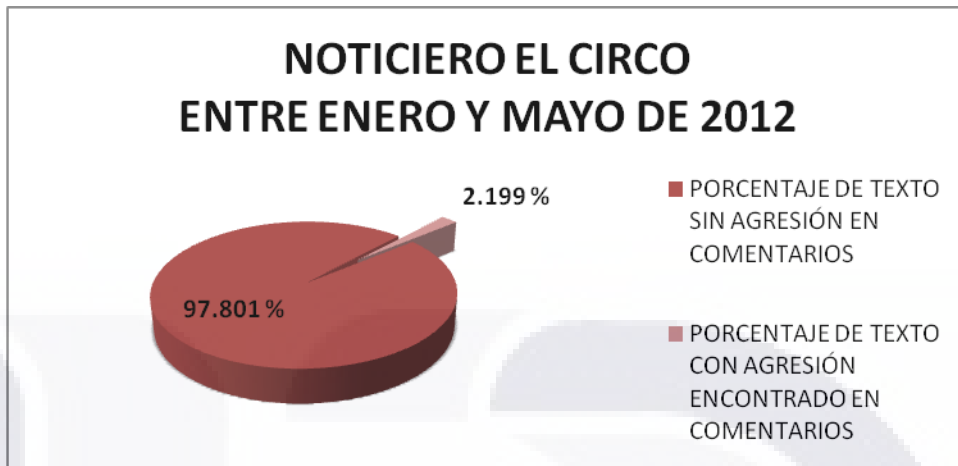
A efecto de contar con una visión clara de la magnitud de los resultados anteriores, se muestran las siguientes gráficas.

- El sitio de noticias online Aguasdigital.com obtuvo los siguientes resultados en porcentaje:



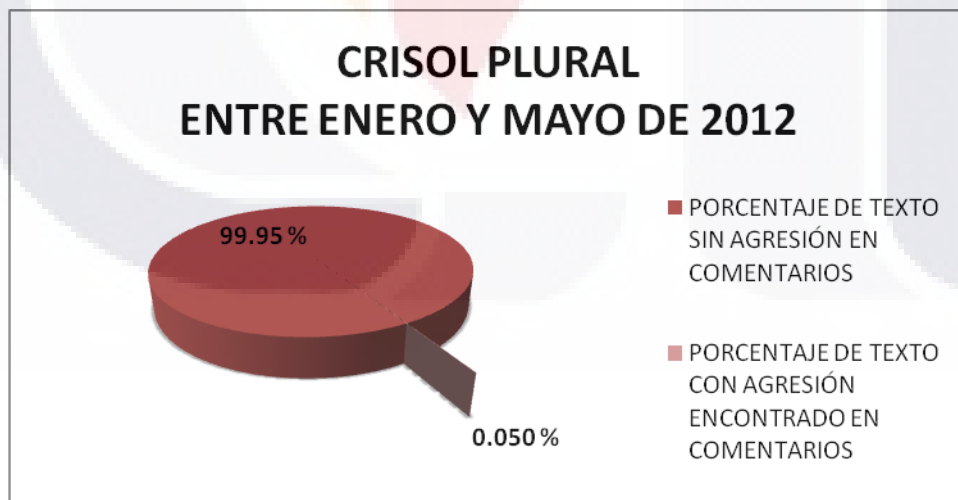
De la gráfica anterior se dedujo que de un texto total de 2,831 palabras encontradas en los comentarios realizados en las noticias online de Aguas digital, publicadas entre enero y mayo de 2012, el 1.448% correspondió a texto agresivo, mientras que el 98.552% de texto no presentó agresión.

- El sitio de noticias online Noticiero El Circo, obtuvo los siguientes resultados:



En esta gráfica se puede apreciar que el texto agresivo encontrado en los comentarios publicados en este sitio de noticias online entre enero y mayo del 2012, correspondió al 2.199 %, con relación a un total de 4,048 palabras que conformaron los comentarios en los que participaron los usuarios, y el porcentaje del texto sin agresión del total de palabras fue del 97.801%.

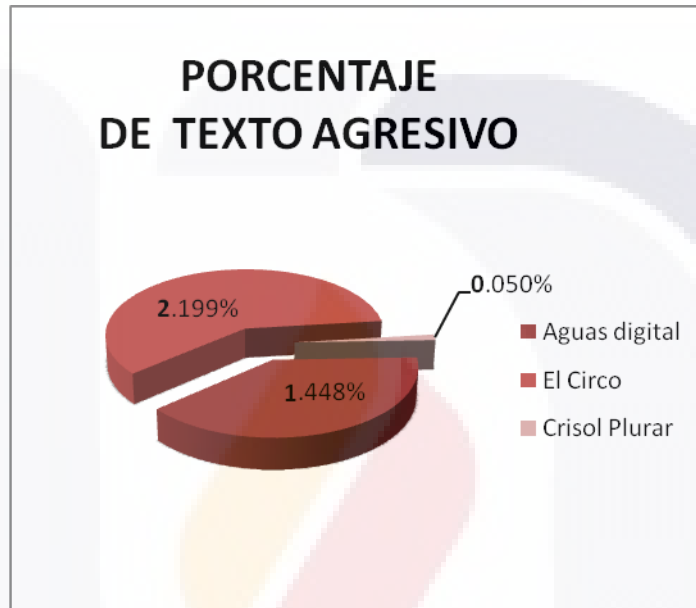
- En el sitio de noticias online Crisol Plural, se obtuvieron los siguientes resultados:



En esta gráfica se muestra que, entre enero y mayo del 2012, el porcentaje de texto agresivo encontrado en los comentarios plasmados por los usuarios entre enero y mayo del 2012, representó tan sólo el 0.050%, en relación a un total de 1989 palabras que conformaron los

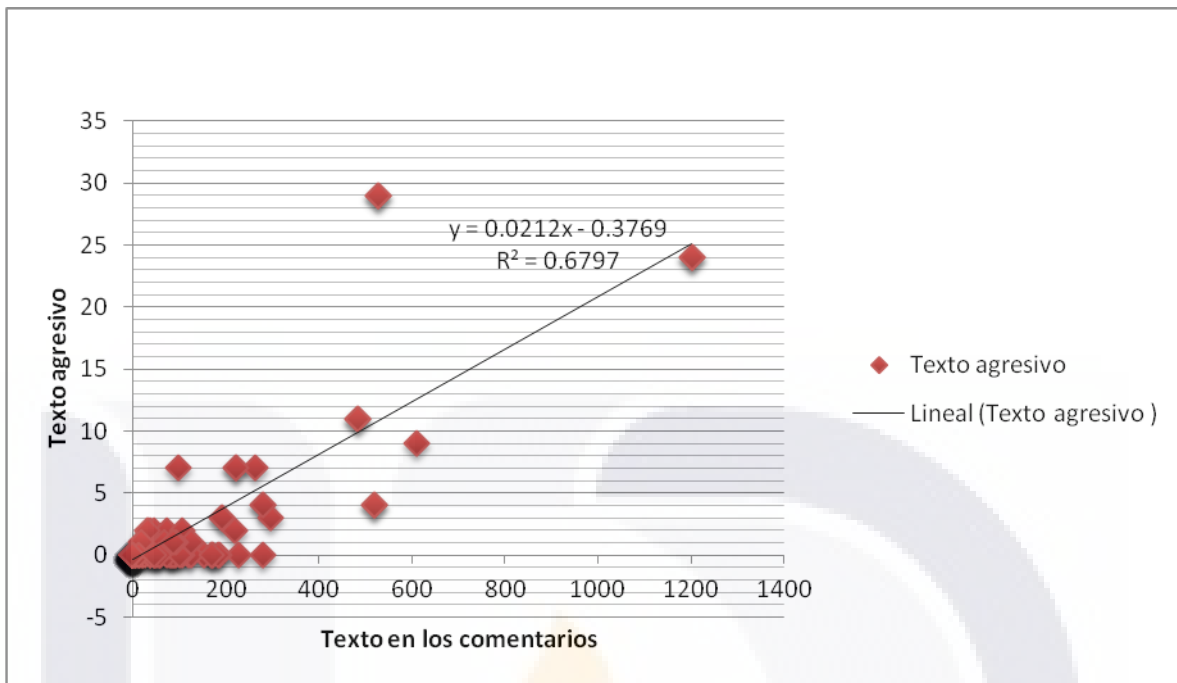
comentarios, mientras que el porcentaje que correspondió a las palabras que no tuvieron agresión es del 99.95%.

Por último, en la siguiente gráfica se plasma el porcentaje que correspondió a cada uno de los sitios online de la muestra, respecto al texto agresivo que presentaron los comentarios de sus noticias entre enero y mayo del 2012.



Por lo tanto, el porcentaje total de texto considerado como agresivo encontrado en los comentarios en las noticias publicadas en el periodo enero-mayo del 2012, corresponde al 3.697% del total de texto en los comentarios.

Los resultados de la correlación que existe entre la el texto agresivo y el texto en los comentarios, se muestran en la siguiente grafica:



Como se puede observar en la grafica anterior, existe una relación entre el texto agresivo con respecto al texto en los comentarios, ya que el coeficiente de determinación R^2 da como resultado 0.6797, lo que indica una relación positiva entre ambas variables, además de que dicha relación va en aumento. Por lo tanto, en base a los resultados obtenidos, se puede inferir que el 67.9% del texto agresivo, puede deberse al incremento en la participación en los comentarios en los sitios de noticias y blogs.

CAPITULO 7.

7. Conclusiones.

La presente investigación se ha dedicado al estudio de la ciber-violencia que se presenta en los sitios de noticias y blogs en el Estado de Aguascalientes, así como la manera en la cual es posible detectarla. Con este fin fue desarrollado un sistema web mediante el cual se detectó texto considerado como agresivo en sitios de noticias y blogs del Estado, los cuales permiten la participación de los lectores o usuarios.

En el desarrollo de la tesis se recabó una serie de conclusiones relevantes, las cuales fueron obtenidas a lo largo del proceso del tema y se muestran a continuación:

- Primero, respecto al objetivo principal de la investigación, consistente en la detección de situaciones de ciber-violencia en sitios de noticias online y blogs del Estado, éste fue alcanzado plenamente, pues del análisis realizado a una muestra representativa de la población que conforma el acervo de sitios de noticias y blogs, se encontró que sí existe ciber-violencia. Dicho hallazgo fue realizado mediante la minería de contenido y comparación de texto; esto último realizado con el apoyo del catálogo de ciber agresión que fue creado para ello.
- Segundo, a través de la investigación bibliográfica se conocieron que los antecedentes de la violencia datan desde la existencia de la humanidad, puesto que por naturaleza el hombre tiende a activar mecanismos de defensa ante la existencia de una amenaza o un peligro al que este expuesto.
- Tercero, de igual manera, mediante la bibliografía se identificó que los tipos de violencia conocidos hasta ahora son:
 - Violencia social
Que a su vez, ésta se divide en: violencia física, violencia emocional, violencia verbal, no verbal, violencia sexual.
 - Violencia económica.
 - Violencia de género.

Así mismo, se encontró que otra clasificación de la violencia es: violencia directa, indirecta, represiva y alienante.

De igual manera, a través de la bibliografía revisada se nombraron y se describieron las causas de la violencia, así como los mecanismos de defensa psicológico utilizados por las personas para protegerse y se concluyó que éstos son activados dependiendo de la perturbación o excitación emocional, la cual es provocada por razones que afectan directamente el comportamiento de los individuos.

Así mismo se definió el concepto de redes sociales, así como sus antecedentes, concluyendo que las redes sociales han venido a revolucionar el concepto de comunicación en internet debido a la cantidad de herramientas que existen y el uso que se les está dando para permitir una comunicación más estrecha entre los cibernautas, además de que permiten una mayor interacción y facilidad en la comunicación entre los usuarios.

De igual manera se encontró que las herramientas que facilitan la comunicación en los sitios de noticias y blogs, son utilizadas con diferentes fines, entre ellas la ciber-violencia.

También se encontró que las herramientas de interacción que contienen los sitios de noticias y blogs no son correctamente utilizadas, y tampoco son correctamente administradas. A pesar de esto en Aguascalientes aun no se ha creado alguna ley que castigue a quienes ciber-agreden por lo que este problema puede seguir creciendo.

Se encontró que el utilizar el algoritmo de la distancia de Levenshtain puede ser una alternativa para descubrir situaciones de ciber-violencia en los sitios de noticias y blogs, en base a la comparación de palabras para detectar aquellas que son consideradas como agresivas. Cabe señalar que no fue posible realizar esta comparación tomando como base al significado de las palabras, ya que es necesaria una investigación que involucre la web semántica.

- TESIS TESIS TESIS TESIS TESIS
- Cuarto, con base a lo anterior, se detectó la existencia de ciber-violencia en los comentarios de los sitios de noticias y blogs en Aguascalientes y se determinó que la misma obedece al tipo de violencia social, manifestada como violencia verbal en su modalidad escrita.

En cuanto a las preguntas que guiaron la presente investigación, se determinó lo siguiente:

- Respecto a si es posible identificar cuáles palabras son las más utilizadas en sitios de noticias y blogs en el Estado, mediante las cuales se ejerce ciber-violencia, se concluyó que sí es posible identificarlas, pues en el caso concreto, ello se realizó con efectividad a través del instrumento de evaluación aplicado en Aguascalientes.
- En lo relativo a si existe ciber-violencia en los sitios de noticias y blogs en Aguascalientes, se concluyó que sí existe, puesto que ello se comprobó a través del Sistema de Detección de Situaciones de Ciber-Violencia, apoyado con el catálogo de agresión creado a partir del instrumento de evaluación.

De acuerdo a lo anterior, y en relación a las hipótesis formuladas en la presente investigación, se derivó lo siguiente:

- La hipótesis de investigación consistente en que los sitios web de noticias y blogs en Aguascalientes contienen ciber-violencia fue comprobada. Lo anterior se afirma ya que con la aplicación del Sistema de Detección de Situaciones de Ciber-Violencia se detectó que existen noticias publicadas que contienen comentarios, en los cuales es posible encontrar texto agresivo, sin embargo el porcentaje de texto agresivo encontrado en las noticias analizadas se encuentra dentro de lo aceptable.
- En lo que respecta a la hipótesis correlativa de causalidad referente a comprobar que a mayor participación en los comentarios de los sitios de noticias y blogs, es mayor la tendencia a generar ciber-violencia, se concluyó que la misma es afirmativa, pues existe una relación positiva entre las dos variables que conforman la misma, y que consisten en la relación entre el texto agresivo y el texto encontrado en los comentarios, de lo cual se concluye que a mayor participación de los usuarios en los comentarios de las noticias, es más amplia la tendencia a generar ciber-violencia.

7.1. Aportación de éste trabajo a investigaciones futuras.

La aportación realizada a través de la presente investigación es el análisis de los sitios de noticias online, mediante técnicas de minería web para la detección de situaciones de ciber-violencia que se presentan en los comentarios de sitios de noticias y blogs de Aguascalientes.

Dicho análisis fue realizado utilizando un catalogo definido como catalogo de agresión, el cual fue alimentado con palabras y frases identificadas como ofensivas dentro de un contexto social. La búsqueda de las situaciones de ciber-violencia fue realizada mediante un sistema web llamado Sistema de Detección de Situaciones de Ciber-violencia, el cual permitió visualizar, extraer, limpiar y analizar el contenido web para realizar los hallazgos

En relación a la aportación para investigaciones futuras, este trabajo representa la base para una nueva línea de investigación que tienda a mejorar las técnicas de minería de contenido web y análisis de texto, y que permita entender la semántica en los comentarios publicados por los usuarios de los sitios de noticias y blogs, para detectar posibles situaciones que sean potencialmente incitadoras de violencia.

GLOSARIO.

AJAX. Acrónimo de Asynchronous JavaScript And XML (Java Script asíncrono y XML), es una técnica de desarrollo web para crear aplicaciones interactivas o RIA (Rich Internet Applications).

API. Es el conjunto de funciones y procedimientos (o métodos, en la programación orientada a objetos) que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.

ASCII. Código Estándar Estadounidense para el Intercambio de Información.

BLOGGER. Es un servicio creado por Pyra Labs, y adquirido por Google en el año 2003, que permite crear y publicar una bitácora en línea.

DOC. Formato de archivo utilizado por el procesador de texto Microsoft Word.

DOM. Es una interfaz de programación de aplicaciones (API) que proporciona un conjunto estándar de objetos para representar documentos HTML y XML.

FLASH. Es un programa informático para edición multimedia.

FRAMES. Es un tag en HTML.

GIF. Es un formato gráfico utilizado ampliamente en la World Wide Web.

HTML. Siglas de HyperText Markup Language (lenguaje de marcado de hipertexto).

INTERNET. Es un conjunto descentralizado de redes de comunicación interconectadas que utilizan la familia de protocolos TCP/IP.

JPG. Es el nombre de un comité de expertos que creó un estándar de compresión y codificación de archivos de imágenes fija.

ONG. Organización No Gubernamental.

PDF. Formato de documento portátil.

RAND CORPORATION. Es un laboratorio norteamericano formado para ofrecer investigación y análisis a las fuerzas armadas norteamericanas.

RICH INTERNET APPLICATION. Son aplicaciones web que tienen la mayoría de las características de las aplicaciones de escritorio tradicionales.

SPSS. Es un programa estadístico informático muy usado en las ciencias sociales y las empresas de investigación de mercado.

BIBLIOGRAFIA.

- [1] Víctor Montoya, «Teorías de la violencia humana». [Online]. Available: <http://sincronia.cucsh.udg.mx/montoya05.htm>. [Accessed: 09-mar-2012].
- [2] Alexander Butchart, «Violence and public health: an integrated, evidence-based approach to preventing domestic violence and child abuse.», *130th international training course visiting experts' papers*.
- [3] Arun Kumar Acharya, «La Dinámica de la violencia de género en el estado de Nuevo León, México», vol. 8, pp. 255–273, 2008.
- [4] Guillermina Díaz Pérez, «La violencia de género en México: reto del gobierno y de la sociedad.», ago. 2009.
- [5] Jamil Salmi, «Violence, Democracy and Education An Analytical Framework», *Department of Human Development*.
- [6] Miquel Domènech i Argemí y Lupicinio Íñiguez Rueda, «La construcción social de la violencia», vol. 2, 2002.
- [7] Augusto Vels, «Los mecanismos de defensa bajo el punto de vista psicoanalítico.», Agrupación de Grafoanalistas Consultivos de España, 1990.
- [8] Martha Híjar Medina, María Victoria López López, y Julia Blanco Muñoz, «La violencia y sus repercusiones en la salud; reflexiones teóricas y magnitud del problema en México», *Salud Pública de México*, vol. 39, 1997.
- [9] Luis Rojas Marcos, «Las semillas de la violencia».
- [10] Guillaume Erétéo, Michel Buffa, Fabien Gandon, Patrick Grohan, Mylène Leitzelman, y Peter Sander, «A state of the art on social network analysis and its applications on a semantic web.», *SDOW2008 Germany (2008)*, feb-2011.

- [11] Gupta Siddharth y Thakur Narina, «Semantic Query Optimisation with Ontology Simulation», *International journal of Web & Semantic Technology*, vol. 1, n.º. 4, oct. 2010.
- [12] Stefano Bortoli, Paolo Bouquet, y Themis Palpanas, «Social Networking: Power to the People».
- [13] Joan Merino Gonzalez, «El acoso escolar-bullying», *Revista de estudios de la violencia*, mar-2008.
- [14] Verena Kantere, Maria Eirini, y Timos Sellis, «Conceptual Synopses of Semantics in Social Networks Sharing Structured Data», *Computer Science*, vol. 5332/2008, pp. 1367–1384, 2008.
- [15] Gayatri Swamynathan, Christo Wilson, Bryce Boe, Kevin Almeroth, y Ben Y. Zhao, «Do Social Networks Improve e-Commerce? A Study on Social Marketplaces», *WOSN'08*, ago. 2008.
- [16] Luz María Velázquez Reyes, «Ciberbullying. El crudo problema de la victimización en línea.», *X congreso nacional de investigación educativa*, 2010.
- [17] Bernhard Debatin, Jennette P. Lovejoy, Ann-Kathrin Horn, M.A., y Brittany N. Hughes, «Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences», *Journal of Computer-Mediated Communication*, pp. 83–108, 2009.
- [18] Ramón Sánchez Carballido, «Perspectives of the Information in Internet: Ciberdemocracy, Social Networks and Semantic Web.», vol. 13, n.º. 25, pp. 61–81, nov. 2008.
- [19] Ralph Gross y Alessandro Acquisti, «Information Revelation and Privacy in Online Social Networks (The Facebook case)», *Workshop on Privacy in the Electronic Society*, 2005.
- [20] Gerardo García, Víctor Manuel Joffre, Gerardo Jesús Martínez, y Arturo Llanes, «Ciberbullying: forma virtual de intimidación escolar», *Rev. Colomb. Psiquiat*, vol. 40, n.º. 1, pp. 115–130, 2011.

- [21] Tiscar Lara, «Weblogs y Periodismo Participativo». [Online]. Available: http://www.tiscar.com/wp-content/Tiscar_Lara_Weblogs_Pparticipativo.pdf. [Accessed: 03-may-2012].
- [22] Jordis Adell, «Wikis en la educación».
- [23] Jesús Alberto Andrade, «Wikipedia: una experiencia mundial de trabajo colaborativo», *Revista Venezolana de Información, Tecnología y Conocimiento*, pp. 81–86, ago-2005.
- [24] Flores Cueto, Juan José, Morán Corzo, Jorge Joseph, y Rodríguez Vila, Juan José, «Las Redes Sociales», *Enlace Virtual*, vol. 1, oct. 2009.
- [25] Calvimontes, Jorge, *El Periódico*, 3.^a ed. México: , 1990.
- [26] M.J. Pérez-Luque y M. Perea, «El Actual Periodismo Online».
- [27] Jeremy B Williams y Joanne Jacobs, «Exploring the use of blogs as learning spaces in the higher education sector», *Australasian Journal of Educational Technology*, pp. 232–247, 2004.
- [28] Bamshad Mobasher, Namit Jain, Eui-Hong(Sam) Han, y Jaideep Srivastava, «Web Mining Pattern Discovery from World Wide Web». 08-mar-1997.
- [29] Ricardo Baeza-Yates, «excavando la web», *El profesional de la informacion*, vol. 13, pp. 4–10, 2004.
- [30] Sergey Brin y Lawrence Page, «The Anatomy of a Large-Scale Hypertextual Web Search Engine», pp. 107–117, 1998.
- [31] Antonio González Torres, «Minería web y personalización: Revisión bibliográfica y propuesta de un marco de referencia».
- [32] Yury Yineth Niño Roa, «Minería de textos médicos web un enfoque hacía la clacificación en categorías.», *Seminario de Investigación*.

- [33] I. Esquivel, «Tratamiento automático de noticias empresariales en la Web, mediante minería textual», *IEEE 4º Congreso Internacional en Innovación y Desarrollo Tecnológico*, 13 octubre 2006.
- [34] Manuel Montes-y-Gómez, Alexander Gelbukh, y Aurelio López-López, «Mining the News: Trends, Associations, and Deviations», 16 2001.
- [35] José E. Medina Pagola, «Estado del Arte del Web Mining», ago. 2007.
- [36] Richard S. Segall y Qingyu Zhang, «Web Mining of Hotel Customer Survey Data».
- [37] Y-F Huang y Zau-NanChen, «Arrangement of traveling plans and web mining of a traveling system», 2001.
- [38] Jamie Murphy, Charles F. Hofacker, y Michelle Bennett, «Website-generated market-research data: tracing the tracks left behind by visitors», vol. 42, pp. 82–91.
- [39] Arno Scharl, Karl W. Wober, y Christian Bauer, «An integrated approach to measure web site Effectiveness in the european hotel industry», *Information Technology & Tourism*, vol. 6, pp. 257–271, 2004.
- [40] Luis Antonio Lucio López, «Agresores escolares en el ciberespacio; el cyberbullying en preparatorias mexicanas», *XI Asamblea General de ALAFEC*, 2009.
- [41] Rishin Haldar y Debajyoti Mukhopadhyay, «Levenshtein Distance Technique in Dictionary Lookup Methods: An Improved Approach».

ANEXO A.

Tabla comparativa de los sitios de noticias online y sus secciones.

PERIODICO ONLINE	NOTICIAS LOCALES	SOCIALES	CIENCIA Y TECNOLOGÍA	CULTURA	DEPORTES	EDUCACIÓN	OTROS
Aguas Digital	*Aguascalientes	*Tendencias		*Tendencias	*El Marcador		*Las Plumas *Actualidad
Blog Aguascalientes	*Fecha *Etiqueta						
Clarinete	*Noticias Locales	*Escenarios		*Artes	*Gramma y Arena		*Inicio *Clarinetazo *En voz de *Contacto
Crisol	*Aguascalientes		*Economía	*Ensayo y Opinión	*Deportes	*Ensayo y Opinión	*Inicio *Ocio *Actualidad
Grilla	*Gobierno del Estado		*Desarrollo Rural... *Economía y Negocios *Ciencia y Tecnología *Medio Ambiente	*Arte, Cultura y Espectáculos	*Salud y Deportes		*Religión *Sindicalismo y Corporativismo
Jornada	*Portada		*Economía	*Cultura	*Deportes	*Educación	*Política *Sociedad y Justicia *Municipios
Noticiero El Circo	*Aguascalientes		*Política	*Espectáculos	*Deportes		*Zacatecas *Jalisco *Última Hora
Picacho	*Ciudad *Familia *Tiempo libre *Recreo		*Familia	*Tiempo libre *Familia	*Tiempo libre *Familia		*Ciudad *Familia *Tiempo libre *Recreo
Revista Cambio		*sociedad		*Cultura y Espectáculos	*Deportes	*Académica	*Salud *Hojas sueltas *Pórtico *Editorial *Portada

Especificación de Requisitos de Software.

Información de la versión	
Proyecto:	Sistema de Detección de Situaciones de Ciber-Violencia.
Número Interno de Versión:	1.0
Fecha:	1 de Mayo del 2012.

INTRODUCCIÓN.

El presente documento especifica los requisitos de software del sistema que se desarrolla: “El sistema de detección de situaciones de ciber-violencia”. Los objetivos de este documento son especificar, identificar, validar y documentar los requisitos de software, es decir determinar las características que deberá tener el sistema, así como las restricciones que deberá cumplir para que sea aceptado por los usuarios.

En el análisis detallado del sistema se utilizarán diagramas gráficos UML, así como la descripción en lenguaje natural de las mismas. La finalidad de los diagramas es ayudar a entender el funcionamiento completo del software y como utilizarlo.

DISEÑO DEL SISTEMA.

Para el análisis y diseño nos auxiliaremos del lenguaje UML. Específicamente, para modelar los requisitos funcionales del sistema utilizaremos los diagramas de casos de uso. Por lo que se realizara una descripción detallada de los requisitos tanto funcionales como no funcionales.

VISIÓN INICIAL DEL SISTEMA.

Encontrar situaciones de ciber-violencia en los contenidos de noticias y blogs implica definir un vocabulario. Para esto es necesario detallar cuáles son las palabras consideradas como ofensivas dentro del contexto en el que se va a poner a prueba el sistema, por lo que es necesario definir el concepto de *catalogo de agresión*.

El catalogo de agresión.

El catalogo de agresión es un conjunto de datos individuales o compuestos obtenidos mediante una encuesta, la cual pretende concentrar todo aquel vocabulario definido por los usuarios de los sitios de noticias y blogs, como vocabulario ofensivo. El contenido de este catalogo tiene una definición dependiendo del contexto en el que se utilice. Por lo que es necesario documentar que únicamente será utilizado con fines de investigación.

ACTORES DEL SISTEMA.



DIAGRAMA DE GRAFOS DE ACTORES.

Detalle de actores.

Actor	Descripción
Usuario Registrado	Actor que representa un usuario registrado por el sistema e interactúa con él. Puede interactuar en los módulos del sistema dependiendo del nivel de acceso al que se tenga autorizado. El usuario registrado interactúa con el sistema a través del teclado y de la pantalla donde se muestra el interfaz del sistema hacia el

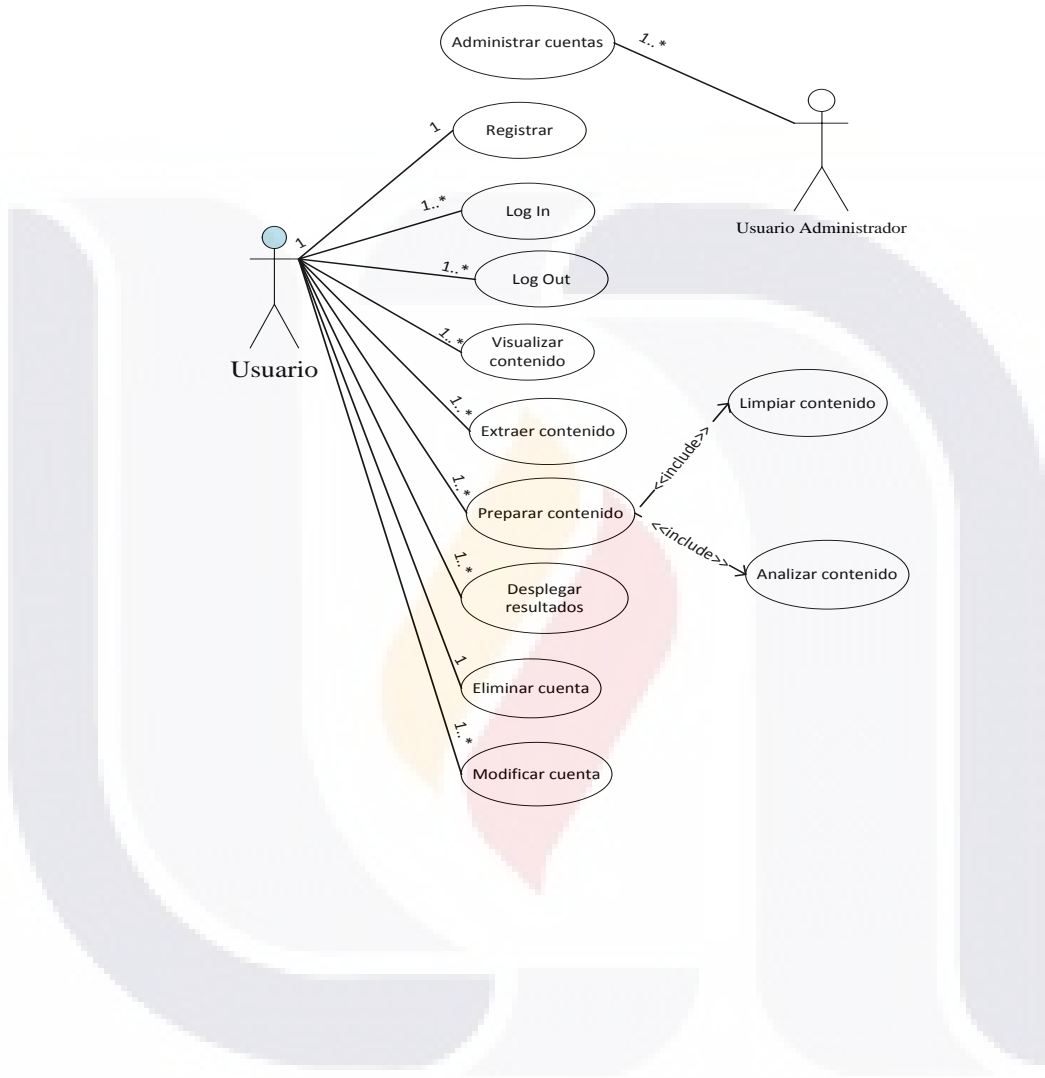
usuario.

Usuario Administrador

Se trata de un usuario registrado que tiene permisos especiales dentro del sistema. Este usuario puede crear modificar y eliminar a los usuarios registrados. El administrador es el primer usuario del sistema y que posee una clave especial. Interactúan con el sistema a través de un teclado o un mando y una pantalla donde se muestra el interfaz del sistema.



CASOS DE



USO.

DIAGRAMA DE CASOS DE USO

DESCRIPCIÓN DE LOS CASOS DE USO DEL SISTEMA DE DETECCIÓN DE SITUACIONES.

Id	RegUsuario	
Caso de uso	Registrar	
Actor	Usuario	
Descripción	La primera vez que una persona desea hacer uso del sistema necesita registrarse. Para ello el sistema solicitará los datos necesarios del usuario y este los introducirá a través de un teclado. El registro será unívoco.	
Precondición	Se cuente con los datos de registro.	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	2.-El usuario no registrado, completamente desconocido para el sistema decide acceder. 4.- El usuario al no estar dado de alta, debe crear un nuevo registro de usuario en el sistema. 6.-El usuario proporciona dicha información al sistema.	1.-El sistema se encuentra en estado de espera de ingreso de un usuario. 3.-El sistema responde preguntando por los datos del usuario. 5.-El sistema solicita la información necesaria para crear el registro. 7.-El sistema valida la información y crea el nuevo registro, notificando al usuario su alta en el sistema.
Extensiones	1.-Si los datos introducidos por el usuario no son correctos, es decir introduce caracteres no validos o por dejar campos vacios, el sistema deberá solicitar los datos de registro faltantes. 2.-Si el usuario no captura los datos completos y cierra o se cierra la aplicación, no se llevara el registro de usuario. 3.- El administrador podrá dar de alta a los usuarios.	

Id	LogInUsuario	
Caso de uso	Log In	
Actor	Usuario	
Descripción	El usuario debe autenticarse ante el sistema para acceder a los servicios que el mismo le ofrece. Para ello debe estar registrado dentro del sistema. Una vez realizada la autenticación, se abre una sesión que se mantiene hasta el momento en que el usuario decide abandonar el sistema.	
Precondición	El usuario se encuentre registrado en el sistema.	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	1.-El usuario introduce los datos necesarios para identificarse.	2.- El sistema valida al usuario. 3.-El sistema le permite alcanzar todas las funcionalidades a las que puede acceder según los permisos de el usuario autenticado
Extensiones	1.-Los datos introducidos no son correctos y el sistema no valida al usuario el cual debe informar con un mensaje de error, y quedarse en el estado anterior. 2.-Los datos introducidos no son correctos. También se debe informar al usuario.	

Id	ElimCuenta	
Caso de uso	Eliminar cuanta	
Actor	Usuario	
Descripción	El usuario desea eliminar su registro en el sistema.	
Precondición	El usuario se encuentre registrado en el sistema.	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	1.-Un usuario registrado desea darse de baja por completo en el sistema para no volver a utilizarlo.	2.-El sistema elimina información relacionadas con el usuario y se lo notifica. 3.-El sistema elimina el registro de este usuario y le hace salir del sistema, terminando así la sesión en curso.
Extensiones	Si el usuario es un administrador de sistema: 1. El usuario es un administrador y desea dar de baja a otro usuario. 3. El sistema permite elegir el usuario a dar de baja. 4. El administrador elige el usuario a dar de baja. 5. El sistema da de baja al usuario seleccionado siguiendo la misma secuencia que en el flujo principal.	

Id	ModCuenta	
Caso de uso	Modificar cuanta	
Actor	Usuario	
Descripción	El usuario desea Modificar su registro de cuenta en el sistema.	
Precondición	El usuario se encuentre registrado en el sistema.	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	<p>1.-Un usuario registrado desea modificar su información en el sistema.</p> <p>3.-El usuario modifica la información.</p> <p>5.- Confirma la acción de cambio de información.</p>	<p>2.-El sistema comprueba la información del usuario.</p> <p>4.-Despliega notificación de cambio de información del usuario.</p> <p>6.- El sistema sustituye la información modificada.</p> <p>7.-Se notifica al usuario el éxito de cambio de información.</p>
Extensiones	Si el usuario no desea cambiar su información podrá cancelar la transacción en cualquier momento.	

Id	LogOutUsuario	
Caso de uso	Log Out	
Actor	Usuario	
Descripción	El usuario del sistema se encuentra con una sesión abierta y debe cerrarla. Esta posibilidad debe estar disponible en cualquier momento al usuario, y debe realizarse sin pérdida de información	
Precondición	El usuario se encuentre registrado en el sistema.	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	1.-El usuario selecciona la opción de cerrar la sesión.	2.-El sistema realiza las operaciones internas necesarias para evitar la pérdida de información. 3.-El sistema cierra la sesión con el usuario y se lo notifica. 4.-El sistema vuelve al estado inicial de espera de autenticación o registro de otro usuario.
Extensiones	1.-Si por algún motivo el usuario no puede cerrar la sesión, el sistema deberá indicar al usuario.	

Id	VisContenido	
Caso de uso	Visualizar contenido	
Actor	Usuario	
Descripción	El usuario desea visualizar el contenido de una página web	
Precondición	Se tenga la URL del sitio que se desea visualizar el contenido	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	1.-Introduce la URL del sitio de noticias o blog del cual se desea visualizar el contenido. 2.-Presiona el botón iniciar para comenzar la visualización.	3.-Realiza la lectura del sitio seleccionado.

		4.-Despliega el sitio en el visualizador. 5.-Notificación de éxito o error al término de la operación.
Extensiones	1.-El sitio a analizar no existe, el sistema debe desplegar un mensaje con información del mismo.	

Id	ExtContenido	
Caso de uso	Extraer contenido	
Actor	Usuario	
Descripción	El usuario desea extraer el contenido de la noticia publicada.	
Precondición	El sitio contenga comentarios textuales generados por los usuarios.	
Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	1.- El usuario presiona el botón “extraer contenido”.	2.-Realiza la extracción del contenido que se encuentra entre las etiquetas HTML o XML de la noticia. 3.- Despliega el contenido extraído de la noticia y lo deberá de presentar en un iframe para que el usuario realice una verificación visual rápida. 4.-Desplegara una notificación de termino de extracción de contenido
Extensiones	Si el texto no se puede extraer el sistema deberá presentar un mensaje notificando al usuario la falla en la extracción de contenido.	

Id	PreContenido
Caso de uso	Preparar contenido
Actor	Usuario
Descripción	El usuario desea preparar el contenido para poder realizar el análisis.
Precondición	Se haya realizado la visualización y extracción del contenido.

Secuencia de interacción entre actor y sistema	ACTOR	SISTEMA
	<p>1.-Presiona el botón “limpiar contenido” para eliminar todos los posibles caracteres que puedan ser no necesarios para el análisis.</p> <p>3.- Presiona el botón “analizar contenido” para realizar la búsqueda y comparación de texto agresivo.</p>	<p>2.-Realiza la limpieza del contenido eliminando los caracteres no necesarios.</p> <p>4.-Realiza el análisis para identificar los textos agresivos que se puedan encontrar en el contenido.</p> <p>5.-Desplegara notificación de termino de análisis.</p>
Extensiones	En la limpieza de contenido debe eliminar todos los caracteres no necesarios incluyendo números, si no encuentra ningún carácter de este tipo el sistema no realizara nada.	

Id	DesResultado	
Caso de uso	Desplegar resultados	
Actor	Usuario	
Descripción	El usuario desea visualizar el resultado general del análisis generado sobre la noticia o blog, con la finalidad de presentar los totales encontrados.	
Precondición	Se haya realizado la limpieza y el análisis del contenido de la noticia para detectar ciber-violencia.	
Curso normal	ACTOR	SISTEMA
	1.-Presiona el botón “mostrar resultados”	2.-Presenta los resultados del análisis.
Pos condición	Si el sistema no detecto ninguna situación de ciber-agresión se desplegara notificación con los resultados.	

REQUISITOS FUNCIONALES DEL SISTEMA.

El sistema a desarrollar deberá cumplir los requisitos indicados a continuación:

RF01 -Registrar usuarios, Estatus – Implementar.

Descripción: El sistema de detección de situaciones debe registrar a cualquier usuario que quiera hacer uso de los servicios ofrecidos, es decir, debe dar de alta a los usuarios para que estos puedan acceder a los servicios. Un usuario registrado es un usuario final real que ha creado un registro en el sistema y puede identificarse y firmarse en él.

RF02 -Autenticar usuario (*log in*), Estatus – Implementar.

Descripción: El sistema debe autenticar a los usuarios registrados antes de que puedan acceder a su dashboard (o al de otros en caso de ser administradores de sistema). En el caso de que un usuario sea completamente nuevo para el sistema debe registrarse antes de acceder a las funcionalidades del mismo. Si un usuario se autentifica correctamente, abre una sesión con el sistema.

RF03 -Cerrar sesión de usuario (*log out*), Estatus – Implementar

Descripción: Los usuarios deben poder abandonar el sistema en cualquier momento.

RF04-Eliminar cuenta, Estatus – Implementar.

Descripción: De la misma manera que se crean registros de nuevos usuarios, deberá ser posible dar de baja a éstos. Esto implica la eliminación de cualquier registro asociado a ese usuario y de toda la información relativa al registro. Un usuario ordinario puede darse de baja a sí mismo. Un administrador puede dar de baja a cualquier usuario del sistema.

RF05- Visualización del sitio, Estatus – Concluido.

Descripción: El sistema de detección de situaciones deberá permitir ingresar una URL en un componente “Text box” de HTML para posteriormente obtener la información a partir de un parseo de texto en el sitio web que deberá realizarse en modo online, una vez ingresado, deberá permitir visualizar el contenido del sitio el cual se quiere analizar. El despliegue deberá hacerse en un iframe dentro de la misma ventana, para no perder el foco en el objeto del análisis.

RF06 - Extraer de contenido, Estatus – Concluido.

Descripción: El sistema deberá extraer y almacenar en una base de datos el contenido de la página web mediante un parseo de datos. El contenido que se almacenara en la base de datos será el texto y los vínculos encontrados.

RF07 – Limpiar contenido, Estatus – Concluido.

Descripción: El sistema realizara una limpieza de datos donde deberá eliminar todo aquel carácter que no sea de utilidad para el análisis.

Los caracteres que deberán ser eliminados se muestran a continuación:

¡ °~?#\$%()=?¡¿?~*+[{]},;-_...~\<>`^^^&(.)[^;]*;/,0123456789
--

RF08 – Preparar contenido, Estatus – Implementar.

Descripción: El sistema deberá identificar los siguientes datos

- *Nombre del sitio:* nombre del sitio y URL del contenido a analizar.
- *Fecha de visita:* fecha de visita del sitio que se está analizando.
- *Título de la noticia:* título de la noticia que se está analizando.
- *Híperuxtualidad:* vínculos hacia otros sitios.
- *Texto del contenido:* palabras detectadas en el sitio analizado.

Los datos deberán ser almacenados en la base de datos llevando un:

- Identificador.

- Fecha de análisis.
- URL del sitio.
- Título de noticia.
- Texto del sitio.

RF09 – Análisis del contenido, Estatus – Implementar.

Descripción: El sistema deberá realizar un análisis del contenido en el cual deberá comparar cada una de las palabras almacenadas en el catálogo de agresión con las palabras extraídas en el parseo de contenido. El algoritmo de comparación deberá hacerse utilizando la distancia de Levensthein. El análisis del contenido deberá realizarse una vez aplicado la limpieza de datos para eliminar los caracteres especiales que no sean de utilidad.

RF10 – Despliegue de resultados, Estatus – Concluido.

Descripción: El sistema desplegara el grado de agresión que se obtuvo en el análisis de contenido, el cual deberá ser almacenado en la base de datos para que permita posteriormente realizar nuevo análisis sobre el mismo sitio y validar si existe incremento de agresión en el contenido en algún rango de fechas.

REQUISITOS NO FUNCIONALES DE LA APLICACIÓN.

En este apartado se incluyen los requisitos no funcionales de los casos de uso descritos anteriormente.

Requisitos de calidad.

1.-Fiabilidad:

- *Recuperación completa del texto.* Asegurar que se extraiga todo el texto contenido en la página web y asegurar que no se pierdan los datos. Esto incluye no perder la información en el envío para análisis al servidor, como manejo de errores entre las transacciones en el servidor.
- *Fiabilidad en las transacciones del sistema.*

2.- Fiabilidad de uso:

- Existencia de una interfaz (GUI) sencilla y amigable de fácil navegación.
- Existencia de un manual de usuario escrito para describir el funcionamiento y el uso del sistema al usuario final.

3.- Soporte:

- Facilidad de mantenimiento, lo que requiere código y diseño documentado.
- Facilidad de actualización hacia nuevas funcionalidades.

4.- Información:

- En todo momento debe informarse al usuario de todas las acciones que puede realizar en el sistema.

Restricciones del sistema.

1. Implementación:

- El desarrollo del sistema se realizará utilizando el paquete XAMPP for Windows Versión 1.7.5 el cual contiene en framework PHP versión 5.3.5, Perl 5.10, J2EE y utilizando para ello el entorno de desarrollo Netbeans IDE 7.0.
- Manejado de base de datos MySQL Version 5.5.
- Interface de desarrollo de base de datos MySQL Workbench 5.2 CE
- La implementación se realizará con fines de desarrollo sobre un terminal de tipo laptop con las siguientes características:
 - Windows 7 Home Premium 32 bits.
 - Procesador Intel® core(TM)2 Duo a 2.40 GHz
 - Memoria RAM 4 GB.
 - HD 320 GB.
- La aplicación deberá ser accedida utilizando el navegador Mozilla Firefox Versión 8+.

ANEXO C.

Listado de noticias analizadas.

Sitio de noticias Aguasdigital.com

Titulo de la noticia	URL
Delatan desvios millonarios en dependencias federales	http://aguasdigital.com/metro/leer.php?idnota=35154
En alerta empresarios del ramo automotriz por subida del dólar	http://aguasdigital.com/metro/leer.php?idnota=35153
Abren vialidades pero no concluyen labores	http://aguasdigital.com/metro/leer.php?idnota=35152
Se llena de tráfico el primer anillo	http://aguasdigital.com/metro/leer.php?idnota=35151
Extienden cierre en H. de Nacozari	http://aguasdigital.com/metro/leer.php?idnota=35150
Confunde semáforo a automovilistas de Av. de Los Maestros	http://aguasdigital.com/metro/leer.php?idnota=35149
Vigila Condusef a las casas de empeño	http://aguasdigital.com/metro/leer.php?idnota=35148
Dependen de las lluvias 150 mil familias campesinas	http://aguasdigital.com/metro/leer.php?idnota=35147
En combate a suicidio urge poner especial atención a jóvenes: SS	http://aguasdigital.com/metro/leer.php?idnota=35141
Diarrea y deshidratación segunda causa de muerte en menores de 5 años	http://aguasdigital.com/metro/leer.php?idnota=35116
Obras en avenida Aguascalientes también desquician fraccionamientos	http://aguasdigital.com/metro/leer.php?idnota=35115
Avanza Aguascalientes en consolidación de Estado Verde	http://aguasdigital.com/metro/leer.php?idnota=35114
12 años edad de inicio en consumo de alcohol: CIJ	http://aguasdigital.com/metro/leer.php?idnota=35113
Sector obrero ya se gastó sus utilidades: CTM	http://aguasdigital.com/metro/leer.php?idnota=35112
Pide Agropecuario más presencia a las Policías	http://aguasdigital.com/metro/leer.php?idnota=35111
Establecen priistas reunión en CMIC	http://aguasdigital.com/metro/leer.php?idnota=35110
También rectores avalan el endeudamiento	http://aguasdigital.com/metro/leer.php?idnota=35108
Crítica JLM declaraciones del dirigente de Canacintra	http://aguasdigital.com/metro/leer.php?idnota=35107
Suman ya 44 suicidios en el 2012	http://aguasdigital.com/metro/leer.php?idnota=35094
Candidatos del PANAL visitan SJG	http://aguasdigital.com/metro/leer.php?idnota=35093
Estudiantes celebran su día	http://aguasdigital.com/metro/leer.php?idnota=35092
Felicita CLT a estudiantes en su día	http://aguasdigital.com/metro/leer.php?idnota=35091
Aguascalientes consolida su posición para inversiones internacionales	http://aguasdigital.com/metro/leer.php?idnota=35090
Comisión de Vigilancia da entrada a petición de endeudamiento	http://aguasdigital.com/metro/leer.php?idnota=35089
Esperan recuperarse comerciantes de Madero de la caída en ventas por la obra	http://aguasdigital.com/metro/leer.php?idnota=35087
Temen vecinos de la Guadalupe Peralta por la construcción de gasera	http://aguasdigital.com/metro/leer.php?idnota=35086
Lamenta AGG que trabajadores se conformen con salarios inferiores al mínimo	http://aguasdigital.com/metro/leer.php?idnota=35085
México no oculta cifras sobre inseguridad: FRR	http://aguasdigital.com/metro/leer.php?idnota=35067
Es hora de trabajar por los jóvenes: SLR	http://aguasdigital.com/metro/leer.php?idnota=35066
Avanzan obras de 3er Anillo de Circunvalación	http://aguasdigital.com/metro/leer.php?idnota=35065
Capacita DIF Estatal en prevención del suicidio	http://aguasdigital.com/metro/leer.php?idnota=34935
Finanzas equilibradas y austeras de Aguascalientes generan perspectivas positivas: CMB	http://aguasdigital.com/metro/leer.php?idnota=34934
Avanza programa Transporte Verde	http://aguasdigital.com/metro/leer.php?idnota=34744
Llama PANAL a los electores a participar	http://aguasdigital.com/metro/leer.php?idnota=34719
Listo sector obrero para desfile del 1 de Mayo	http://aguasdigital.com/metro/leer.php?idnota=34670
Asegura AGE que funcionario municipal tiene amenazada empresa	http://aguasdigital.com/metro/leer.php?idnota=34612
MRM se pronunció en favor de reforma energética	http://aguasdigital.com/metro/leer.php?idnota=34586
Incrementar presupuesto para educación: Paty Muñoz	http://aguasdigital.com/metro/leer.php?idnota=34585
Improvisan estacionamientos en el perímetro ferial	http://aguasdigital.com/metro/leer.php?idnota=34528
Sólida la alianza de GPPAN y GE	http://aguasdigital.com/metro/leer.php?idnota=34299
Pagan a profesores jubilados prima de antigüedad	http://aguasdigital.com/metro/leer.php?idnota=34262
Aumenta amibiasis en menores	http://aguasdigital.com/metro/leer.php?idnota=34168
Apremiante, renovar esquemas de salud: MRM	http://aguasdigital.com/metro/leer.php?idnota=34142
Siguen abusos por parte de "guarros" en entros	http://aguasdigital.com/metro/leer.php?idnota=34063
Continuarán actividades del Municipio	http://aguasdigital.com/metro/leer.php?idnota=34016
IMSS sancionará y boletinará a contratistas que incumplieron en HGZ2 y 3	http://aguasdigital.com/metro/leer.php?idnota=33902
Aguascalientes de los mejores 5 en servicios periciales: PGJE	http://aguasdigital.com/metro/leer.php?idnota=33867
Licencia de funcionario estatales solo por la campaña: CLT	http://aguasdigital.com/metro/leer.php?idnota=33866
3 mil casas afectadas por fallas: OLV	http://aguasdigital.com/metro/leer.php?idnota=33808
CLT inauguró la Segunda Feria Estatal de Orientación Vocacional	http://aguasdigital.com/metro/leer.php?idnota=33770
10 boletos de visita del Papa para Aguascalientes	http://aguasdigital.com/metro/leer.php?idnota=33695
Agradece CLT apoyo del gobernador del Banco de México	http://aguasdigital.com/metro/leer.php?idnota=33644
Aumentan vigilancia en RR	http://aguasdigital.com/metro/leer.php?idnota=33533
Sella SS establecimiento de venta de medicamentos a granel: FEP	http://aguasdigital.com/metro/leer.php?idnota=33462
Realiza personal del HH reimplante de mano	http://aguasdigital.com/metro/leer.php?idnota=33440
Llama Reyes Velázquez a panistas a que ya se pongan a trabajar	http://aguasdigital.com/metro/leer.php?idnota=33433
Nuevo vuelo Aguascalientes - Cancún	http://aguasdigital.com/metro/leer.php?idnota=33379
Prioritaria infraestructura carretera en municipios del interior: SICOM	http://aguasdigital.com/metro/leer.php?idnota=33378
Prepara SEBIDESO blindaje electoral	http://aguasdigital.com/metro/leer.php?idnota=33331
Censura CCEA cambios en delegaciones federales	http://aguasdigital.com/metro/leer.php?idnota=33327
Poco a poco, van cediendo niñas en el oriente	http://aguasdigital.com/metro/leer.php?idnota=33306
Desmiente ARA uso político de cuentas publicas	http://aguasdigital.com/metro/leer.php?idnota=33257
Caso MOS todavía no esta cerrado: JCRR	http://aguasdigital.com/metro/leer.php?idnota=33256
En Aguascalientes PVEM solo respaldará al PRI	http://aguasdigital.com/metro/leer.php?idnota=33123
Michoacán y Aguascalientes estrechan vínculos	http://aguasdigital.com/metro/leer.php?idnota=33121
Presentan auditoria a Ccapama	http://aguasdigital.com/metro/leer.php?idnota=33102
A la alza divorcios en Aguascalientes: INEGI	http://aguasdigital.com/metro/leer.php?idnota=33094
Que la FEPADE no persiga solo a candidatos de oposición al gobierno federal: PRI	http://aguasdigital.com/metro/leer.php?idnota=33082
Suman esfuerzos gobiernos Federal y Municipal para acercar agua a localidades rurales	http://aguasdigital.com/metro/leer.php?idnota=33081

Listado de noticias analizadas.

Sitio de noticias Noticieroelcirco.com

Titulo de la noticia	URL
Ministeriales de Aguascalientes abaten a "el c	http://noticieroelcirco.blogspot.mx/2012/05/ministeriales-de-aguascalientes-abaten.html
Fallece Nuestro Entrañable Amigo Rubén Neg	http://noticieroelcirco.blogspot.mx/2012/05/fallece-nuestro-entranable-amigo-ruben.html
se debe trabajar por el bienestar de Aguascalie	http://noticieroelcirco.blogspot.mx/2012/05/se-debe-trabajar-por-el-bienestar-de.html
separemos la lucha sindical de la toma de decis	http://noticieroelcirco.blogspot.mx/2012/05/separemos-la-lucha-sindical-de-la-toma.html
desencanta enrique iglesias a sus fans en Agu	http://noticieroelcirco.blogspot.mx/2012/05/desencanta-enrique-iglesias-sus-fans-en.html
las mujeres son capaces de administrar hasta e	http://noticieroelcirco.blogspot.mx/2012/05/las-mujeres-son-capaces-de-administrar.html
La actriz Yadira Carrillo recibe "Pitos" y le grita	http://noticieroelcirco.blogspot.mx/2012/05/la-actriz-yadira-carrillo-recibe-pitos.html
Carlos Lozano, Gobernador del Estado recibe	http://noticieroelcirco.blogspot.mx/2012/05/carlos-lozano-gobernador-del-estado.html
El crimen organizado apoya las candidaturas d	http://noticieroelcirco.blogspot.mx/2012/05/el-crimen-organizado-apoya-las.html
El crimen organizado apoya las candidaturas d	http://noticieroelcirco.blogspot.mx/2012/05/el-crimen-organizado-apoya-las.html
Debate lo ganó Josefina Vázquez Mota: Toño	http://noticieroelcirco.blogspot.mx/2012/05/debate-lo-gano-josefina-vazquez-mota.html
Soy una candidata aguerrida y quiero ser la vo	http://noticieroelcirco.blogspot.mx/2012/05/soy-una-candidata-aguerrida-y-quiero.html
Presuntos "Zetas", atrás de la reventa en la F	http://noticieroelcirco.blogspot.mx/2012/05/presuntos-zetas-atras-de-la-reventa-en.html
PROPUESAS DE EPN EN EL DEBATE SON	http://noticieroelcirco.blogspot.mx/2012/05/propuestas-de-epn-en-el-debate-son.html
Reformar la Ley para la prevención y gestión d	http://noticieroelcirco.blogspot.mx/2012/05/reformar-la-ley-para-la-prevencion-y.html
125 mil aguascalentenses se beneficiarán con	http://noticieroelcirco.blogspot.mx/2012/05/125-mil-aguascalentenses-se.html
http://noticieroelcirco.blogspot.mx/2012/05/cla	http://noticieroelcirco.blogspot.mx/2012/05/claudio-gonzalez-acude-debate-en.html
ley de Protección contra Periodistas", no es su	http://noticieroelcirco.blogspot.mx/2012/05/ley-de-proteccion-contr-periodistas-no.html
Se requiere una reforma laboral que no polanic	http://noticieroelcirco.blogspot.mx/2012/05/se-requiere-una-reforma-laboral-que-no_05.html
La educación elemento fundamental para vivir	http://noticieroelcirco.blogspot.mx/2012/05/la-e-educacion-elemento-fundamental-para.html
SE MATAN DOS PARAMEDICOS DE LA CRUZ	http://noticieroelcirco.blogspot.mx/2012/05/se-matan-dos-paramedicos-de-la-cruz.html
Les quitan una quincena en Sagarpa para las c	http://noticieroelcirco.blogspot.mx/2012/05/les-quitan-una-quincena-en-sagarpa-para.html
El incremento de la red universitaria que propo	http://noticieroelcirco.blogspot.mx/2012/04/el-incremento-de-la-red-universitaria.html
Legislar a favor del campo propone Salvador F	http://noticieroelcirco.blogspot.mx/2012/04/legislar-favor-del-campo-propone.html
En Aguascalientes, Peña firma promesas ante	http://noticieroelcirco.blogspot.mx/2012/04/en-aguascalientes-pena-firma-promesas.html
CAPTURAN A "EL MANE", ASESINO Y CARBONIZO LA	http://noticieroelcirco.blogspot.mx/2012/04/capturan-el-mane-asesino-y-carbonizo-la.html
Es necesario que prevalezca el estado de derec	http://noticieroelcirco.blogspot.mx/2012/04/es-necesario-que-prevalezca-el-estado.html
Los diputados deben otorgar mayores recurso	http://noticieroelcirco.blogspot.mx/2012/04/los-diputados-deben-otorgar-mayores.html
Gregorio Zamampa nefasto y desobligado coo	http://noticieroelcirco.blogspot.mx/2012/04/gregorio-zamampa-nefasto-y.html
Seré el Senador de los municipios y de las muj	http://noticieroelcirco.blogspot.mx/2012/04/sere-el-senador-de-los-municipios-y-de.html
Paty Muñoz trabajará en el Congreso de la Uni	http://noticieroelcirco.blogspot.mx/2012/04/paty-munoz-trabajara-en-el-congreso-de.html
CAPTURAN A PADRES INFERNALES; "EL F	http://noticieroelcirco.blogspot.mx/2012/04/capturan-padres-infemales-el-padre.html
Elementos de seguridad "Omega" vuelven a h	http://noticieroelcirco.blogspot.mx/2012/04/elementos-de-seguridad-omega-vuelven.html
Programas urgentes para apoyar madres solter	http://noticieroelcirco.blogspot.mx/2012/04/programas-urgentes-para-apoyar-madres.html
"Soy un candidato a Senador por convicción,	http://noticieroelcirco.blogspot.mx/2012/04/soy-un-candidato-senador-por-conviccion.html
Recibirán miles de maestros en activo y jubilac	http://noticieroelcirco.blogspot.mx/2012/04/recibiran-miles-de-maestros-en-activo-y.html
SIN ESCUELA DE TIEMPO COMPLETO LA E	http://noticieroelcirco.blogspot.mx/2012/04/sin-escuela-de-tiempo-completo-la.html
Programas urgentes para apoyar madres solter	http://noticieroelcirco.blogspot.mx/2012/04/programas-urgentes-para-apoyar-madres.html
Jesús Santillán ya "domesticó" a sus regidore	http://noticieroelcirco.blogspot.mx/2012/03/jesus-santillan-ya-domestico-sus.html
35 años de cárcel para cuatro secuestradores d	http://noticieroelcirco.blogspot.mx/2012/03/35-anos-de-carcel-para-cuatro.html
Gran número de panistas se sumarán a la izqui	http://noticieroelcirco.blogspot.mx/2012/03/gran-numero-de-panistas-se-sumaran-la.html
Ana Cristina Wheelock López es la Reina de la	http://noticieroelcirco.blogspot.mx/2012/03/ana-cristina-wheelock-lopez-es-la-reina.html
Detiene por ebrios y escandalosos a la diputad	http://noticieroelcirco.blogspot.mx/2012/03/detiene-por-ebrios-y-escandalosos-la.html
Se busca crear órgano que sancione el acoso l	http://noticieroelcirco.blogspot.mx/2012/03/se-busca-crear-organ-que-sancione-el.html
LOCALIZAN EL VEHICULO UTILIZADO POR	http://noticieroelcirco.blogspot.mx/2012/03/localizan-el-vehiculo-utilizado-por-los.html

Listado de noticias analizadas.

Sitio de noticias Crisoplural.com

Titulo de la noticia	URL
Contra ira, propuestas	http://crisoplural.com/2012/05/27/contra-ira-propuestas/
Yo también vi el debate	http://crisoplural.com/2012/05/10/yo-tambien-vi-el-debate/
Conexión México – Buenos Aires – Madrid (Tercera y última parte)	http://crisoplural.com/2012/05/06/conexion-mexico-buenos-aires-madrid-tercera-y-ultima-parte/
El debate	http://crisoplural.com/2012/05/06/el-debate/
#Yo soy 132	http://crisoplural.com/2012/05/23/yo-soy-132/
Japoneses en Aguascalientes (II)	http://crisoplural.com/2012/05/21/japoneses-en-aguascalientes-ii/
Carlos Fuentes	http://crisoplural.com/2012/05/18/carlos-fuentes/
La debacle de Vázquez Mota	http://crisoplural.com/2012/04/11/la-debacle-de-vazquez-mota/
También en Puertecito de la Virgen...	crisoplural.com/2012/04/09/tambien-en-puertecito-de-la-virgen/
La Semana Santa	crisoplural.com/2012/04/08/la-semana-santa-2/
Breve historia de un juicio oral	http://crisoplural.com/2012/04/04/breve-historia-de-un-juicio-oral/
Visión imperial del Vaticano	http://crisoplural.com/2012/03/30/vision-imperial-del-vaticano/
Tu es Petrus, et súper hanc petram aedificabo Ecclesiam meam	http://crisoplural.com/2012/03/28/tu-es-petrus-et-super-hanc-petram-aedificabo-ecclesiam-meam/
La industrialización de Aguascalientes (III)	http://crisoplural.com/2012/03/26/la-industrializacion-de-aguascalientes-iii/
La verdad por detrás	http://crisoplural.com/2012/03/25/la-verdad-por-detras/
Zaldivar, el debido proceso legal y el caso Cassez	http://crisoplural.com/2012/03/21/zaldivar-el-debido-proceso-legal-y-el-caso-cassez/
Raphael: aquí nos sabemos todas sus canciones	http://crisoplural.com/2012/03/12/raphael-aqui-nos-sabemos-todas-sus-canciones/
Pitágoras y yo	http://crisoplural.com/2012/03/04/pitagoras-y-yo/
Por los caminos de Dios y del diablo	http://crisoplural.com/2012/02/26/por-los-caminos-de-dios-y-del-diablo/
Vialidades, parques y jardines	http://crisoplural.com/2012/02/24/vialidades-parques-y-jardines/
Actualizar el estudio del Derecho	http://crisoplural.com/2012/02/22/actualizar-el-estudio-del-derecho/
La industrialización de Aguascalientes	http://crisoplural.com/2012/02/20/la-industrializacion-de-aguascalientes/
Saber hortográfico*	http://crisoplural.com/2012/02/19/saber-hortografico/
La elección panista	http://crisoplural.com/2012/02/12/la-eleccion-panista/
Carta a un amigo	http://crisoplural.com/2012/01/22/carta-a-un-amigo/
Las y los herederos	http://crisoplural.com/2012/01/13/las-y-los-herederos/
De sociología y sociólogos	http://crisoplural.com/2012/01/09/de-sociologia-y-sociologos/
Dos historias para contarse	http://crisoplural.com/2012/01/01/dos-historias-para-contarse/
¿Y ahora quien podrá avergonzarse?	http://crisoplural.com/2011/12/26/%C2%BFy-ahora-quien-podra-avergonzarse/

Cuestionario.***Instrucciones:***

Contesta todas las preguntas que se listan a continuación, encerrando en un círculo la respuesta que elijas y complementándola donde se te indique.

-
1. ¿Visitas sitios de noticias o blogs en internet?
 - a) Sí (*pase a la pregunta 2*)
 - b) No(*pase a la pregunta 4*)

 2. ¿Qué secciones del sitio visitas ?(*selecciona al menos 3*)
 - a) Deportes
 - b) Sociales
 - c) Economía
 - d) Ciencia y tecnología
 - e) Religión
 - f) Opinión
 - g) Educación
 - h) Política nacional e internacional
 - i) Cultural
 - j) Local
 - k) Otras secciones, ¿cuál? _____

 3. ¿Con que frecuencia visitas estos sitios?
 - a) 1 vez a la semana
 - b) De 2 a 5 veces por semana
 - c) Más de 5 veces por semana

 4. ¿Utilizas las herramientas de “comentarios” o “wikis” que contienen los sitios de noticias o blogs para participar en los comentarios de los contenidos publicados?

- a) Sí
- b) No
- c) Algunas veces

5. ¿Alguna vez te ha agredido algún usuario de internet cuando publicas algún comentario en alguna noticia utilizando la herramienta de “comentarios” o “wikis” las cuales te permiten participar en los temas publicados en los sitios de noticias o blogs?

- a) Sí
- b) No

6. ¿Alguna vez has agredido a algún usuario de internet cuando publica algún comentario en alguna noticia utilizando la herramienta de “comentarios” o “wikis” que te permiten participar en los temas publicados en los sitios de noticias o blogs?

- a) Sí
- b) No

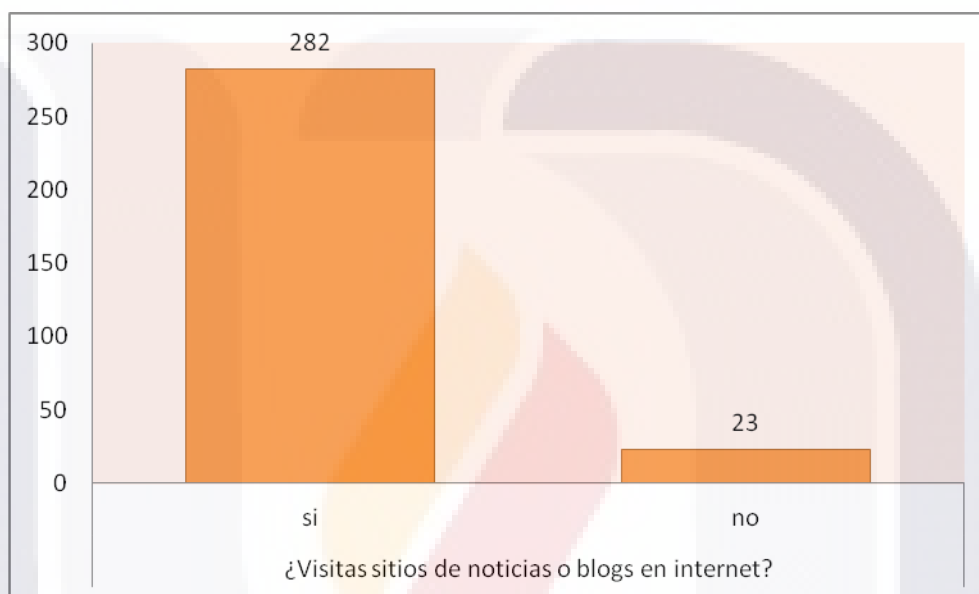
7. Describe qué sientes y qué haces cuando te agreden en internet.

8. Escribe una lista de 20 palabras, frases o ideas agresivas (sin importar que tan ofensivas sean) que has leído en los comentarios realizados por los usuarios en los sitios de noticias y blogs.

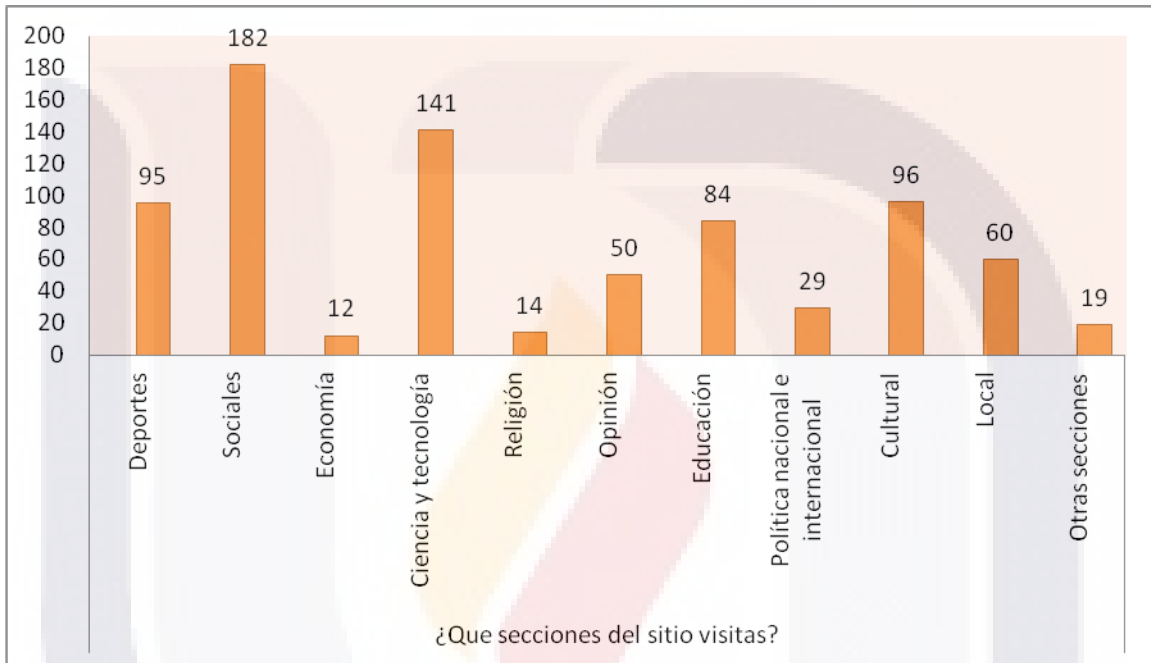
ANEXO F.

Resultados obtenidos de la aplicación del instrumento de evaluación o cuestionario.

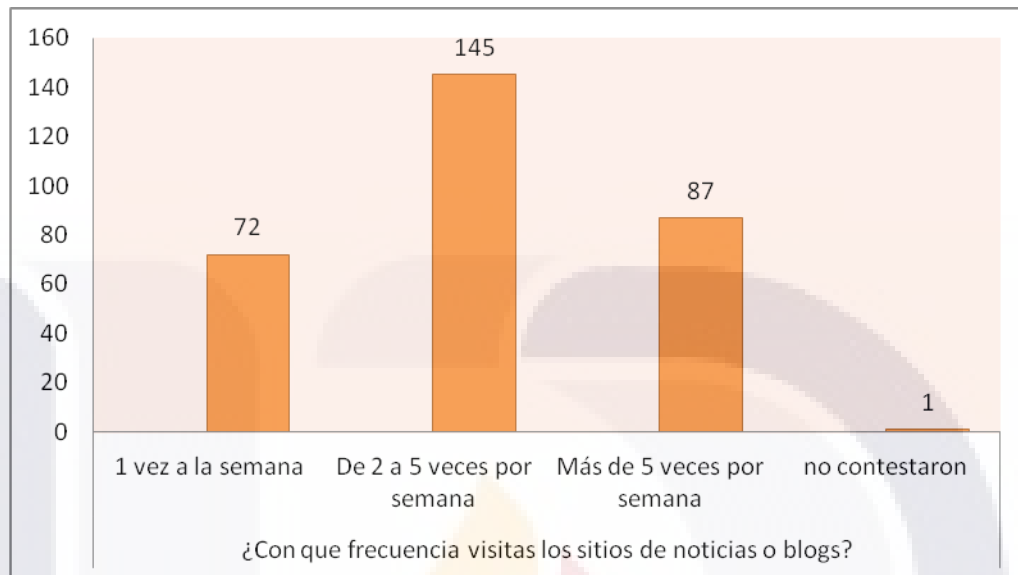
- A la pregunta 1 que dice: ¿visitas sitios de noticias o blogs en internet?, del total de encuestados, 282 respondió que sí visitan sitios de noticias o blogs.



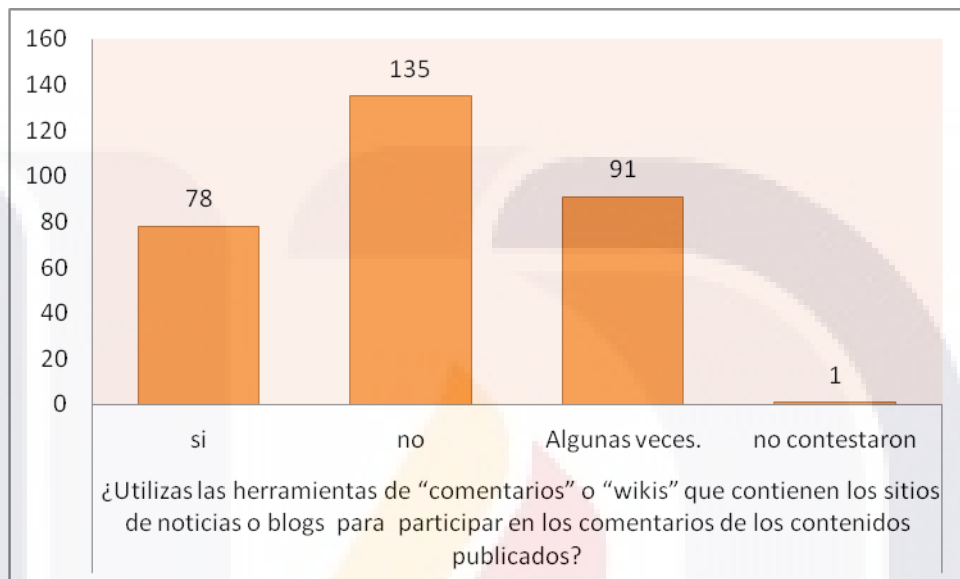
- A la pregunta 2 que dice: ¿qué secciones del sitio visitas?, del total de encuestados, 95 dijo que deportes; 182 visita sociales; 12, economía; 141, ciencia y tecnología; 14, religión; 50, opinión; 84, educación; 29, política nacional e internacional; 96, cultural; 60, local; y 19, otras secciones.



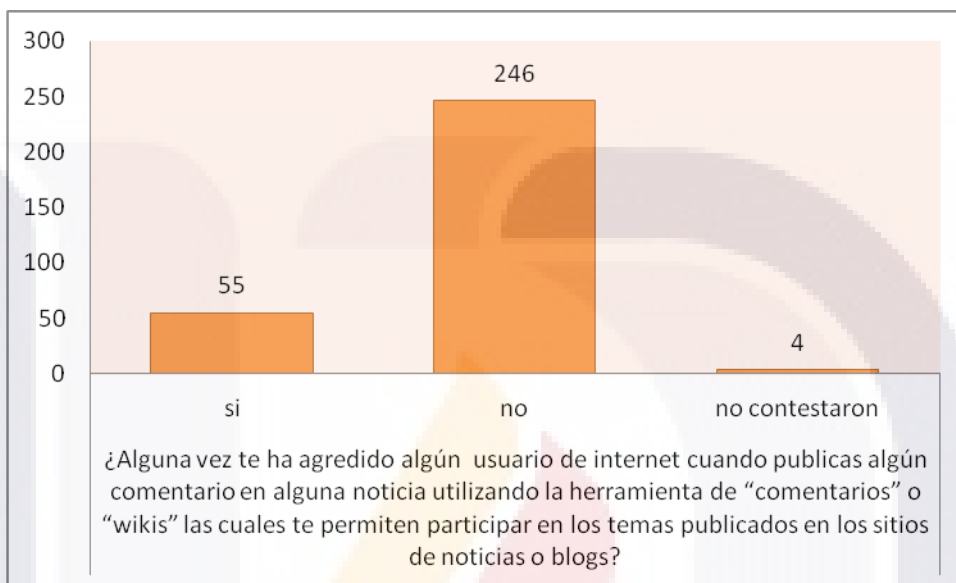
- A la pregunta 3 que dice: ¿con qué frecuencia visitas los sitios de noticias o blogs?, 72 dijeron que una vez a la semana; 145, de dos a cinco veces; y 87, más de cinco veces.



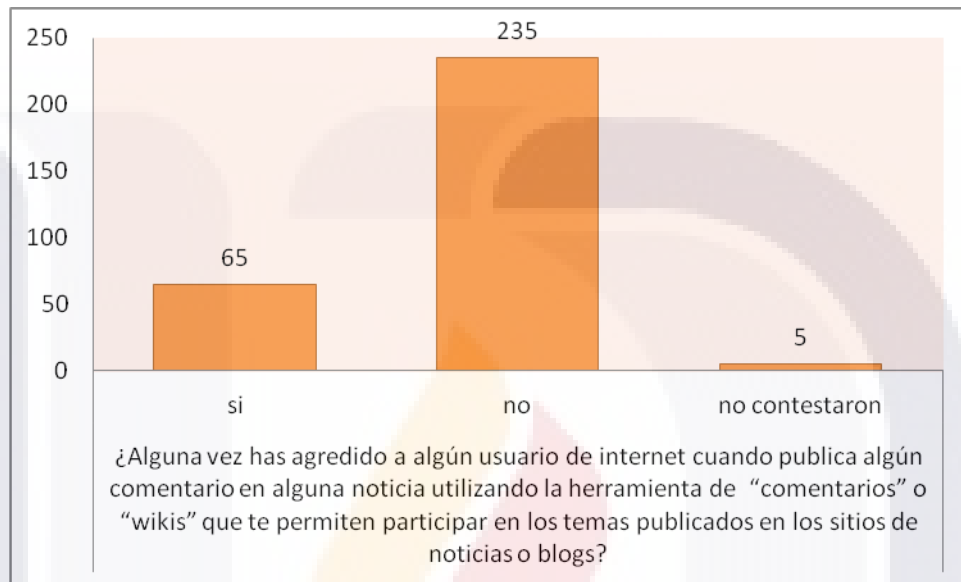
- En la pregunta 4 que dice: ¿utilizas las herramientas de “comentarios” o “wikis” que contienen los sitios de noticias o blogs para participar en los comentarios de los contenidos publicados?, del total de encuestados, 78 dijeron que sí; 135 dijeron que no; y 91 que algunas veces.



- Respecto a la pregunta 5 que dice: ¿alguna vez te ha agredido algún usuario de internet cuando publicas algún comentario en alguna noticia utilizando la herramienta de “comentarios” o “wikis” las cuales te permiten participar en los temas publicados en los sitios de noticias o blogs? 55 dijeron que sí y 246 dijeron que no.



- En la pregunta 6 que dice: ¿alguna vez has agredido a algún usuario de internet cuando publica algún comentario en alguna noticia utilizando la herramienta de “comentarios” o “wikis” que te permiten participar en los temas publicados en los sitios de noticias o blogs? del total de los encuestados, 65 dijo que sí y 235 dijo que no.



- En la pregunta número 7, en la que se pidió que se describiera qué siente la persona y qué hace cuando le agreden en internet, la mayoría señaló que se mostraba indiferente ante las agresiones.

- La pregunta número 8 en la cual se solicitó que se escribiera una lista de 20 palabras, frases o ideas agresivas (sin importar que tan ofensivas fueran) que se han leído en los comentarios realizados por los usuarios en los sitios de noticias y blogs, el cual corresponde al catálogo utilizado en el sistema de detección de situaciones de ciber-agresión. Cabe precisar que no todos los encuestados contestaron ésta pregunta.

Total de palabras y frases obtenidas: 2, 725.

Palabras y frases duplicadas: 706.

Las 10 palabras y frases ofensivas más nombradas por los encuestados se listan a continuación representadas numéricamente, considerando la primera como la más mencionada:

Texto	Total
1	118
2	112
3	105
4	98
5	71
6	69
7	61
8	59
9	57
10	56