



**Social media como medio de análisis y prevención del acoso cibernético,
aplicado a los perfiles de Facebook, utilizando técnicas de procesamiento de
lenguaje natural con base en Levenshtein**

Jácome Andino, Danny Alexander

Departamento de Ciencias de la Computación

Carrera de Ingeniería de Sistemas e Informática

Trabajo de Titulación previo a la obtención del título de Ingeniero en Sistemas e Informática

Ing. Tapia León, Freddy Mauricio.

21 de Agosto del 2020

Document Information

Analyzed document tesis_danny_16-07-2020_final.docx (D78043977)
Submitted 8/21/2020 6:29:00 PM
Submitted by
Submitter email fmtapia@espe.edu.ec
Similarity 3%
Analysis address fmtapia.espe@analysis.arkund.com



Sources included in the report

| | | |
|-----------|---|---|
| W | URL: https://repositorio.espe.edu.ec/bitstream/21000/14165/1/T-ESPE-057697.pdf Fetched: 11/19/2019 11:36:24 AM |  11 |
| SA | CALDERON QUISHPE 29-06-2020..pdf Document CALDERON QUISHPE 29-06-2020 .pdf (D75924915) |  1 |
| W | URL: https://core.ac.uk/download/pdf/76599844.pdf Fetched: 4/28/2020 1:05:25 AM |  5 |
| W | URL: https://riunet.upv.es/bitstream/handle/10251/16576/Memoria.pdf?sequence=1 Wohlin, Fetched: 8/21/2020 6:30:00 PM |  3 |
| SA | Documento Final Farez Catalina.docx Document Documento Final Farez Catalina docx (D63380019) |  2 |
| W | URL: https://journal.espe.edu.ec/ojs/index.php/geeks/article/download/280/259 Fetched: 9/30/2019 11:14:00 PM |  1 |



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

CERTIFICACIÓN

Certifico que el trabajo de titulación, “**Social media como medio de análisis y prevención del acoso cibernético, aplicado a los perfiles de Facebook, utilizando técnicas de procesamiento de lenguaje natural con base en Levenshtein**” fue realizado por el señor **Jácome Andino, Danny Alexander** el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 21 de agosto de 2020.

Firma:



.....
Ing. Tapia León, Freddy Mauricio MSc.

C. C. 1714745690



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

RESPONSABILIDAD DE AUTORÍA

Yo, **Jácome Andino, Danny Alexander**, con cédula de ciudadanía n° 1719945394, declaro que el contenido, ideas y criterios del trabajo de titulación: **“Social media como medio de análisis y prevención del acoso cibernético, aplicado a los perfiles de Facebook, utilizando técnicas de procesamiento de lenguaje natural con base en Levenshtein”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 21 de agosto de 2020

Firma

Jácome Andino, Danny Alexander

C.C. 1719945394



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

AUTORIZACIÓN DE PUBLICACIÓN

Yo **Jácome Andino, Danny Alexander**, con cédula de ciudadanía n° 1719945394, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Social media como medio de análisis y prevención del acoso cibernético, aplicado a los perfiles de Facebook, utilizando técnicas de procesamiento de lenguaje natural con base en Levenshtein”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 21 de agosto de 2020

Firma

Jácome Andino, Danny Alexander

C.C. 1719945394

DEDICATORIA

A Dios por darme la sabiduría y fortaleza para concluir este camino, alcanzando uno de mis más grandes anhelos.

A mis padres Oswaldo y Gladys por el amor y esfuerzo que han realizado para que nada me falte, gracias por inspirar en mi respeto y valentía para alcanzar mis metas.

A mis hermanas Nataly, Gabriela y Andrea por su cariño y apoyo incondicional, durante todo este proceso, por estar conmigo en todo momento gracias.

A mi querido sobrino Esteban Xavier por darme tantas alegrías con sus gestos y demostraciones de cariño inocentes.

Danny Alexander Jácome Andino.

AGRADECIMIENTO

Agradezco a Dios por la vida y sabiduría entregada para culminar este pequeño camino.

A mis padres, hermanas y tíos que fueron un apoyo durante la etapa universitaria y más aún en los momentos difíciles pero inolvidables, por la dedicación, enseñanzas y motivación para ser cada día una mejor versión de mí.

Al sorprendente grupo de amigos, por su amistad, lealtad y todas las vivencias, gracias por hacer de esta etapa algo imperecedero.

A mi director de tesis Ing. Freddy Tapia por el tiempo dedicado, su experiencia, motivación y paciencia para orientarme durante todo el proceso.

Y por supuesto a mi querida Universidad, los docentes y autoridades, por permitirme formarme profesionalmente en esta gran institución.

Danny Alexander Jácome Andino.

INDICE DE CONTENIDO

| | |
|--|----|
| Índice de tablas | 11 |
| Índice de figuras | 12 |
| Resumen | 13 |
| Abstract | 14 |
| Capítulo I | 15 |
| Introducción | 15 |
| Antecedentes | 15 |
| Problemática | 20 |
| Justificación..... | 22 |
| Objetivos | 23 |
| <i>Objetivo general</i> | 23 |
| <i>Objetivos específicos</i> | 23 |
| Alcance | 24 |
| Descripción general del proyecto..... | 25 |
| Capítulo II | 26 |
| Revisión de literatura | 26 |
| Acoso Cibernético..... | 27 |
| Patrones de personalidad..... | 29 |
| Componentes de la personalidad | 30 |
| Redes Sociales..... | 37 |
| Procesamiento de Lenguaje Natural (PLN) | 38 |
| Técnicas para el Procesamiento de Lenguaje Natural (PLN) | 39 |
| Técnicas basadas en análisis de deletreo y distancia | 39 |
| Técnicas Compuestas | 52 |
| Variaciones al Algoritmo de Levenshtein | 52 |
| Trabajos similares con PLN..... | 58 |
| Capítulo III | 61 |
| Desarrollo de la propuesta | 61 |
| Planteamiento de un corpus de palabras | 62 |

| | |
|--|----|
| Recursos..... | 64 |
| Recursos humanos | 64 |
| Recursos de hardware | 64 |
| Análisis..... | 65 |
| Requerimientos de software | 65 |
| Diagrama de casos de usos | 67 |
| Casos de usos detallado | 67 |
| Diseño..... | 72 |
| Diagrama de arquitectura | 72 |
| Modelo de base de datos | 73 |
| Arquitectura de componentes. | 74 |
| Implementación | 74 |
| Selección de Herramientas..... | 74 |
| Procesos separados | 77 |
| Descarga de datos (Comentarios) | 77 |
| Limpieza de Datos | 79 |
| Almacenamiento de datos | 81 |
| Implementación Front - End | 83 |
| Implementación Back-End..... | 83 |
| Funcionalidad específica del aplicativo | 84 |
| Despliegue de la aplicación..... | 85 |
| Capitulo IV | 87 |
| Evaluación | 87 |
| Diseño de la actividad | 88 |
| Desarrollo de la actividad | 89 |
| Inducción sobre el acoso cibernético (Cyberbullying). | 91 |
| Aplicación del test de personalidad | 91 |
| Valoración pesos semánticos..... | 92 |
| Uso de CERO_ACOSO y desarrollo de la actividad | 93 |
| Resultados..... | 93 |
| Valoración de pesos semánticos..... | 94 |
| Impacto en función del peso semántico..... | 95 |

| | |
|--|-----|
| | 10 |
| Similitudes encontradas en comentarios propios con CERO_ACOSO | 95 |
| Similitudes encontradas en comentarios recibidos con CERO_ACOSO..... | 98 |
| Capítulo V | 102 |
| Conclusiones y recomendaciones y trabajos futuros | 102 |
| Conclusiones | 102 |
| Recomendaciones | 103 |
| Trabajos Futuros..... | 103 |
| REFERENCIAS BIBLIOGRÁFICAS | 104 |

Índice de tablas

| | | |
|----------|---|----|
| Tabla 1 | <i>Búsqueda en repositorios digitales</i> | 26 |
| Tabla 2 | <i>Tipos de acoso cibernético (Cyberbullying)</i> | 27 |
| Tabla 3 | <i>Temperamento y carácter</i> | 32 |
| Tabla 4 | <i>Factores de primer orden de la personalidad</i> | 34 |
| Tabla 5 | <i>Factores de segundo orden de la personalidad</i> | 36 |
| Tabla 6 | <i>Redes sociales y tipo de servicios</i> | 37 |
| Tabla 7 | <i>Algoritmo de Guth</i> | 40 |
| Tabla 8 | <i>Algoritmo Soundex</i> | 43 |
| Tabla 9 | <i>Código Soundex</i> | 44 |
| Tabla 10 | <i>Algoritmo Metaphone</i> | 45 |
| Tabla 11 | <i>Código Metaphone</i> | 46 |
| Tabla 12 | <i>Algoritmo NYSIIS</i> | 48 |
| Tabla 13 | <i>Algoritmo Phonex</i> | 50 |
| Tabla 14 | <i>Distancia Damereau-Levenshtein</i> | 53 |
| Tabla 15 | <i>Cálculo de la distancia Damereau-Levenshtein</i> | 53 |
| Tabla 16 | <i>Algoritmo Soundex-SP</i> | 55 |
| Tabla 17 | <i>Corpus de palabras</i> | 63 |
| Tabla 18 | <i>Recursos humanos</i> | 64 |
| Tabla 19 | <i>Recursos de hardware</i> | 65 |
| Tabla 20 | <i>Requerimiento 1</i> | 65 |
| Tabla 21 | <i>Requerimiento 2</i> | 65 |
| Tabla 22 | <i>Requerimiento 3</i> | 66 |
| Tabla 23 | <i>Requerimiento 4</i> | 66 |
| Tabla 24 | <i>Caso de uso: Listar participantes / requerimiento_001</i> | 68 |
| Tabla 25 | <i>Caso de uso: Listar y analizar información / Requerimiento_002_003</i> | 68 |
| Tabla 26 | <i>Caso de uso: Generar Reporte / Requerimiento_004</i> | 70 |
| Tabla 27 | <i>Recursos de software</i> | 74 |
| Tabla 28 | <i>Conectores y señalizadores</i> | 80 |
| Tabla 29 | <i>Ejemplo proceso de limpieza</i> | 81 |
| Tabla 30 | <i>Descripción del diagrama de base de datos</i> | 82 |
| Tabla 31 | <i>Cantidad de similitudes encontradas vs distancia</i> | 85 |
| Tabla 32 | <i>Proceso de la actividad</i> | 90 |

Índice de figuras

| | |
|--|-----|
| Figura 1. <i>Frecuencia de uso diario de Internet en el Ecuador.</i> | 15 |
| Figura 2. <i>Uso de redes sociales desde su smartphone.</i> | 16 |
| Figura 3. <i>Edades con mayor presencia de (Cyberbullying) en Ecuador.</i> | 21 |
| Figura 4. <i>Componentes de la personalidad.</i> | 30 |
| Figura 5. <i>Diseño de trabajo para el algoritmo fonético.</i> | 48 |
| Figura 6. <i>Porcentaje de coincidencias en apellidos.</i> | 57 |
| Figura 7. <i>Porcentaje de coincidencias en nombres propios.</i> | 57 |
| Figura 8. <i>Proceso obtención corpus de palabras.</i> | 63 |
| Figura 9. <i>Diagrama de arquitectura.</i> | 67 |
| Figura 10. <i>Diagrama de arquitectura.</i> | 72 |
| Figura 11. <i>Base de datos modelo lógico.</i> | 73 |
| Figura 12. <i>Base de datos modelo físico.</i> | 74 |
| Figura 13. <i>Logo NetBeans.</i> | 75 |
| Figura 14. <i>Logo MySQL / phpMyAdmin</i> | 76 |
| Figura 15. <i>Logo RestFB.</i> | 77 |
| Figura 16. <i>Petición del aplicativo al API.</i> | 78 |
| Figura 17. <i>Código extracción datos de Facebook.</i> | 78 |
| Figura 18. <i>Flujograma limpieza de datos.</i> | 79 |
| Figura 19. <i>Identificación de comentarios.</i> | 82 |
| Figura 20. <i>Diagrama de la tabla comentario.</i> | 82 |
| Figura 21. <i>Front – End de la aplicación.</i> | 83 |
| Figura 22. <i>Primera evaluación CERO ACOSO.</i> | 85 |
| Figura 23. <i>Primera evaluación CERO ACOSO.</i> | 88 |
| Figura 24. <i>Capacitación sobre acoso cibernético (Cyberbullying).</i> | 91 |
| Figura 25. <i>Matriz pesos semánticos.</i> | 92 |
| Figura 26. <i>Pesos semánticos.</i> | 94 |
| Figura 27. <i>Impacto vs Palabras.</i> | 95 |
| Figura 28. <i>Frecuencia X Palabra C.P.</i> | 96 |
| Figura 29. <i>Probabilidad X Impacto C.P.</i> | 97 |
| Figura 30. <i>Nivel de Agresividad Activa.</i> | 97 |
| Figura 31. <i>Frecuencia X Palabra C.R.</i> | 98 |
| Figura 32. <i>Probabilidad X Impacto C.R.</i> | 99 |
| Figura 33. <i>Nivel de Agresividad Pasiva.</i> | 100 |
| Figura 34. <i>Evaluación del factor E.</i> | 101 |

Resumen

El uso de teléfonos inteligentes y otros dispositivos electrónicos han permitido un crecimiento y acceso masivo a las redes sociales, permitiendo la interconexión entre personas de diferentes lugares, culturas, etnias, afinidades y gustos en particular; los cuales pueden crear y compartir diferentes contenidos a través de las mismas. Se podrían destacar que estas interconexiones han permitido revolucionar la forma de compartir conocimientos, sentimientos, emociones y sin duda también la forma de agredir a nuestros iguales. En consecuencia, este trabajo propone prevenir el ciberacoso en redes sociales particularmente en Facebook, para esto se utiliza métricas para la identificación de patrones de agresividad, asociado con el procesamiento de lenguaje natural – PLN; todo esto basado en el algoritmo de Levenshtein he integrado con Visual Studio como herramienta de entorno de desarrollo para configurar las funcionalidades del aplicativo en sí; finalmente la obtención de los datos por medio de una API de Facebook. Para esta investigación se analizaron trabajos previos en el campo de acción, se desarrolló un aplicativo de escritorio, se llevaron a cabo pruebas, se evaluaron y homologaron los resultados a través de matrices de probabilidad e impacto en función de pesos semánticos; permitiendo evidenciar la existencia de ciberacoso en la sociedad ecuatoriana y la posibilidad de identificar los presuntos actores de esta actividad mediante el uso de herramientas tecnológicas y el procesamiento de la información bajo el algoritmo de distancias de edición, destacando así las bondades del presente trabajo y su aporte a la sociedad.

PALABRAS CLAVE:

- **PROCESAMIENTO DE LENGUAJE NATURAL – PNL**
- **REDES SOCIALES**
- **ACOSO CIBERNÉTICO**

Abstract

The use of smartphones and other electronic devices has permitted an increase and massive access to social networks, allowing the interconnection between people from different places, cultures, ethnicities, affinities, and tastes in particular; which can create and share different content through them. It could be highlighted that these interconnections have allowed revolutionizing the shape of share knowledge, feelings, emotions, and also the way of attacking our equals. Emphasizing that one of the best-known attacks is the (Cyberbullying); which is a phenomenon that affects all user, regardless of their age or gender, which is gaining more and more space and due to certain culture conditions is taking like common or nature facts. Consequently, this research proposes to prevent (Cyberbullying) on social networks, especially on Facebook, for this, metrics will be used to identify aggressively patterns associated with the Natural Language Processing - NLP; all of these based on the Levenshtein algorithm integrated with Visual Studio as a development tool to configure the functionalities of the application; finally, to obtain the data through a Facebook API. For this research, previous researches in this field were analyzed, an application desktop was developed, tests were carried out, the results were evaluated and standardize through probability and impact matrices based on semantic weights; allowing evidence of the (Cyberbullying) in the Ecuadorian society and the possibility to identify the possible actors from this activity through the use of technological tools and the natural language processing, emphasizing the benefits of this research and its contribution to the society.

KEYWORDS:

- **INFORMATION SECURITY**
- **CRITICAL INFRASTRUCTURES**
- **FRAMEWORKS**

Capítulo I

Introducción

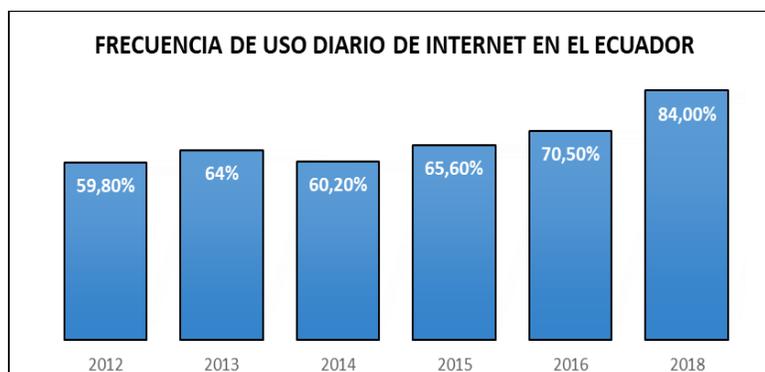
Antecedentes

El uso de las redes sociales o “social media” sigue en crecimiento, esencialmente en adultos jóvenes, sin embargo es un fenómeno que no distingue edad, estrato social o género, las redes sociales han permitido que las personas creen contenido y sean parte de un grupo social afín, en tendencias o rasgos de personalidad, (Correa, Hinsley, & De Zuniga, 2010), expresa que el uso de redes sociales crea nuevas experiencias, afecta la estabilidad emocional y es invasiva, el papel que juegan los Smartphone es muy importante, ya que estos permiten el acceso, en EE.UU. el 95% de los jóvenes tiene acceso a un Smartphone, y el 45% de ellos lo utiliza para estar conectado durante todo el día. Mientras que en Ecuador el 94.57% de personas utilizan su Smartphone para acceder a una red social (Anderson & Jiang, 2018) (Encuesta Multipropósitos, Septiembre 2019).

En la Encuesta de Seguimiento al Plan Nacional de Desarrollo “Encuesta Multipropósitos” realizada en el 2019, refleja un crecimiento en el uso diario del Internet en el Ecuador, como se puede apreciar en la Figura 1, el uso diario de Internet en el 2018 es de 84,00% de la población.

Figura 1.

Frecuencia de uso diario de Internet en el Ecuador.

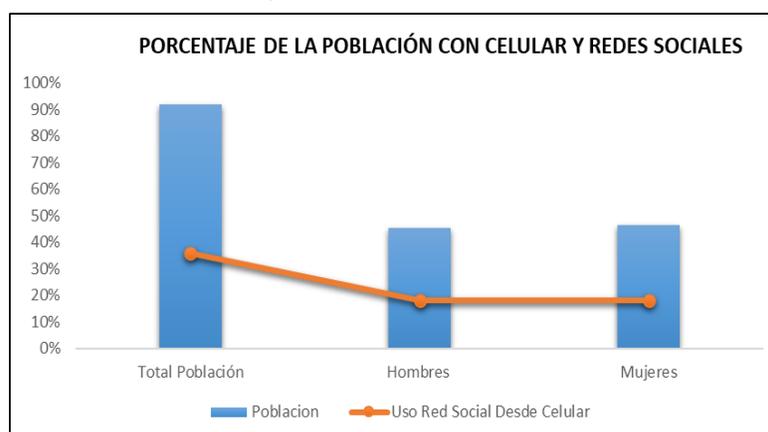


Nota: Tomado de (*Encuesta de Seguimiento al Plan Nacional de Desarrollo –Encuesta Multipropósitos*)

Según la última encuesta realizada por el Instituto Nacional de Estadísticas y Censos del Ecuador en el año 2018, teniendo en cuenta una población del 92% de la totalidad, comprendida entre una edad mayor o igual a los 5 años, el porcentaje de personas que utilizan al menos una red social desde su teléfono inteligente o “Smartphone”, es del 36% de la totalidad, dentro de los resultados de esta encuesta se puede evidenciar que el género no es un variable de importancia ya que no existe una mayor diferencia entre hombres y mujeres.

Figura 2.

Uso de redes sociales desde su smartphone.



Nota: Tomado de (*Encuesta de Seguimiento al Plan Nacional de Desarrollo – Encuesta Multipropósitos*)

Las redes sociales se han convertido y siguen siendo cada vez más populares entre las personas de todo el mundo ya que ayudan a mantenerse en contacto. (J. Haucap y U. Heimeshoff, 2013). Al ver desde una perspectiva mundial, Facebook se encuentra en primer lugar dentro de las redes sociales con mayor número de usuarios.

Facebook, desde su creación en el 2004 ha tenido un crecimiento exponencial, ahora cuenta con más de seis millones de cuentas solo de estudiantes universitarios en USA, diariamente en todo el mundo se crean alrededor de 20.000 nuevas cuentas. (Samuel D. Gosling, 2007), Según la encuesta realizada por Pew Research Center, el uso de Facebook ha disminuido; pero más de la mitad, el 51% de los jóvenes estadounidenses en el rango de edad entre los 13 y

17 años mantiene una cuenta activa en esta red social, en el Ecuador cerca de un 97,9% de personas que tiene una edad mayor o igual a 12 años, forman parte como usuarios de la red social Facebook, mientras que un 96,8% de esta población ha utilizado un Smartphone para conectarse a la red social Facebook (Hootsuite & We Are Social, 2020), convirtiéndola en la red social más utilizada del país, en comparación con otras como Twitter, Instagram, YouTube y otras. (Encuesta Multipropósitos, Septiembre 2019)

La OMS en su portal define a la violencia como: “El uso intencional de la fuerza física, amenazas contra uno mismo, otra persona, un grupo o una comunidad que tiene como consecuencia o es muy probable que tenga como consecuencia un traumatismo, daños psicológicos, problemas de desarrollo o la muerte.”, en un estudio realizado por la OMS en el 2014, los resultados indican que el 91% de violencia registrada corresponde a maltrato infantil, seguido del 71% de violencia juvenil en la región de las Américas.

La violencia penetra en todos los ámbitos sociales, tradicionales (escuela, familia, iglesia, política, etc.) y emergentes, como es el caso de la sociedad virtual promovida por las TIC's. (Solarte Lindo, 2002) y (Tokunaga, 2010). Un estudio muy amplio realizado en el sistema educativo de España en la ciudad de Murcia, a los estudiantes de la ESO “Educación Secundaria Obligatoria”, muestra que el ciberacoso a través de las redes sociales influye significativamente en las posibilidades de fracaso escolar (González Calatayud, 2017), esto se genera ya que los niños y jóvenes adolescentes durante la época escolar necesitan mantener su autoestima, concentración e interés estable para logran un buen desempeño. La UNESCO en un estudio realizado en el año 2017 reportó que el acoso afecta el proceso de aprendizaje, y que la prevalencia del acoso psicológico es el doble que el acoso físico, teniendo en el Ecuador un porcentaje de acoso físico aproximando del 15% y más del 30% de acoso psicológico. (UNESCO, 2017)

Según la UNICEF en un estudio realizado en el 2015, expresa que el acoso escolar está presente entre el 57% y 60% de su totalidad, por otro lado, la agresión a través de medios electrónicos tiene un porcentaje de 9 y 10, en un rango de edad de 11 y 18 años, y se mantiene en aumento debido a la interacción que tienen los jóvenes en las redes sociales.

El acoso digital o (Cyberbullying) surge por el rápido avance de las nuevas tecnologías de la comunicación como Internet, telefonía móvil, videojuegos, etc. En el artículo de la revista RISE “International Journal of Sociology of Education” en el 2015 se menciona que el (Cyberbullying) puede ser definido, en pocas palabras, como “el uso, envío y acción de colgar textos o imágenes dañinas o crueles en Internet u otros medios digitales de comunicación de manera reiterada para el acoso y amenaza de otro individuo o grupo de individuos”. (Fernández, C. S., & Hernández, L. L., 2015)

Las redes sociales almacenan una gran cantidad de información “Social media big data”, la misma que se ha explotado y es clave de las ideas cruciales sobre el manejo y comportamiento humano dentro de la sociedad, esto ha sido estudiado y analizado por académicos, corporaciones, políticos, periodistas y gobiernos (Boyd y Crawford 2012, Lazer et al, 2009).

(Jácome, Danny; Tapia, Freddy; Lascano, Jorge; Fuertes, Walter, 2019, February) Manifiestan que la información que se puede encontrar en las redes sociales específicamente en Facebook, es muy amplia pero la calidad de la misma depende mucho del enfoque o ámbito en el cual se lo analice, en otras palabras, la información existente puede ser de utilidad o no lo puede ser, en este artículo los autores utilizan la información obtenida de Facebook para determinar la confiabilidad que existe en FanPage dedicadas al comercio en Ecuador, para dicho ámbito los autores expresan que más del 40% del texto analizado no tiene información de valor.

Según (Samuel D. Gosling, 2007) y (Andreassen, Pallesen, & Griffiths, 2017). Los comentarios, notificaciones y otro tipo de expresión escrita o visual, son formas en las cuales un

sujeto intenta transmitir como le gustaría ser visto o tratado por los demás, las características van desde pistas sutiles que se encuentran en la elección de la ropa de un individuo a demandas más directas como declaraciones verbales. Otro tipo, es el residuo de la conducta, esta se refiere a las pistas dadas por el comportamiento de una persona. Por ejemplo, unas colecciones de películas, discos o libro bien organizadas reflejan la tendencia de un individuo a ser organizado, un mínimo detalle puede ser utilizado, como pista para determinar un rasgo o patrón de personalidad, comportamiento.

Los comentarios que son expresados como texto escrito por los usuarios de las redes sociales en forma de lenguaje humano pueden ser procesados y analizados para obtener información de gran valides, utilizando para esto procesamiento de lenguaje natural (PLN), el mismo que se entiende como la habilidad de la máquina para procesar información, y no solo identificar letras y/o sonidos del lenguaje humano. (Vásquez, A. C., Quispe, J. P., & Huayna, A. M. 2009).

Una de las técnicas más utilizadas en el procesamiento de lenguaje natural (PLN) es el estudio léxico o morfológico, esto se debe a que todas las técnicas de PLN recuren de un modo u otro a teorías semánticas (Vásquez, 2009), el mismo que se encarga de analizar texto de una forma ordenada, dividiendo todo el texto en oraciones para luego fraccionar a palabras e identificar o extraer señales, rasgos flexivos, unidades léxicas compuestas. Para realizar esta tarea o proceso se puede utilizar el algoritmo de Levenshtein que se basa en técnicas de comparación alfabética y utiliza métricas de distancia de edición, que se calcula mediante la asignación de pesos a cada uno de los cambios realizados ya sean de inserción, eliminación o sustitución al momento comparar la primera cadena con la segunda.

Problemática

Un estudio realizado por la (UNICEF, Violencia entre pares en el sistema educativo: Una mirada en profundidad al acoso escolar en el Ecuador., 2015) en el Ecuador, revelo que en las instituciones escolares la violencia entre iguales es notable, en los últimos años de educación general básica y el bachillerato (desde octavo de básica a tercero de bachillerato), dentro de las edades de 11 a 18 años, el 58,8% de los encuestados dijo que ha recibido al menos un acto de violencia en los últimos 5 meses, esto quiere decir que al menos 6 de cada 10 estudiantes han sido víctimas de acoso escolar, entre las victimas los casos más frecuentes de violencia son las ofensas con un 38,4%, los rumores un 27,8% y agresión por medio de TIC's 9,7%.

La tecnología y el maltrato entre iguales, se mantienen en un cambio permanente, resultando de esto nuevas maneras de irrumpir contra los iguales. El ciberacoso o (Cyberbullying) es una nueva modalidad de maltrato, agresión que se realiza a través de medios tecnológicos TIC's, teniendo como principal medio de interacción la Internet, para intimidar y molestar a las diferentes víctimas. (Monelos Muñoz, 2015)

Es claro que el acoso cibernético (Cyberbullying) está presente en Ecuador y que cada vez es más común encontrar este tipo de comportamientos en los estudiantes de colegios, universidades, profesionales, etc. las afecciones de este tipo de acoso puede provocar diferentes problemas dentro del desenvolvimiento de las personas, los cuales pueden ser reflejados en el aprovechamiento académico, disciplinario, laboral y/o en el comportamiento dentro de la sociedad. Esto se ve reflejado en la encuesta multipropósito realizada por el INEC "Instituto Nacional de Estadística y Censos"

En el acoso cibernético el proceso de victimización es constante, causando perturbaciones psicológicas de gran escala en las víctimas, ocasionadas por la forma en que la información permanece visible durante todo el día, por un tiempo indefinido y en diferentes

ocasiones es muy difícil de eliminar, debido a que los agresores permanecen en el anonimato. (Feinberg & Robey, 2009)

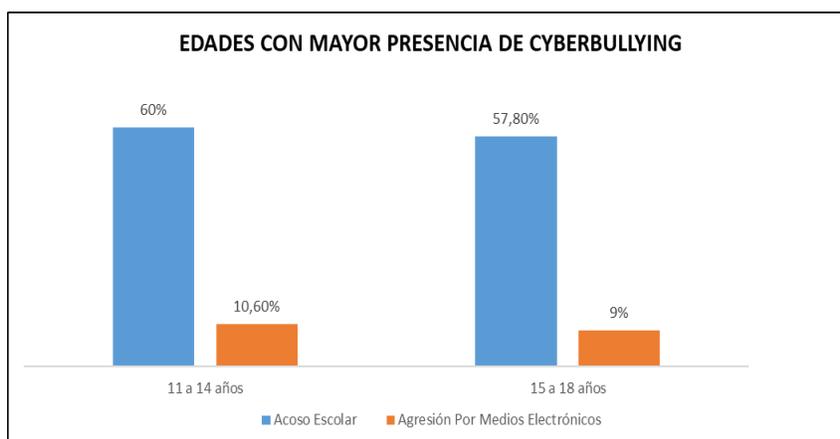
Las afecciones que el (Cyberbullying) provoca se manifiestan en todos los involucrados, víctimas, agresores y observadores, los cuales mantienen diferentes niveles de angustia, afectando el estado de ánimo, bienestar, desarrollo y la salud, teniendo como situación más extrema el suicidio o muerte de la víctima. (Limber, S. P., Olweus, D., Wang, W., Masiello, M., & Breivik, K., 2018).

Según las Estadísticas Vitales de la UNICEF, el Ecuador en el año 2011, hubo 204 casos de personas en el rango de edad de 10 a 19 años, que murieron por homicidio, pero aún más alarmante es que en el mismo año hubo 223 personas que murieron por suicidio en el mismo rango de edad. (UNICEF, Violencia contra los niños , s.f).

De acuerdo a las estadísticas de la UNICEF en el 2015, los niños y adolescentes son más propensos ante un ataque violento en lugares donde ellos deberían estar y sentirse más seguros, sus hogares, escuelas, colegios y comunidades. Las escuelas son uno de los lugares donde existe mayor probabilidad que un niño o adolescente sea víctima de diferentes actos y tipos de violencia, como se puede observar en la Figura 3, los niños y adolescentes entre las edades de 11 a 18 años, en su mayoría han sido víctimas de acoso escolar. (UNICEF, Violencia entre pares en el sistema educativo: Una mirada en profundidad al acoso escolar en el Ecuador., 2015)

Figura 3.

Edades con mayor presencia de (Cyberbullying) en Ecuador.



Nota: Tomado de (UNICEF, Violencia entre pares en el sistema educativo: Una mirada en profundidad al acoso escolar en el Ecuador., 2015)

Con lo mencionado anteriormente en la formulación del problema se expone la siguiente pregunta de investigación:

¿Es posible analizar y prevenir el acoso cibernético (Cyberbullying), mediante el estudio de perfiles en redes sociales como Facebook, utilizando el Procesamiento de Lenguaje Natural (PLN) con base en el algoritmo de Levenshtein?

Justificación

En la actualidad la violencia por medio de las tecnologías de la información y la comunicación TIC's representa un difícil problema, las similitudes entre los mecanismos básicos del acoso (bullying) y el acoso digital (Cyberbullying), como por ejemplo, aislamiento de un grupo, hostigamiento, daño psicológico (Martínez, 2013), permiten determinar que se trata de un mismo fenómeno, teniendo en cuenta que la mayor diferencia entre estos es el uso de medios tecnológicos como el Internet, computador y dispositivos móviles en el caso del (Cyberbullying) para cometer abusos a través de sus diferencias modalidades. Permitiendo la identificación de componentes comunes o similares e identificar y detallar los específicos, que sin duda existen y enmarcan las consecuencias que estos provocan.

La forma de actuar de los agresores en el caso del (Cyberbullying) es recurrente e intencional, los abusadores han identificado la diferencia de poder que tienen sobre sus víctimas, la falta de juicio moral y ético no les permite mantener una conducta proactiva, no pueden diferenciar entre el “deber hacer moralmente”, y el “querer hacer éticamente”, ignorando las normas sociales por falta de conciencia, siendo una de sus salidas la agresión hacia los demás. (Martínez, 2013)

Según un estudio realizado por (Bachrach, 2015) la red social Facebook puede proporcionar información de alto interés para identificar el actuar de una persona, los perfiles de usuario de Facebook permiten incluir pensamientos, opiniones, sentimientos y emociones. Mediante el proceso y análisis de esta información se puede identificar niveles o patrones de (Cyberbullying) a través del contenido publicado en la red social.

Objetivos

Objetivo general

Utilizar la social media como medio para el análisis y prevención del acoso cibernético, aplicado a los perfiles de Facebook, empleando técnicas de procesamiento de lenguaje natural con base en Levenshtein permitiendo determinar un presunto perfil de personalidad el cual sirva de apoyo a los psicólogos.

Objetivos específicos

- i. Identificar perfiles de usuario en Facebook donde el (Cyberbullying) esté presente, usuarios con una franja de edad mayor a los 11 años y menor a los 18 años.
- ii. Realizar una investigación exploratoria para identificar y plantear un diccionario de datos de acuerdo con el (Cyberbullying), con las palabras más influyentes de acuerdo con las directrices de un psicólogo.

- iii. Identificar la interfaz de programación de aplicaciones (API) que permita extraer la información de Facebook.
- iv. Realizar un aplicativo de escritorio para el procesamiento de los datos obtenidos de las redes sociales.
- v. Analizar y evaluar los patrones de (Cyberbullying) identificados mediante matrices de probabilidad e impacto y el apoyo de un psicólogo con conocimientos en el ámbito del ciberacoso para sugerir un perfil de personalidad.

Alcance

La investigación comprende el diseño e implementación de un prototipo para el análisis e identificación de patrones relacionados con el ciberacoso o (Cyberbullying), mediante el algoritmo de Levenshtein basado en distancias de edición, con el fin de analizar y prevenir el maltrato entre pares.

El proyecto se llevará a cabo bajo el cumplimiento de las siguientes fases:

- i. Planteamiento de un diccionario de datos o corpus de palabras personalizado con relación al (Cyberbullying).
- ii. Recolección de datos de usuarios de Facebook por medio de una API que permita la extracción de los datos.
- iii. Almacenamiento de los datos en una base digital.
- iv. Limpieza de datos, eliminación de las palabras de conexión dentro de los comentarios extraídos tales como “el, la, los, las, de, y, con, etc.”.
- v. Realizar un aplicativo “prototipo” de escritorio para el procesamiento de los datos obtenidos.
- vi. Analizar, evaluar y sugerir un perfil de personalidad.

El estudio se realizará con estudiantes de la unidad educativa “Sociedad Educativa Futuro SEF” que estén dentro del rango de edad de los 11 a los 18 años, por motivos de autorización de acceso a la información de los estudiantes se trabajara con aquellos que están cursando segundo y tercero de bachillerato, en donde la media de edad de los estudiantes es de 17 años, para obtener una muestra lo más homogénea posible y numéricamente representativa al momento de recolectar los datos, para de esta manera tener un menor rango de error o sesgo al identificar patrones de comportamiento, frente al acoso cibernético (Cyberbullying).

Descripción general del proyecto

El presente proyecto de investigación tiene como propósito desarrollar un sistema para analizar perfiles de redes sociales e identificar patrones de comportamiento apegados al (Cyberbullying), basados en los comentarios y publicaciones realizadas en cada uno de los perfiles de usuario y de esta manera prevenir el acoso cibernético en la unidad educativa “Sociedad Educativa Futuro SEF” tomada como referencia para el caso de estudio.

Para lo cual se utilizará un algoritmo especializado en métricas de distancias, basado en la comparación de caracteres específicamente en el Algoritmo de Levenshtein, y poder llegar a una presunción de perfiles de personalidad agresivos o pasivos “sumisos”.

Capítulo II

Revisión de literatura

Para redactar este capítulo se revisaron proyectos realizados que tienen una relación con el tema planteado, identificación de (Cyberbullying) a través de Procesamiento de Lenguaje Natural – PLN, mediante la identificación de patrones de la personalidad con una base en el algoritmo de Levenshtein utilizando como banco de datos las diferentes redes sociales, dichos proyectos fueron obtenidos al realizar un estudio de antecedentes o también conocido como mapeo sistemático. (Wohlin & Runeson, 2013). Este análisis se realizó con el objetivo de conocer los ambientes actuales y disponibles que existen sobre la temática planteada.

Como resultado del mapeo sistemático, se obtuvieron las palabras más influyentes para la investigación, las cuales se detallan en la Tabla 1, por medio de las palabras clave se formaron cadenas de búsqueda, para identificar de forma adecuada los trabajos con mayor relevancia en los diferentes repositorios digitales.

Tabla 1

Búsqueda en repositorios digitales

| Idioma | Palabras Claves |
|---------|--|
| Español | Procesamiento de lenguaje natural PNL Redes sociales Acoso cibernético Ciberacoso Levenshtein Perfil personalidad Técnicas de emparejamiento Minería de datos |

| | |
|--------|---------------------------------|
| Ingles | Natural language processing PNL |
| | Social media |
| | Cyberbullying |
| | Levenshtein |
| | Personality profile |
| | Matching techniques |
| | Text mining |

Acoso Cibernético

El acoso cibernético o (Cyberbullying) es una manera de realizar actos de violencia a través de medios electrónicos especialmente mediante redes sociales, ya que estas permiten y tiene un enfoque y efecto positivo dentro de la sociedad, sin embargo no solo existen beneficios al momento de utilizar las redes sociales ya que estas exponen nuestra información o parte de ella, lo cual puede ser contraproducente, dentro de una sociedad en la cual no todos están de acuerdo con ciertas creencias ya sean religiosas, filosóficas y otras. Y eso puede convertir a los usuarios en un blanco de burlas y maltratos por parte de los demás usuarios, los cuales pueden llegar a afectar nuestra salud mental o la de otros usuarios.

Según (Menay-López & de la Fuente-Mella, 2014) el (Cyberbullying), es un modo disimulado de acoso verbal y/o escrito. Willard (2004) y otros autores han identificado algunas categorías de violencia verbal y escrita a través de las nuevas tecnologías:

Tabla 2

Tipos de acoso cibernético (Cyberbullying)

| Categoría | Descripción |
|----------------------|---|
| Flameante (Flaming). | Envío de mensajes vulgares o que muestran enfado sobre una persona a un grupo en línea o a esa persona vía email o SMS. |

| | |
|---------------------------------------|--|
| Acoso Online. (Hostigamiento) | Envío prolongado y repetido de mensajes ofensivos vía email o SMS a una persona. |
| Ciber-persecución (Cyberstalking). | Acoso online que incluye amenazas de daño o intimidación excesiva. |
| Denigración. | Envíos perjudiciales, falsas y crueles afirmaciones sobre una persona o comentarios en lugares online, es muy común que se lo realice a través de redes sociales como Facebook, WhatsApp, Instagram, etc. |
| Suplantación de personalidad. | Hacerse pasar por otra persona, en este caso por la víctima y enviar o colgar archivos de texto, video o imagen que hagan quedar mal al agredido u otras personas. |
| Excursión (Outing). | Enviar o colgar material sobre una persona que contenga información sensible, privada o embarazosa, incluido respuestas de mensajes privados o imágenes, hasta llegar a ser compartidas por otras personas. |
| Exclusión. | Cruel expulsión de un individuo de un grupo online o red social, quitándole el derecho a participar de la misma. |
| Paliza Feliz (Happy Slapping) | Realizar agresión física a una persona, mientras es grabada por algún medio tecnológico, para posteriormente compartir dicho material en redes sociales, con el fin de que más personas puedan acceder o visualizar el material. |

Nota: Tomado de (Menay-López & de la Fuente-Mella, 2014)

Al momento de identificar casos de acoso cibernético, también se pueden identificar las partes involucradas, las mismas que pueden ser parte activa o pasiva de la agresión, entre estas tenemos: (Marroquin A. B., 2017)

El agresor. – individuo que realiza actos de violencia u hostigamiento a otra persona o grupos de personas, utilizando como medio para conseguir y divulgar información, las redes sociales y la tecnología. A menudo presentan problemas psicológicos debido a situaciones conflictivas en el pasado.

La víctima. – se le conoce también como afectado, individuo que es vulnerado, las víctimas no se consideran con la capacidad de defenderse y escogen el aislamiento físico y social frente a una agresión.

Los testigos. – también conocidos como observadores, presencian la agresión, existen tres tipos: los “reforzadores”, no actúan, pero aprueban e incitan a más agresiones, los “neutros” se les considera parte de la violencia al no denunciar y los “defensores” que buscan ayudar a la víctima. (Constante Portero, 2018).

Patrones de personalidad

La personalidad de un individuo se puede determinar en función de su comportamiento o actitudes frente a diferentes situaciones, las mismas que prevalecen y marcan ciertas similitudes entre algunas de ellas.

La personalidad como una estructura psíquica dinámica según (Balarezo Chiriboga, 2015), permite identificar singularidades en cuanto al pensar, el sentir y el actuar de un sujeto, dichas manifestaciones al ser comunes permiten identificar rasgos de personalidad dentro de un grupo de sujetos.

En el artículo (Bhardwaj, Atrey, Saini, & El Saddik, 2016) se explica que la conducta social en línea y el uso de las redes sociales están relacionadas con la personalidad y que las personas se comportan de la misma manera que lo hacen en Internet, debido a que en las redes sociales están asociadas por la interacción humana. Este estudio determinó que las personas no son conscientes a la hora de usar las redes sociales, por ejemplo, dar “Like” o un “me gusta” a una

página de su agrado, esta función se la hace inconscientemente de acuerdo al gusto que define la personalidad del usuario.

Según los estudios de (Ghavami, S. M; Asadpour, M; Hatami, J; Mahdavi, M, 2015, May), la personalidad de un usuario puede predecirse con exactitud a través de la información disponible públicamente. En ese estudio, se construyó un sistema con características lingüísticas y estructurales utilizadas para predecir la personalidad de las personas en las redes sociales. Este sistema puede predecir puntuaciones de rasgos de personalidad dentro del 11% -18% de sus valores reales.

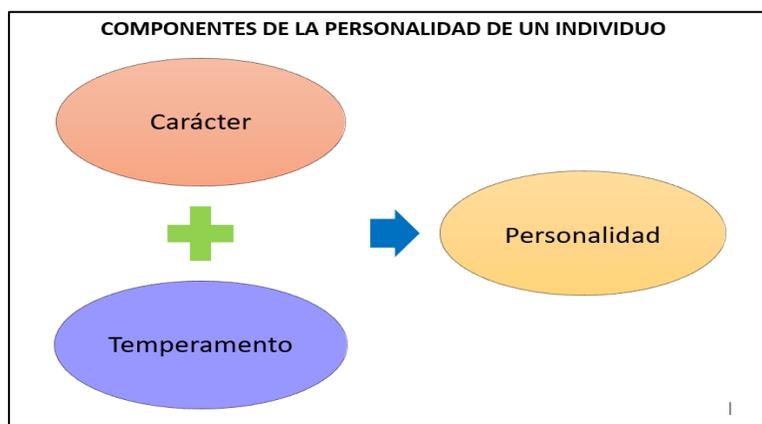
Existen diferentes tipos de personalidades de acuerdo a las expresiones comportamentales de un sujeto o individuo, las mismas que marcan similitudes entre sí.

Componentes de la personalidad

La personalidad de un sujeto se puede clasificar o identificar basándose en patrones de similitud, los cuales dependen de sus componentes, el carácter y el temperamento.

Figura 4.

Componentes de la personalidad.



Nota: Tomado de (Sánchez, H; Benedetti, M; Luzardo, M; Boggio, K; Bagnato, M; Martín, A; Premuda, P, 2017)

Según un estudio realizado al modelo de Cloninger¹ de personalidad, (Bayón, 2016) expresa que el carácter tiene un porcentaje muy pequeño de ser heredado ya que este depende del proceso de socialización, a diferencia del temperamento que presenta mayor porcentaje.

Carácter

El carácter es parte de la personalidad, el mismo que es innato de cada persona, cuando una persona nace ya tiene su carácter definido² este se puede ver manifestado en sus sentimientos y como los expresa hacia los demás.

La forma de llevar las normas sociales, el comportarse dentro de una sociedad y el lenguaje que utiliza también constituyen el carácter y por ende la personalidad de un individuo.

Temperamento

Según varios estudios realizados el temperamento es una característica que tienen las personas, forma parte de la personalidad de cada uno, en dichos estudios se menciona que el temperamento de cada persona puede ser en parte heredado, es un componente de la personalidad porque determina de alguna manera, ciertas características de cognición, comportamiento y emocionales. (Kolak & Volling, 2013).

Como resumen de algunos autores, definen como temperamento a diversas cosas como:

- La respuesta a los cambios del entorno, incluidas las reacciones somáticas y autónomas
- El temor e inhibición ante lo novedoso
- La impulsividad
- El ánimo positivo o negativo
- El nivel general de actividad
- La atención constante

¹ Modelo de personalidad psicobiológica, basado en el temperamento y el carácter.

² El carácter de una persona está definido bajo tres dimensiones: Autodirección; Cooperatividad; y Autotrascendencia.

- La autorregulación

Bajo aporte de varios autores y orientaciones psicodinámicas “salud mental”, psicobiológicas “conducta humana”, Bayón, Svrakic y Cloninger (2010), plantean un esquema básico de personalidad, Tabla 3. A través de las variables y su interpretación se podrá observar cuál de los componentes, carácter o temperamento, es más importante para el análisis y la creación de patrones de personalidad asociados al (Cyberbullying) para la investigación en curso.

Tabla 3

Temperamento y carácter

| Variable | Temperamento | Carácter |
|---------------------------------------|--|---|
| Tipo de Aprendizaje | Procedural | Proposicional |
| Nivel de Conocimientos | Automáticos | Intencionales |
| Tipo de Memoria | Preceptos, conceptos | Procedimientos, Proposiciones |
| Principios de Aprendizaje | Condicionamiento asociativo | Conceptual, Intuición |
| Rasgos Manifiestos | Hábitos, Emocional | Conceptos adaptativos, socialización |
| Emociones Asociadas | Primarias (ira, temor, perseverancia, apego) | Secundarias (orgullo, compasión, ética, altruismo) |
| Rol del sujeto en la actividad mental | Pasivo, reproductivo | Activo, Constructivo |
| Forma de representación mental | Estimulo-respuesta que varían en intensidad | Redes interactivas (esquema conceptual) que varían cualitativamente |

Nota: Tomado de (Seelbach González, 2013)

Como se puede observar en la Tabla anterior se exponen algunas variables las cuales ayudan a identificar en un individuo la personalidad que este puede llegar a tener o tiene, el modelo propone que las deficiencias en el carácter podrían encausar en algún tipo de desorden mental, mientras que los diferentes temperamentos pueden llevar a una posible patología. (Sánchez, H; Benedetti, M; Luzardo, M; Boggio, K; Bagnato, M; Martín, A; Premuda, P, 2017).

Tipos de personalidad (Clasificación)

La personalidad o rasgos de personalidad que presenta o tiene un sujeto, nos permite saber cómo actúa, los rasgos se los pueden considerar como tendencias usuales a ciertas reacciones, estos son parámetros básicos de la personalidad, los rasgos se pueden clasificar en rasgos comunes y rasgos únicos.

Entre los rasgos comunes están: La inteligencia, la afiliación, extroversión, etc. Los mismos que determinan la eficiencia o esfuerzo para alcanzar una meta en otras palabras la capacidad, dinamismo y/o temperamento al momento de realizar una actividad.

Los rasgos únicos son aquellos que reflejan los gustos de una persona: asertividad, la afabilidad, la irritabilidad, etc. Lo cual es en forma general un estilo y el tono emocional de la conducta, también considerados como motor del comportamiento, debido a que enmarcan los intereses, ambiciones y motivaciones que tiene un individuo. (Marroquin A. B.).

Según (Cattell, R. B., & Cattell, M. D., 1997)³, los rasgos también se pueden clasificar por constitución y modelado, teniendo en cuenta que los rasgos por constitución se generan de forma

³ Se hace necesario mencionar que, a pesar de la antigüedad de la publicación de Cattell, es necesario el uso, debido a la prevalencia de los conocimientos utilizados, en el planteamiento del test HSPQ, que aún se maneja en la actualidad con la única variación, traducción al idioma español.

biológica y los de modelado se deben a influencia o cambios sociales y físicos en el entorno del individuo, en palabras más sencillas son características y conductas aprendidas.

Con base en todo anterior expuesto (Cattell, R. B., & Cattell, M. D., 1997), propone el test HSPQ por sus siglas en ingles “High School Personality Questionnaire” es un cuestionario de Personalidad para Adolescentes entre los 12 y 18 años de edad, el cual tiene como finalidad ayudar a los profesionales que trabajan con este grupo de personas a entender o identificar problemas emocionales y de conducta basados en la personalidad de un individuo, mediante el test se puede identificar posibles problemas en la escolaridad y el entorno social.

En la siguiente (Tabla 4), se puede observar las 14 escalas del HSPQ, que permiten identificar rasgos de personalidad o factores de primer orden.

Tabla 4

Factores de primer orden de la personalidad.

| Escala | Puntuación alta | Puntuación baja | Descripción |
|---------------|--------------------------------|------------------------|--|
| Factor A | Reservado. | Abierto. | Mide el grado de respuesta favorable ante diferentes situaciones. |
| Factor B | Inteligencia baja. | Inteligencia alta. | Es una medida de aspectos intelectuales. |
| Factor C | Afectado por los sentimientos. | Emocionalmente estable | Mide el control emocional de un sujeto frente a los demás. |
| Factor D | Calmoso. | Excitable. | Esta escala mide la tendencia a exhibir excitación ante las provocaciones o estímulos. |

| | | | |
|-----------|--------------------|----------------------|--|
| Factor E | Sumiso. | Dominante. | Esta medida permite identificar si un sujeto es activo, dogmático, agresivo o dócil. |
| Factor F | Sobrio. | Entusiasta. | Esta medida se basa en afectos recibidos a lo largo de su vida. |
| Factor G | Despreocupado. | Consciente. | Refleja el grado en que un sujeto ha adoptado los valores del mundo de los adultos. |
| Factor H | Cohibido. | Emprendedor. | Mide el grado de sociabilidad, el relacionarse libremente o el alejarse. |
| Factor I | Sensibilidad dura. | Sensibilidad blanda. | Este rasgo permite identificar el grado de dependencia de un sujeto. |
| Factor J | Seguro. | Dubitativo. | Permite identificar si un sujeto gusta de las actividades grupales o individuales, si es crítico o no. |
| Factor O | Sereno. | Aprensivo. | Este rasgo permite diferenciar los sujetos neuróticos de los normales. |
| Factor Q2 | Sociable. | Autosuficiente. | Introversión-Extraversión, que tan dominante es en las relaciones con los demás. |
| Factor Q3 | Menos integrado. | Más integrado. | Mide el control sobre sus emociones y conducta, en cuando a despreocupación y negligencia. |

| | | | |
|-----------|-----------|--------|--|
| Factor Q4 | Relajado. | Tenso. | Permite identificar niveles de irritabilidad, mal humor o compostura social. |
|-----------|-----------|--------|--|

Nota: Tomado de (Cattell, R. B., & Cattell, M. D., 1997)

En la Tabla 5 a continuación se puede observar las 4 escalas del HSPQ, que permiten identificar rasgos de personalidad o factores de segundo orden.

Tabla 5

Factores de segundo orden de la personalidad.

| Escala | Puntuación alta | Puntuación baja | Descripción |
|---------------|------------------------|-------------------------|--|
| Factor QI | Ansiedad baja. | Ansiedad alta. | Permite identificar la ansiedad en un sujeto, el estar satisfecho con sus respuestas ante situaciones de la vida. |
| Factor QII | Introversión. | Extraversión. | Permite identificar que tan abierto es un individuo ante las demás personas y situaciones que se presentan. |
| Factor QIII | Calma. | Excitabilidad / Dureza. | Esta medida sirve para identificar sujetos agresivos o sumisos. |
| Factor QIV | Dependencia. | Independencia. | Este factor al igual que el QIII permite identificar la agresividad de un sujeto, y la dependencia de un sujeto para realizar actividades. |

Nota: Tomado de (Cattell, R. B., & Cattell, M. D., 1997)

Es necesario aclarar que las descripciones utilizadas en las Tablas 4 y 5 son una breve explicación de la naturaleza de los factores, que son útiles para identificar diferentes tipos de patrones de personalidad, una descripción más detallada se puede encontrar en (Cattell, R. B., & Cattell, M. D., 1997), en la versión original o en la versión en español.

Redes Sociales

Una red social es una herramienta o plataforma que permite la interacción de los involucrados permitiéndoles crear, publicar y compartir contenido ya sea texto, imágenes, videos, etc. Los involucrados pueden ser personas u organizaciones las cuales se comunican con el fin de establecer vínculos de acuerdo a los intereses comunes.

Entre las redes sociales más conocidas a nivel mundial se encuentran las siguientes, las cuales han sido clasificadas de acuerdo al tipo de servicio que prestan a los usuarios. (Hootsuite & We Are Social, 2020)

Tabla 6

Redes sociales y tipo de servicios

| Tipo de servicio | Red social | Ámbito |
|-------------------------|-------------------|---------------------------|
| Medios Sociales | Facebook | Publico/ocio ⁴ |
| | LinkedIn | Empresarial |
| | Instagram | Publico/ocio |
| | Tumblr | Microblogging |
| | YouTube | Musical |
| | Twitter | Microblogging |
| | Douyin/Tiktok | Publico/ocio |
| | Qzone | Publico/ocio |

⁴ Experiencia o fenómeno humano, que busca beneficio propio como la libertad, la creatividad, la satisfacción, el disfrute y el placer característica de tipo personal o social, en otras palabras, buena vida y bienestar.

| | | |
|---------------|-----------|--------------|
| Mensajería | WhatsApp | Personal |
| Instantánea / | Messenger | Personal |
| VOIP | Snapchat | Publico/ocio |
| | WeChat | Personal |
| | QQ | Publico/ocio |
| | Telegram | Personal |

Nota: Tomado de (Hootsuite & We Are Social, 2020)

Existen varias redes sociales con diferentes enfoques, lo cual permite que los usuarios interactúen entre sí de acuerdo a la necesidad y decisión de cada uno de ellos, ya sea de forma grupal o personal como se puede observar en la Tabla 6, tanto en redes sociales dedicadas al contenido social “social media” o a la mensajería instantánea “chats” en todo el mundo. En cuanto al Ecuador según el último reporte de DATAREPORTAL (Hootsuite & We Are Social, 2020), las redes sociales más visitadas de acuerdo al número de usuarios contabilizados en millones son: Facebook con 12, Instagram con 3.9, Snapchat con 1.75 mientras que Twitter cuenta con 1.11 millones de usuarios.

Procesamiento de Lenguaje Natural (PLN)

Desde hace miles de años atrás el humano ha construido su mayor tesoro, el conocimiento, y lo más impórtate que ha hecho con este es reproducirlo, almacenarlo y compartirlo a las siguientes generaciones, esto no quiere decir que se lo haya manejado de la mejor manera y como lo expresa (Gelbukh, 2010). El manejo que se le da no siempre es el adecuado, se puede decir que es negligente y falto de responsabilidad, muchas de las veces al momento de transmitir dicho conocimiento no se tiene cuidado y puede ser mal interpretado cambiando el sentido de la información.

La forma de comunicación normal o natural utilizada por los seres humanos es oral, escrita y/o las señales corporales, al momento de comunicarse entre humanos no es una tarea sencilla tanto para la persona que intenta comunicar un mensaje “emisor” como para la persona que intenta entender dicho mensaje “receptor”, se necesita de un complejo proceso de captación, para un humano transmitir o comprender un mensaje se realizan varias o miles de conexiones neuronales antes, durante y después de ser transmitido el mensaje. (Ríos, 2017)

Técnicas para el Procesamiento de Lenguaje Natural (PLN)

Establecer una comparación entre dos textos de diferente origen con la finalidad de determinar si ambos textos corresponden a una misma instancia de una entidad, es una tarea compleja que requiere de la participación de una o más técnicas asociadas con un proceso de reconocimiento y búsqueda de coincidencia de nombres, distintos en su composición fonética, morfológica y sintáctica pero semánticamente iguales y referentes a la misma entidad. Tenemos tres grupos de técnicas: (Delgado, J., Galarraga, F., Fuertes, W., Toulkeridis, T., Villacís, C., & Castro, F., 2016, March)

A continuación, se realiza un análisis de algunos algoritmos de distancia y deletreo, fonéticos y mixtos para determinar más adelante cual es algoritmo adecuado para esta investigación.

Técnicas basadas en análisis de deletreo y distancia

Identifican la diferencia que existe entre dos cadenas de caracteres, diferencias que se generan por errores cometidos al momento de la digitación, errores tipográficos de inserción, de supresión, o modificación de caracteres. No necesitan diccionarios de variantes, ni reglas morfológicas o sintácticas, ya que establecen mediciones basadas en la sola diferencia de

caracteres en la cadena. (Delgado, J., Galarraga, F., Fuertes, W., Toulkeridis, T., Villacís, C., & Castro, F., 2016, March)

Algoritmo de Guth

Es un método que no toma en cuenta consideraciones etimológicas ni de lengua alguna. Revisa la secuencia de caracteres de izquierda a derecha y hace el chequeo de coincidencias de letras más cercanas. Para obtener un valor de coincidencia, se aplican nueve reglas para cada posición de las letras en la cadena de caracteres indicadas en la Tabla 7. Si una letra en una posición del primer nombre corresponde a misma letra en la misma posición del segundo nombre, se pasa a la siguiente letra para seguir comparando. Si se ha llegado al final y si todas las comparaciones de todas las posiciones han resultado equivalentes, entonces se considera que ambos nombres son coincidentes. En el ejemplo de la Tabla 7, se considera que hay semejanza entre las dos cadenas de caracteres, pues para letras no coincidentes “valor = 0”, se cumple que al menos una vez y por lo menos en una de las nueve reglas existe coincidencia “valor = 1”. (Loyola, 2015)

Tabla 7
Algoritmo de Guth

| Regla | Posición | Posición | E R I K S O N | | | | | | | |
|-------|----------|----------|---------------|---|---|---|---|---|---|---|
| | en el | en el | | | | | | | | |
| | Nombre 1 | Nombre 2 | E | R | I | C | S | S | O | N |
| 1 | X | X | 1 | 1 | 1 | 0 | 1 | 0 | 0 | - |
| 2 | X | X + 1 | | | | 0 | | 1 | 1 | - |
| 3 | X | X + 2 | | | | 0 | | | | - |

| | | | | |
|---|-------|-------|---|---|
| 4 | X | X - 1 | 0 | 1 |
| 5 | X - 1 | X | 0 | |
| 6 | X + 1 | X | 0 | |
| 7 | X + 2 | X | 0 | |
| 8 | X + 1 | X + 1 | 1 | |
| 9 | X + 2 | X + 2 | | |

Nota: Tomado de (Loyola, 2015)

El algoritmo de Guth es simple para codificar, portable y provee resultados confiables. Sin embargo, es muy débil al comparar cadenas de caracteres cortas, por ejemplo, confunde “León” con “Noel”, el algoritmo no es adecuado para comparar cadenas de caracteres en el lenguaje español, razón por la cual es limitado para la presente investigación ya que toda la investigación se basa en cadenas de caracteres escritas en lenguaje español.

Algoritmo de Levenshtein.

El algoritmo realiza comparaciones alfabéticas mediante distancia de edición, el algoritmo pondera valores para cada una de las operaciones que se realizan ya sea de inserción, sustitución y eliminación, para tener una coincidencia exacta o similitud entre las dos cadenas de caracteres que son comparadas. Para determinar la similitud de las cadenas de caracteres se calcula la distancia mínima basada en la edición de la cadena. La Ecuación 1 simboliza el cálculo de la mínima distancia $distld(s, t)$ entre las cadenas s y t . (Lubis, A. H., Ikhwan, A., & Kan, P. L. E. , 2018)

$$distld(s, t) = \min \sum_{i=1}^N W_i (|x|, |y|) \quad (\text{Ecuación 1})$$

En donde s y t son las cadenas de caracteres a comparar, s es la cadena origen y t es la cadena destino. N simboliza el número de transformaciones realizada para que las cadenas s y t sean similares entre sí, para identificar el momento o posición en la cual se realizó una transformación se utiliza el índice i . Mientras W es el valor de pesos mínimos para cada

transformación y equiparar los fonemas, por último $|x|$ como $|y|$ son los fonemas involucrados en la edición. En un estudio realizado por (Jácome, Danny; Tapia, Freddy; Lascano, Jorge; Fuertes, Walter, 2019, February) expresan que la distancia entre dos cadenas de caracteres es siempre simétrica, cumpliéndose las siguientes desigualdades:

$$0 \leq distld(s, t) \leq \max(|s|, |t|), \text{ y } abs(|s| - |t|) \leq distld(s, t) \text{ (Ecuación 2)}$$

También se menciona como calcular la similitud entre dos cadenas s y t . Ecuación 3:

$$simld(s, t) = 1,0 - \left(\frac{distld(s,t)}{\max(|s|,|t|)} \right) \text{ (Ecuación 3)}$$

Algunos autores como Cohen y Christen, mantienen que la distancia debido al valor que se asigna a las operaciones de inserción, eliminación y sustitución son “el costo de mejor secuencia de operaciones de edición para convertir una cadena s en t ”.

El algoritmo de Levenshtein basado en distancias es adecuado para la investigación, ya que permite encontrar similitud entre cadenas idénticas y no idénticas debido a la pluralización, edición o mala digitación. (Jácome, Danny; Tapia, Freddy; Lascano, Jorge; Fuertes, Walter, 2019, February)

Técnicas basadas en sonidos y fonética

Se fundamentan en la percepción que se tiene de los sonidos representados en una letra o en una sílaba. Se asigna un valor a una secuencia de caracteres basados en el sonido que producen. Como lo expresa (Villaverde, Aranda, & Hernandez-Orallo, 2012), “se permite un cierto error en la búsqueda, dando lugar a cadenas que, si no son iguales, se parecen bastante al patrón de búsqueda”. El proceso de estas técnicas se expresa así:

“Dada una cadena sacada de un conjunto ‘S’ de posibles cadenas (compuestas por caracteres de un alfabeto ‘A’, el objetivo es encontrar una cadena ‘T’ que se aproxime a ‘S’. La tarea consiste en obtener el subconjunto ‘T’ de cadenas de ‘S’ que sean

suficientemente parecidas o encontrar las 'N' cadenas de 'T'" (Villaverde, Aranda, & Hernandez-Orallo, 2012)

Según (Shah & Singh, 2014), todas las coincidencias aproximadas se denominan "positivos", y el resto son "negativos", Si son relevantes, se llaman "verdaderos positivos". Aquellos no relevantes se consideran "falsos positivos".

Algoritmo Soundex.

Convierte cada nombre a una secuencia de cuatro caracteres para identificar nombres equivalentes. Las reglas para codificar un nombre en el algoritmo Soundex pueden verse en la Tabla 8. Mediante estas reglas, por ejemplo, el código S530, se asigna a Smith, Smeath, Smeith, Smiyth, Smitte, y Smett (Chakkrit & Bruckner, Thailand, 2009);

Tabla 8
Algoritmo Soundex

| Regla o Paso | Descripción |
|--|--|
| 1. Obtener la Letra Prefija | Recuperar la primera letra del nombre de entrada y utilizarla como letra prefija inicial |
| 2. Ignorar sonidos vocales, la H y la W. | Si las letras A, E, I, O, U, Y, H, W no son iniciales, ignorarlas completamente |
| 3. Aplicar el Código Soundex | Mientras el nombre de entrada no se haya convertido código Soundex de salida, y si es menor de 4 caracteres, convertir las otras letras. |
| 4. Generar la Salida Codificada | Obtener el nombre traducido a código Soundex |
| 5. Adecuar la Salida Codificada | Si el código Soundex de salida es menor de cuatro caracteres, agregar y anteponer ceros. De lo contrario, quitar los caracteres más a la |

derecha, (ignorando las letras que quedan a continuación) para darle el formato en secuencia: letra, dígito, dígito.

Nota: Tomado de (Chakkrit & Bruckner, Thailand, 2009)

La conversión a la forma canónica⁵ de los nombres se muestra en la Tabla 9, en la misma que se especifica la codificación numérica de acuerdo al grupo alfabético.

Tabla 9

Código Soundex

| Carácter | Codificación |
|------------------------|--------------|
| A, E, I, O, U, H, Y, W | 0 |
| B, P, F, V | 1 |
| C, G, J, K, Q, S, X, Z | 2 |
| D, T | 3 |
| L | 4 |
| M, N | 5 |
| R | 6 |

Nota: Tomado de (Russell, 1918)

Es evidente que este algoritmo empareja palabras distintas por la generalización de sonidos, “Soundex es únicamente útil en aplicaciones en que se pueda tolerar un alto número de falsos positivos y falsos negativos”, llegando a la conclusión que no es apto para la investigación, durante este proceso se necesita minimizar el número de falsos positivos. (Shah & Singh, 2014).

Algoritmo Metaphone

⁵ Modelo de características perfectas.

Codifica letras por las reglas más comunes de pronunciación en inglés: Para la letra ‘C’, distingue la pronunciación como ‘S’ y como ‘K’. Busca similitudes de palabras que se pronuncian igual, ignora las vocales después de la primera letra y reduce el alfabeto a dieciséis sonidos consonantes: B, X, S, K, J, T, F, H, L, M, N, P, R, Q, W, Y. Cada nombre se convierte a un código de 4 letras. (Chakkrit & Bruckner, Thailand, 2009), (Shanker Singh, 2003).

Tabla 10
Algoritmo Metaphone

| Regla o Paso | Descripción |
|---|--|
| 1. Depuración de caracteres | Eliminar caracteres no alfanuméricos y cambiar todas las letras a mayúsculas. |
| 2. Eliminar la letra inicial en algunos casos | Si el nombre comienza con KN, GN, PM, AE ó WR, eliminar la primera letra |
| 3. Reemplazo de una letra inicial | Si la letra X es inicial, cambiarla por S |
| 4. Reemplazo de una secuencia de letras iniciales | Si se presenta la secuencia de letras ‘WH’ como inicial, cambiarla a W. |
| 5. Aplicar el Código Metaphone | Convertir cada palabra del nombre a un código Metaphone de cuatro caracteres: <ul style="list-style-type: none"> 5.1 Ignorar letras duplicadas. 5.2 Si la A, E, I, O, U y la Y son iniciales del nombre, mantenerlas como parte del código, de otro modo eliminarlas. Y convertir a código las letras en dieciséis sonidos consonantes: B X S K J T F H L M N P R Q W Y según la Tabla de Códigos. |
| 6. Generar la salida codificada. | Obtener el nombre traducido a código Metaphone. |

Nota: Tomado de (Shanker Singh, 2003).

En vista de que el algoritmo original Metaphone contenía muchos errores, se crearon dos versiones: Doublé Metaphone y Metaphone-3. Esta última corrige miles de codificaciones erradas. (Haunts, 2014)

Aunque Metaphone-3, puesta en producción en octubre de 2009, introduce codificación adicional para palabras no escritas en inglés y de uso común en los Estados Unidos, aún está orientada a reglas de pronunciación inglesa. (Parmar & Kumbharana, 2014).

Las reglas del código fonético Metaphone se muestran en la Tabla 11.

Tabla 11
Código Metaphone

| Letra | Codificación | Regla / Restricción |
|-------|-----------------------|---|
| B | B | A menos que la letra B aparezca al final de una palabra después de una “m” como en “dumb”. |
| C | X, S, K | X: Si suena como ‘sh’, ‘cia’, ó ‘ch’ S: Si suena como ‘ci’, ‘ce’, ó ‘cy’ K: Si suena diferente, incluyendo el sonido de ‘sch’. |
| D | J, T | J: Si forma parte de ‘dge’, ‘dgy’, ‘dgi’ T: Si suena diferente a los casos mencionados |
| F | F | Mantener la misma letra |
| G | Dejar en blanco, J, K | Dejar en Blanco: Si es silenciada por ‘gh’ y no se encuentra al final o antes de una vocal. En ‘gn’, o en ‘gned’. En ‘dge’, como en la regla anterior J: Si aparece antes de la ‘i’, ‘e’, ó la ‘y’. Si no es una doble ‘gg’. K: Si suena diferente a los casos mencionados. |
| H | Dejar en blanco, H | Dejar en Blanco: Si sigue a una vocal o no vocal. |

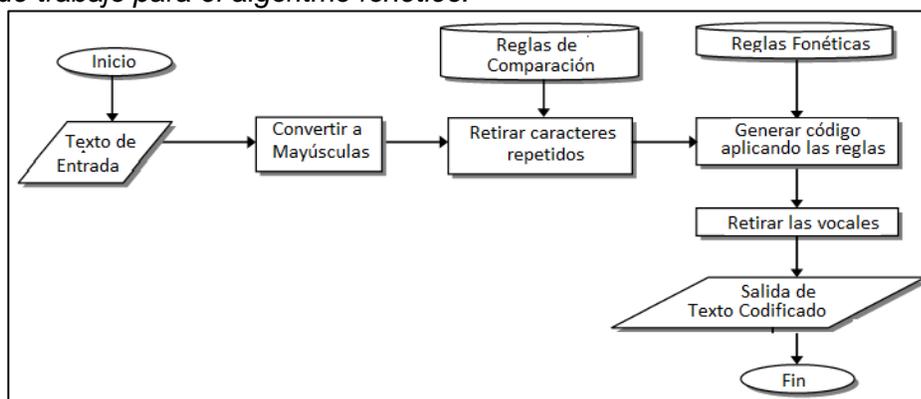
| | | |
|---|---------------------------------|---|
| | | H: En casos diferentes al anterior. |
| J | J | Mantener la misma letra |
| K | Dejar en blanco, K | Dejar en blanco: Si se produce un silencio luego de 'c' K: Si suena diferente al caso anterior. |
| L | L | Mantener la misma letra |
| M | M | Mantener la misma letra |
| N | N | Mantener la misma letra |
| P | F, P | F: si aparece antes de la 'h' P: En otros casos |
| Q | K | Cambiar |
| R | R | Mantener la misma letra |
| S | X, S | X: Si aparece antes de la 'h', como en 'sh', 'sio' ó 'sia' S: De otro modo, mantener la misma letra |
| T | X, 0 (cero), Dejar en blanco, T | X: Si suena como en 'tia' o 'tio' 0 (cero): Si aparece antes de la h, como en 'th' Dejar en Blanco: Si es silenciosa, como en 'tch' T: En otros casos. |
| V | F | Cambiar |
| W | Dejar en blanco, W | Dejar en Blanco: Si no está precedida de una vocal W: Si está seguida de una vocal. |
| X | KS | Cambiar |
| Y | Dejar en blanco, Y | Dejar en blanco: Si hay silencio y no es seguida por una vocal. Y: Si es seguida por una vocal |

Z S Cambiar.

Nota: Tomado de (Chakkrit & Bruckner, Thailand, 2009)

El flujo de trabajo para la codificación Metaphone se muestra en la Figura 5.

Figura 5.
Diseño de trabajo para el algoritmo fonético.



Nota: Tomado de (Parmar & Kumbharana, 2014).

El Algoritmo Metaphone y sus variaciones están diseñados para el idioma inglés, o pronunciación inglesa, un limitante enorme, esta diferencia entre las dos lenguas no permite una comparación de cadena de caracteres correcta cuando la palabra está escrita en lenguaje español, debido al código fonético utilizado.

- **Algoritmo NYSIIS**

El Algoritmo “New York State Identification and Intelligence System” (NYSIIS) se diferencia del método Soundex en cuanto a que retiene la información de vocales en una palabra codificada y convierte todas las vocales a letra “A”. La versión original trunca a 6 caracteres al final. (Stephens J, 2013)

Tabla 12
Algoritmo NYSIIS

| Regla o Paso | Descripción |
|--------------------|--|
| 1 Depurar el final | Quitar los caracteres ‘S’ y ‘Z’ del final del apellido |

| | | |
|----|--|---|
| 2 | Recodificar caracteres iniciales | Ejemplos de recodificación: MAC → MCC; PH → F; PF → FF |
| 3 | Recodificar caracteres en posiciones intermedias | Ejemplos de recodificación: IX → IC; EX → EC; YE, EE, IE → Y; NT, ND → D |
| 4 | Recodificar caracteres que no están al inicio del nombre | Ejemplos de recodificación EV → EF |
| 5 | Usar el primer carácter como clave | La primera letra del nombre sirve como clave de índice de búsqueda |
| 6 | Quitar un carácter en el nombre según el caso | Quitar la letra 'W' si esta sigue a una vocal |
| 7 | Remplazar vocales | Reemplazar todas las vocales por la 'A' |
| 8 | Recodificar | GHT → GT; DG → G; PH → F |
| 9 | Eliminar un carácter | Si no es el primer carácter, eliminar todas las letras 'H' que precedan o sigan a una vocal. |
| 10 | Recodificar con condición | Cambiar KN → N; o K → C Si no es el primer carácter, cambiar M → N Si no es el primer carácter, cambiar Q → G |
| 11 | Recodificar | SH → S; SCH → S; YW → Y |
| 12 | Cambiar con condición | Si no es el primer o último carácter, cambiar Y → A |
| 13 | Recodificar | WR → R |
| 14 | Cambiar con condición | Si no es el primer carácter, cambiar Z → S |
| 15 | Recodificar al final | Si aparece al final del nombre: AY → Y |
| 16 | Quitar las vocales | Quitar las vocales que sigan a continuación |

| | | |
|----|--------------------|---|
| 17 | Reducir caracteres | Reducir las cadenas con caracteres repetidos |
| 18 | Agregar caracteres | Si el primer carácter del apellido original fue una vocal, agregarlo al código. |

Nota: Tomado de (Black, 2009); (Ameri, S., Vahdati, S., & Lange, C., 2017, September)

Es un algoritmo planteado únicamente para el idioma inglés, debido a la codificación de sonidos actual el algoritmo no se lo puede aplicar en palabras escritas en el idioma español, la pronunciación entre español e inglés es muy diferente.

Algoritmo Phonex

Es una combinación de los métodos usados en los algoritmos Soundex y Metaphone. Según (Chakkrit & Bruckner, Thailand, 2009), mantiene un buen rendimiento cuando es aplicado a nombres escritos en inglés. Fue desarrollado por A.J. Lait y B. Randell y convierte cada palabra en un código equivalente a cuatro caracteres usando los pasos de la Tabla 13.

Tabla 13
Algoritmo Phonex

| Regla o Paso | Descripción |
|--------------|--|
| 1 | <p>Pre-procesar el nombre</p> <p>1) Quitar todos caracteres 'S' al final del nombre</p> <p>2) Convertir los pares de letras al inicio: KN→N; WR→ R; PH → F</p> <p>3) Convertir las letras solas a su equivalente fonético</p> <p>H → quitar la H muda</p> <p>E, I, O, U, Y → A; K, Q → C</p> <p>P → B; J → G; V → F; Z → S</p> |

- | | | |
|---|---|---|
| 2 | Codificar el nombre pre-procesado según reglas | <p>4) Mantener la primera letra del nombre, y quitar las ocurrencias de A, E, H, I, O, U, W, Y en otras posiciones.</p> <p>5) Asignar los siguientes números a las letras que quedan después de la letra inicial:</p> <p>B, F, P, V → 1</p> <p>C, G, J, K, Q, S, X, Z → 2</p> <p>D, T → 3 (Si no está seguida por la C)</p> <p>L → 4 (Si no está seguida por vocal o está al final)</p> <p>M, N → 5 (Ignorar la próxima letra si es D o G)</p> <p>R → 6 (Si no está seguida por vocal o está al final)</p> <p>6) Ignorar la letra actual si tiene el mismo dígito que el último carácter del código</p> |
| 3 | Convertir el código a la forma: letra, dígito, dígito | <p>Añadir ceros antes de los dígitos (si hay menos de tres dígitos), o eliminando los dígitos más a la derecha si es que son más de tres.</p> |

Nota: Tomado de (Chakkrit & Bruckner, Thailand, 2009)

La codificación que mantienen los algoritmos fonéticos, no están planteados para el idioma español, razón por la cual no son aplicables en forma directa a procesos donde el idioma utilizado es el español, para un manejo adecuado de la información con estos algoritmos se deberían plantear nuevas codificaciones, con el fin de satisfacer las necesidades fonéticas en el lenguaje español.

Técnicas Compuestas

Se basan en variaciones que se hacen al interior de los algoritmos de las técnicas anotadas en el apartado anterior; o bien, en combinar dos técnicas a la vez.

Variaciones al Algoritmo de Levenshtein

Distancia TFIDF: Aplica una medida de distancia general a dos documentos que contienen nombres a emparejar. Se basa en el grado de relevancia que tiene un nombre dentro de cada documento y mide la proximidad que existe entre dos documentos para cada nombre o término que se compara. Una simplificación de la fórmula de Distancia TF-IDF es la siguiente:

$TF_{(t,i)}$ (Del inglés “term frequency”) mide el grado de relevancia que tiene el término t en todo el documento i (d_i). Equivale a la frecuencia de t en d_i .

$IDF_{(t,i)}$ (Del inglés “inverse document frequency”) mide que tan común o que tan raro es el término t dentro del documento i (d_i), donde $|D|$ corresponde al valor absoluto del total de términos en d_i .

$$IDF_{(t,i)} = \log \frac{|D|}{|\{d_i \in D: t \in d_i\}|} \quad (\text{Ecuación 4})$$

$$TFIDF_{t,i} = TF_{t,i} \times IDF_t \quad (\text{Ecuación 5})$$

Para los documentos d_i y d_j , la proximidad respecto a t , se calcula como:

$$proximidad(d_i, d_j) \text{ para } t = \text{sum}(TFIDF(t, i) * TFIDF(t, j)) \quad (\text{Ecuación 6})$$

En (Cohen, Ravikumar, & Fienberg, 2003, August) se hace un análisis comparativo a detalle entre este y otros algoritmos de distancia.

Damereau-Levenshtein: Calcula la distancia más corta entre una palabra y otra independientemente de las alteraciones que se hayan hecho para equiparar caracteres y

| | | | | | | | |
|---|---|---|---|----------------------------|-----|---|-----|
| 4 | S | 4 | - | Inserción | 1 | T | 0 |
| 5 | T | 4 | T | Coincidencia (Sustitución) | 0 | R | 0,5 |
| 6 | E | 5 | R | Transposición | 0,5 | E | 0,5 |
| 7 | R | 6 | E | Transposición | 0,5 | | |

Nota: Tomado de (Erikson K., 1999, pp. 20)

Erikson (1999), ilustra con un ejemplo el cálculo de distancia de Damereau-Levenshtein, resumido en la Tabla 15. La mínima distancia de edición para el ejemplo es 4. Nótese que la transposición tiene un peso menor que las otras operaciones puesto que es un caso especial de eliminación e inserción.

Damereau-Levenshtein, permite identificar palabras que tienen el mismo significado, pero su escritura es algo diferente, por conceptos como género o pluralización, por ejemplo: “bueno”, “buena”, “buenos” y “buenas”, son 4 palabras que mantienen un mismo significado, pero la escritura varia, mediante el algoritmo y el truncamiento de distancias mínimas se identifica las coincidencias o similitudes entre dos palabras.

Jaro: Calcula una medida de similitud usando el número de caracteres comunes (los caracteres que son iguales dentro de la mitad de la longitud de la cadena más larga) y el número de transposiciones. (Schay, 2011).

Jaro & Winkler: Mejora el algoritmo de Jaro, basado en estudios que concluyen que la menor cantidad de errores de equiparación se producen al inicio de los nombres. (Schay, 2011).

Es importante mencionar que los algoritmos de Jaro y su variación tiene gran capacidad para comparar similitud en cadenas de caracteres, pero, no alcanzan el rango de los algoritmo Damereau-Levenshtein, al limitar la comparación únicamente a las letras iniciales y dentro de la mitad de la palabra, tomando en cuenta que las palabra en el

lenguaje español puede cambiar su estructura al final, pero aún mantiene su significado, por ejemplo: “neutral”; “neutrales”, mientras Damereau-Levenshtein, se enfoca en la palabra completa.

Editex: Combina técnicas de distancia con fonéticas. Aplica la distancia Damereau-Levenshtein y agrupan letras según los códigos Soundex y Phonix. Asigna un valor de cero si encuentra dos letras iguales; 1 si las letras están en el mismo grupo, y de 2 en otro caso. También trata a las letras h y w como casos especiales, ya que en el idioma inglés estas letras no se pronuncian. (Schay, 2011).

Distancia por alineación de Sílabas: Busca coincidencias yendo de sílaba en sílaba, antes que de carácter en carácter. Usa las transformaciones del algoritmo Phonix y luego aplica reglas para encontrar el comienzo de las sílabas. La distancia entre dos cadenas de sílabas las calcula usando un algoritmo de distancia del tipo Levenshtein.

Algoritmo SOUNDEX-SP: Es una adaptación del algoritmo Soundex propuesto y mejorado por (Fernández L., 2010) en la Web para el idioma español. Tiene como pasos básicos los mostrados en la Tabla 16.

Tabla 16
Algoritmo Soundex-SP

| Regla | Descripción |
|--------------|---|
| 1 | Retener la primera letra de la cadena. Tener en cuenta las letras dobles como CH y LL. |
| 2 | Remover todas las ocurrencias de las letras siguientes a partir de la segunda posición: a, e, i, o, u, h, w, y (cuando suena como vocal i) |
| 3 | Asignar números a las siguientes letras (luego de la primera): <ul style="list-style-type: none"> • b, f, p, v = 1 • c, g, j, k, q, s, x, z = 2 |

- d, t = 3
- l = 4
- m, n = 5
- r = 6
- ll, y, ch = 7

4 Si hay números consecutivos, dejar solamente uno en la serie

5 Retornar los cuatro primeros caracteres, si son menos de cuatro completar con ceros

Nota: Tomado de (Fernández L, 2010)

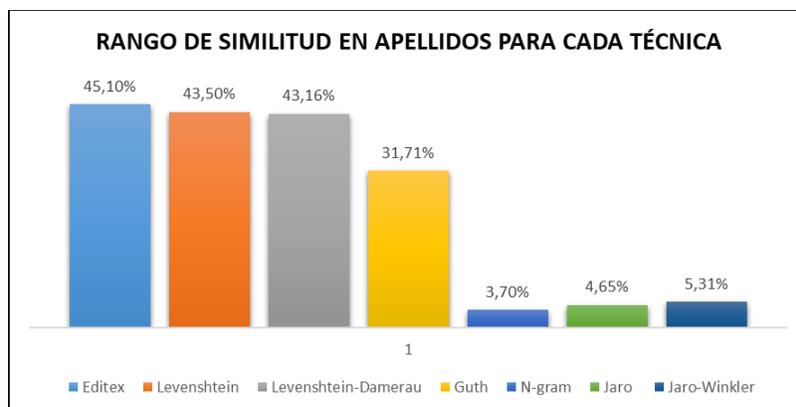
Una adaptación más reciente, propuesta por (Mazariegos O, 2012), asigna el número 7 a las letras Q y J y reduce la 'CH' a 'V' y la 'LL' a 'J'.

Es necesario y relevante mencionar que los algoritmos compuestos, tiene como base el mejoramiento de sus predecesores, pero el enfoque principal está en satisfacer, la utilización de los mismo es diferentes idiomas, existe una infinidad de variación en cuando a pronunciación entre el lenguaje español y el inglés, y más aun con otros idiomas.

Como se pudo observar algunos de los algoritmos son muy eficientes para determinar la coincidencia entre cadenas de caracteres en el idioma inglés, y basados en la investigación realizada por (Loyola, 2015) y (Jácome, Danny; Tapia, Freddy; Lascano, Jorge; Fuertes, Walter, 2019, February), donde se evalúa diferentes tipos de algoritmos de deletreo y distancia, fonéticos y mixtos para identificar coincidencia entre cadenas de caracteres, expresan que los algoritmos más eficientes para la identificación de coincidencias son Editex y Levenshtein con sus variaciones, en el estudio de Loyola, los porcentajes de similitud para estas técnicas sobre pasan

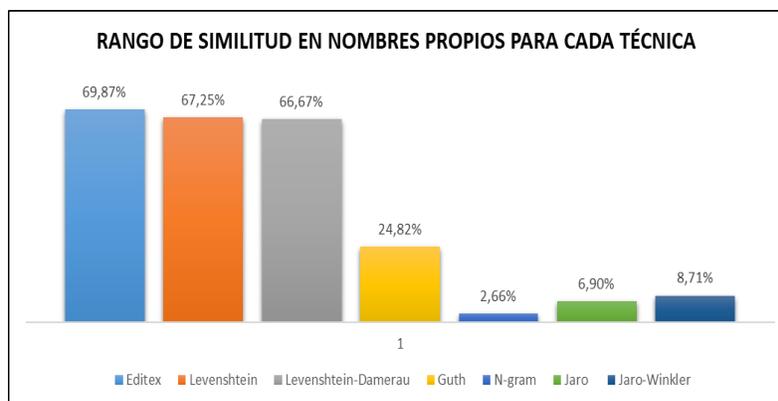
el 95%, en la Figura 6 y 7, se pueden observar los valores de similitud obtenidos por las diferentes técnicas en estudio.

Figura 6.
Porcentaje de coincidencias en apellidos.



Nota: Tomado de (Loyola, 2015)

Figura 7.
Porcentaje de coincidencias en nombres propios.



Nota: Tomado de (Loyola, 2015)

Es importante mencionar que la técnica Editex es una técnica combina o mixta la cual introduce una verificación de sonidos a las operaciones de distancia realizadas por las técnicas de Levenshtein y Dumerau, con esta aclaración y lo mencionado en el trabajo de (Jácome, 2019), para analizar información en idioma español, las técnicas más adecuadas son Levenshtein y Guth, por medio de los cuales se puede encontrar coincidencias “Exactas y Parciales” después de realizar operaciones de edición, sustitución, eliminación y transposición, lo cual es muy

importante para la presente investigación ya que el estudio se enmarca en el análisis de coincidencias en el idioma castellano.

Trabajos similares con PLN

Según (Jácome, Danny; Tapia, Freddy; Lascano, Jorge; Fuertes, Walter, 2019, February), las redes sociales como Facebook, son una fuente con mucha información la cual puede ser útil para diferentes campos de trabajo, en este artículo se hace un estudio sobre el grado de confiabilidad que tienen las FanPages⁶ de Facebook dirigidas a ecuatorianos, mediante el procesamiento de lenguaje natural (PLN), buscan patrones para identificar comentarios positivos o negativos hacia dichas FanPages y de esta manera atribuir cualidades efectivas y desfavorables, para llegar a determinar si es confiable o no.

En el artículo planteado por (Martel, Carranco, & Cevallos, 2016), se analiza la agresividad que se genera en las redes sociales, para este estudio se utiliza el contenido de Facebook como banco de datos, mediante la herramienta R se realiza el análisis a través del logaritmo incorporado "stringsim", que pertenece al paquete "stringdist" que está basado en el algoritmo de Levenshtein, teniendo como resultado del algoritmo un valor de probabilidad de coincidencia o similitud entre las dos cadenas de caracteres comparadas, lo que no se menciona en este trabajo es la cantidad de arreglos máximos para cada comparación entre cadenas y el valor que se asigna a cada tipo de arreglo como inserción, eliminación, etc.

Varias técnicas de coincidencia de caracteres, así como fonéticas, las cuales están fundamentadas en "tokens" e "híbridas", fueron comparadas y puestas a pruebas, durante un estudio realizado por (Bilenko, Mooney, Cohen, Ravikumar, & Fienberg, 2003), en este estudio se menciona el valor del umbral con el que se trabaja, pero sin mencionar como fue calculado

⁶ Se asocia a empresas o personas que crean y diseñan estrategias de marketing a través de Facebook

dicho valor para cada técnica, además no se tomaron en cuenta los errores tipográficos, la tasa de error y el tamaño de la muestra. Al finalizar la evaluación ellos afirman que Monge-Elkan tiene mejor promedio, pero Soft-IDF tuvo mejor desempeño que el resto de las técnicas desde una mirada general.

En un análisis de precisión para diferentes técnicas de comparación de palabras, basadas en tokens, se utilizan distintos grupos de datos para observar su comportamiento, para este análisis se tomaron en cuenta los errores de edición y reemplazo, (Hassanzadeh, Sadoghi, & Miller, 2007, September), mencionan que el valor del umbral si afecta el rendimiento de cada una de las técnicas, pero al igual que en el estudio realizado por Bilenko, no se tomó en cuenta los errores tipográficos y la cantidad de datos en cada uno de los grupos de prueba.

(Peng, Li, & Kennedy, 2012) Presenta una comparativa de técnicas de coincidencia de cadenas de caracteres, durante este trabajo se evaluó el rendimiento de las diferentes técnicas, para lo cual se tomaron en cuenta algunos factores que no fueron mencionados en los estudios realizados por Hassanzadeh y Bilenko, factores que tiene influencia en los resultados de cada técnica. Al finalizar el estudio Peng menciona que no hay una técnica mejor, en general, Jaro-Winkler y Jaro funcionan mejor que los demás en la coincidencia de nombres. El trabajo también afirmó que la tasa de error en el conjunto de datos tiene efecto sobre los valores del umbral. Sin embargo, no consideran nombres de pila e igual que Hassanzadeh no consideraron los errores tipográficos o de digitación.

Basados en una amplia revisión de las investigaciones existentes en el campo de estudio, coincidimos con varios autores que no existe una técnica que abarque todos los problemas o tareas en la comparación de cadenas de caracteres, más aún cuando se trata de palabras del

idioma castellano, teniendo en cuenta que este trabajo de investigación se lo va a realizar en un país de habla hispana Ecuador, también se debe mencionar que existen ciertas palabras las cuales son muy utilizadas pero no son parte de un diccionario, neologismos y modismos los cuales pueden cambiar el sentido de una oración, su uso y significado depende del lugar geográfico, por ejemplo: en la costa ecuatoriana “Cana” es utilizada para hacer referencia a la cárcel, mientras que en la región sierra es utilizada generalmente para referirse a un cabello de color blanquecino.

Al igual que no existe una técnica exclusiva para coincidencia de caracteres, tampoco hay una única forma de calificar o evaluar la personalidad de un individuo dentro de las redes sociales, tomando en cuenta que un perfil dentro de una red social es un avatar que cada persona crea con ciertos rasgos únicos, por medios de los cuales expresa la forma en que quiere ser tratado dentro de la sociedad en la cual convive. Lo cual permite identificar patrones en su actuar, el tipo de carácter o temperamento, esto a su vez posibilita el análisis de uno de ellos para la identificación de una presunta personalidad.

Capítulo III

Desarrollo de la propuesta

La investigación propuesta para el trabajo planteado es de tipo exploratorio ya que consiste en proveer un análisis bibliográfico referencial que permita determinar la situación actual y posibles trabajos futuros en relación a la problemática planteada. lo que permitirá llegar a las conclusiones basándose en experimentos donde se reflejarán los hechos de una forma cuantitativa, estos experimentos se llevarán a cabo con varios grupos homogéneos de estudiantes pertenecientes a la unidad educativa “Sociedad Educativa Futuro SEF” del cantón Rumiñahui, lo que permitirá contar con resultados fiables, este proceso se apoyara en la opinión de un psicólogo con conocimientos y un test de conducta, lo que validara la propuesta desde la perspectiva del campo ocupacional relacionado al tema.

La metodología que se utilizó para el desarrollo del aplicativo está basada en prototipos, cumpliendo con las actividades de las metodologías tradicionales como son las actividades de análisis, diseño, implementación y evaluación, es importante mencionar que el proceso es repetitivo y en el cual se realizan modificaciones al prototipo, llegando a tener prototipos útiles lo cual se puede logra manteniendo comunicación, organización y control. (Mayhew, 1992).

Es necesario mencionar ciertas consideraciones tomadas en cuenta y los recursos con los cuales se cuenta para el desarrollo del aplicativo, seguido de esto se detalla el análisis, diseño e implementación del aplicativo propuesto, que a partir de ahora se le denominará CERO_ACOSO.

Planteamiento de un corpus de palabras

Es necesario plantear un corpus como base para el esquema, diseño y análisis de patrones relacionados con el ciberacoso, para alimentar el corpus se realizó un proceso de observación sistemático, basado en una investigación empírica, en este corpus de palabras estarán presentes silogismos y neologismo los cuales tendrán un significado diferente de acuerdo a la región o ubicación geográfica, es necesario aclarar que la investigación se enfoca en la región sierra del Ecuador específicamente en la provincia de Pichincha.

En la Figura 8, se puede observar el proceso de selección de palabras influyentes para el estudio, como primera fase se identificaron 10 perfiles de usuarios de Facebook al azar, pero que cumplan con las siguientes especificaciones; Edad: 11-18 años, País de origen (Nacionalidad): Ecuador/ecuatoriana y Ubicación geográfica: Pichincha-Quito. En la segunda fase bajo la colaboración del encargado del DECE del colegio “SEF”, psicólogo especializado en consejería estudiantil se observaron los perfiles y se identificaron palabras agresivas utilizadas por los jóvenes en la red social, en la tercera fase se realiza un cotejamiento de las palabras encontradas por los observadores (34 en total y 8 repeticiones como máximo) y por medio de un análisis del índice de aceptación “*indacep*” con un valor mayor o igual al 24% de apariciones entre todos los perfiles, la palabra es admitida, Ecuación 8, y como fase final se obtiene un corpus de palabras con las palabras agresivas más ocupadas por los jóvenes en la red social, el total de palabras admitidas es de 17 las cuales se detallan en la Tabla 17.

$$indacep = \frac{Mi}{Pt} * 100\% \quad (\text{Ecuación 8})$$

En donde: “*M*” es el número máximo de incidencias de una palabra y “*P*” es la muestra total de palabras agresivas encontradas.

Figura 8.

Proceso obtención corpus de palabras.

**Tabla 17**

Corpus de palabras

| Palabras o | Significado |
|--------------------|---|
| Abreviatura | |
| Idiota | Tonto o corto de entendimiento. |
| Estúpido | Necio, falta de inteligencia. |
| Mudo | Falto de entendimiento o de razón. |
| Tonto | Que padece una deficiencia mental. / Que carece de sentido o de motivo. |
| Bruto | Necio, incapaz. |
| Verga | Utilizada para expresar sorpresa, protesta, disgusto o rechazo. |
| Pendejo | Tonto, estúpido. / Cobarde, pusilánime. |
| Animal | Persona ruda y de poco entendimiento. |

| | |
|------------|---|
| Marica | Dicho de un hombre: Apocado, falta de coraje, pusilánime o medroso. |
| Gay | Dicho de una persona, especialmente de un hombre: homosexual. |
| Lesbiana | Dicho de una mujer: homosexual. |
| Homosexual | Dicho de una persona: Inclined sexualmente hacia individuos de su mismo sexo. |
| Puta | Forma de abreviar la palabra prostituta |
| Prostituta | Persona que mantiene relaciones sexuales a cambio de dinero. |
| Mmv | Mama verga |
| Hp | Hijo de puta |
| Vcv | Ve cara de verga |

Recursos.

Recursos humanos

Para el desarrollo del aplicativo de escritorio se cuenta con un estudiante egresado de la Carrera de Ingeniería en Sistemas e Informática de la Universidad de las Fuerzas Armadas ESPE descritos en la Tabla 18:

Tabla 18
Recursos humanos

| Nombres | Cargo | Estudios |
|-------------------------------|---------------|--------------------------------------|
| Danny Alexander Jácome Andino | Desarrollador | Ingeniería en Sistemas e Informática |

Recursos de hardware

Para el desarrollo del aplicativo se cuenta con un computador personal con las especificaciones detalladas en la Tabla 19.

Tabla 19*Recursos de hardware*

| Código | Marca | Procesador | Sistema Operativo | RAM | Capacidad |
|---------------|--------------|-------------------|--------------------------|------------|------------------|
| PC01 | Asus | Intel Core i7 | Windows | 16GB | 1TB |

Análisis.

A continuación, en el apartado 3.1.1 se presentan los requerimientos identificados los cuales establecen la esencia de la aplicación.

Requerimientos de software

Los requerimientos de software fueron licitados en base a las necesidades que se presentaron durante el proceso de la presente investigación.

Tabla 20*Requerimiento 1*

| Código | Requerimiento_001 |
|---------------|--|
| Nombre | Listar los sujetos de la muestra. |
| Descripción | Se extrae de la base de datos todos los registros de persona existentes. <ul style="list-style-type: none"> • Datos de extraídos: <ul style="list-style-type: none"> ○ Código de persona. ○ Nombre de persona. (seudónimo) |

Tabla 21*Requerimiento 2*

| Código | Requerimiento_002 |
|---------------|--|
| Nombre | Extracción de comentarios por persona. |

Descripción Una persona es un registro de la base datos y tiene asociada la tabla comentarios donde se almacena la descripción de los comentarios hecho o recibidos.

- Datos de extraídos:
 - Código de persona.
 - Código de comentario.
 - Descripción de comentario.

Tabla 22

Requerimiento 3

| Código | Requerimiento_003 |
|-------------|--|
| Nombre | Análisis de información por persona. |
| Descripción | Se debe realizar un análisis detallado de los comentarios extraídos, comparándolos con el corpus de palabras. <ul style="list-style-type: none"> • Como resultado del análisis se informa si el sujeto analizado es un presunto agresor o no. |

Tabla 23

Requerimiento 4

| Código | Requerimiento_004 |
|-------------|--|
| Nombre | Generar reporte |
| Descripción | Al finalizar el análisis de la información se debe generar un reporte con las similitudes encontradas. <ul style="list-style-type: none"> • Como resultado del análisis se debe reportar cuantas coincidencias se encontraron a lo largo de todo el análisis. |

Diagrama de casos de usos

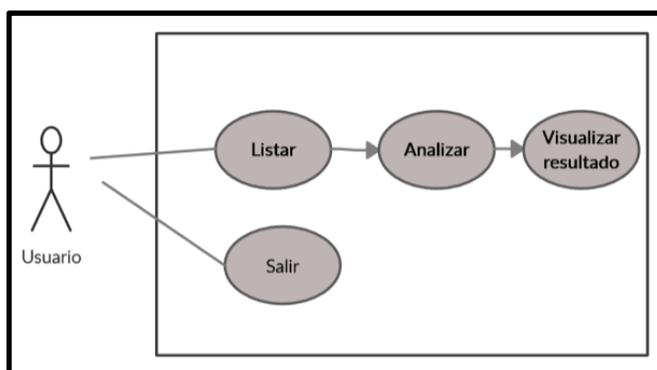
Un diagrama de casos de uso es un conjunto de escenarios que identifican la naturaleza de los usos para el sistema que se va a construir. Los diagramas de casos de usos consisten en describir la interacción del usuario con la aplicación.

La propuesta de la presente investigación señala un único actor, este puede realizar una comparación de cadenas de caracteres para identificar similitudes entre la información descargada y el corpus de palabras planteado, para lo cual deberá listar los participantes existentes, seleccionar uno de ellos, empezar con la comparación de caracteres, visualizar los resultados obtenidos y salir de la aplicación.

Es importante mencionar que algunas de las acciones se deben realizar de forma secuencial, mientras que otras no, tal como se indica en la figura 9.

Figura 9.

Diagrama de arquitectura.



Casos de usos detallado

En la siguiente sección se puede visualizar el diagrama de casos de uso detallado para cada Requerimiento, en los cuales se puede observar el proceso que se llevara a cabo para cumplir con cada uno de los requerimientos antes identificados, los casos de uso detallados

tienen gran similitud de funcionamiento esto debido a que son secuenciales como se pudo observar en el apartado anterior.

Tabla 24

Caso de uso: Listar participantes / requerimiento_001

Características

| | |
|---------------------|---|
| Resumen | EL usuario podrá visualizar la lista de participantes, que se encuentra en la base de datos |
| Pre-Condición | Al menos un participante debe ser parte de la B.D. |
| Post-Condición | Podría seleccionar a los participantes |
| Actor Principal | Usuario |
| Actores Secundarios | |

Curso Normal

| Paso | Actor | Descripción |
|------|------------|--|
| 1 | Usuario | El usuario lista los participantes, botón "Listar" |
| 2 | Aplicación | El sistema muestra todos los sujetos ingresados en BD. |

Tabla 25

Caso de uso: Listar y analizar información / Requerimiento_002_003

Características

| | |
|---------------------|---|
| Resumen | El usuario podrá analizar la información descargada de uno de los sujetos de la muestra |
| Pre-Condición | El usuario deberá listar los sujetos |
| Post-Condición | Se activará el botón para visualizar los resultados |
| Actor Principal | Usuario |
| Actores Secundarios | |

Curso normal

| Paso | Actor | Descripción |
|------|------------|---|
| 1 | Usuario | El usuario lista los sujetos. |
| 2 | Aplicación | El sistema muestra todos los sujetos ingresados en BD. |
| 3 | Usuario | El usuario debe seleccionar uno de los sujetos de la lista. |
| 4 | Usuario | El usuario debe iniciar el proceso de análisis con el botón "Empezar" |
| 5 | Aplicación | Deberá procesar la información realizando comparaciones de edición de distancia entre la información del sujeto y el corpus planteado mediante el algoritmo de Levenshtein. |
| 6 | Aplicación | Informa por medio de una caja de texto, si el sujeto que estuvo bajo análisis, es un presunto agresivo o pasivo. |

Curso alternativo y extensiones del curso normal

| Paso | Actor | Descripción |
|------|---------|--|
| 7 | Usuario | Si el sujeto seleccionado no tiene información no se activará el botón de reporte, el usuario deberá listar nuevamente a los sujetos. Ir al paso 3 |

Tabla 26

Caso de uso: Generar Reporte / Requerimiento_004

Características

| | |
|---------------------|---|
| Resumen | El usuario podrá visualizar el número de repeticiones que se encontraron por cada ítem del corpus |
| Pre-Condición | El usuario deberá empezar el análisis de coincidencias |
| Post-Condición | Ninguna/ podrá regresar al Requerimiento_001 |
| Actor Principal | Usuario |
| Actores Secundarios | |

Curso normal

| Paso | Actor | Descripción |
|------|------------|--|
| 1 | Usuario | El usuario lista los sujetos. |
| 2 | Aplicación | El sistema muestra todos los sujetos ingresados en BD. |

- | | | |
|---|------------|---|
| 3 | Usuario | El usuario debe seleccionar uno de los sujetos de la lista. |
| 4 | Usuario | El usuario debe iniciar el proceso de análisis con el botón "Empezar" |
| 5 | Aplicación | Deberá procesar la información realizando comparaciones de edición de distancia entre la información del sujeto y el corpus planteado mediante el algoritmo de Levenshtein. |
| 6 | Aplicación | Se activará el botón para generar el reporte con el número de coincidencias encontradas. |
| 7 | Usuario | El usuario podrá acceder al reporte de coincidencias al dar clic en el botón "Reporte", como resultado la aplicación le mostrará el número de coincidencias encontradas por cada ítem del corpus. |

Curso alternativo y extensiones del curso normal

| Paso | Actor | Descripción |
|------|---------|--|
| 8 | Usuario | Si el sujeto seleccionado no tiene información no se activará el botón de reporte, el usuario deberá listar nuevamente a los sujetos. Ir al paso 3 |

Diseño

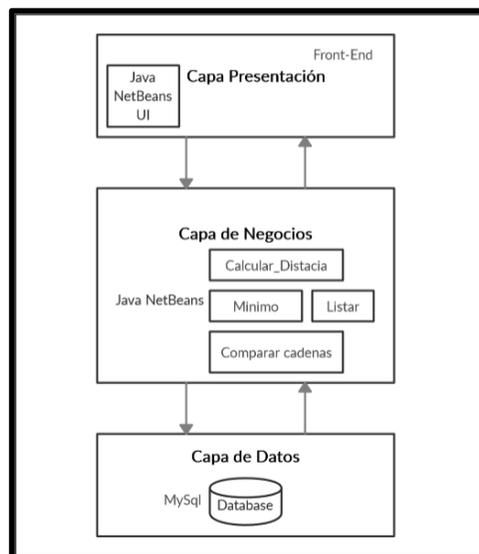
Una vez identificados los requerimientos y después de definir el diagrama de casos de usos para el usuario, se presentan los diagramas de arquitectura, con el fin de mostrar la operatividad del aplicativo.

Diagrama de arquitectura

Un diagrama de arquitectura ayuda a plantear una vista completa del sistema que se va a construir (Pressman, 2010), es decir permite organizar desde una visión general cada proceso en la creación de código.

En la figura 10 se puede observar la arquitectura general con la que fue diseñada la aplicación bajo la arquitectura por capas, el front-end que es en la interfaz de usuario de NetBeans en la capa presentación, se comunica con la capa negocios en donde tenemos el back-end de la aplicación en donde se desarrolla las diferentes funciones, y por último tenemos la capa datos mediante la cual se puede realizar la conexión entre la base de datos y el aplicativo.

Figura 10.
Diagrama de arquitectura.

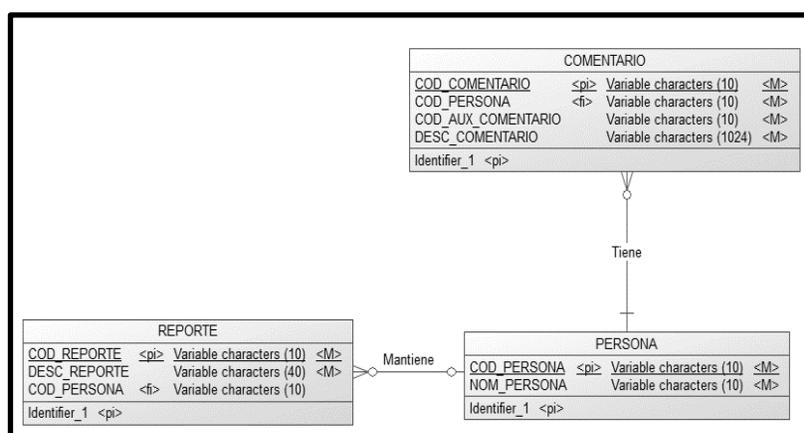


Modelo de base de datos

El modelado de la base de datos ayuda a estructurar de forma adecuada los datos, determinando el modo organizar la información y como tratarla desde y hacia la base de datos. A continuación, en la figura 11 se puede observar el modelo lógico de la base de datos donde se detalla las entidades y la relación que existe entre ellas al igual que los atributos y las características que cada uno de ellos tiene.

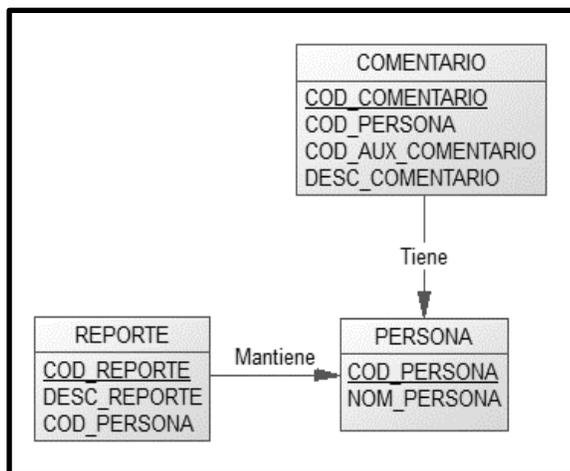
Figura 11.

Base de datos modelo lógico.



En la figura 12 se puede observar el modelo físico de la base de datos mediante el cual se puede entender la cardinalidad que tiene las tablas (comentario, persona y reporte), mismas que conforman la base de datos.

Figura 12.
Base de datos modelo físico.



Arquitectura de componentes.

Implementación

Es necesario aclarar que después de realizar el estudio del arte, se pudo observar que el Procesamiento de Lenguaje Natural es muy amplio, por lo cual únicamente se trabajó con el algoritmo de Levenshtein basado en la distancia de edición.

Selección de Herramientas.

El software utilizado para el desarrollo, su descripción y versión son detallados en la siguiente tabla.

Tabla 27
Recursos de software

| Software | Descripción | Versión |
|-----------|---|---------|
| Apache | Servidor web | 2.4.41 |
| Graph Api | Aplicación web propia de Facebook, gestión de servicios | 5.0 |

| | | |
|-----------------------|--|-------|
| MySQL / phpMyAdmin | Base de datos / administración de MySQL | 4.9.2 |
| NetBeans | IDE | 8.0.2 |
| RestFB | Librería que permite la conexión de Api Graph de Facebook con Java NetBeans. | 5.02 |
| XAMPP | Software libre que actúa como servidor | 3.2.4 |

- **IDE de desarrollo NetBeans**

NetBeans (Oracle, NetBeans IDE, 2017) es una herramienta de desarrollo integrado libre, su principal lenguaje de programación es Java, con esta herramienta es posible conectarse a bases de datos. Por lo cual se utilizará para el desarrollo de la aplicación que realizará el procesamiento del lenguaje natural.

Figura 13.
Logo NetBeans



- **XAMPP**

XAMPP (Apache, Friends, 2019) software libre, que actúa como servidor web, el mismo que está disponible para diferentes tipos de sistema operativo como Microsoft Windows, GNU/Linux, Solaris y Mac OS X, permite la gestión de bases de datos MySQL, servidor web Apache y la interpretación de lenguajes de programación script PHP y Perl. En este proyecto

XAMPP se utilizará para la conexión entre el aplicativo de escritorio y la base de datos creada en MySQL.

- **MYSQL**

MySQL (Oracle, MySQL, 2017) es un sistema de gestión de bases de datos relacional y que se la puede conseguir de manera Open Source. Esta herramienta permite guardar los datos a obtener de los perfiles de Facebook para posteriormente ser analizados con procesamiento de lenguaje natural. Los datos obtenidos de un perfil de Facebook se pueden guardar de una forma ordenada y con relación a diferentes tablas.

Figura 14.

Logo MySQL / phpMyAdmin



- **API Graph Facebook**

API Graph (Facebook, 2019) es una herramienta propia de Facebook que permite a otras aplicaciones establecer conexión con los perfiles de usuario, permitiendo interactuar bajo las acciones de lectura y escritura, para acceder a esta herramienta se debe ser parte de Facebook y registrarse como desarrollador. En la investigación se la utiliza como puente entre la aplicación de escritorio y los perfiles de usuario y a su vez los comentarios o post de cada usuario.

- **Librería RestFB**

RestFB (Restfb, 2019) es un cliente de API Graph de Facebook, simple y flexible para aplicaciones de escritorio realizadas en Java. Se trata de software de código abierto liberado bajo los términos de la Licencia MIT.

RestFB sólo necesita usar un método para obtener información y publicar nuevos elementos en Facebook. Por medio de esta librería obtendremos la conexión directa al API Graph de Facebook y a la información “comentarios” existente dentro de los perfiles de Facebook.

Figura 15.
Logo RestFB.



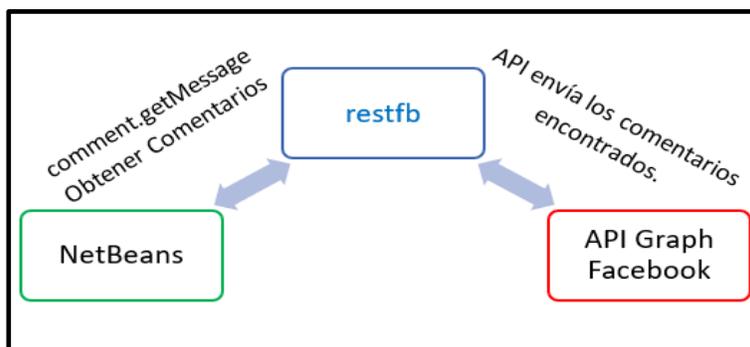
Procesos separados

Tres procesos separados de la función principal fueron esenciales para la implementación del aplicativo: la descarga de los datos desde Facebook, la limpieza y el almacenamiento de los mismos.

Descarga de datos (Comentarios)

Proceso que permite la comunicación entre la API de Facebook y la herramienta de desarrollo NetBeans, se lo realiza a través de la librería *restfb*, creando un puente para poder utilizar las funciones de la API. Para conectarse con la API es necesario generar un *AccesToken*, el mismo que permitirá mantener la conexión abierta y consecuentemente descargar los comentarios desde Facebook.

Figura 16.
Petición del aplicativo al API.



En la figura 18 se puede observar el código que se utilizó para la extracción de los datos de los diferentes perfiles de Facebook, esta acción se realizó desde NetBeans es importante mencionar que los datos confidenciales fueron reemplazados con “XXXXX” por seguridad.

Figura 17.
Código extracción datos de Facebook.

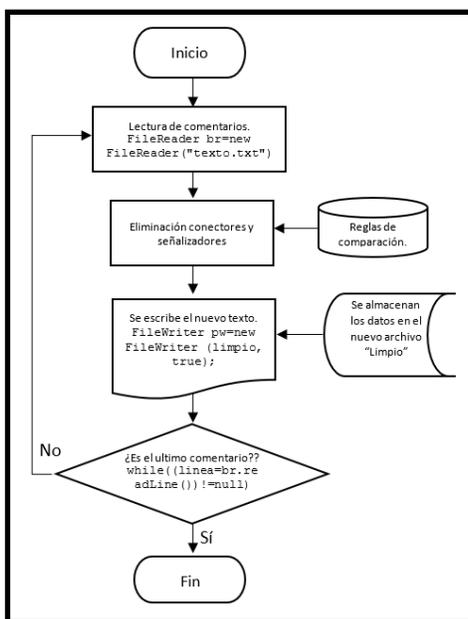
```

public class Extraer_Datos_Facebook {
    public static void main(String[] args) {
        String accesToken="xxxxxx";
        FacebookClient fbClient= new DefaultFacebookClient(Version.LATEST).createClientWithAccessToken(accesToken);
        User me = fbClient.fetchObject("me", User.class);
        String b=me.getBirthDay();
        Page page = fbClient.fetchObject("xxxxxx", Page.class);
        System.out.println(page.getFanCount());
        ScopeBuilder scopeBuilder = new ScopeBuilder();
        scopeBuilder.addPermission(FacebookPermissions.USER_EVENTS);
        String postId = "xxxxxx";
        Connection<Comment> commentConnection = fbClient.fetchConnection(postId + "/comments",
            Comment.class, Parameter.with("limit", 10));
        int personalLimit = 50;
        int n=1;
        for (List<Comment> commentPage : commentConnection){
            for (Comment comment : commentPage){
                System.out.println(" N:" +n+" Id: "+comment.getId() +" Comentario:"+ comment.getMessage());
                personalLimit--;
                if (personalLimit == 0) {
                    return;}
                try{
                    File archivo=new File("Datos_extraidos.txt");
                    FileWriter escribir=new FileWriter(archivo,true);
                    try (BufferedWriter bw = new BufferedWriter(escribir)) {
                        bw.write(comment.getMessage());
                        bw.newLine();}
                    }
                catch(Exception e){
                    System.out.println("Error al escribir");}
                n++;}
        }
        Post post = fbClient.fetchObject("xxxxxx", Post.class,
            Parameter.with("fields", "comments.limit(2){from(id),message}"));
    }
}
  
```

Limpieza de Datos

La limpieza de datos permite la eliminación de los conectores y señalizadores más comunes dentro de un comentario. Como se puede observar en el flujograma de la Figura 18, el proceso de limpieza de datos inicia con la carga de los datos “Comentarios”, para continuar con la comparación e identificación de coincidencias exactas entre las palabras que componen el comentario y los conectores y señalizadores Tabla 25, una vez encontradas las coincidencias se procede a eliminar y reescribir el comentario ya sin conectores y señalizadores en un nuevo archivo de texto.

Figura 18.
Flujograma limpieza de datos.



Este proceso permite realizar el análisis de los datos, de manera más rápida, ya que el algoritmo (Funcionalidad Principal) no tendrá que comparar las directrices con los conectores y señalizadores, en la Tabla 26 se puede observar un ejemplo de todo lo antes mencionado.

Tabla 28
Conectores y señalizadores

| Nombre | Descripción | Nombre | Descripción |
|------------|-------------|--------|-------------|
| A | Señalizador | Al | Conector |
| Así | Señalizador | Aún | Conector |
| El | Señalizador | De | Conector |
| En | Señalizador | Del | Conector |
| Es | Señalizador | En | Conector |
| Esta/ Esto | Señalizador | Por | Conector |
| La | Señalizador | Pero | Conector |
| Las / Los | Señalizador | Que | Conector |
| Más | Señalizador | Se | Conector |
| Para | Señalizador | Un | Conector |
| Si | Señalizador | Y | Conector |
| Tal | Señalizador | | |

Como se puede observar en la tabla anterior se detallan algunos de los conectores y señalizadores gramaticales existentes, para la selección de los conectores y señalizadores se tomó en cuenta una singularidad, que su estructura está formada por una sola palabra.

Tabla 29
Ejemplo proceso de limpieza

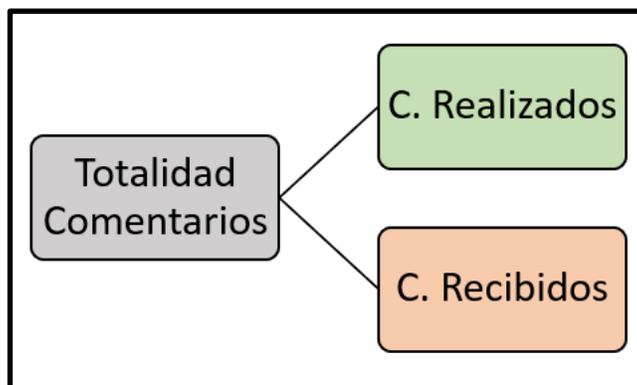
| | Descripción del comentario |
|----------------|---|
| Con conectores | “Va a ser duro seguir esta carrera, pero así es la vida así que tengo que ponerle mucho empeño a esto para así ser el mejor chef” |
| Sin conectores | “Va ser duro seguir carrera vida tengo ponerle mucho empeño ser mejor chef” |

En la Tabla 30 tenemos un ejemplo de cómo se manejó el proceso de limpieza, en la primera fila tenemos el comentario tal como se lo descargo de Facebook el mismo que cuenta con 27 palabras, mientras que el segundo ya ha pasado por el proceso de limpieza y tiene la mitad de palabras 13, disminuyendo la cantidad de comparaciones a la mitad al igual que el tiempo de procesamiento.

Almacenamiento de datos

Para el almacenamiento de la información en la base de datos se realiza un proceso de clasificación, en un perfil de Facebook se pueden encontrar, comentarios propios del usuario y comentarios publicados por otros a modo de respuesta o participación de información, motivo para separar los comentarios de acuerdo a la procedencia, al momento de la descarga cada comentario contara con la información de su propietario (User_Id).

Figura 19.
Identificación de comentarios.



En la Figura 16, se observa el criterio que se manejó para la separación de comentarios, este proceso es necesario, al analizar por separado los comentarios realizados y recibidos, permite identificar si el individuo es víctima o agresor.

Durante el proceso de separación, se lleva a cabo el almacenamiento en la base de datos, para la tabla comentario se maneja dos códigos un primario propio de Facebook y un código auxiliar para identificar la procedencia (Realizado o Recibido) como se puede observar en la Figura 20.

Figura 20.
Diagrama de la tabla comentario.

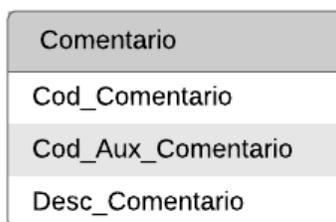


Tabla 30
Descripción del diagrama de base de datos

| Nombre | Tipo | Descripción |
|----------------|---------|-------------------------|
| Cod_Comentario | Varchar | El identificador único. |

| | | |
|--------------------|---------|--|
| Cod_Aux_Comentario | Varchar | Identificador auxiliar para procedencia. |
| Desc_Comentario | Text | Contenido exacto del comentario. |

Implementación Front - End

NetBeans tiene incluido en su paquete componentes que permiten crear una interfaz gráfica muy sencilla de utilizar, la herramienta permite crear la interfaz por medio de arrastre de los componentes a ser utilizados. La integración de las funciones, para cada componente se las realiza de acuerdo a las necesidades. Para este proyecto la interfaz de usuario es sencilla y se utiliza para mostrar los resultados, en la Figura 22, se puede observar el diseño final de la interfaz, que se presentara al usuario.

Figura 21.

Front – End de la aplicación.



Implementación Back-End

El Back-End de la aplicación se desarrolla bajo el patrón de diseño Modelo Vista Controlador-MVC, el mismo que permite dividir en tres módulos bien definidos y con

funcionalidades muy claras. El primero módulo (Modelo), permite realizar la conexión entre la base datos y el aplicativo, además permite la interacción con la base de datos mediante operaciones de inserción, extracción y eliminación de registros. El segundo módulo (Vista), como su nombre lo indica, en este módulo se encuentra la interfaz de usuario, es la pantalla con la que va a interactuar el usuario. Y como tercer módulo (Controlador), esta capa es la encargada de conectar la capa Modelo con la capa Vista, en esta capa se realiza la parte más relevante para el aplicativo, el análisis de los datos mediante el algoritmo de distancias de Levenshtein.

Funcionalidad específica del aplicativo

La función principal del aplicativo es identificar patrones mediante la comparación de cadenas de caracteres en función de la distancia de edición, a través del algoritmo de Levenshtein, la programación se realiza en NetBeans IDE que permite un uso sin restricciones, a continuación, se detalla el proceso para la implementación.

El proceso empieza con la selección de un individuo de la muestra, a continuación, se realiza una comparación entre los comentarios asociados al individuo y el corpus de palabras y finalmente el resultado es un reporte de las coincidencias encontradas.

Para la comparación de cadenas de caracteres se realizaron dos funciones “Mínimo” y “Calcular_Distancia”; las cuales permiten calcular la distancia que existe entre una cadena de caracteres respecto de otra, se utiliza para encontrar similitudes entre los comentarios realizados y el corpus de palabras planteado. Es necesario aclarar que la función “Calcular_Distancia”, realiza procesos de eliminación, transposición e inserción. Para esta investigación el rango de aceptación para determinar si una cadena de caracteres es similar a otra es: $0 \leq \text{distancia} \leq 2$.

Este rango se planteó después de realizar pruebas de ensayo y error en donde se utilizaron textos en español (Anexo 2), para medir la cantidad de coincidencias que se encuentran con los diferentes rangos, es importante mencionar que si el umbral es muy limitado no se podrán

encontrar similitudes, por lo contrario, si es muy abierto se generan inconsistencias y falsas similitudes. En la siguiente tabla se pueden observar los resultados de las pruebas realizadas.

Tabla 31
Cantidad de similitudes encontradas vs distancia.

| Distancia | Coincidencias | Texto 1 | Texto 2 | Razones |
|-------------------|-----------------------|---------|---------|--|
| | Reales T1 – T2 | | | |
| $0 \leq x \leq 1$ | 4 – 3 | 3 | 1 | Se pierden positivos |
| $0 \leq x \leq 2$ | 4 – 3 | 9 | 4 | Se agregan plurales / MÁS ADECUADO. |
| $0 \leq x \leq 3$ | 4 – 3 | 76 | 45 | Presencia exponencial falsos positivos |

Despliegue de la aplicación

La aplicación se desplegará como se puede observar en la figura 23. CERO_ACOSO permitirá interactuar al usuario directamente con el aplicativo, los requerimientos identificados en el análisis, se pueden apreciar mediante las acciones que permite el prototipo final de la aplicación.

Figura 22.
Primera evaluación CERO_ACOSO.



Como se puede observar las funcionalidades se desarrollan de forma secuencial como en el Diagrama de casos de usos, donde tenemos:

- Listar: Permite al usuario visualizar todos los sujetos de la muestra por seguridad de las identidades se ocultan los nombres.
- Empezar: Una vez que el usuario ha seleccionado a uno de los sujetos listados, el aplicativo comienza a buscar coincidencias en los datos del sujeto “comentarios”, una vez que finaliza el proceso se muestra una ventana flotante informando el estado del análisis. En la caja texto inferior se le muestra al usuario una presunta personalidad agresiva o pasiva.
- Reporte: Presenta al Usuario un reporte del número de coincidencias encontradas por cada una de las palabras del corpus.

Partiendo del análisis, donde se identificó los requerimientos en los que CERO_ACOSO trabajaría, a continuación se realizó el diagrama de casos de uso, se desarrolló el diseño, se empezó con el diagrama de arquitectura con el que se trabajaría, una vez identificado todos los aspectos, fue la implementación del aplicativo donde IDE NetBeans y MySql fueron las herramientas seleccionadas, en la primera herramienta se desarrolló el back-end basada en el patrón M.V.C. y el front-end de la aplicación la cual se desarrolló por medios de formularios de la IU de la herramienta, mientras que en la segunda herramienta se desarrolló la base de datos, mientras que las otras herramientas fueron de apoyo para completar las diferentes actividades.

Capítulo IV

Evaluación

El nombre del prototipo para identificar ciberacoso o (Cyberbullying) en los comentarios realizados en la red social Facebook es CERO_ACOSO, el cuál ha sido probada en la Unidad Educativa Sociedad Educativa Futuro “SEF”, ubicada en el cantón Rumiñahui provincia de Pichincha – Ecuador.

En esta sección se describe la evaluación del aplicativo de escritorio CERO_ACOSO desarrollado con el fin de identificar actividad de agresión en la red social Facebook. El objetivo de la evaluación es valorar el contenido de los perfiles de Facebook, como un medio para evitar actos de agresividad por medio de tecnologías de la información.

CERO_ACOSO buscan ser una herramienta de apoyo para la identificación de presuntas agresiones dentro de la red social, la cual a través de las presunciones encontradas sirva como un punto partida para un estudio amplio de la personalidad del sujeto en estudio.

Durante el período de evaluación se realizó un experimentó para evaluar el aplicativo, con los estudiantes de la unidad educativa “SEF” que cursan el bachillerato, donde se evaluó CERO_ACOSO como una herramienta para identificar presuntos patrones de agresividad. Tras la finalización, como método de contraposición o afirmación de los resultados obtenidos se aplicó un test especializado.

Al finalizar el experimento no se presentaron problemas con el funcionamiento de la aplicación, pero se pudo identificar limitaciones para acceder a la información de ciertos perfiles de Facebook, esto se debe al tipo de configuraciones que mantienen ciertos usuarios para mantener a salvo su información.

Figura 23.
Primera evaluación CERO_ACOSO.



Diseño de la actividad

Se realizó una inducción referente al acoso cibernético y la funcionalidad de CERO_ACOSO, la temática que se abordó fue, (Cyberbullying) como nuevo tipo de agresión, tipos de (Cyberbullying) y cómo reaccionar frente a un hecho de (Cyberbullying), al finalizar los estudiantes serán capaces de identificar este tipo de agresiones y cómo reaccionar.

Objetivo: Evaluar el funcionamiento y la predictibilidad que tiene CERO_ACOSO, ante un comentario ofensivo, mediante la obtención de resultados que validen la funcionalidad y la predicción de presuntos sujetos agresivos.

Objetivo pedagógico: La identificación de este tipo de agresión, es importante para prevenir futuras agresiones por medio de TIC'S.

Participantes: Esta actividad fue planificada para 25 estudiantes de Unidad Educativa Sociedad Educativa Futuro "SEF", que cursan segundo y tercero de bachillerato.

Sesión: Durante la Evaluación solo se necesitó una reunión con los participantes en la misma que se realizó la capacitación sobre (Cyberbullying) y al finalizar se les evaluó mediante un test Basado en el test HSPQ especializado en la identificación de patrones de personalidad en adolescentes con un rango de edad de 12 a 18 años, el cual fue interpretado con el apoyo de un psicólogo de profesión “DECE”, el mismo que se utilizó para contraponer los resultados obtenidos a través de CERO_ACOSO. Además, cada participante realizó una valoración personal de las palabras descritas en el (Planteamiento del Corpus de Palabras), en función de la agresividad percibida.

Cabe aclarar que el conocimiento adquirido durante la capacitación no afectará el resultado del test planteado, ya que el tipo de preguntas y respuestas del test son de selección múltiple y el contexto de las preguntas abarca temas sobre gustos o preferencias, por ejemplo.

- ¿Qué preferirías hacer?

Posibles respuestas: A) visitar un zoológico B) no estoy seguro C) viajar en avión

- ¿Te molesta cierta clase de películas?

Posibles respuestas: A) sí B) quizás C) no

- ¿Qué materias preferirías estudiar?

Posibles respuestas: A) matemáticas B) no estoy seguro C) un idioma extranjero o arte dramático

Desarrollo de la actividad

En la siguiente tabla se detalla cada una de las fases de la actividad y que se realizó en cada una de ellas.

Tabla 32
Proceso de la actividad

| | Actividad | Descripción | Finalidad |
|-----------------------------------|---|---|---|
| Inducción acoso cibernético |  | Información sobre el acoso cibernético y sus implicaciones. | Prevenir futuras agresiones. |
| Test HSQP |  | Desarrollo de un test, de personalidad | Contrastar los resultados de CERO_ACOSO |
| Encuesta pesos semánticos |  | Ponderación de pesos semánticos. | Establecer nivel de impacto para matriz probabilidad e impacto. |
| Uso de CERO_ACOSO |  | Analizar los datos obtenidos de los perfiles. | Encontrar similitudes mediante el algoritmo de Levenshtein y analizarlos mediante las matrices de probabilidad x impacto. |

La sesión de evaluación conto con la presencia de los 25 estudiantes de la Unidad Educativa Sociedad Educativa Futuro “SEF”, y para un manejo más adecuado de esta actividad el psicólogo del establecimiento educativo estuvo presente. El tiempo que se designó para esta actividad fue de 120 minutos. Durante la actividad un estudiante tuvo que salir por motivos personales, y 2 estudiantes mencionaron que no mantiene actualmente una cuenta de Facebook activa, por lo cual al finalizar la sesión se tuvo un total de 22 personas encuestadas. El test planteado se encuentra como ANEXO 3 en este documento.

Inducción sobre el acoso cibernético (Cyberbullying).

En esta parte de la actividad, se les dio a conocer los diferentes tipos de acoso haciendo hincapié en el acoso cibernético, el tipo de afecciones que produce en una víctima y como se puede evitar este tipo de abusos. Se les dio procesos de ayuda a seguir cuando no están en la capacidad de solventar solos un problema propio o de un ajeno, en referencia a cualquier tipo de agresión entre pares.

Figura 24.

Capacitación sobre acoso cibernético (Cyberbullying).



Como se puede observar en la Figura 24, se está impartiendo la capacitación de (Cyberbullying), manteniendo una participación activa de los estudiantes para poder despejar todas las dudas planteadas durante esta sesión.

Aplicación del test de personalidad

En esta actividad de la sesión se les expuso a los participantes la importancia de realizar el test de manera consciente e individual, teniendo en cuenta que el test no mide conocimientos o inteligencia, razón por la cual no existen respuestas erróneas. Es necesario aclarar que el tiempo para responder el test no fue delimitado, pero no excedió los 20 minutos hasta que todos

finalicen. Como ya se mencionó el test planteado es parte del test HSPQ, pero únicamente se tomaron en cuenta los factores E, Q III y Q IV en los cuales se puede analizar agresividad.

Valoración pesos semánticos.

Durante el proceso de ponderación de pesos semánticos a las palabras del corpus de elementos se pudo evidenciar que la mayoría de los participantes tenían familiaridad con los términos. Fue necesario explicar algunos conceptos, con la finalidad de tener un resultado exitoso: **Jerarquía:** Se tamo como el reflejo o recuerdo hacia un contexto, en este caso, la jerarquía simboliza el porcentaje en el que una palabra definidora le recuerda el ciberacoso (Cyberbullying) al encuestado. **Peso semántico:** Valor que se le otorga a una palabra definidora en función de la ponderación Jerárquica asignada por todos los participantes.

Figura 25.

Matriz pesos semánticos.

| Definidoras Jerarquías | I d i o t a | E s t ú p i d o | M u d o | T o t o | B r u t o | V e r g a | P e n d e j o | A n i m a l | M a c a | G a y | L e s b i a n a | H o s e x u a l | P u t a | P r o s t i t u t a | M m v | H p | V c v |
|---------------------------|----------------------------|--------------------------------------|------------------|------------------|-----------------------|-----------------------|---------------------------------|----------------------------|------------------|-------------|--------------------------------------|--------------------------------------|------------------|--|-------------|--------|-------------|
| Jerarquía 1 | | | | | | | | | | | | | | | | | |
| Jerarquía 2 | | | | | | | | | | | | | | | | | |
| Jerarquía 3 | | | | | | | | | | | | | | | | | |
| Jerarquía 4 | | | | | | | | | | | | | | | | | |
| Jerarquía 5 | | | | | | | | | | | | | | | | | |
| Jerarquía 6 | | | | | | | | | | | | | | | | | |
| Jerarquía 7 | | | | | | | | | | | | | | | | | |
| Jerarquía 8 | | | | | | | | | | | | | | | | | |
| Jerarquía 9 | | | | | | | | | | | | | | | | | |
| Jerarquía 10 | | | | | | | | | | | | | | | | | |

En la Figura 25, se muestra una matriz, en donde los participantes de esta actividad colocarán una X, de acuerdo a la jerarquía que le quieran dar a cada palabra, teniendo en cuenta que la jerarquía 1 es la más alta y tiene un valor de 10, y la jerarquía 10 ocupa el escalón más bajo y su valor es 1, para la valoración de los términos está basado en la temática “Agresión verbal o escrita”. Es importante mencionar que los pesos semánticos son utilizados, para medir

el impacto que causa dicha palabra en una persona, lo cual se más adelante se utilizara como un cuantificador para las matrices de probabilidad e impacto.

Para finalizar la actividad, los participantes compartieron el nombre de usuario con el cual están registrados en Facebook con el investigador, para poder analizar los comentarios de cada perfil de Facebook y realizar procesos de identificación de patrones.

Uso de CERO_ACOSO y desarrollo de la actividad

Esta actividad no fue detallada en la planificación de la actividad, debido a que los participantes no son parte activa de la misma, y se la realizo después de obtener la información “comentarios”, de cada uno de los participantes. Una vez con la información almacenada se procedió a realizar la actividad para lo cual se necesitó de un tiempo aproximado de 10 minutos por cada perfil de Facebook. Durante la ejecución del programa se pudo evidenciar.

Resultados

Los resultados se basan en la información que fue recolectada de los 22 participantes.

Durante el proceso de evaluación los participantes realizaron dos actividades, con el fin obtener un factor de personalidad más aproximado a cada participante.

En primer cuestionario que fue realizado por cada uno de los participantes el objetivo fue medir el peso semántico de una palabra, a través de lo cual se pudo medir el impacto que causa cada una de las palabras al hablar sobre agresión. Como se puede observar en la siguiente Figura 26. estos fueron los pesos que cada participante dio a las palabras.

Valoración de pesos semánticos.

En primer cuestionario que fue realizado por cada uno de los participantes el objetivo fue medir el peso semántico de una palabra, a través de lo cual se pudo medir el impacto que causa cada una de las palabras al hablar sobre agresión. Como se puede observar en la siguiente Figura 26. estos fueron los pesos que cada participante dio a las palabras.

Figura 26.
Pesos semánticos.

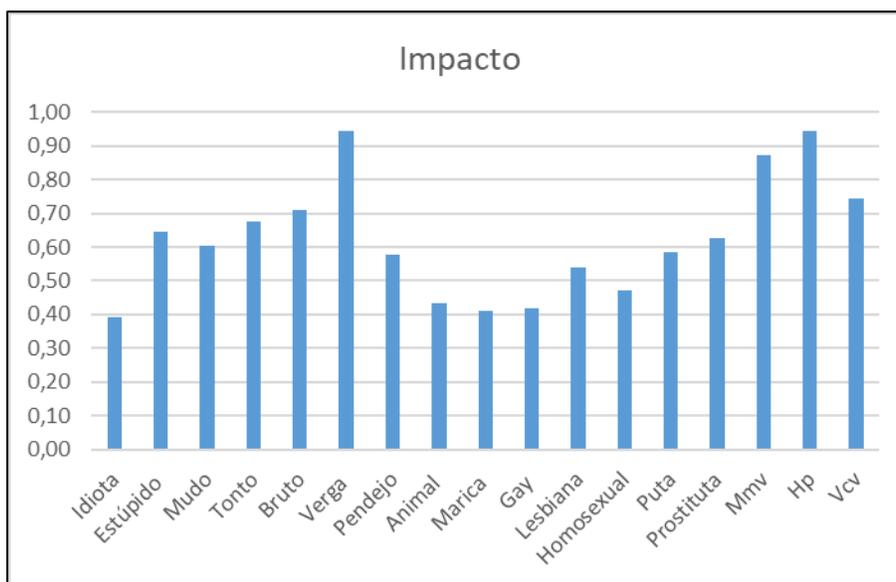
| Definidoras/ Jerarquías | Idiota | Estúpido | Mudo | Tonto | Bruto | Verga | Pendejo | Animal | Marica | Gay | Lesbiana | Homosexual | Puta | Prostituta | Mmx | Hp | Xcv |
|----------------------------|--------|----------|------|-------|-------|-------|---------|--------|--------|-----|----------|------------|------|------------|-----|----|-----|
| Jerarquía 1 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 12 | 15 | 8 |
| Jerarquía 2 | 0 | 8 | 0 | 5 | 8 | 3 | 0 | 0 | 1 | 1 | 0 | 1 | 2 | 2 | 0 | 6 | 5 |
| Jerarquía 3 | 0 | 3 | 3 | 4 | 1 | 1 | 1 | 0 | 1 | 1 | 2 | 1 | 1 | 2 | 5 | 0 | 4 |
| Jerarquía 4 | 0 | 0 | 4 | 5 | 1 | 1 | 1 | 0 | 1 | 1 | 5 | 1 | 6 | 6 | 4 | 0 | 0 |
| Jerarquía 5 | 6 | 0 | 8 | 1 | 11 | 1 | 14 | 9 | 0 | 1 | 6 | 3 | 5 | 4 | 0 | 0 | 0 |
| Jerarquía 6 | 5 | 6 | 5 | 5 | 0 | 0 | 5 | 4 | 5 | 5 | 0 | 5 | 4 | 1 | 0 | 0 | 0 |
| Jerarquía 7 | 4 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 8 | 7 | 5 | 7 | 1 | 1 | 1 | 1 | 0 |
| Jerarquía 8 | 1 | 2 | 0 | 2 | 1 | 0 | 1 | 5 | 0 | 0 | 4 | 1 | 1 | 1 | 0 | 0 | 0 |
| Jerarquía 9 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 0 | 3 | 2 | 3 | 0 | 0 | 2 |
| Jerarquía 10 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| N.º Participantes | 22 | | | | | | | | | | | | | | | | |
| Peso Semántico | 86 | 142 | 133 | 149 | 156 | 208 | 127 | 95 | 90 | 92 | 119 | 104 | 129 | 138 | 192 | 20 | 164 |

Los resultados de la valoración semántica, a pesar de ser individual y de carácter personal, tienen similares valoraciones, como se puede observar en la Figura anterior que; 16, 12 y 15 personas les dieron una jerarquía alta a las palabras ubicada en las posiciones 6 (verga), 15 (Mvm) y 16 (Hp) respectivamente, mientras que la palabra que recibieron menor valoración se ubica en la posición 1 (Idiota). Mediante este proceso se pueden obtener el impacto que produce cada una de esas palabras en un individuo.

Impacto en función del peso semántico.

Para definir cuanto impacto produce una palabra agresiva en un individuo, se utilizó la valoración semántica. El impacto se valora en un rango de 0 a 1, siendo la máxima afección 1 y 0 representa que no afecta. En la Figura 27, se puede observar las palabras con mayor impacto hablando de Agresividad.

Figura 27.
Impacto vs Palabras.



Las palabras que recibieron mayor peso semántico, también tienen el mayor impacto, el valor promedio en el cual afecta una palabra de la lista anterior es del 62%.

Similitudes encontradas en comentarios propios con CERO_ACOSO

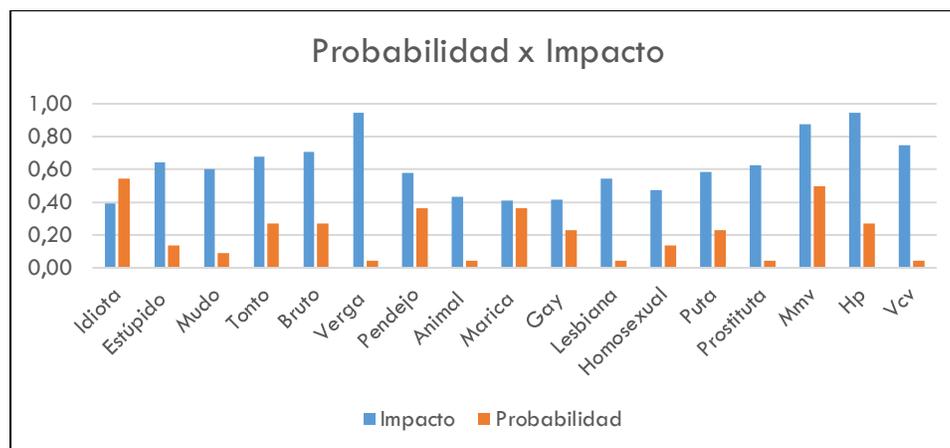
Para los siguientes análisis a los participantes se les asignó un identificador "S" para mantener la privacidad de su identidad, durante el proceso de análisis y comparación de datos con el prototipo CERO_ACOSO, se obtuvieron los resultados parciales por cada participante. Durante este proceso se pudo determinar que uno de los participantes (S16), mantiene una cuenta activa, pero sin actividad (sin contenido).

Figura 28.
Frecuencia X Palabra C.P.

| Palabra | Sujetos En Estudio | | | | | | | | | | | | | | | | | | | | | | Total |
|--------------|--------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-------|
| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 | S19 | S20 | S21 | S22 | |
| Idiota | 1 | | | | | | | 1 | | | 2 | 1 | 1 | | 1 | | 1 | 1 | | 1 | | 2 | 12 |
| Estúpido | | | | | | | 3 | | | | | | | | | | | | | | | | 3 |
| Mudo | | | 2 | | | | | | | | | | | | | | | | | | | | 2 |
| Tonto | | | | 2 | | | | | | 2 | | | | | | | | | | 2 | | | 6 |
| Bruto | | | | | 1 | | | 2 | | | 2 | 1 | | | | | | | | | | | 6 |
| Verga | | | | | | | | | 1 | | | | | | | | | | | | | | 1 |
| Pendejo | | 2 | | | | 1 | | | | | | | | | 1 | | | | | 1 | 3 | | 8 |
| Animal | | | | | | | | | | | | 1 | | | | | | | | | | | 1 |
| Marica | | | | | | 3 | | | | | | | | | | | | | 5 | | | | 8 |
| Gay | | | | | | | 1 | | | | 3 | | | | | | | | 1 | | | | 5 |
| Lesbiana | | | | | | | | | | | | | | 1 | | | | | | | | | 1 |
| Homosexual | | | | | | | | | | | | | | | | | | | | 3 | | | 3 |
| Puta | | | | | | | | | | 2 | | | | | 3 | | | | | | | | 5 |
| Prostituta | | | | | | | | | | | | | | | | | | 1 | | | | | 1 |
| Mmv | | | 3 | | | | 2 | | | | | | | | | | 1 | | | | | 5 | 11 |
| Hp | | | | | | | | | | | 3 | | | | | | | | | | | | 3 |
| Vcv | 1 | | | | | | | | | | | | | | | | | | | | | | 1 |
| Total | 2 | 2 | 5 | 2 | 1 | 4 | 6 | 3 | 1 | 4 | 10 | 3 | 1 | 1 | 5 | 0 | 3 | 1 | 6 | 7 | 3 | 10 | |

En la Figura 28, se expresan la cantidad de coincidencias que el prototipo CERO_ACOSO, encontró en los comentarios de cada uno de los participantes, como ya se mencionó uno de los participantes tiene cero coincidencias, debido al cese de sus actividades en Facebook, la frecuencia más alta en incidencias por individuo es de 10, con los participantes (S11 y S22) y la frecuencia por palabra es de 12 para (Idiota), lo cual indica que los sujetos mencionados son presuntos promotores de (Cyberbullying).

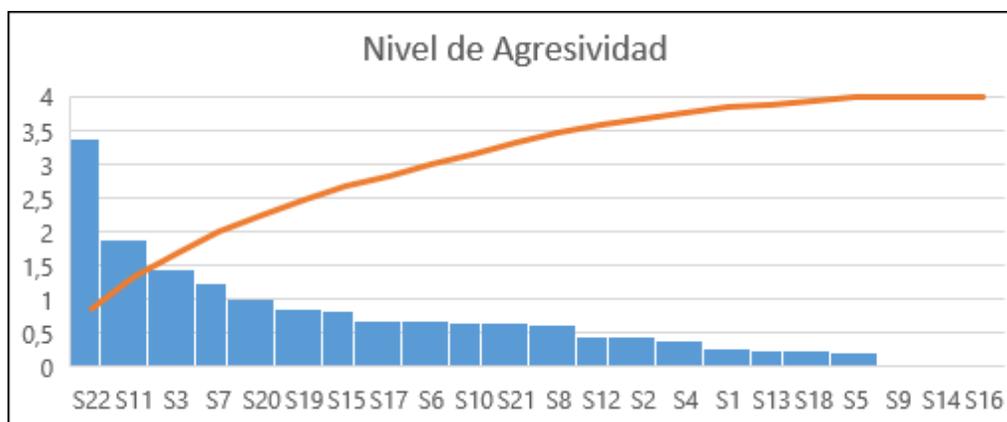
Figura 29.
Probabilidad X Impacto C.P.



La matriz de probabilidad x impacto Figura 29, nos ayuda a identificar, la posibilidad de que ocurra un evento y el costo “grado de afección” que tendría, en nuestro caso utilizamos para identificar posibles eventos de agresividad mientras más coincidencias encontradas, la posibilidad de ser un agresor aumenta gradualmente de acuerdo al impacto que marcan dichas coincidencias.

En la Figura 30, que mostramos a continuación se observa el grado de agresividad de los sujetos, teniendo en cuenta que 10 es el nivel más alto de agresividad.

Figura 30.
Nivel de Agresividad Activa.



Similitudes encontradas en comentarios recibidos con CERO_ACOSO

Con la finalidad de identificar si un sujeto es víctima de acoso cibernético (Cyberbullying), se analizó por separado los comentarios que son realizados por terceros en los perfiles de cada participante el resultado se puede observar en la Figura 31.

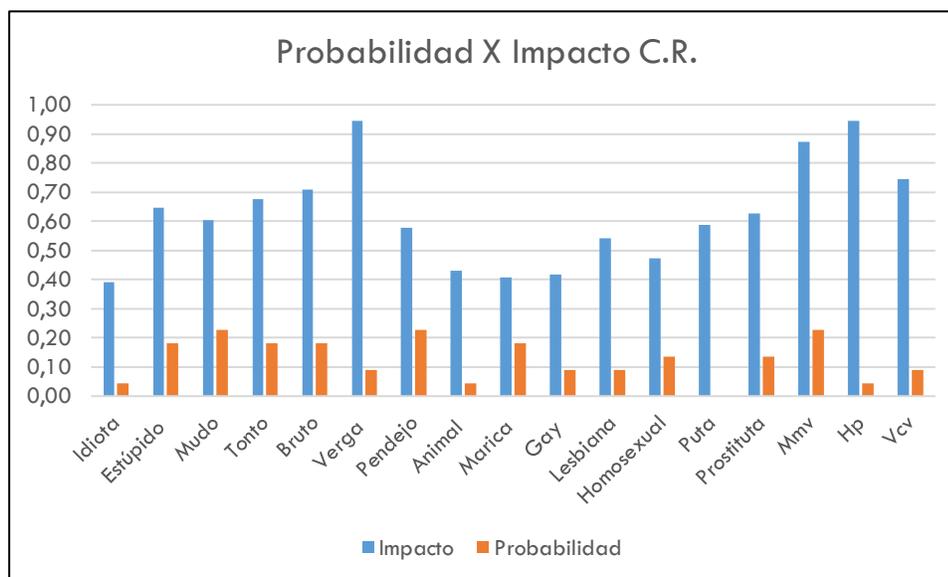
Figura 31.
Frecuencia X Palabra C.R.

| Palabra | Sujetos En Estudio | | | | | | | | | | | | | | | | | | | | | | Total |
|--------------|--------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-------|
| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 | S19 | S20 | S21 | S22 | |
| Idiota | | | | | | | | | | | | 1 | | | | | | | | | | | 1 |
| Estúpido | | 1 | | | | | | | | | 1 | | | | 2 | | | | | | | | 4 |
| Mudo | | | 1 | | | | 1 | | | | | | | 2 | | | 1 | | | | | | 5 |
| Tonto | | | | | | | | 1 | | 1 | | | | | | | | | | | | 2 | 4 |
| Bruto | | | | 1 | | | | | | | | | 2 | | | | 1 | | | | | | 4 |
| Verga | | | | | | | | | | | 2 | | | | | | | | | | | | 2 |
| Pendejo | | | | | | | | 2 | | | | | | | | | 1 | | | | | 2 | 5 |
| Animal | | | | | 1 | | | | | | | | | | | | | | | | | | 1 |
| Marica | | | | | | | | | | 2 | 1 | | | | | | 1 | | | | | | 4 |
| Gay | | | | | | | | | | | | | | | | | | | | | | 2 | 2 |
| Lesbiana | | | | | | | | | 2 | | | | | | | | | | | | | | 2 |
| Homosexual | | | | | | 1 | | | | | | | | | | | | | | | 2 | | 3 |
| Puta | | | | | | | | | | | | | | | | | | | | | | | 0 |
| Prostituta | | | 1 | | | | | 2 | | | | | | | | | | | | | | | 3 |
| Mmv | | | | | | | | | | | 1 | | 1 | | | | | 1 | 2 | | | | 5 |
| Hp | | | | | | | | | | | | | | | | | | | | | | 1 | 1 |
| Vcv | 1 | | | | | | | | | | | | | | | | | | | | 1 | | 2 |
| Total | 1 | 1 | 2 | 1 | 1 | 1 | 3 | 3 | 2 | 3 | 5 | 1 | 3 | 2 | 2 | 0 | 4 | 1 | 2 | 3 | 0 | 7 | |

Dentro del rango de mayor coincidencia por sujeto, aun cuando el porcentaje es menor, se mantiene una actividad en los participantes (S11 y S22), mientras los demás participantes, mantienen valores similares y de baja frecuencia.

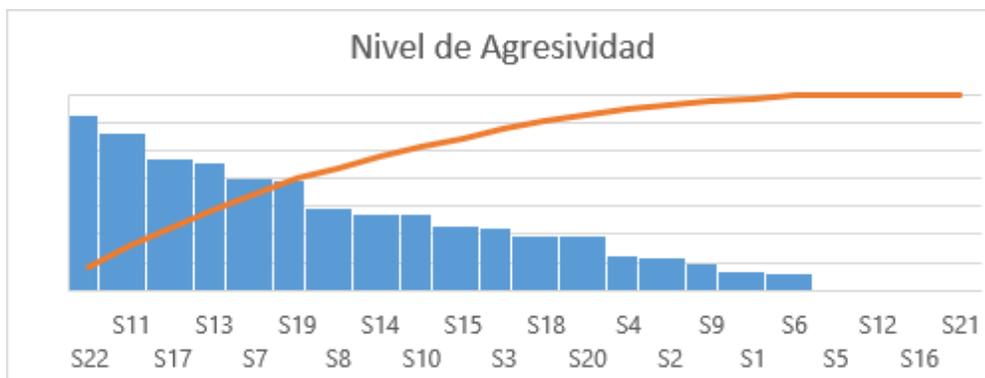
Utilizamos matriz de probabilidad x impacto, para identificar presuntos eventos de agresividad hacia los participantes, mientras más coincidencias encontradas, la posibilidad de ser una víctima aumenta. Es notable en la Figura 32, que la frecuencia de actividad agresiva es baja.

Figura 32.
Probabilidad X Impacto C.R.



Para identificar presuntas víctimas se realizó un cruce entre la matriz de probabilidad e impacto y los datos de cada participante para determinar, la existencia de una presunta víctima. A continuación, se puede observar los resultados obtenidos.

Figura 33.
Nivel de Agresividad Pasiva.



La Figura 34, refleja los resultados mostrados en la Figura 33, donde es claro que el índice de coincidencias es muy bajo, razón por la cual se puede presumir que no existen presuntas víctimas dentro de la muestra.

Discusión de resultados

Los resultados obtenidos durante el proceso de evaluación, permitieron identificar patrones y presuntos perfiles de con agresividad, es necesario realizar una comparación de los resultados arrojados por la aplicación y la mirada de un psicólogo mediante un test adecuado. Es importante aclarar que solo se toma en cuenta los dos sujetos que tuvieron como resultado una presunta agresividad (S11, S22), los cuales alcanzaron un nivel de agresividad de 3,38 y 1,87 respectivamente, en una escala de 1 a 10, siendo 10 el más alto.

Los resultados obtenidos del test aplicado a los dos sujetos fueron los siguientes:

Figura 34.
Evaluación del factor E.

| Sujeto / Preguntas | Pr1 | Pr2 | Pr3 | Pr4 | Pr5 | Pr6 | Pr7 | Pr8 | Pr9 | Pr10 | Total |
|--------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-------|
| S11 | 1 | 2 | 2 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 12 |
| S22 | 1 | 1 | 2 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 8 |

| | Rango | | | | | | | | | | |
|--------|-------|---|---|-----|---|------|-----|-------|-------|-------|----------|
| Sumiso | 0-3 | 4 | 5 | 6-7 | 8 | 9-10 | 11 | 12-13 | 14-15 | 16-20 | Agresivo |
| | | | | S22 | | | S11 | | | | |

El participante S11, esta sobre el rango (12-13), que está sobre la medida promedio apuntando, a un perfil de personalidad agresivo, marcado por el factor E (+); el individuo que mantiene un valor alto superior a la media es un individuo: Agresivo, dominante, dogmático, obstinado. Lo que se traduce en un individuo que tiene problemas de conducta.

En el caso del sujeto S22, tiene un valor menor perteneciente a la media, con una tendencia al factor E (-); el individuo que presenta valores bajos tiende a ser: Sumiso, Dócil, Obediente.

Lo cual permite llegar a concluir que la presunción que se obtuvo del aplicativo en el primer caso es efectiva.

Capítulo V

Conclusiones y recomendaciones y trabajos futuros

Conclusiones

- En este trabajo de investigación para la titulación se diseñó, implemento y evaluó una aplicación que permite identificar abuso cibernético a través de una red social en estudiantes de bachillerato, lo cual permitió una intervención adecuada, con los estudiantes que tienen presunta agresividad.
- La identificación de rasgos o factores de personalidad “patrones de personalidad”, fue solventada mediante la utilización del HSQP y la orientación de un psicólogo, para generalizar en dos grandes grupos, agresores y víctimas de acoso cibernético (Cyberbullying).
- CERO_ACOSO mediante la utilización del algoritmo de Levenshtein, fue capaz de identificar patrones y coincidencias en el lenguaje español, y determinar presuntos patrones de agresividad en dos sujetos, de los cuales uno de ellos después de realizar un test profesional coincide con la presunción de agresividad.
- Las palabras que alimentaron el corpus de elementos, fueron adecuadas, al momento de realizar la comparación de cadenas, se identificaron varias coincidencias en los diferentes perfiles de usuario de Facebook, las cuales permitieron posteriores análisis.
- El porcentaje de acoso cibernético o (Cyberbullying), encontrado en la muestra es del 6,81%, el porcentaje se traduce en; 1 de cada 10 estudiantes es parte del (Cyberbullying), sea de forma activa o pasiva.

- El uso de la matriz de probabilidad e impacto para cuantificar el nivel de agresividad de una persona, permitió incluir características como el peso semántico. Y de esta manera se identificó rangos, donde un individuo marca puntos para ser considerado como presunto agresor o víctima de (Cyberbullying).

Recomendaciones

- Se sugiere realizar un proceso de verticalización de datos, para mejorar los tiempos de respuesta, al momento de analizar un comentario.
- Se sugiere realizar el mismo estudio, pero utilizando una muestra mucho más grande y variada en cuanto al ámbito socio económico de la población.
- El uso de Facebook en los adolescentes está disminuyendo debido a las nuevas redes sociales del momento, lo cual ha disminuido la interacción de los usuarios, para futuras investigaciones se sugiere utilizar como banco de datos otro tipo de red social.

Trabajos Futuros

A lo largo de la investigación, durante la revisión del estado del arte y aún más en el periodo de evaluación, se presentaron ideas que podrían mejorar el desempeño de la propuesta principal, actuando en diferentes campos como el laboral, económico, financiero, de mercado o político. De este modo nacen nuevas ideas de trabajos futuros:

- Se propone desarrollar una aplicación web la cual pueda ser utilizada como una extensión en Facebook, la misma que permitirá identificar acoso cibernético (Cyberbullying) en un perfil de Facebook, permitiendo una intervención más pronta.
- Se propone trabajar en el análisis y crecimiento de la aceptación de una Figura público – política mediante “marketing dirigido”, para la selección adecuada de un representante o candidato a una elección popular.

REFERENCIAS BIBLIOGRÁFICAS

- Ameri, S., Vahdati, S., & Lange, C. (2017, September). Exploiting interlinked research metadata. . In International Conference on Theory and Practice of Digital Libraries . Springer, Cham., pp. 3-14.
- Anderson, M., & Jiang, J. (2018). Teens, social media & technology 2018. Pew Research Center. 31.
- Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive behaviors*, 64, 287-293.
- Apache, Friends. (2019). Xampp. Obtenido de <https://www.apachefriends.org/es/index.html>
- Bachrach, Y. (2015). Human Judgments In Hiring Decisions Based On Online Social Network Profiles. IEEE.
- Balarezo Chiriboga, L. A. (2015). Psicoterapia integrativa focalizada en la personalidad.
- Bayón, C. (2016). Modelo psicobiológico de personalidad de Cloninger: Aproximación integradora en la evaluación de los trastornos de personalidad y proceso psicoterapéutico. *Psiquiatría.com*, 10(1).
- Bhardwaj, S., Atrey, P. K., Saini, M. K., & El Saddik, A. (2016). Personality assessment using multiple online social networks. . *Multimedia Tools and Applications*, 75(21), 13237-13269.
- Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P., & Fienberg, S. (2003). Adaptive name matching in information integration. . *IEEE Intelligent Systems*, 18(5), 16-23.

- Black, P. E. (2009). NYSIIS. US National Institute of Standards and Technology.
Dictionary of algorithms and data structures. Gaithersburg, MD, USA: NIST.
- Cattell, R. B., & Cattell, M. D. (1997). HSPQ: cuestionario de personalidad para adolescentes (12-18 años). Tea.
- Chakkrit, S., & Bruckner, M. (Thailand, 2009). Novel Phonetic Name Matching Algorithm with a Statistical Ontology for Analyzing Names Given in Accordance with Thai Astrology. Faculty of Science, Naresuan University, Phitsanulok, Pages 497-515.
- Cohen, W., Ravikumar, P., & Fienberg, S. (2003, August). A comparison of string metrics for matching names and records. In Kdd workshop on data cleaning and object consolidation, Vol. 3, pp. 73-78.
- Constante Portero, D. D. (2018). Las redes sociales como plataforma para el grooming en los jóvenes de 12 a 17 años de la ciudad de Ambato. (Bachelor's thesis, Universidad Técnica de Ambato. Facultad de Diseño, Artes y Arquitectura. Carrera de Diseño Gráfico Publicitario.).
- Correa, T., Hinsley, A. W., & De Zuniga, H. G. (2010). Who interacts on the Web? The intersection of users' personality and social media use. *Computers in human behavior*, 26(2), 247-253.
- Delgado, J., Galarraga, F., Fuertes, W., Toulkeridis, T., Villacís, C., & Castro, F. (2016, March). A proposal of an entity name recognition algorithm to integrate governmental databases. In 2016 Third International Conference on eDemocracy & eGovernment.
- Encuesta Multipropósitos. (Septiembre 2019). Encuesta de Seguimiento al Plan Nacional de Desarrollo . Obtenida de:
https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/2018/201812_Boletin_Multiproposito_Tics.pdf.

- ENEMDU. (2017). Tecnologías de la Información y Comunicación. Obtenido de:
https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/2017/Tics%202017_270718.pdf.
- Facebook. (2019). API Graph. Obtenido de
<https://developers.facebook.com/docs/graph-api>
- Feinberg, T., & Robey, N. (2009). (Cyberbullying): Intervention and prevention strategies. *National Association of School Psychologists*, 38(4), 22-24.
- Fernández, C. S., & Hernáez, L. L. (2015). Factores de riesgo en el Cyberbullying. Frecuencia y exposición de los datos personales en Internet. *International Journal of Sociology of Education*, 1-25.
- Gelbukh, A. (2010). Procesamiento de lenguaje natural y sus aplicaciones. . *Komputer Sapiens*, 1, 6-11.
- Ghavami, S. M; Asadpour, M; Hatami, J; Mahdavi, M. (2015, May). Facebook user's like behavior can reveal personality. In 2015 7th Conference on Information and Knowledge Technology (IKT), (pp. 1-3). IEEE.
- González Calatayud, V. (2017). Prevalencia del ciberacoso en las aulas de Educación Secundaria Obligatoria y la percepción del profesorado. Proyecto de investigación.
- Hassanzadeh, O., Sadoghi, M., & Miller, R. J. (2007, September). Accuracy of Approximate String Joins Using Grams. In QDB , 11-18.
- Haunts, S. (2014). Phonetic String Matching: Soundex. Retrieved from:
<https://stephenhaunts.com/2014/01/17/phonetic-string-matching-soundex/>.
- Hootsuite & We Are Social. (2020). Digital 2020 Global Digital Overview. Recuperado de: <https://datareportal.com/reports/digital-2020-ecuador>.

- INEC. (2014). «Encuesta de Condiciones de vida realizada por el Instituto Nacional de Estadísticas y Censos, Quito».
- Jácome, Danny; Tapia, Freddy; Lascano, Jorge; Fuertes, Walter. (2019, February). Contextual Analysis of Comments in B2C Facebook Fan Pages Based on the Levenshtein Algorithm. International Conference on Information Technology & Systems (págs. 528-538). Springer, Cham.
- Kolak, A. M., & Volling, B. L. (2013). Coparenting moderates the association between firstborn children's temperament and problem behavior across the transition to siblinghood. *Journal of Family Psychology*, 27(3), 355.
- Letelier, P. (15 de 01 de 2006). Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP). Obtenido de http://www.cyta.com.ar/ta0502/b_v5n2a1.htm
- Limber, S. P., Olweus, D., Wang, W., Masiello, M., & Breivik, K. (2018). Evaluation of the Olweus Bullying Prevention Program: A large scale study of US students in grades 3–11. . *Journal of school psychology*,, 69, 56-72.
- Loyola, J. (2015). Hacia un Algoritmo Optimo de Emparejamiento de Nombres. *Analítika: revista de análisis estadístico*, (9), 95-109.
- Lubis, A. H., Ikhwan, A., & Kan, P. L. E. . (2018). Combination of levenshtein distance and rabin-karp to improve the accuracy of document equivalence level. *International Journal of Engineering & Technology*, 7(2.27), 17-21.
- Marroquin, A. B. (2017). Rasgos de personalidad que presentan los agresores de bullying que sonestudiantes de nivel secundario en centro educativo de la ciudad deguatemala.

- Marroquin, A. B. (s.f.). Rasgos de personalidad que presentan los agresores de bullying que son estudiantes de nivel secundario en centro educativo de la ciudad de Guatemala.
- Martel, W., Carranco, D., & Cevallos, D. (2016). Determinación de niveles de agresividad en comentarios de la red social Facebook por medio de Minería de Texto. GEEKS DECC-REPORTS, 6(1).
- Martínez, J. M. (2013). Análisis psicosocial del cyberbullying: claves para una educación moral. Papeles del psicólogo, 34(1), 65-73.
- Menay-López, L., & de la Fuente-Mella, H. (2014). Plataformas comunicacionales del (Cyberbullying): Una aplicación empírica en dos colegios de la quinta región, Chile. Estudios pedagógicos (Valdivia), 40(2), 117-133.
- Monelos Muñoz, M. E. (2015). Análisis de situaciones de acoso escolar o bullying en centros de Educación Secundaria de La Coruña.
- Oracle, C. (2017). MySQL. Obtenido de <https://www.mysql.com/>
- Oracle, C. (2017). NetBeans IDE. Obtenido de <https://netbeans.org/features/index.html>
- Parmar, V. P., & Kumbharana, C. K. (2014). Study Existing Various Phonetic Algorithms and Designing and Development of a working model for the New Developed Algorithm and Comparison by implementing it with Existing Algorithm (s). Comparison by implementing it with Existing Algorithm (s). International Journal of Computer Applications, 98(19), 45-49.
- Peng, T., Li, L., & Kennedy, J. (2012). A Comparison of Techniques for Name Matching. GSTF journal on Computing, 2(1).
- Restfb. (2019). restfb. Obtenido de <https://restfb.com/>
- Ríos, M. C. (2017). Evaluación de procesos de Comunicación para el Desarrollo. Commons. Revista de Comunicación y Ciudadanía Digital.

- Russell, R. (1918). Specifications of Letters. US patent number 1, 261,167.
- Samuel D. Gosling, S. G. (2007). Personality Impressions Based on Facebook Profiles. ICWSM'2007 Boulder, Colorado, USA.
- Sánchez, H; Benedetti, M; Luzardo, M; Boggio, K; Bagnato, M; Martín, A; Premuda, P. (2017). Adaptación a población adulta montevideana de la Escala de Temperamento y Carácter Revisada (TCI-R): resultados preliminares. *Psicología, Conocimiento y Sociedad*, 7(1),169-198.
- Schay, W. (2011). A Generic Framework for the Matching of Similar Names. . Faculty of Engineering and the Built Environment. University of Witwatersrand, pages 11–203.
- Seelbach González, G. A. (2013). Teorías de la personalidad. Primera). Retrieved from http://www.affiliat.org.mx/BibliotecasDigitales/Psicologia/Teorias_de_la_personalidad.pdf.
- Shah, R., & Singh, D. K. (2014). Analysis and comparative study on phonetic matching techniques. *International Journal of Computer Applications*, 87(9).
- Shanker Singh, B. (2003). Search Algorithms. DRTC Workshop on Digital Libraries: Theory and Practic. March 2003. DRTC, Bangalore. Disponible en: https://drtc.isibang.ac.in/bitstream/1849/32/2/E_Searchalgo_brijesh.pdf.
- UNESCO. (2017). Decidamos cómo medir la violencia en las escuelas. . Recuperado de <http://unesdoc.unesco.org/images/0024/002469/246984s.pdf>.
- UNICEF. (2015). Violencia entre pares en el sistema educativo: Una mirada en profundidad al acoso escolar en el Ecuador. World Vision.
- UNICEF. (s.f). Violencia contra los niños . Obtenido de: https://www.unicef.org/ecuador/media_28087.html.

- Vásquez, A. C. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*, 6(2), 45-54.
- Villaverde, A. L., Aranda, C. M., & Hernandez-Orallo, J. (2012). Recuperado el 01 de 2015, de <https://riunet.upv.es/bitstream/handle/10251/16576/Memoria.pdf?sequence=1>
- Wohlin, C., & Runeson, P. (2013). On the reliability of mapping studies in software engineering. *The Journal of Systems and Software*, 2594-2610.