



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación
y Desarrollo Tecnológico

Tesis de Maestría

Procesamiento de Lenguaje Natural para el Análisis de la
Escritura Emocional Autorreflexiva (EEA) en mujeres víctimas de
violencia

presentada por

Ing. Mario Iván Ortiz Velázquez

como requisito para la obtención del grado de
Maestría en Ciencias la Computación

Director de tesis

Dr. Noé Alejandro Castro Sánchez

Codirectora de tesis

Dra. Leonor Rivera Rivera

Cuernavaca, Morelos, México. marzo de 2022.

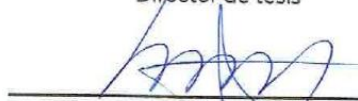
Cuernavaca, Morelos, 18/marzo/2022

OFICIO No. DCC/023/2022
Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

DR. CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial del C. MARIO IVÁN ORTIZ VELÁZQUEZ, con número de control M19CE061, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado "PROCESAMIENTO DE LENGUAJE NATURAL PARA EL ANÁLISIS DE LA ESCRITURA EMOCIONAL AUTORREFLEXIVA EN MUJERES VÍCTIMAS DE VIOLENCIA", y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.


DR. NOÉ ALEJANDRO CASTRO SÁNCHEZ
Director de tesis


DR. JUAN GABRIEL GONZÁLEZ SERNA
Revisor


DRA. LEONOR RIVERA RIVERA
Codirectora de tesis


DR. DANTE MÚJICA VARGAS
Revisor

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante
JGGS/ibm

Resumen

Este trabajo de investigación se enfoca en la propuesta de una solución computacional para el análisis de algunas características de los textos obtenidos a partir de la aplicación de una terapia de escritura narrativa, con la finalidad de proporcionar herramientas y una base sólida, a partir de la cual, se pueda analizar la utilidad de dicho enfoque para el análisis de esta terapia, aplicando el estudio a una muestra de mujeres víctimas de algún evento de violencia. Una parte central en este trabajo es la implementación de técnicas de Procesamiento de Lenguaje Natural para llevar a cabo el análisis de los textos, así como la extracción de in conjunto de características representativas de dichos datos de entrada, enfocándose en los niveles morfológico y sintáctico.

Si bien las tareas de análisis de textos y extracción de características se utilizan de forma común para abordar las problemáticas relacionadas con el análisis del lenguaje, la traducción de textos o la corrección ortográfica, lo que destaca a esta investigación es su aplicación a una problemática actual y real, en donde no se parte de conjuntos de datos existentes para el análisis, sustituyendo dicho proceder por la construcción misma del conjunto de datos, y su posterior análisis. Debido a lo anterior durante el desarrollo de esta tesis se desarrollaron dos herramientas independientes, pero complementarias entre sí, siendo una de ellas una interfaz para la aplicación de la terapia de escritura narrativa, también conocida como terapia de Escritura Emocional Autorreflexiva (EEA).

El proyecto desarrollado servirá como base para futuras investigaciones relacionadas con la aplicación de Procesamiento de Lenguaje Natural para el estudio de datos textuales procedentes de la aplicación de la terapia EEA a diferentes poblaciones de estudio.

Abstract

This research work focuses on the proposal of a computational solution for the analysis of some characteristics of the texts obtained from the application of a narrative writing therapy, in order to provide tools and a solid base, from the which, the usefulness of said approach for the analysis of this therapy can be analyzed, applying the study to a sample of women victims of some event of violence. A central part in this work is the implementation of Natural Language Processing techniques to carry out the analysis of the texts, as well as the extraction of a set of representative characteristics of said input data, focusing on the morphological and syntactic levels.

Although text analysis and feature extraction tasks are commonly used to address problems related to language analysis, text translation or spelling correction, what stands out in this research is its application to a current problem. and real, where it is not based on existing data sets for the analysis, substituting said procedure for the construction of the data set itself, and its subsequent analysis. Due to the above, during the development of this thesis, two independent but complementary tools were developed, one of them being an interface for the application of narrative writing therapy, also known as Self-Reflective Emotional Writing therapy (SEE).

The developed project will serve as a basis for future research related to the application of Natural Language Processing for the study of textual data from the application of EEA therapy to different study populations.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por su Programa Nacional de Posgrados de Calidad (PNPC) por medio del cual me fue otorgada una beca para ser estudiante de tiempo completo y desarrollar este trabajo de investigación. Así como al Tecnológico Nacional de México, Campus Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET).

Doy las gracias a los profesores que contribuyeron con mi formación como Maestro en Ciencias de la Computación, especialmente a mi director, Dr. Noé Alejandro Castro Sánchez y a mi codirectora Dra. Leonor Rivera Rivera. Así mismo agradezco a mis profesores y revisores Dr. Dante Mújica Vargas y Dr. Juan Gabriel González Serna.

Dedicatoria

A mi madre con mucho cariño, por su apoyo incondicional en mi preparación como profesionalista. Por su amor, comprensión y motivación, y por enseñarme a no conformarme y siempre buscar mejores oportunidades de desarrollo personal y profesional.

Índice de contenido

Capítulo I	11
Introducción	11
1.1 Introducción	12
1.2 Descripción del problema	13
1.3 Justificación	15
1.4 Objetivos.....	17
1.5 Alcances y Limitaciones	18
Limitaciones	18
Capítulo II	19
Marco teórico	19
2.1 Contexto social	20
2.2 Contexto computacional	21
Capítulo III	23
Estado del arte	23
3.1 Trabajos relacionados con el Procesamiento de Lenguaje Natural	24
3.2 Trabajos relacionados con la violencia y los trastornos psiquiátricos asociados a eventos de esta naturaleza	30
3.3 Tabla comparativa de los trabajos relacionados.....	36
Capítulo IV	41
Metodología de solución	41
Capítulo V	49
Desarrollo de la solución.....	49
Capítulo VI	89
Pruebas	89
6.4 Precisión de los módulos de análisis de textos	105
Capítulo VI	106
Resultados y conclusiones.....	106
6.1 Resultados obtenidos	107
6.2 Conclusiones.....	109

6.3 Trabajos futuros	111
Aportaciones y productos académicos	112
Referencias	113
Anexos	119

Índice de tablas

Tabla 1. Resumen de los elementos clave en los que se justifica este trabajo.	16
Tabla 2. Comparativa de trabajos relacionados al Procesamiento de Lenguaje Natural (PLN), parte 1.....	37
Tabla 3. Comparativa de trabajos relacionados al Procesamiento de Lenguaje Natural (PLN), parte 2.....	38
Tabla 4. Comparativa de trabajos relacionados con la violencia y el tratamiento de trastornos, parte 1.....	39
Tabla 5. Comparativa de trabajos relacionados con la violencia y el tratamiento de trastornos, parte 2.....	40
Tabla 6. Porcentaje de aparición y no aparición de los datos dentro del corpus.....	55
Tabla 7. Principales características del corpus.....	56
Tabla 8. Listado de las variables que se obtienen utilizando la herramienta HAAT y una breve descripción de estas.....	59
Tabla 9. Descripción del puntaje de la prueba de Flesch.....	62
Tabla 10. Composición del corpus NVS2019 y porcentaje de error de procesamiento.....	90
Tabla 11: Características de la estructura del texto del corpus NVS2019	91
Tabla 12: Características de sentimiento extraídas del corpus NVS2019.....	91
Tabla 13: Resultados obtenidos del análisis de emociones en el corpus NVS2019.....	93
Tabla 14: Comparativa de precisión obtenida en las pruebas del corpus NVS2019.....	94
Tabla 15: Comparativa de precisión obtenida en las pruebas del corpus de revisión de artículos... ..	95
Tabla 16: Características relacionadas con la estructura sintáctica del corpus de revisión de artículos.....	96
Tabla 17: Precisión obtenida en el análisis de sentimientos y emociones con el corpus de revisión de artículos	97
Tabla 18: Porcentaje de error en el análisis de los textos obtenidos de la prueba piloto de la terapia de escritura	100
Tabla 19: Estructura de los textos obtenidos de la prueba de la terapia de escritura	101
Tabla 20: Características de sentimiento de los textos obtenidos de la terapia de escritura	102
Tabla 21: Etiquetado de emociones de los textos obtenidos de la terapia de escritura.....	103
Tabla 22: Resumen de la precisión obtenida por cada uno de los módulos de análisis de la herramienta HAAT	105

Índice de figuras

Figura 1. Actividades por realizar para el desarrollo de la herramienta de análisis de textos.	42
Figura 2. Procedimiento para la recopilación y análisis de los textos procedentes de la terapia de EEA.	46
Figura 3. Relato en forma textual obtenido de Facebook, relacionado con la iniciativa Cuéntalo.	54
Figura 4. Uno de los registros que componen el corpus NVS2019.	56
Figura 5. Proceso de análisis de texto individual.	58
Figura 6. Modelo de valencia – activación (izquierda) y modelo de polaridad – subjetividad (derecha).	65
Figura 7. Interfaz web de la herramienta HAAT.	67
Figura 8. Servidores de análisis en ejecución.	68
Figura 9. Encabezado de la página de inicio del portal web.	69
Figura 10. Esquema que muestra la relación entre las interfaces de la página web.	70
Figura 11. Menú de inicio de la ventana para realizar la terapia de EEA	72
Figura 12. Ventana para la aplicación del ejercicio de escritura.	74
Figura 13. Relación de las interfaces incorporadas en el sistema web para aplicar la terapia.	77
Figura 14. Estructura y relación de la base de datos de la terapia web	81
Figura 15. Aspecto de la pantalla de navegación principal de la aplicación móvil	83
Figura 16. Diagrama de las interfaces que componen la aplicación móvil.	86
Figura 17. Estructura y relación de la base de datos de la aplicación móvil para la aplicación de la terapia.	87
Figura 18. Puntaje de sentimiento de cada texto del corpus NVS2019	92
Figura 19. Etiquetado de emociones en el corpus NVS2019	94
Figura 20. Polaridad de sentimiento con textos positivos corpus de revisión de artículos	98
Figura 21. Polaridad emocional con textos positivos corpus de revisión de artículos	99
Figura 22. Interfaz web de aplicación de la terapia de escritura	100
Figura 23. Aplicación móvil para la aplicación de la terapia de escritura	101
Figura 24. Polaridad de sentimiento de los textos obtenidos de la terapia de escritura	103
Figura 25. Etiquetado de emociones de los textos obtenidos de la terapia de escritura	104

Capítulo I

Introducción

1.1 Introducción

El procesamiento de Lenguaje Natural (PLN) se enmarca dentro del campo de estudio de la Inteligencia Artificial (IA) como la rama enfocada en la ejecución de procesos de análisis de datos textuales redactados a partir del uso de lenguajes naturales, con la finalidad de realizar la interpretación y manipulación de los datos, aportando información relevante en el contexto de un análisis de la estructura del texto o permitiendo el desarrollo de aplicaciones e interfaces con la capacidad de satisfacer una tarea concreta, como los traductores automáticos, correctores ortográficos, motores de búsqueda o en tareas de generación de lenguaje natural.

El uso de técnicas de PLN en campos ajenos a la computación actualmente es escaso en México. Uno de los campos que representa una gran oportunidad para la inclusión de la Inteligencia Artificial es el del tratamiento psicológico de pacientes o para el estudio de problemas asociados a la salud pública, específicamente en lo referente al fenómeno de la violencia experimentado por una población considerable de mujeres.

La terapia de Escritura Emocional Autorreflexiva (EEA) es un tipo de intervención alternativa aplicada para el tratamiento de trastornos ocasionados por uno, o más, eventos detonantes, los cuales ocasionan la aparición de síntomas dañinos para la salud psicológica del paciente. De entre todos los tipos de síntomas presentes en un individuo expuesto a este tipo de violencia, el principal es la reexperimentación, malestar que se presenta en la víctima en forma de conductas de evitación, o llevándole a presentar la sensación de indefensión. Estos síntomas se asocian al Trastorno de Estrés Postraumático (TEPT). Por medio de la escritura sobre el evento traumático acontecido se puede producir una mejora en la salud y bienestar del paciente, por medio de la erradicación o disminución de los síntomas negativos aparecidos a raíz del evento experimentado.

En este trabajo se plantea la utilización del Procesamiento de Lenguaje Natural, por medio de la aplicación del procesamiento de textos, análisis de sentimientos, modelado de tópicos y análisis de la estructura narrativa, con la finalidad de generar una herramienta de análisis del lenguaje utilizado durante la aplicación de la terapia de Escritura Emocional Autorreflexiva (EEA) en una población de mujeres expuestas a eventos de violencia, tomando como base la información en formato escrito como vía de transmisión de información, ideas y sentimientos hacia otros individuos. El objetivo del estudio es desarrollar una herramienta para el análisis de los textos generados por medio de la aplicación de la terapia EEA en mujeres víctimas de violencia, y que presenten síntomas de TEPT.

1.2 Descripción del problema

La violencia que se ejerce en contra de las mujeres, predominantemente de parte de los hombres, se describe por parte de la Organización Mundial de la Salud (OMS) como la ejecución de actos dañinos en contra de la salud física o emocional, realizados por una persona cercana la víctima, o por un individuo externo. Los hechos se pueden dar en forma de un evento único o con un desarrollo progresivo, provocando lesiones o daños corporales, mediante golpes, contacto físico o haciendo uso de un objeto como arma, también se incluyen una serie de comportamientos en contra de la mujer, como lo son humillaciones, insultos, críticas o amenazas.

Los tratamientos que en la actualidad son suministrados a víctimas de violencia se enfocan en la reparación de los daños a la salud física, psicológica y reproductiva de la víctima, poniendo un menor énfasis en subsanar el daño en la salud mental del afectado. Una técnica con la que actualmente se experimenta como medida terapéutica alternativa a la aplicación en mujeres que han sido víctimas de violencia es la técnica de la Escritura Emocional Autorreflexiva (EEA).

La terapia EEA es un proceso utilizado en el campo de la psicología y que consiste en solicitar a un paciente, previamente diagnosticado con la sintomatología propia de algún trastorno psicológico, la redacción de su experiencia traumática en una serie de escritos individuales, los cuales se redactan de forma separada a través de un rango de tiempo previamente establecido por el terapeuta. El objetivo del procedimiento es generar una reestructuración cognitiva y la toma de conciencia acerca de los eventos experimentados, generando por medio del proceso de redacción una mejora en el estado emocional del paciente al fomentar la reducción de los síntomas negativos.

Procesar los textos redactados y entrevistas que se aplican para obtener información procedente de las mujeres víctimas puede arrojar información de gran valor médico, al permitir el conocimiento de la evolución emocional del paciente y los cambios narrativos asociados a su reestructuración cognitiva. Una ventaja adicional es el estudio del fenómeno de forma más detallada, conllevando a la adquisición de nuevos conocimientos sobre la efectividad de la terapia y mejoras en su aplicación.

En la actualidad el análisis de los textos que se obtienen a través de la EEA se basa en el uso principal de herramientas de tipo test, haciendo uso de escalas Likert para puntuar una serie de preguntas y respuestas predefinidas, lo que da como resultado un puntaje general que se compara con la escala correspondiente a dicho instrumento de evaluación. El diagnóstico entonces se basa en los puntajes obtenidos por medio de pruebas como la Escala de Síntomas de Estrés Postraumático (PSS) o la Escala de Evaluación Global de Estrés Postraumático (EGEP).

Los textos que se redactan como parte de la terapia no son muy utilizados para el diagnóstico final por parte de los especialistas del área, lo cual principalmente se debe a la incapacidad por parte de los terapeutas de obtener información relevante del texto, ya que la escritura realizada por una persona contiene una estructura propia, debido a factores como el nivel educativo, el lenguaje utilizado, así como factores sociales y culturales propios de cada individuo. Una consideración adicional es que la terapia de EEA se considera una intervención rápida, lo que implica un número mínimo de sesiones entre el paciente y el terapeuta, que en el paradigma original de Pennebaker son 4 sesiones, aplicadas en un lapso de tiempo variable que depende de las condiciones iniciales determinadas por el terapeuta y por el tiempo disponible para realizar el tratamiento, así como de la propia disponibilidad del paciente.

Existe un nicho de oportunidad para el desarrollo de una solución que permita realizar un análisis implementando Inteligencia Artificial para el estudio del lenguaje utilizado por los pacientes al redactar los textos en la terapia EEA. En este trabajo de investigación se propone el desarrollo de una herramienta para el análisis de los textos recopilados a partir de la aplicación de la terapia de EEA, realizando de manera automática el proceso a través del análisis de sentimientos, el modelado de tópicos y el análisis de la estructura narrativa, permitiendo la disminución en el tiempo de estudio de los textos y proporcionando una descripción estadística del contenido de los textos.

1.3 Justificación

La violencia es un fenómeno social que afecta a mujeres de distintas clases sociales, orígenes, estado civil, nivel educativo, raza u orientación sexual. Es una de las formas principales de violación de los derechos humanos, con lo cual se afecta la vida de las mujeres y su integridad física y psicológica (Trigueiro et al., 2017). Este fenómeno tiene consecuencias negativas que se manifiestan tanto físicamente como a nivel psicológico, desembocando en síntomas provocados por el evento altamente estresante.

Los eventos altamente estresantes son alteraciones en la historia de vida de un individuo, que generan la necesidad de adaptarse para sobrellevar las secuelas de un episodio violento. Estos eventos son potencialmente traumáticos cuando su exposición implica una amenaza para su integridad física o la de los demás (He, Veldkamp, Glas & de Vries, 2016). En los procesos psicológicos que se desencadenan tras un evento altamente estresante resulta relevante la percepción subjetiva de la amenaza, conocida como severidad subjetiva del evento, lo que puede ser un indicador de la aparición de Trastorno de Estrés Postraumático (TEPT), entre otros trastornos.

El TEPT es una modificación en la personalidad de una persona que vive o experimenta un trauma, los síntomas pueden aparecer en los meses posteriores al evento. Si una persona no consigue recuperarse de los síntomas después de un tiempo se considera entonces que padece este trastorno psicológico, el cual es común entre veteranos de guerra, personas accidentadas y víctimas de violencia sexual. La agresión psicológica y física está asociado con un alto riesgo psicológico y con una mayor probabilidad de padecimientos psiquiátricos, entre los cuales se encuentra el TEPT. Aunque los eventos potencialmente traumáticos como la agresión sexual son asociados con una amplia gama de trastornos psiquiátricos y traumas, en el caso de las víctimas de violencia la ocurrencia del evento es el detonante de los síntomas.

Los patrones de habla y escritura de los individuos pueden proporcionar una ventana a sus perspectivas emocionales y cognitivas. Se sugiere que las palabras y expresiones de los pacientes pueden aportar detalles para la detección temprana de enfermedades). En tal caso el lenguaje se convierte en un medio importante para que los psicólogos y terapeutas puedan entender a la persona que muestra síntomas por haber sido expuesta a un evento altamente estresante.

Con el desarrollo de este trabajo de investigación se buscó desarrollar una herramienta para realizar de forma automática el análisis de lingüístico de la estructura narrativa de textos obtenidos mediante la aplicación de la técnica de Escritura Emocional Autorreflexiva, enfocando la aplicación de la terapia a una población de mujeres víctimas de violencia.

En la siguiente tabla se resumen las partes más relevantes en las que se basa la justificación para llevar a cabo este trabajo de investigación:

Tabla 1. Resumen de los elementos clave en los que se justifica este trabajo.

¿Cuál es el problema?	La aplicación de la terapia de EEA se realiza de forma manual y con resultados superficiales.
¿Por qué es un problema?	Requiere mucho tiempo y los resultados no se basan en el análisis del texto.
¿Que se ha hecho para resolverlo?	PLN y ML para la clasificación de textos asociados a trastornos psiquiátricos.
Propuesta de solución	Desarrollar una herramienta para el análisis automático de textos procedentes de la aplicación de la terapia de EEA, con la finalidad de analizar su estructura lingüística.

1.4 Objetivos

Objetivo general

Desarrollar una herramienta para el análisis automático de textos que permita el análisis de la estructura y elementos lingüísticos presentes en los textos obtenidos por medio de la aplicación de la terapia de Escritura Emocional Autorreflexiva (EEA) en mujeres víctimas de violencia, haciendo uso de Procesamiento de Lenguaje Natural.

Objetivos específicos

- Diseñar y desarrollar una herramienta para el análisis de la estructura narrativa, análisis de sentimientos y características de los textos obtenidos a partir de la implementación de la terapia de EEA.
- Recolectar los textos generados por medio de la aplicación de la terapia a mujeres víctimas de violencia, conformando un corpus.
- Identificar las características lingüísticas relevantes de los textos generados a partir de la terapia de EEA mediante la comparativa de los resultados obtenidos del análisis individual de cada uno de los textos.
- Conformar y analizar un corpus lingüístico que agrupe todos los textos analizados.

1.5 Alcances y Limitaciones

Alcances

- Se desarrollará una herramienta que permita procesar los textos obtenidos por medio de la técnica de EEA, realizando un proceso de recopilación y análisis, para generar un estudio de los elementos lingüísticos y la estructura narrativa presente en cada uno de los textos.
- Se realizará el análisis de textos procedentes de mujeres víctimas de violencia y con síntomas de Trastorno de Estrés Postraumático (TEPT) o sin ellos.
- La determinación de sintomatología relacionada a TEPT en las mujeres a intervenir se realizará por medio de la aplicación de la Lista Checable de Trastorno por Estrés Postraumático (PCL) para DSM-5.
- La herramienta que se desarrollará funcionará independientemente de la población a la que se le aplique la técnica, permitiendo el análisis de la terapia de EEA aplicada a otras poblaciones de estudio.

Limitaciones

- La herramienta que se desarrollará realizará el análisis únicamente en el idioma español.
- No se mostrará ningún tipo de diagnóstico médico.
- La herramienta llevará a cabo un análisis descriptivo de los textos, por lo tanto, la interpretación de los datos generados dependerá de los terapeutas, psicólogos o especialistas en el tratamiento de mujeres víctimas de violencia.

Capítulo II

Marco teórico

En el presente capítulo se muestran los conceptos teóricos relacionados con las temáticas tratadas en esta investigación, con la finalidad de proporcionar una definición concreta que permita abordar la lectura del documento con una base teórica.

2.1 Contexto social

En esta sección se muestran los conceptos relacionados a la problemática social en la que se contextualiza el problema planteado y que sirven como base para el entendimiento del fenómeno social en el que se enmarca este trabajo de investigación.

2.1.1 Trastorno de Estrés Postraumático (TEPT)

De acuerdo a la American Psychiatric Association (Ferrer & R. Delgado, 2018) el Trastorno de Estrés Postraumático (TEPT) se define como un trastorno que se caracteriza por la exposición a un evento de violencia, presentando síntomas intrusivos como pesadillas, reacciones disociativas en las que el sujeto siente o actúa como si se diera una repetición del evento, malestar psicológico intenso y prolongado, así como reacciones fisiológicas intensas. La sintomatología relacionada a este padecimiento también incluye la evitación de estímulos asociados al suceso traumático y las alteraciones en el estado anímico tras el evento.

En el Manual Diagnóstico y Estadístico para las Enfermedades Mentales (DSM-V5) (Dimsdale et al., 2019) el TEPT se menciona que el diagnóstico de TEPT se realiza con base en la aparición y presentación de la sintomatología relacionada a este trastorno, adicionalmente se menciona que el individuo tiende a experimentar síntomas recurrentes, o persistentes de despersonalización y desrealización.

2.1.2 Terapia

De acuerdo con la American Psychiatric Association (Ferrer & R. Delgado, 2018) el concepto de terapia es definido como el conjunto de técnicas, medios y métodos que tienen como finalidad el alivio o curación de la sintomatología relacionada a enfermedades o padecimientos fisiológicos y psiquiátricos, velando por el bienestar del paciente.

2.1.3 Escritura Emocional Autorreflexiva (EEA)

Escribir sobre eventos traumáticos o estresantes produce una mejora en la salud física y psicológica.

Por medio del paradigma de la escritura reflexiva, también referido como escritura emocional, terapia de escritura emocional, terapia de escritura narrativa o Escritura Emocional Autorreflexiva (EEA), las personas víctimas de eventos de naturaleza traumática cuentan con una herramienta terapéutica para el control, manejo y superación de las secuelas ocasionadas por el evento disparador de la sintomatología (Baikie & Wilhelm, 2005).

Para Pennebaker (Pennebaker, 2004) enfrentar un trauma al escribir sobre el evento y reconocer las emociones asociadas conlleva a la reducción gradual del estrés psicológico y físico de la persona. La confrontación del evento ocurrido requiere la traducción de los pensamientos y sentimientos a palabras, generando en el proceso una integración cognitiva y una comprensión con mayor detalle de la situación, contribuyendo con la disminución de los síntomas fisiológicos y emocionales.

2.2 Contexto computacional

En esta sección se muestran los conceptos relacionados al Procesamiento de Lenguaje Natural y a las técnicas computacionales utilizadas para el análisis de textos desde el punto de vista de la ingeniería.

2.2.1 Procesamiento de Lenguaje Natural (PLN)

La manipulación de lenguajes naturales, los cuales se han desarrollado a lo largo del tiempo para permitir la comunicación entre los seres humanos con la finalidad de razonar y proporcionar un medio de comunicación efectiva, es una de las áreas de estudio de la Inteligencia Artificial (IA). La manipulación de dichos lenguajes se realiza mediante el uso de métodos, técnicas, lenguajes de programación y herramientas computacionales, proporcionándose de esa forma los mecanismos necesarios para el procesamiento y análisis de la información lingüística que se genera por medio del proceso de interacción.

El uso de PLN permite el desarrollo de aplicaciones que permitan la interacción humano-computadora entre un usuario y sistemas computacionales e informáticos, siendo una alternativa a interfaces gráficas, permitiendo el desarrollo de interacciones intuitivas que proporcionan una ventaja en la comunicación.

2.2.2 Preprocesamiento de textos

El procesamiento de textos hace referencia a una etapa inicial que contempla la labor de estructurar el texto en un formato simplificado, eliminando caracteres no relevantes y palabras no representativas, así como signos de puntuación y otros elementos

no esenciales para el análisis del texto. En esta etapa se parte de un conjunto de datos textuales no estructurados y se agrupan en un corpus lingüístico, entendiéndose este último término como un conjunto representativo de textos referentes a una temática o fenómeno específico.

2.2.3 Análisis de sentimientos

El análisis de sentimientos hace referencia al estudio subjetivo de las opiniones subjetivas que son generadas por los usuarios de distintos portales de internet, incluyendo redes sociales y plataformas de comercio electrónico (Sobrino Sande, 2018). Dichas opiniones son referentes a productos, servicios o temas de interés y proporcionan una fuente de datos analizable para determinar la polaridad emocional, o las emociones y sentimientos a las que se apela por parte de los usuarios. Formalmente este campo de estudio se define como el tratamiento de opiniones y sentimientos subjetivos representados en forma de textos y procesados por medio de técnicas computacionales.

Capítulo III

Estado del arte

En esta sección se presentan los trabajos de investigación relacionados con el trabajo de tesis desarrollado en este documento, presentando el objetivo, desarrollo y conclusión. Al final del capítulo se incluye una tabla comparativa con la relación entre de los trabajos que componen el estado del arte y esta propuesta de tesis.

El estado del arte se divide en dos partes, la primera consta de investigaciones relacionadas con al Procesamiento de Lenguaje Natural (PLN), mientras que la segunda parte contiene un conjunto de trabajos de investigación relacionados con la violencia y con los trastornos psicológicos que son ocasionados por este fenómeno social.

3.1 Trabajos relacionados con el Procesamiento de Lenguaje Natural

En esta sección se incluyen los trabajos de investigación enfocados en el estudio, uso y aplicación de técnicas de PLN, y que se relacionan con este trabajo.

Evaluación automatizada de las auto narrativas de los pacientes con trastorno de estrés postraumático mediante procesamiento del lenguaje natural y minería de textos

En el trabajo de investigación de (He, Veldkamp, Glas & de Vries, 2016) se presenta un sistema de evaluación automático para la detección de pacientes con trastorno de estrés postraumático (TEPT), haciendo uso de técnicas de Procesamiento de Lenguaje Natural (PLN) y minería de textos. Se utilizaron cuatro algoritmos de aprendizaje automático, en combinación con modelos de representación para la identificación de patrones con características verbales, haciendo uso de escritura emocional y auto narrativa para la recopilación de los datos necesarios.

Dados los esfuerzos para el desarrollo de un método de evaluación textual basado en palabras clave, este estudio buscó aplicar técnicas de minería de datos en el diagnóstico de TEPT, y evaluar su eficacia en conjunto con diferentes modelos de clasificación de texto, siendo necesario que los textos auto narrativos, utilizados para la recopilación de información, fueran escritos por la persona participante en el estudio, no por familiares o amigos, así mismo, el texto debió describir eventos traumáticos y su impacto en la vida diaria de la persona víctima. El resultado mostrado en este trabajo es un sistema automático para el diagnóstico de estrés postraumático, basado en escritos auto narrativos, y utilizando técnicas de PLN y minería de texto, concluyendo que el método planteado podría mejorar el procedimiento de diagnóstico de TEPT.

Construcción de un corpus marcado con emociones para el análisis de sentimientos en twitter en español

El objetivo de este trabajo de investigación fue la recopilación y evaluación de un corpus lingüístico conformado por tweets y hashtags con un significado o interpretación emocional o sentimental. Con base en lo anterior se realiza la evaluación de la herramienta SentiStrength2 para detectar la fuerza del sentimiento en textos cortos de redes sociales, en este caso twitter. De acuerdo a (Sidorov, Galicia & Camacho, 2019), un sentimiento es definido como una actitud, o un juicio, provocado por un hecho.

A partir de la definición anterior, se realizó un pre procesamiento de los tweets, detectando la posible relación entre las preposiciones y conjunciones con expresiones de sentimientos. Durante el mismo proceso se aplicó el diccionario Linguistic Inquiry and Word Count (LIWC), el cuál mide emociones negativas, positivas y específicas, así como palabras y raíces de palabras dentro de setenta categorías.

En este trabajo se realizó la aplicación de un pre procesamiento de características, polaridad y desequilibrio contextual, obteniendo las raíces de cada palabra. Con base en el uso de clasificadores se evaluaron las características ambigüedad, polaridad, inesperado y escenarios emocionales. Como resultado de este trabajo se generó un corpus lingüístico basado en las emociones extraídas de la red social twitter.

Clasificación de roles semánticos usando características sintácticas, semánticas y contextuales

El objetivo de este trabajo, desarrollado por (Reyes et al., 2013), fue identificar, mediante la tarea de clasificación, el tipo de rol semántico existente entre un evento y sus actantes; para ello se presenta un análisis de características para seleccionar un subconjunto que mejore el desempeño de la tarea. El interés principal del trabajo fue determinar qué rol semántico lleva a cabo cada entidad relacionada a un evento, esto en documentos periodísticos en español. El trabajo se basó en la teoría de L. Tesniere y M.A.K. Holliday, quienes consideran que los elementos fundamentales de la oración son: Los actores (o actantes), la acción (verbo) y el decorado (circunstantes). Adicionalmente se consideran cinco categorías de roles semánticos.

Se realizó un proceso de selección de características para observar su impacto en la clasificación de roles semánticos, determinando el tipo de rol existente (agente, objeto, beneficiario, circunstante u otro), en pares evento-entidad. Por lo tanto, se buscó construir un clasificador semántico de roles, basado en el conjunto de pares de entrenamiento. Como resultado se logró la clasificación de las categorías de roles semánticos.

Análisis de sentimientos a nivel de emoción: la evaluación de la opinión efectiva

Investigación llevada a cabo por (Almashraee, Monett & Paschke, 2016). El planteamiento del problema se basó en que una clasificación de sentimientos extraídos de un texto, abarcando únicamente dos categorías: una positiva y una negativa, arroja una conclusión incompleta, lo cual lleva a una investigación a evaluación insuficiente. Por lo tanto, este trabajo se centró en determinar una metodología más completa para realizar la extracción de opiniones y evaluar las emociones presentes en textos, basándose plenamente en los métodos existentes en el análisis de sentimientos, una rama del Procesamiento de Lenguaje Natural.

En este trabajo se propone la extracción y procesamiento de palabras referentes a la opinión de las personas acerca de un producto publicitado por internet. Con base en la opinión original, se procede a realizar una serie de tareas de pre-procesamiento, una vez contando con una serie de sentencias etiquetadas y con un tipo definido, se procedió a determinar la intensidad del sentimiento presente, por medio de la detección de los intensificadores, los cuales son elementos que cambian el grado de intensidad de un sentimiento. El resultado de la investigación muestra el proceso de asignación de la polaridad emocional a una palabra, por medio del enfoque de Plutchik, y obteniendo como valores de salida puntajes complementarios al etiquetado de palabras en categorías simples.

Reconociendo emociones en texto usando un conjunto de clasificadores

En este trabajo (Isidoros & Ioannis, 2016) se presentó un sistema de análisis de sentimientos para el reconocimiento automático de emociones en el texto, utilizando un conjunto de clasificadores. El trabajo se centró en el uso de tres clasificadores: Dos del tipo estadístico (Naive Bayes y Máxima Entropía), y un tercero de tipo herramienta de conocimiento, basada en el análisis de las sentencias del lenguaje natural. El flujo masivo y constante de datos textuales en la web puede reflejar los sentimientos, opiniones y pensamientos de las personas sobre una variedad de fenómenos con distintas temáticas, o de distinta naturaleza.

El análisis de la información textual puede proporcionar factores altamente representativos sobre el estado emocional de una persona o comunidad. En ese sentido, los modelos emocionales se emplean con la finalidad de determinar la postura del individuo acerca de una entidad particular. En este artículo se presenta un conjunto de clasificadores aplicados al análisis de sentimientos en textos.

Se hace uso de más de un clasificador con el fin de integrar de manera conjunta las

ventajas de cada uno, así como minimizar sus vulnerabilidades, mejorando el rendimiento general de la clasificación de los sentimientos. El resultado obtenido fue el reconocimiento de la polaridad emocional con unos resultados lo suficientemente buenos como para considerar que la implementación de múltiples clasificadores puede ser un camino viable para realizar un análisis profundo, obteniendo mejores resultados.

Reconociendo las emociones en el texto utilizando redes neuronales y lógica difusa

Trabajo de investigación desarrollado por (Kanger & Bathla, 2017), en donde se realizó el análisis de sentimiento utilizando lógica difusa y redes neuronales, para la extracción de emociones en textos. El método propuesto buscó mejorar los resultados obtenidos en términos de precisión, sensibilidad y especificidad. Uno de los aspectos principales aplicados en este trabajo fue la implementación de tres clasificadores: Naive Bayes, Máxima Entropía y una base de conocimiento. La finalidad de realizar dicha integración se debió a que la tarea de análisis de sentimientos es compleja, por lo tanto se buscó obtener las ventajas que cada uno puede ofrecer de manera independiente.

La metodología propuesta por los autores se basó en seis etapas, separadas en dos sub procesos independientes, siendo el primero el responsable de realizar una descarga y procesamiento inicial de los textos, seguido del proceso de análisis de sentimientos para posteriormente realizar el almacenamiento de los datos en una base de conocimiento. El segundo proceso se hizo cargo de realizar las pruebas del modelo previamente entrenado, para determinar el nivel de precisión del algoritmo. Como resultado se logró concluir que la combinación de técnicas de minería de datos y aprendizaje automático permite la obtención de resultados con mayor precisión, en comparación con el uso individual de ambas técnicas.

Extracción de texto a emoción utilizando técnicas supervisadas de aprendizaje automático

En este trabajo (Azim & Bhuiyan, 2018) se presentó un enfoque centrado en el contenido textual producido en las diferentes plataformas y medios informáticos, abordando técnicas multidisciplinarias para automatizar el proceso de análisis de las emociones que forman parte de la estructura de los textos y opiniones compartidas. Los autores de este trabajo realizaron un proceso exploratorio, concluyendo que todos los sistemas de análisis afectivo asignan a cada palabra un valor de intensidad, esto dependiendo de las diferentes emociones.

El sistema propuesto en este trabajo buscó eliminar la necesidad de realizar manualmente la creación de un léxico de intensidad emocional, y para el etiquetado de

emociones se utilizaron nueve categorías. Este trabajo contribuyó, en primer lugar, al proponer una aproximación basada en el aprendizaje automático para extraer las categorías básicas de las emociones a partir de textos, eliminando la necesidad de realizar la construcción manual de un léxico de intensidad emocional. En segundo lugar, se buscó realizar una comparación entre tres clasificadores para evaluar el desempeño en la tarea de extracción emocional de textos, información que se extrajo de redes sociales. Como resultado se obtuvo una aproximación basada en el aprendizaje automático para la extracción de categorías emocionales a partir de textos.

Análisis de sentimientos multilingüe: estado del arte y comparación independiente de técnicas

Trabajo en el que (Dashtipour et al., 2016) compara las investigaciones existentes sobre el análisis de sentimiento, tratando de validar si estos trabajos permiten una implementación precisa y una reproducción confiable de los resultados informados. En este documento se discuten los enfoques existentes y se informan los resultados de la reproducción experimental de los enfoques analizados, lo que permitió una comparación directa de los resultados.

Las principales técnicas utilizadas para el análisis de sentimientos, así como para el procesamiento de la información, recursos léxicos, herramientas y características se pueden clasificar en enfoques basados en aprendizaje automático, enfoques basados en léxicos y enfoques híbridos.

Los resultados obtenidos en este trabajo, aplicando una variedad de técnicas y herramientas de análisis de sentimientos, permitieron una comparación directa de propuestas y llegar a la conclusión de que no siempre es posible la reproducción de las técnicas utilizadas previamente en otros trabajos de investigación, debido a que existen variables con naturaleza cambiante que dificultan la reproducción exacta de los experimentos que se validan o evalúan.

Una breve revisión de los desafíos éticos en el Procesamiento de Lenguaje Natural clínico

En este trabajo se realizó el análisis sobre los desafíos fundamentales que se dan a partir de la implementación de Procesamiento de Lenguaje Natural (PLN) en el estudio de textos clínicos. Para (Suster et al., 2017), la implementación de estas técnicas representa un nicho de oportunidad en el sentido de permitir la manipulación de grandes conjuntos de datos, sin embargo, a nivel ético se presentan problemas relacionados con el difícil acceso a los datos clínicos y a posibles problemas de privacidad, los cuales representan un obstáculo para el progreso del PLN en el ámbito clínico.

Como resultado de este trabajo se obtuvieron los desafíos fundamentales para la implementación de PLN en los ámbitos médico y clínico. Se realizó la señalización de los requisitos de consentimiento que pueden afectar el desarrollo de una investigación, teniendo a consideración el hecho de que los encuentros médico-pacientes no solo muestran problemas de salud, ya que también incluyen las opciones de estilo de vida de un individuo y posibles condiciones socialmente estigmatizadas. El consentimiento informado al paciente debe estar presente en aquellos actos que cuentan una protección legal, siendo necesaria su presencia para el desarrollo de una investigación con PLN con límites éticos y morales establecidos.

Marcadores lingüísticos del lenguaje natural moderado y absoluto

En este estudio presentado por (Al-Mosaiwi & Johnstone, 2018) se propuso un enfoque para la identificación de marcadores lingüísticos en textos extraídos de internet, en forma de reseñas, con la finalidad de realizar la clasificación alternativa a la escala Likert, por medio de la identificación de los tokens más representativos de dos categorías: Textos con reseñas absolutas y textos con reseñas moderadas. El objetivo de la investigación se basa en la hipótesis de que los marcadores podrían ampliar la comprensión del lenguaje y cognición relacionados con las respuestas absolutas y moderadas. Las respuestas absolutas se relacionan con una serie de factores cognitivos, sociales y culturales. Un menor coeficiente intelectual y un menor nivel educativo se asocian a respuestas absolutas, así como poca tolerancia a la ambigüedad y pensamiento simplista.

Una conclusión a la que se llega en este estudio es que los marcadores lingüísticos no pueden depender de un sujeto en específico, ya que esto puede variar. Por lo tanto, la selección de palabras se debe restringir a aquellas que transmiten poco o ningún contenido. Como resultado se obtuvo un conjunto de 31 tokens principales que representan la mayor capacidad de discriminación entre las clases absoluta y moderada. Haciendo uso de un clasificador bayesiano ingenio se realizó el análisis del texto, tratando y evaluando individualmente cada uno de los tokens. Del grupo de características lingüísticas, 11 fueron específicas para calificaciones absolutas y las 20 restantes para opiniones moderadas.

3.2 Trabajos relacionados con la violencia y los trastornos psiquiátricos asociados a eventos de esta naturaleza

En esta sección se incluyen los trabajos de investigación enfocados en el estudio de la violencia, así como de trastornos psiquiátricos y tratamientos.

Sintomatología del trastorno de estrés postraumático en una muestra de mujeres víctimas de violencia sexual

El trabajo desarrollado por (Lozano-Oyola et al., 2017) analiza la sintomatología del trastorno por estrés postraumático en mujeres víctimas de violencia sexual. La violencia de carácter sexual es un gran estresor y un acontecimiento traumático que puede provocar trastorno de estrés postraumático (TEP). Es frecuente que existan síntomas que son consecuencia de la victimización sexual. De entre todos los tipos de síntomas presentes en un individuo expuesto a este tipo de violencia, el principal es la reexperimentación, malestar que se presenta en la víctima en forma de conductas de evitación, o llevándole a presentar la sensación de indefensión. Este tipo de síntomas son considerados como los mejores predictores de la posterior aparición de TEP.

No hay un patrón único de respuesta a un hecho tan traumático como es la violencia sexual, por ello el objetivo de esta investigación se centra en comparar los síntomas presentes, asociados al TEP, en una muestra de mujeres abusadas sexualmente, en función del momento en el que se produjeron dichos abusos y del tipo de violencia recibida. Adicionalmente se plantea un método de selección de las participantes, se realiza una recopilación de datos a partir de una entrevista semiestructurada, y se realiza un diagnóstico con base en la escala de gravedad de síntomas del TEP, así como un análisis estadístico de los datos. Como resultado se aportó información de base para abordar el análisis de problemáticas relacionadas con el fenómeno de violencia sexual en contra de la mujer.

Rumiación, crecimiento y sintomatología postraumática en personas que han vivido experiencias altamente estresantes

El estudio (García et al., 2018) analizó la influencia de distintos tipos de rumiación (negativa, reflexiva, intrusiva y deliberada), y la severidad subjetiva del evento sobre el crecimiento postraumático (CPT), y la sintomatología postraumática (SPT). El término rumiación hace referencia a un proceso cognitivo que consiste en pensar repetidamente sobre el evento vivido o sus consecuencias. Siendo una clase de pensamiento consciente que gira en torno a un tema fundamental común y se repite en ausencia de demandas ambientales inmediatas. De acuerdo a este trabajo, los eventos altamente estresantes son alteraciones en la historia de vida de un individuo, que generan la necesidad de adaptarse.

Para los autores de este trabajo, surge la necesidad de estudiar los caminos que conducen al CPT y a la SPT, y discriminar cuales estrategias llevan a alguno de los posibles resultados. Esta situación se aborda a partir de la comparación diferencial de los procesos de rumiación, infiriendo que un tipo negativo o intrusivo sería un factor de vulnerabilidad, mientras que las rumiaciones deliberadas y reflexivas tendrían una influencia más fuerte. Haciendo uso de instrumentos, escalas de medición y cuestionarios, se realizó un proceso de análisis de datos, generándose un modelo que pretendió analizar el efecto que producen distintos tipos de rumiación en el desarrollo de CPT o de la SPT. Como resultado se obtuvo un modelo hipotetizado para analizar los efectos que producen distintos tipos de premiación en el desarrollo del Trastorno de Estrés Postraumático.

El sufrimiento psicológico en la vida cotidiana de las mujeres que han experimentado violencia sexual: un estudio fenomenológico

El objetivo de los autores fue comprender las acciones cotidianas de las mujeres que han sufrido violencia sexual, para generar una perspectiva fenomenológica que pudiera revelar elementos importantes para ampliar las posibilidades de reflexión y abordar este problema en el contexto de la asistencia sanitaria, la enseñanza y la investigación. Se utilizó el marco teórico y filosófico de Alfred Schütz, utilizando un método cualitativo, siendo desarrollado en un centro especializado en la atención a mujeres con experiencias en situaciones de violencia sexual (Trigueiro et al., 2017).

En este trabajo se realizó el agrupamiento de las víctimas en dos categorías: La primera mostro una vida cotidiana marcada por el miedo resultante del evento. El impacto de la violencia trascendió a distintos aspectos de la vida de la víctima, reflejándose negativamente en las actividades y relaciones sociales de la persona.

La segunda categoría se integró por mujeres que hicieron referencia a la superación del sufrimiento psicológico causado por la violencia sexual, apoyándose en familiares, amigos, actividades laborales, académicas y por la posibilidad de acudir a un centro de asistencia especializado en brindar apoyo a víctimas de este tipo de violencia. Los resultados obtenidos buscaron comprender como la violencia sexual, y los factores asociados a este fenómeno, influyo en la vida cotidiana de las mujeres que experimentaron un evento de este tipo.

Perfiles de estrés postraumático en adolescentes abusados sexualmente

Investigación realizada por (Guerra, Plaza & Farkas, 2017) con el objetivo de identificar distintos perfiles de víctimas adolescentes de acuerdo a su sintomatología de estrés postraumático, ya que existe una alta variabilidad en las reacciones emocionales

presentes en víctimas de abuso sexual, por lo tanto, es necesario conocer los factores que pueden explicar dicha variabilidad y utilizar esa información para generar tratamientos y políticas públicas más específicas.

La violencia sexual ha sido reconocida como un factor de riesgo para la salud mental, existiendo una diversidad de síntomas asociados a estas experiencias de vulneración, entre ellos: Estrés postraumático, depresión, bajo desempeño académico, agresividad, aislamiento y conducta sexualizada. El trastorno de Estrés Postraumático (TEPT) es el reportado con mayor frecuencia, tanto a nivel de sintomatología intrusiva (recuerdos involuntarios del acontecimiento traumático, pesadillas), síntomas evitativos (dificultades para enfrentar estímulos asociados al trauma) y de aumento de activación fisiológica.

Los resultados obtenidos en este trabajo contribuyen a la comprensión de la alta variabilidad existente en la sintomatología experimentada por distintas víctimas de abuso sexual. A diferencia de estudios previos que abordan esta problemática, y que se centran en el comportamiento de variables específicas, este estudio se centró en el estudio de las características o perfiles de los sujetos desde el enfoque de la persona. El objetivo fue evaluar, de modo empírico, distintos perfiles de adolescentes que han sufrido abusos sexuales de acuerdo con su sintomatología, esperando encontrar grupos con distinto nivel de sintomatología de TEPT.

Prevención del estrés postraumático relacionado con la agresión sexual a través de la intervención temprana: una revisión sistemática

De acuerdo con (Dworkin & Schumacher, 2016), comprender los elementos clave de las intervenciones tempranas que se dan entre las víctimas y los respondedores, tanto formales (el respondedor es una institución o autoridad) como informales (amigos, familiares, etc.) podría ayudar a mejorar las respuestas ante agresiones de carácter sexual y promover el bienestar del superviviente. Se expone la existencia de un mecanismo causal entre respuestas posteriores a la agresión y la aparición de TEPT, por lo tanto, se plantea que si la víctima tiene experiencias negativas inmediatamente después de la agresión sexual, esto podría prolongar la experiencia del trauma.

Se deduce en este estudio que el primer mes después del ataque es el momento de mayor oportunidad para una intervención, debido a la mayor disponibilidad de los respondedores formales durante este tiempo, creándose múltiples oportunidades para minimizar la probabilidad de que los pacientes desarrollen TEPT. Los resultados obtenidos en este trabajo son mixtos, evidenciando que algunas intervenciones son más significativas que otras y siendo imposible comprender el papel de esas interacciones en la recuperación de las víctimas, siendo necesario conocer la efectividad y bajo qué condiciones se puede

mitigar la aparición o los síntomas de TEPT.

Escritura expresiva para mejorar la resistencia al trauma: un ensayo de viabilidad clínica

En este estudio, llevado a cabo por (Glass et al., 2019), se evaluó la viabilidad y efectividad potenciales de la aplicación de un programa de escritura expresiva para generar resiliencia en una muestra compuesta por personas con experiencias traumáticas. El marco más estudiado de la escritura expresiva es el paradigma de Pennebaker, ya que se ha comprobado en otros estudios como una terapia efectiva para mejorar la salud emocional, psicológica y fisiológica en aquellos individuos con traumas ocasionados por un evento detonante.

En este trabajo se señaló la importancia de modificar el protocolo tradicional de escritura expresiva para mejorar la eficacia del tratamiento. El programa propuesto en este estudio expande el paradigma de Pennebaker en un marco terapéutico que hace uso de seis escritos diferentes. Se destaca que la secuencia de la escritura expresiva puede influir en la efectividad, debido a que los pacientes se pueden beneficiar de un espectro más amplio de estilos de escritura. Cada estilo ayuda al participante a encontrar nuevos puntos de vista desde los cuales explorar sus traumas y cultivar la curación.

Los resultados obtenidos en este trabajo mostraron que la implementación de este programa fue altamente aceptada por los participantes. Las variables psicológicas medidas mostraron una mayor capacidad de recuperación y adaptación a experiencias adversas. De acuerdo a los autores, la expansión del paradigma de Pennebaker puede proporcionar un tratamiento no farmacológico a corto plazo, sirviendo como estrategia para aumentar la resiliencia de los pacientes tratados.

Personalidad y uso del lenguaje en las auto narrativas

El objetivo de este trabajo de investigación se basó en examinar la relación existente entre los patrones de uso de lenguaje natural y los fenómenos psicológicos, buscando la asociación entre los rasgos de personalidad y el uso del lenguaje producido por medio de escrituras auto-narrativas. Para los autores (B. Hirsh & B. Peterson, 2009), la frecuencia individual de aparición de cada palabra se correlaciona con las medidas de los cinco grandes rasgos de personalidad.

Las narraciones sobre la trayectoria de vida de un individuo, incluyendo descripciones subjetivas del pasado y futuro, parecen ser fundamentales para definir la identidad. Es razonable suponer, de acuerdo a los autores, que los patrones de uso de las palabras dentro de los textos auto-narrativos podrían tener relaciones más fuertes con rasgos

de personalidad que los encontrados utilizando otros medios o herramientas.

Este estudio busco combinar el uso de palabras y enfoques narrativos, con el análisis de los rasgos de personalidad y su relación con el uso de palabras durante la redacción de textos auto-narrativos. Como resultado de este trabajo se presenta la correlación entre frases y rasgos de personalidad, concluyendo que el conteo de palabras y el género del redactor no fueron variables significativas para la distinción de rasgos. Sin embargo, la elección de palabras en la tarea de escritura tuvo una significativa asociación con la personalidad.

Escribir sobre discapacidad y enfermedad, y el caso de la narración autorreflexiva

En este artículo (Sutherland, 2016) se exploran las ventajas y desventajas de la escritura auto reflexiva. La narración personal permite a los lectores identificarse y conectarse con el escritor, sin embargo, es arriesgado debido a que el autor puede temer a la crítica, sintiéndose vulnerable, aunque desde el punto de vista de la etnografía las preocupaciones se consideran como una dimensión crucial de la investigación, ya que proporciona información sobre la mente del autor durante el proceso de redacción.

La auto reflexividad y la vulnerabilidad en la escritura pueden abarcar una perspectiva de salud y enfermedad, planteando preguntas ontológicas y preocupaciones fenomenológicas que los datos médicos tradicionales pueden pasar por alto. La enfermedad y discapacidad van más allá de la fisiología de una persona, y abarcan sus relaciones sociales, laborales y familiares, por medio de una amplia gama de emociones y sentimientos. Un individuo es el mejor representante de sus propios sentimientos y experiencias.

La autora llega a la conclusión de que la escritura no es solo de captar particularidades de una experiencia personal, también permite comunicarse con médicos, investigadores, víctimas y con cualquier persona en general. La narrativa puede ofrecer una exploración terapéutica de cuestiones cruciales tanto para lectores, como para escritores.

Uso del análisis de cuantificación de recurrencia para examinar las asociaciones entre los cambios en la estructura del texto en una intervención de escritura expresiva y la reducción de los síntomas de angustia en mujeres con cáncer de mama

En este trabajo (Skovgaard et al., 2019) presentó un análisis exploratorio del uso del Análisis de Cuantificación de Recurrencia (RQA por sus siglas en ingles), para el análisis de textos obtenidos a partir de Intervención de Escritura Expresiva (EWI), a partir de la

selección de mujeres tratadas por cáncer de mama. Según los autores, el EWI se ha utilizado como un medio potencialmente efectivo para reducir las consecuencias psicológicas y físicas negativas ocasionadas por experiencias estresantes o traumáticas, con un costo relativamente bajo dada la poca participación de los especialistas en la aplicación del procedimiento.

En el trabajo se sugiere que una intervención efectiva consiste en la creación de una narrativa coherente ya que. De acuerdo a investigaciones previas, la estructura narrativa del texto es un indicador de un proceso de reestructuración cognitivo, que se puede estimular por medio de EWI, pudiendo conducir a efectos benéficos. Los resultados obtenidos en este trabajo muestran una diferencia entre los participantes con mayor o menor mejoría. Para investigar si la estructura del texto estaba asociada con el cambio de síntomas a los 9 meses de seguimiento, se utilizaron correlaciones de Pearson.

El estudio mostró que el cambio en el porcentaje de determinismo se relacionó con la mejoría de la angustia relacionada al cáncer, y experimentada por parte de los pacientes, a los 3 meses posteriores de la intervención, así como una correlación significativa a los 9 meses.

Transformando transcripciones en historias: Un método múltiple de aproximación al análisis narrativo

En este trabajo de investigación (Nasheeda et al., 2019) se describe una metodología para la realización de un proceso de investigación narrativa por medio de la transcripción de entrevistas realizadas a una población de participantes. Se propone que las historias son representaciones esenciales del pasado y presente de un ser humano y se pueden transcribir a narrativas por medio de la aplicación de procesos analíticos. En este artículo se sugiere la implementación de un método para el análisis narrativo, iniciando con la selección de los participantes del estudio la transcripción de las entrevistas grabadas en audio, el análisis y familiarización del contenido de la lectura, el trazado cronológico de los elementos de la historia, el uso de entrevistas de seguimiento y, finalmente, se concluye con el desarrollo de una historia haciendo uso del análisis estructural.

Como resultado de este estudio se concluye que es muy importante la extracción de la esencia de las experiencias vividas por el participante. Mantener el sentido del texto es de vital importancia en el análisis narrativo, teniendo como resultado un proceso basado en el análisis estructural del texto. El uso de eventos para contar historias se considera un buen indicador de eventos narrativos importantes, concluyendo también que los picos narrativos evocan interés en la historia. Finalmente se destaca el uso de una secuencia lógica para el análisis de una narrativa, re-organizando el texto en una secuencia cronológica para la identificación del tiempo y lugar de los eventos descritos en la historia.

3.3 Tabla comparativa de los trabajos relacionados

A continuación, se muestran las tablas comparativas correspondientes a los trabajos relacionados con el Procesamiento de Lenguaje Natural (Tabla 2 y 3) y los trabajos relacionados con violencia y trastornos psiquiátricos relacionados a la experimentación de estos eventos (Tabla 4 y 5).

Tabla comparativa de trabajos relacionados con Procesamiento de Lenguaje Natural

Se realizó una comparativa de las herramientas y resultados obtenidos en la literatura revisada para el desarrollo de este trabajo de investigación. En la columna “Resultado” se refleja la apreciación por parte de cada autor.

Tabla 2. Comparativa de trabajos relacionados al Procesamiento de Lenguaje Natural (PLN), parte 1.

Trabajo	Objetivo	Herramientas	Resultado	Año
Extracción de texto a emoción utilizando técnicas supervisadas de aprendizaje automático (Azim & Bhuiyan, 2018)	Desarrollo de un sistema para eliminar la necesidad de realizar manualmente la creación de un léxico de intensidad emocional	Clasificadores: Naive Bayes, Máxima Entropía y SVM	Aproximación basada en el aprendizaje automático para extraer las categorías básicas de las emociones a partir de textos	2018
Marcadores lingüísticos del lenguaje natural moderado y absoluto (Al-Mosaiwi & Johnstone, 2018)	Identificación de marcadores lingüísticos y alternativa a la escala de Likert	Clasificadores	Conjunto de 31 características lingüísticas detectadas para la correcta clasificación	2018
Reconociendo las emociones en el texto utilizando redes neuronales y lógica difusa (Kanger & Bathla, 2017)	Análisis de sentimiento utilizando lógica difusa y redes neuronales, para la extracción de emociones	Software: MATLAB	Metodología para realizar el análisis de sentimientos combinando lógica difusa y redes neuronales	2017
Una breve revisión de los desafíos éticos en el Procesamiento de Lenguaje Natural clínico (Suster et al., 2017)	Análisis sobre las consideraciones éticas, morales y legales de implementar Procesamiento de Lenguaje Natural con datos clínicos.	No especificado	Recopilación de los desafíos fundamentales a considerar en la investigación clínica y médica, haciendo uso de PLN	2017
Reconociendo emociones en texto usando un conjunto de clasificadores (Isidoros & Ioannis, 2016)	Desarrollo de un sistema de análisis de sentimientos para el reconocimiento automático de emociones	Software: WordNet Affect, SentiWordNet 3.0, Stanford Parser	Reconocimiento de la polaridad emocional textual con base en el uso de un conjunto de clasificadores	2016

Tabla 3. Comparativa de trabajos relacionados al Procesamiento de Lenguaje Natural (PLN), parte 2.

Trabajo	Objetivo	Herramientas	Resultado	Año
Análisis de sentimientos a nivel de emoción: la evaluación de la opinión efectiva (Almashraee, Monett & Paschke, 2016)	Detección de opiniones, por medio de la extracción y procesamiento de palabras emocionales	No especificado	Se determina el proceso de asignación emocional de opiniones	2016
Análisis de sentimientos multilingüe: estado del arte y comparación independiente de técnicas (Dashtipour et al., 2016)	Comparación de las investigaciones existentes sobre el análisis de sentimiento, tratando de validar la precisión	Software: WEKA y LIBSVM	Método para el procesamiento de textos y la aplicación de técnicas de aprendizaje automático	2016
Construcción de un corpus marcado con emociones para el análisis de sentimientos en twitter en español (Sidorov, Galicia & Camacho, 2019)	Recopilación y evaluación de un corpus lingüístico conformado por tweets y con un significado emocional	Software: SentiStrength, LIWC, WEKA y Freeling	Se genera un corpus lingüístico emocional de tweets	2016
Evaluación automatizada de las auto-narrativas de los pacientes con trastorno de estrés posttraumático mediante procesamiento del lenguaje natural y minería de textos (He, Veldkamp, Glas & de Vries, 2016)	Desarrollo de un sistema de para la detección de pacientes con TEPT, haciendo uso de técnicas de PLN	Software: Python Natural Language Toolkit (NLTK)	Sistema automático para el diagnóstico de estrés posttraumático, basado en escritos auto-narrativos	2015
Clasificación de roles semánticos usando características sintácticas, semánticas y contextuales (Reyes et al., 2013)	Identificar clasificación, el tipo de rol semántico existente entre un evento y sus actantes	Software: Freeling y clasificadores	Clasificación de las categorías de roles semánticos, y detección del mejor clasificador	2013

Tabla comparativa de trabajos relacionados con violencia y padecimientos psicológicos

Se realizó una comparativa de las terapias, objetivos, y resultados obtenidos en la literatura revisada para el desarrollo de este trabajo de investigación.

Tabla 4. Comparativa de trabajos relacionados con la violencia y el tratamiento de trastornos, parte 1.

Trabajo	Objetivo	Herramienta	Terapia	Resultado	Año
Escritura expresiva para mejorar la resistencia al trauma: un ensayo de viabilidad clínica (Glass et al., 2019)	Modificar el paradigma de Pennebaker para terapia de EEA	No especificado	Escritura Emocional Autorreflexiva	Terapia basada en la aplicación de 6 tipos de escritos narrativos	2019
Uso del análisis de cuantificación de recurrencia para examinar las asociaciones entre los cambios en la estructura del texto en una intervención de escritura expresiva y la reducción de los síntomas de angustia en mujeres con cáncer de mama (Skovgaard et al., 2019)	Análisis exploratorio del uso del Análisis de Cuantificación de Recurrencia (RQA) para el análisis de textos obtenidos de intervenciones basadas en EEA	Entrevistas	Escritura Emocional Autorreflexiva	Detección de que el cambio en la estructura del texto se asocia con cambios en la sintomatología presentada por el paciente. Uso de RQA puede ser una herramienta para el monitoreo del bienestar psicológico	2019
Transformando transcripciones en historias: Un método múltiple de aproximación al análisis narrativo (Nasheeda et al., 2019)	Propuesta de un método para el análisis de la estructura narrativa de un texto	No especificado	No especificado	Propuesta de un método para el análisis del texto, a través del estudio de su estructura	2019
Sintomatología del trastorno de estrés postraumático en una muestra de mujeres víctimas de violencia sexual (Lozano-Oyola et al., 2017)	Análisis de la sintomatología TEPT en mujeres víctimas de violencia sexual	Software: SPSS	No especificado	Establecimiento de los criterios de evaluación para determinar la propensión a desarrollar TEPT	2018
Rumiación, crecimiento y sintomatología postraumática en personas que han vivido experiencias altamente estresantes (García et al., 2018)	Analizar la influencia de la rumiación y la severidad subjetiva del evento sobre TEPT	Software: SPSS y Amos V18	No especificado	Análisis del efecto que producen los distintos tipos de rumiación en el desarrollo de un trastorno estresante	2018

Tabla 5. Comparativa de trabajos relacionados con la violencia y el tratamiento de trastornos, parte 2.

Trabajo	Objetivo	Herramienta	Terapia	Resultado	Año
El sufrimiento psicológico en la vida cotidiana de las mujeres que han experimentado violencia sexual: un estudio fenomenológico (Trigueiro et al., 2017)	Detectar las acciones cotidianas de las mujeres que han sufrido violencia sexual	Entrevistas	No especificado	Método para la realización de un estudio cuantitativo, enfocado en víctimas de violencia sexual	2017
Perfiles de estrés postraumático en adolescentes abusados sexualmente (Guerra, Plaza & Farkas, 2017)	Identificar perfiles de víctimas adolescentes de acuerdo a su sintomatología de estrés postraumático	Software: SPSS	No especificado	Perfiles de adolescentes que han sufrido abusos sexuales	2017
Prevención del estrés postraumático relacionado con la agresión sexual a través de la intervención temprana: una revisión sistemática (Dworkin & Schumacher, 2016)	Revisión sistemática para comprender los elementos particularmente importantes en la prevención de TEPT	No especificado	No especificado	Mecanismos de respuesta para la atención de síntomas asociados al TEPT	2016
Escribir sobre discapacidad y enfermedad, y el caso de la narración autorreflexiva (Sutherland, 2016)	Exploración de las ventajas y desventajas de la terapia de escritura emocional	No especificado	Escritura Emocional Autorreflexiva	Análisis de la técnica de Escritura Emocional Autorreflexiva	2016
Personalidad y uso del lenguaje en las auto narrativas (B. Hirsh & B. Peterson, 2009)	Estudio de los patrones de uso del lenguaje y las características psicológicas de un individuo	No especificado	No especificado	Relación entre los patrones de uso de lenguaje natural y los fenómenos psicológicos	2009

Capítulo IV

Metodología de solución

En esta sección se explica y desarrolla cada uno de los componentes de la metodología de solución propuesta para dar cumplimiento a los objetivos planteados en este proyecto de investigación.

Ya que el desarrollo de este proyecto exploró una aplicación del Procesamiento de Lenguaje Natural (PLN) poco explorada en la literatura, fue necesario construir la metodología a partir de los componentes esenciales del PLN, adaptándolos para construir una herramienta capaz de satisfacer los requerimientos propios de este trabajo de investigación. En la Figura 1 se presentan las actividades contempladas para el desarrollo de la herramienta:

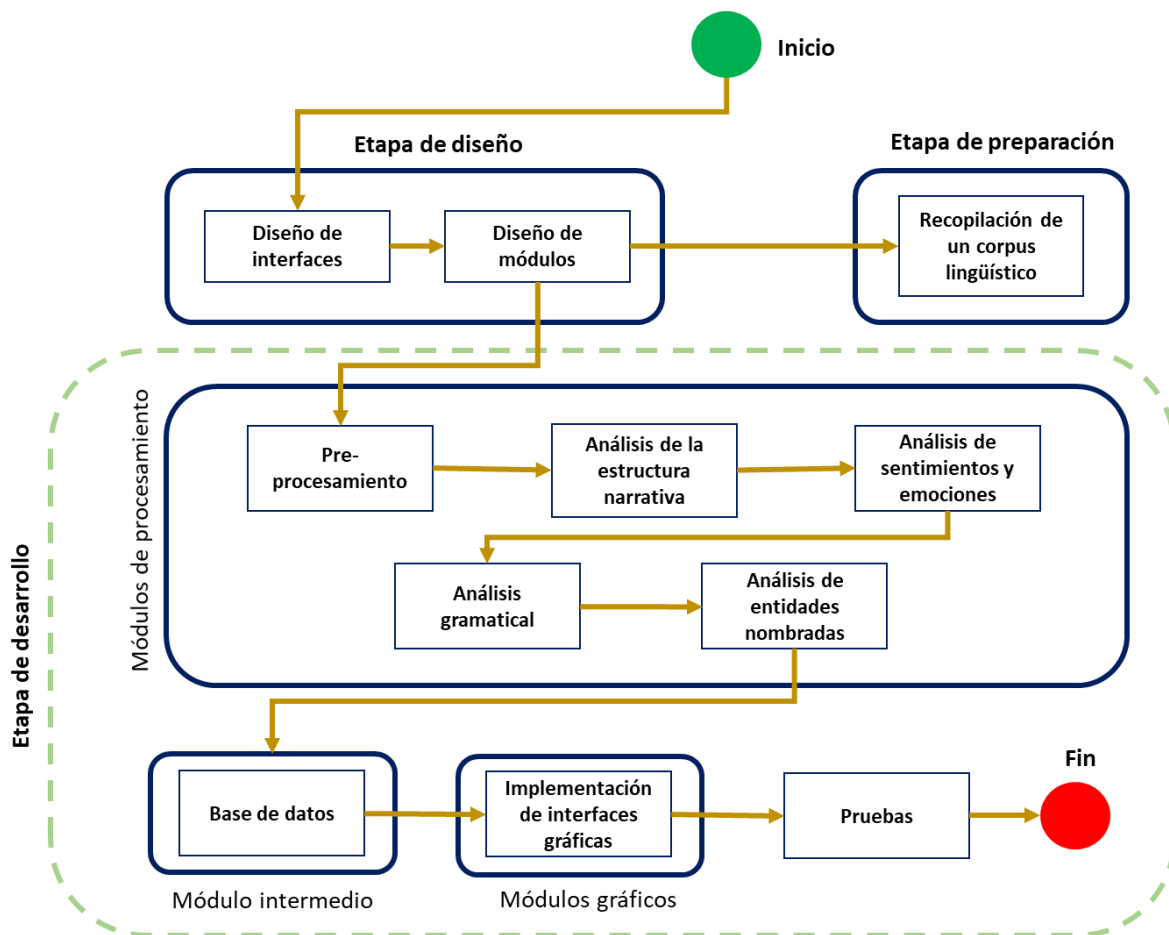


Figura 1. Actividades por realizar para el desarrollo de la herramienta de análisis de textos.

A continuación, se detallan las actividades que se muestran en la Figura anterior, y que corresponden a las actividades contempladas para el óptimo funcionamiento de la herramienta de software desarrollada:

Etapa de diseño: La etapa de diseño contempla el estudio del problema y la

formalización de este, de tal forma que se pueda generar una estructura conceptual a partir de la cuál se pueda realizar el desarrollo de la herramienta de software. Las tareas que se contemplan en esta etapa se dividen en dos categorías, separando el diseño de la parte gráfica (interfaces), de la parte lógica (programación), siendo estas explicadas a continuación:

- a) Diseño de interfaces: El objetivo de esta tarea es el diseño de la distribución de los elementos que serán parte de las interfaces gráficas de la herramienta, así como determinar la distribución más adecuada de los componentes y, en conjunto, generar un concepto gráfico general de las vistas que se incorporarán más adelante a la herramienta.
- b) Diseño de módulos: Esta actividad se enfoca en el diseño de las funcionalidades con las que contará la herramienta de análisis de textos, representando a manera de módulos independientes cada una de las técnicas de Procesamiento de Lenguaje Natural requeridas para el proceso de análisis de textos y para la obtención de la salida esperada, y que se describen en el apartado “Módulos de procesamiento”.

Etapa de preparación: Esta etapa contempla las tareas no relacionadas directamente con el desarrollo de la herramienta, pero si necesarias para complementar el desarrollo general de este trabajo de investigación. Esta etapa de preparación incluye la creación de un corpus lingüístico recopilatorio de textos escritos por mujeres víctimas de violencia, y difundidos a través de redes sociales, así como la definición de herramientas de diagnóstico psicológico utilizadas en el tratamiento convencional de las mujeres pertenecientes a este segmento poblacional y que, por lo tanto, son la población de estudio de esta tesis de investigación.

Etapa de desarrollo: Es la etapa clave en el desarrollo de la solución. Esta afirmación se justifica debido a que el objetivo general de este trabajo de investigación consiste en obtener como producto final una herramienta de software. Debido a lo anterior esta etapa a su vez se divide en dos: por un lado, el desarrollo de los módulos de procesamiento y, en segundo lugar, la generación de los módulos gráficos.

Módulos de procesamiento: En esta categoría se incluyen los módulos que se harán cargo de la recepción de los textos, su manipulación, tratamiento y análisis.

A continuación, se detalla cada uno de los módulos definidos para realizar las tareas de procesamiento:

- a) Pre-procesamiento: Inicialmente la herramienta debe realizar la normalización de los textos que se ingresen por parte del usuario, para proceder a su posterior manipulación. Es un proceso realizado en tres pasos secuenciales:
 - Normalización: Eliminación de elementos atípicos o lingüísticamente no relevantes, siendo signos de puntuación, palabras vacías (*stopwords*), etiquetas, íconos o símbolos.
 - Tokenización: División de las cadenas textuales en piezas o unidades más pequeñas, con la

finalidad de segmentar el texto en palabras que puedan analizarse de forma individual, o a nivel de oración.

- Lematización: Colocación de los elementos del texto en igualdad de condiciones semánticas, por medio de la conversión de cada palabra a su forma base o raíz semántico - morfológica.
- b) Análisis de la estructura narrativa: Para realizar el análisis de la estructura narrativa se requiere la ejecución de las siguientes tareas: Conteo de elementos, tokens y evaluación de la complejidad de las oraciones escritas.
- c) Análisis de sentimientos y emociones: La finalidad de esta tarea es identificar la polaridad emocional a nivel de documento, párrafo u oración. Así cómo la detección y etiquetado de las palabras en alguna categoría emocional.
- d) Análisis gramatical: En esta tarea se incluye el etiquetado de la categoría gramatical de los tokens (*POS tagger*, por sus siglas en inglés).
- e) Análisis de entidades nombradas: Extracción y conteo de frecuencia de aparición de las palabras pertenecientes a alguna categoría de entidades lingüísticas, como nombres propios, organizaciones, cantidades o lugares.

Módulo intermedio: Adicionalmente se complementan los módulos gráficos y de análisis, con un módulo de almacenamiento de datos, físicamente a ser implementado como una base de datos que sirva como un elemento intermedio, o un “puente”, entre los elementos gráficos y los elementos de procesamiento. Este paso es necesario para que los elementos visuales sean independientes de los elementos lógicos, y con ello se puedan implementar cambios o mejoras futuras de manera más eficiente, al modularizar los elementos de la herramienta.

Módulos gráficos: Desarrollo de interfaces: En esta tarea se contempla el desarrollo de las interfaces gráficas que permitirán la interacción del usuario con la herramienta.

Pruebas: Periodo de pruebas contemplado para la comprobación de las funcionalidades de la herramienta, así como su integración en un equipo de cómputo que sirva como servidor de aplicación, ejecutando permanentemente el software y procesando las solicitudes que se reciben, a través de la base de datos que funge como elemento intermediario entre la interfaz y las funciones de análisis de textos.

El resultado esperado al finalizar las etapas de diseño y desarrollo es la obtención de una herramienta de software capaz de realizar el análisis automático de textos, utilizando técnicas de procesamiento de lenguaje natural para la extracción de características lingüísticas

relevantes, siendo una herramienta probada y aplicada al estudio preliminar de una aplicación de la terapia de Escritura Emocional Autorreflexiva (EEA) en una población de mujeres víctimas de violencia.

Justamente es debido a la necesidad de probar la herramienta a desarrollar en una muestra compuesta por miembros de la población objetivo que se propone, en la Figura 2, el procedimiento para la prueba de la herramienta en el análisis automático de textos obtenidos al aplicar la terapia de EEA en una muestra compuesta por participantes voluntarias.

Ya que se trata de un trabajo en el que se involucran participantes de un estudio psicológico, y por lo tanto médico, es necesario contar con la adecuada asesoría y supervisión para cumplir con lo concerniente a estándares, reglamentos y procesos de realización propios de este tipo de intervenciones terapéuticas. Para el caso de este proyecto de titulación se contó con el apoyo del Instituto Nacional de Salud Pública, a través de la Dra, Leonor Rivera y de la Mtra. Marina Seris, ambas investigadoras de esa institución.

En este trabajo se propone dividir el software a desarrollar en este trabajo de tesis, implementando por un lado la herramienta de análisis de textos, enfocada a la extracción de características lingüísticas a partir de textos, y por el otro la herramienta para la aplicación de la terapia. Por lo tanto, se deben desarrollar medios para aplicar la terapia de EEA que sean independientes de la terapia de análisis, siendo herramientas de software independientes pero complementarias entre sí para el caso específico de este trabajo de investigación.

A continuación, se muestran y describen las actividades necesarias para poder aplicar la terapia de Escritura Emocional Autorreflexiva en una muestra de mujeres víctimas de violencia, para el posterior análisis de los datos recopilados haciendo uso de la herramienta de análisis previamente descrita:

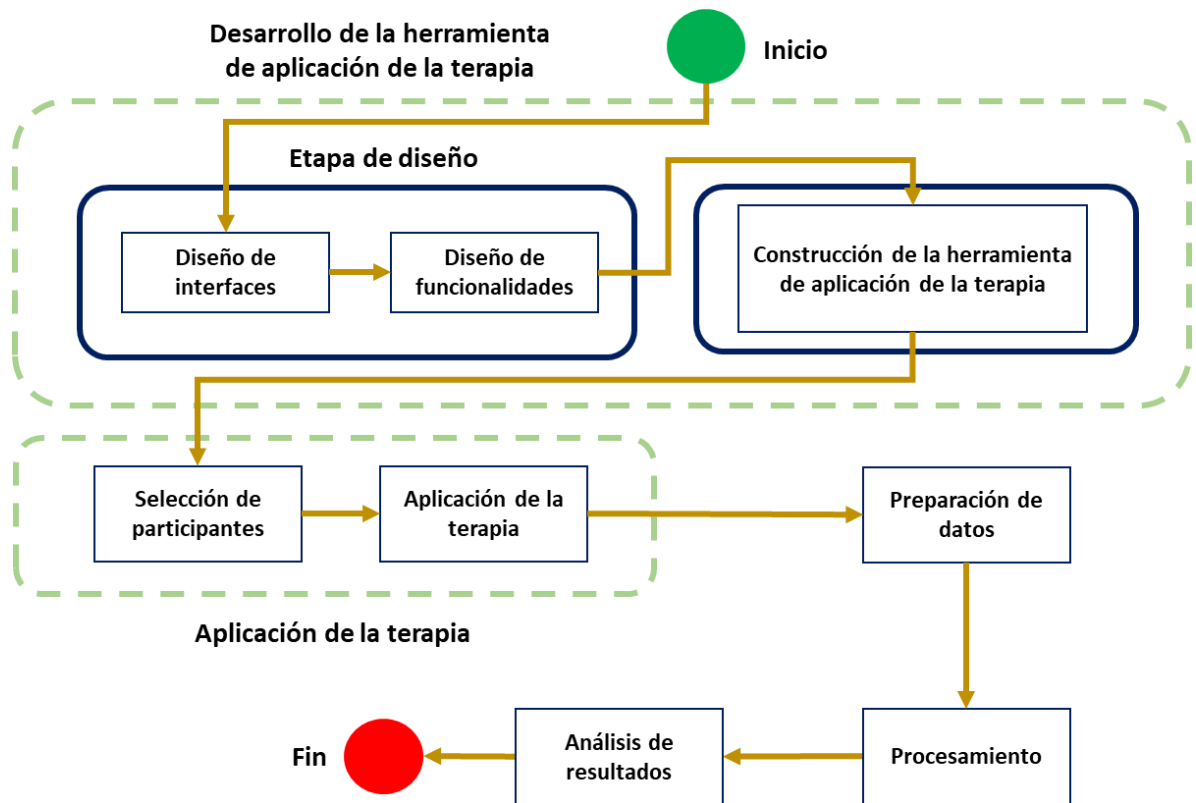


Figura 2. Procedimiento para la recopilación y análisis de los textos procedentes de la terapia de EEA.

A continuación se detallan las actividades que se muestran en la Figura 2:

Etapa de diseño: Se parte de una etapa de diseño para la especificación y formalización del mecanismo, y medios, necesarios para la implementación de la terapia de Escritura Emocional Autorreflexiva (EEA), teniendo dos actividades principales:

- a) **Diseño de interfaces:** Diseño de los elementos y vistas a mostrar a las usuarias participantes en la terapia de EEA, y a través de las cuales deben desarrollar las actividades planificadas para el cumplimiento de esta terapia.
- b) **Diseño de funcionalidades:** Definición de las funcionalidades de las interfaces de aplicación de la terapia, así como de la conectividad necesaria entre los diferentes elementos lógicos.

Etapa de desarrollo de la herramienta de aplicación de la terapia: Programación y estructuración de las interfaces y funcionalidades lógicas, para obtener una herramienta de software capaz de permitir la aplicación de la terapia de EEA. Esta etapa debe incluir la implementación de una base de datos para el almacenamiento de los resultados obtenidos por

cado una de las participantes, y en un formato que pueda ser interpretado y leído por la herramienta de análisis de textos, recordando que está última es una segunda herramienta, independiente de los medios de aplicación de la terapia.

Etapa de aplicación de la terapia EEA: Esta etapa corresponde a la aplicación de la terapia de Escritura Emocional Autorreflexiva (EEA), utilizando la herramienta de aplicación a desarrollarse en la etapa anterior, y considerando las siguientes actividades:

- a) Selección de participantes: Con la finalidad de seleccionar a las mujeres participantes se debe prescindir de aquellas mujeres que hayan sido participes en algún otro tratamiento psicológico con anterioridad, y solo se deben seleccionar aquellas participantes voluntarias.
- b) Aplicación de la terapia: Es necesario que sean mujeres que hayan experimentado algún evento de violencia, realizando las actividades correspondientes a la terapia a través de la herramienta de software desarrollada para tal fin.

Esta intervención de EEA se debe basar en el paradigma de Pennebaker, considerando que para este trabajo de investigación se ha de generar una adaptación de dicho método de aplicación de la terapia, contemplando la aplicación de las sesiones de escritura terapéutica a través de 4 escritos narrativos individuales aplicados en intervalos de un escrito por semana, o día, con una duración por sesión no menor a 15 minutos y no mayor a 45 minutos, escritos a computadora a través de los medios puestos a disposición de las participantes para tal fin, durante 4 - 6 semanas hasta completar el tratamiento.

Etapa de preparación de datos: Como paso previo para el procesamiento de los textos en la herramienta de análisis, se realizará la descarga, revisión y estandarización de formato de los datos obtenidos.

Etapa de procesamiento: Etapa en la que se hace uso de la herramienta de análisis de textos desarrollada con anterioridad para el análisis y extracción de las características lingüísticas obtenidas. La herramienta de análisis tendrá que recibir como entrada el conjunto de textos obtenidos durante la aplicación de la terapia y generar un conjunto de datos con las características lingüísticas extraídas.

Etapa de análisis de resultados: Como paso final se ha de realizar el análisis de los resultados obtenidos, para determinar las características de la muestra de datos obtenida y el desempeño de la herramienta.

Se debe aclarar que, tanto la herramienta de análisis de textos, como la herramienta de aplicación de la terapia son dos soluciones totalmente individuales, pero complementadas en el caso de este trabajo para dar solución al objetivo general propuesto en esta tesis.

Sin embargo, se ha pensado como una solución adaptable y aplicable a otros problemas que requieran el análisis y extracción de características lingüísticas a partir de textos, de ahí que la metodología de solución esté basada en módulos, los cuales se pueden modificar, eliminar, o se pueden agregar módulos con nuevas funcionalidades.

En los siguientes capítulos se desarrolla cada una de las fases mostradas, así como la forma utilizada para materializar las tareas definidas en una solución tangible al problema planteado para la resolución en este trabajo de investigación.

Capítulo V

Desarrollo de la solución

5.1 Generación de un corpus lingüístico

En el presente capítulo se describe el procedimiento llevado a cabo para la búsqueda y recopilación de un corpus lingüístico compuesto por textos en español, escrito por mujeres víctimas de violencia y publicados a través de redes sociales. Esta actividad se realizó como paso previo al inicio del desarrollo de la herramienta, debido principalmente a dos motivos: El primero de ellos se debe a que no se encontró en la literatura un conjunto de textos construido con relatos o narraciones de mujeres víctimas de violencia sexual, mientras que el segundo motivo fue la necesidad de contar con un conjunto de datos para utilizarlos como datos de entrenamiento al momento de realizar algunas tareas de clasificación que se proponen como parte de la solución a la problemática abordada en este trabajo de tesis.

Por los anteriores motivos fue necesario crear un corpus lingüístico compuesto por narrativas compartidas en forma de comentarios a través de la red social Facebook, los cuales fueron extraídos manualmente para ser recopilados y posteriormente procesados, sirviendo en primera instancia como un contacto inicial con textos escritos por mujeres con el perfil objetivo de esta investigación, así como para proporcionar datos de prueba que permitan realizar pruebas durante el desarrollo de la herramienta, con datos aproximados a los que se esperan obtener durante las futuras tareas de desarrollo de la tesis.

Para realizar la generación del corpus se realizaron una serie de actividades, las cuales se describen en a continuación:

5.1.1 Criterios a considerar para la generación de un corpus

De acuerdo con (Parodi, 2008), al menos se pueden señalar doce características necesarias para la construcción de los corpus lingüísticos. Esas características se pueden resumir en 5, las cuales se mencionan a continuación:

1. Recolección de textos en entornos naturales: Es preferible que los datos sean recolectados en un entorno en donde se encuentre presente el fenómeno que se desea estudiar, con la finalidad de tratar de capturar e incorporar la mayor cantidad posible de rasgos distintivos del lenguaje utilizado en dicho entorno.
2. Explicitud de los rasgos definitorios: Los textos recopilados en el corpus han de ser muestras reales que plasmen el uso de la lengua en la población de estudio, debe existir una alta probabilidad de que los textos sean una muestra representativa fiel al objeto de estudio del lenguaje.

3. Formato final de tipo digital: El formato final debe ser en formato electrónico, preferentemente como texto plano, para su manipulación o visualización a través de técnicas y procedimientos computacionales.
4. Tamaño: Los corpus abarcan un tamaño finito, aunque subordinado al objetivo de la investigación, a los recursos disponibles para su recopilación y al tamaño de la muestra objetivo. Por lo general se considera que mientras un corpus sea más grande proporcionara un conjunto representativo mayor, por lo que es un objetivo prioritario alcanzar el mayor tamaño posible.
5. Sustento o procedencia inicial especificada: En este punto se pueden incluir varios conceptos, iniciando con el sustento ético y legal, tanto de la procedencia de los datos como del uso y tratamiento que se les dé. Así mismo, también se puede considerar que se debe especificar el tipo de texto, género o temáticas que incluirá, siendo necesario que los textos sean seleccionados de acuerdo con una serie de criterios lingüísticos especificados previo a su construcción.

5.1.2 Método para la generación del corpus

Se empleó un método compuesto por 5 etapas para realizar la construcción del corpus, el cual se muestra a continuación, y cuyo uso se enfoca en la generación de un corpus textual:

- a) Etapa de preparación: Se definen los criterios que deberán tener los textos para ser incluidos en el corpus, se define la población de estudio y objetivo de este, alcances y limitaciones, duración, idioma, zona geográfica, herramientas y, en general, cualquier aspecto que requiera de una previa planificación.
- b) Búsqueda de los textos: Consiste en localizar la fuente de información más adecuada para realizar la recolección de datos, teniendo un enfoque principal en la búsqueda de textos acordes al objeto de estudio, y obteniéndose como resultado de esta etapa la fuente de información seleccionada, la cual deberá contar con una considerable cantidad de textos útiles que permitan la composición de una muestra de lenguaje.
- c) Filtrado de los textos: Al localizar la plataforma e interfaz ideal para la extracción de los textos, se deben seleccionar los textos que cuenten con las características definidas por el estudio, además de descartar la inclusión de aquellos datos que no se consideren un aporte para el corpus o que se encuentren en un formato distinto al texto (Imágenes, audios o videos). Esta etapa permite comprender con un alto grado de detalle los componentes del corpus, los textos de entrada y salida, así como las características de la fuente de información seleccionada.

- d) Extracción y recopilación de los textos: Con base en la etapa de filtrado y los criterios de inclusión impuestos durante la etapa de búsqueda, se procede a la recopilación, manual o automática, de los textos seleccionados. Los textos extraídos se deberán almacenar en un formato de texto o en forma de tabla, tratando de incluir la mayor cantidad de datos adicionales que sean considerados útiles, como la procedencia, autor, edad, etc. La salida de esta etapa ya se puede considerar un corpus.
- e) Análisis: El procedimiento de análisis se puede aplicar como parte de la misma construcción del corpus, o de forma separada a él, como parte de un estudio mayor o si se tiene la intención de contar con una mayor cantidad de características representativas. Se pueden considerar diversos tipos de análisis, como los del tipo estadístico, análisis de sentimientos o enfocados en la gramática, solo por poner algunos ejemplos.

5.1.3 Construcción del corpus

Haciendo uso del método propuesto en la sección anterior, se realizó la generación de un corpus construido con base en relatos de mujeres agredidas sexualmente, siendo mostrado el procedimiento a continuación:

Etapa de preparación

El proceso de planificación se enfocó en textos compartidos por mujeres, que hubiesen sido en algún momento víctimas de algún evento de violencia sexual. Como eventos de violencia sexual se consideraron los criterios establecidos por parte de la Organización Mundial de la Salud (WHO, 2020), descritos a continuación:

- Agresión sexual: Actos sexuales forzados, realizados por una persona cercana la víctima, o por un individuo externo. Incluye el forzar a la mujer a mantener relaciones sexuales o realizar cualquier acto sexual en contra de su voluntad, lastimarla durante el acto u obligarla a mantener relaciones sexuales sin protección.
- Violencia de pareja: Maltrato cometido por una persona relacionada afectiva y sentimentalmente con la víctima.

- **Violencia física:** Lesiones o daños corporales ocasionados a la mujer, mediante golpes, contacto físico o haciendo uso de un objeto como arma.
- **Maltrato emocional o psicológico:** Incluye una serie de comportamientos en contra de la mujer, como lo son: Humillaciones, insultos, críticas o amenazas.
- **Comportamientos controladores:** Incluye el aislamiento de la víctima con su grupo social cercano (familia y amigos), vigilancia o incriminaciones injustificadas.

Una vez definidas las clases de textos adecuados, se definió como zona geográfica a Hispanoamérica y como idioma objetivo el español (en sus diferentes dialectos). Como plataforma de extracción se seleccionó la red social Facebook, teniendo en cuenta únicamente el uso de datos difundidos públicamente y debiendo ser un procedimiento de extracción manual, debido a la cantidad de contenido descartable tan alto existente en la plataforma.

Búsqueda de los textos

La búsqueda se realizó utilizando palabras clave, para tratar de localizar paginas o grupos de Facebook en donde se pudiese encontrar la información buscada. Fue entonces cuando se encontró el denominado “proyecto #Cuéntalo”, consistente en una iniciativa surgida a inicios del año 2018 en redes sociales y cuyo propósito se encontraba en la creación de una memoria colectiva compuesta por relatos de mujeres con alguna experiencia de agresión sexual, invitando a la participación voluntaria de mujeres a través de las redes sociales, Facebook y twitter principalmente, y proporcionando un espacio para compartir públicamente las narrativas personales de las víctimas (Proyecto Cuéntalo, 2019). Los textos encontrados en la cuenta de Facebook de esta iniciativa se convirtieron en el objetivo de las siguientes etapas.

Un punto que se tuvo en cuenta para este trabajo de tesis, fue que al momento de realizado este trabajo no se cuenta con acceso al conjunto de datos, por lo tanto, se debió de realizar el proceso manual de recopilación de algunos de los textos del “proyecto #Cuéntalo”.

Filtrado de los textos

La selección de los textos narrativos se realizó con base en los siguientes criterios de inclusión:

- El relato debía de encontrarse en formato textual, excluyendo capturas de pantalla de textos e imágenes de todo tipo.
- El encabezado del texto debía incluir una numeración, siguiendo un orden consecutivo. El contar con una clave numérica permitió identificar los textos y evitar la duplicidad de estos dentro del corpus.
- El relato tuvo que ser escrito en primera persona, o en tercera persona, y siempre haciendo referencia a un evento de violencia sexual. Se descartaron textos periodísticos y comentarios realizados sobre el texto original.

De acuerdo con los tres criterios anteriores, el formato de los textos con potencial de inclusión en el corpus lingüístico se debió de visualizar como se muestra en la Figura 3:

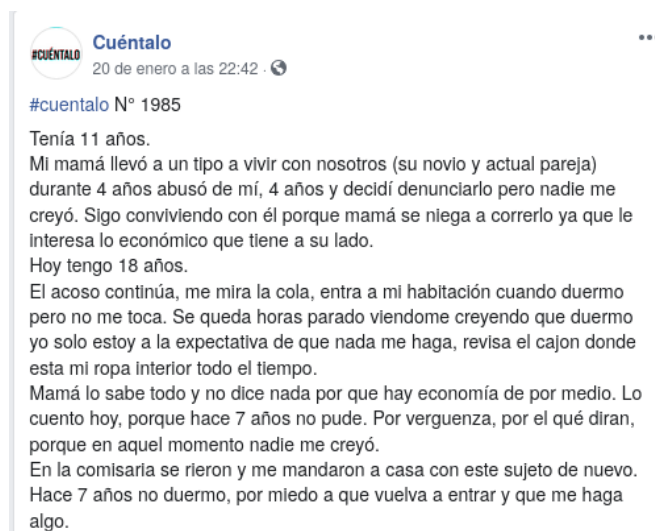


Figura 3. Relato en forma textual obtenido de Facebook, relacionado con la iniciativa Cuéntalo.

Extracción y recopilación de los textos

Partiendo del formato definido como el adecuado para su recopilación, se procedió a realizar una búsqueda manual de los relatos apropiados, recopilando un total de 581 textos con diferentes extensiones, de los cuales 553 se redactaron en primera persona, y los restantes 28 en tercera persona.

Adicionalmente se incluyó el nombre, edad y nacionalidad del autor del texto, del 100% de los textos obtenidos, el porcentaje de aparición de las otras tres características se

muestra en la Tabla 6:

Tabla 6. Porcentaje de aparición y no aparición de los datos dentro del corpus.

Característica	Dato incluido	Dato no incluido
Nombre	12.39% (72 casos)	87.60% (509 casos)
Edad	57.83% (336 casos)	42.16% (245 casos)
Nacionalidad	73.66% (428 casos)	26.33% (153 casos)

En cuanto a las edades de las personas que expresaron su narrativa, el promedio de edad de las participantes es de 22 años, y el rango de edad se encuentra entre 13 y 59 años. El promedio de edad se determinó con base en las 336 narrativas que incluyeron este dato, por lo que el resto se descartó para la determinación del promedio.

Análisis

No se realizaron análisis adicionales al corpus en esta etapa.

5.1.4 Corpus obtenido

Aplicando los pasos explicados anteriormente se obtuvo como resultado el corpus lingüístico, al cual se le denominó “Corpus NVS2019” (Corpus de Narrativas de Violencia Sexual 2019), del cual se muestra un fragmento en la Figura 4:

Relato	Nombre	Edad	Nacionalidad	Punto de vista
<p>Tengo 19 años y he sido víctima varias veces de abuso:</p> <p>1- A mis cinco años fui abusada por el hijo de un vecino, yo tenía cinco y el quince. Nunca lo conté a mis papás por miedo a q no me creyeran.</p> <p>2- A mis 8 años volví a ser abusada por parte de un tío, y tampoco no dije nada.</p> <p>3- A los 13 un hombre se subió al carro que yo me trasportaba para mi colegio y me tocó las piernas, era un vecino, ese día me llené de coraje y lo insulté.</p> <p>4- A los 15, un hombre me molestaba mucho, me decía cosas asquerosas, cuando me veía me llamaba cada rato, (mis papás se daban cuenta y no decían nada solo pensaban que era parte de una broma).</p> <p>5- A los 17, un chico de mi colegio me siguió toda la tarde hasta mi casa, no dije nada porque no tenía como probar que el me seguía.</p>	Anónima	19	Anónimo	Primera persona

Figura 4. Uno de los registros que componen el corpus NVS2019.

En la siguiente tabla se muestran las características principales del corpus generado, con base en la clasificación mostrada en (Llisterri & Torruella, 1999), y con los datos obtenidos durante su construcción:

Tabla 7. Principales características del corpus.

Nombre	Corpus NVS2019
Tipología y clasificación	
Origen	Textual
Codificación y Anotación	Anotado (Estructura textual)
Especificidad de los Textos	Especializado-informativo (Científico)
Autoría	Genérico
Tiempo	Sincrónico (Periódico)
Propósito	Específico (De estudio)
Lenguaje	Monolingüe
Cantidad de texto	Pequeño
Características	
Número de textos	581
Idioma	Español
Datos incluidos	Texto, nombre, edad, nacionalidad y punto de vista

5.2 Desarrollo de herramienta de análisis de textos

En esta sección se detallan las actividades realizadas para construir la Herramienta para el Análisis Automático de Textos que cumpla con el objetivo principal del trabajo de tesis presentado en este documento.

5.2.1 Proceso de análisis de textos

La herramienta de análisis desarrollada se denominó “Herramienta para el Análisis Automático de Textos” o HAAT. Dicho software permite la extracción de características lingüísticas de textos escritos, tanto de textos individuales como de varios textos a la vez (4 textos simultáneamente). La entrada que solicita el sistema es un Título (opcional, aunque recomendable), y el ingreso del texto que se desea analizar (obligatorio).

La herramienta permite la interacción a través de una interfaz web, la cuál puede ser alojada en un servidor web local o remoto. En cuanto al procesamiento de los textos, este se realiza en un programa de análisis alojado en un equipo de cómputo que realiza la función de servidor de análisis. Ambos componentes se comunican entre sí utilizando la base de datos como entidad intermedia para la interacción, tal y como se muestra en la Figura 5. Para el caso de las pruebas mostradas en este capítulo la interfaz se alojó de manera local en un equipo de cómputo tipo laptop, al igual que el servidor de análisis de textos.

En la Figura 5 se pueden apreciar las actividades y mecanismos de control que determinan el flujo de trabajo de la herramienta HAAT. Para ilustrar dicho procedimiento se ha utilizado un diagrama de tipo BPMN (*Business Process Model and Notation*, por su nombre en inglés). La selección de este tipo de diagrama se debe a que se trata de un estándar de graficado de procesos que permite el modelado de procesos de negocio en un formato de flujo de trabajo (Silver, 2012), proporcionando un lenguaje común para el modelado e interpretación de procesos.

Se utilizó un conjunto de corpus lingüísticos para complementar el trabajo de extracción de características que realiza la HAAT, todos ellos aplicados a las tareas de análisis de sentimientos y de emociones. Es importante la distinción entre ambos conceptos, ya que, de acuerdo con (Rodríguez et al., 2011) las emociones hacen referencia a reacciones de tipo fisiológico que ocurren de manera espontánea debido a ciertos factores de estímulo externo al individuo.

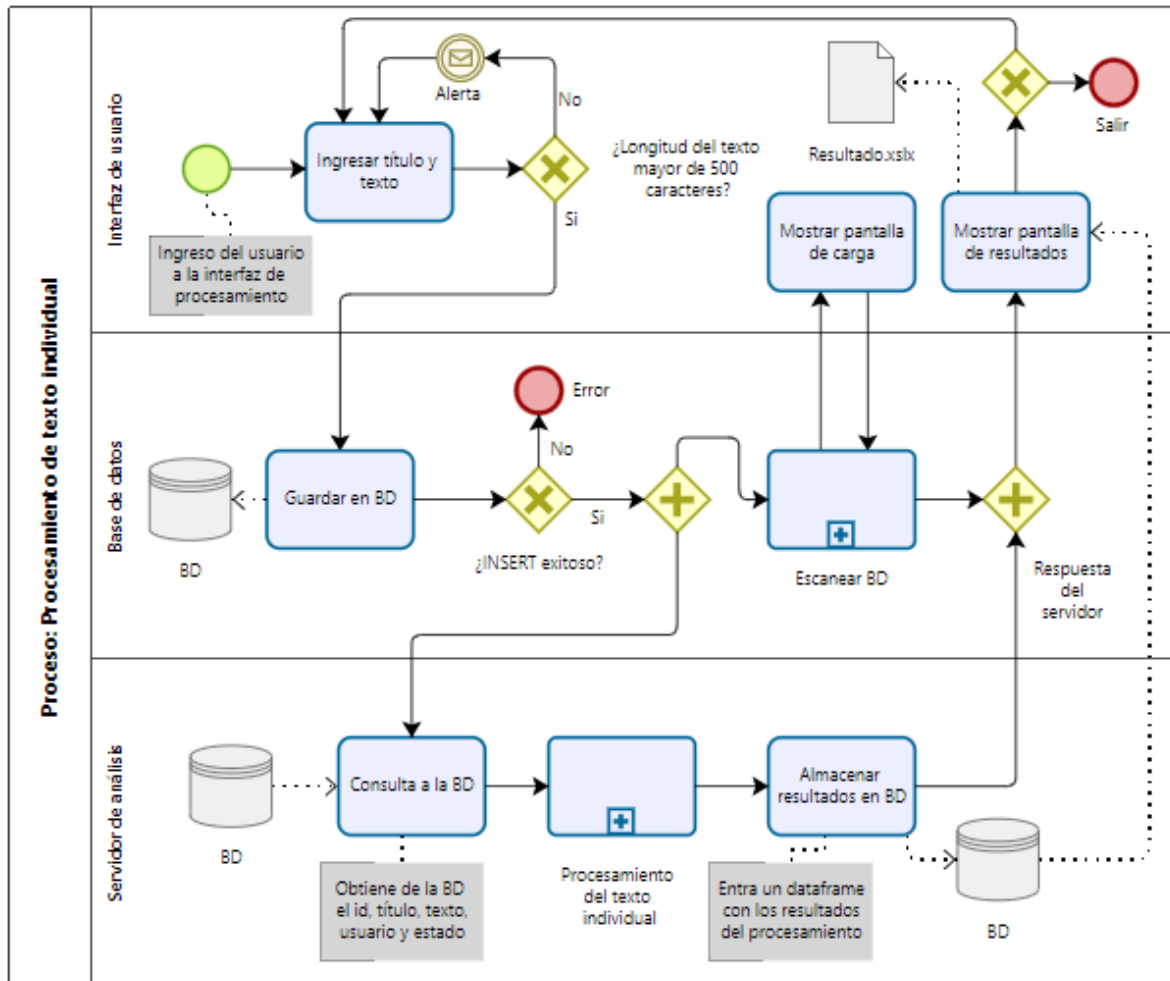


Figura 5. Proceso de análisis de texto individual.

Mientras que los sentimientos están más relacionados con factores psicológicos, es decir, se tratan de la interpretación de los pensamientos, emociones u opiniones propios del individuo. Por lo tanto, podemos decir que las emociones son estímulos fisiológicos de ocurrencia espontánea y de corta duración, mientras que los sentimientos son interpretaciones enteramente cognitivas del pensar del individuo y que pueden ser de mediana o larga duración.

Por supuesto que, al trabajar con textos como dato de entrada al sistema, lo cual se realizó en este trabajo, el concepto de sentimiento y emoción parecieran no encajar, ya que para medir las emociones sería necesario que se midiera la reacción física del individuo, algo que no se contempló en este trabajo de investigación. La solución para poder incorporar tanto el concepto de emoción como de sentimiento en este desarrollo fue utilizar corpus lingüísticos específicamente etiquetados para el análisis de emociones en textos, y para el análisis de sentimientos en textos.

Para cubrir con la tarea de análisis de sentimientos se utilizaron los léxicos de palabras positivas y negativas de (Bing, 2010), así como el léxico vader que es utilizado por la librería (VaderSentiment, 2021).

Para el caso del análisis de emociones se hace uso del léxico de subjetividad para adjetivos y verbos en inglés, incorporado en la librería (TextBlob, 2021). Lo anterior se debe a que este léxico asigna un puntaje positivo o negativo basado en los adjetivos y verbos etiquetados, recordando que las emociones son de naturaleza fisiológica y los verbos son palabras que expresan acciones de un individuo, se ha considerado un léxico coherente para tratar de analizar las emociones encontradas en el texto, lo que sería representado principalmente por los verbos que representan acciones de un individuo.

5.2.2 Descripción de las variables extraídas

La herramienta de análisis HAAT recibe como entrada un título (dato opcional) y un texto, en su función de análisis de texto individual. En caso de requerir un análisis múltiple (cuatro textos), como es el caso de los textos que se obtienen de las cuatro sesiones de la terapia de Escritura Emocional Autorreflexiva (EEA), la herramienta recibe como entrada un título (opcional) y un texto por cada una de las sesiones, recibiendo en total 8 entradas (4 títulos y 4 textos). En cualquiera de las dos opciones de análisis la herramienta realiza un flujo de trabajo prácticamente igual, dando como salida 30 variables (120 variables para el análisis múltiple).

En la Tabla 8, mostrada a continuación, se describen las características que se extraen durante el análisis y que se obtienen en forma de tabla para su consulta (archivo .xlsx):

Tabla 8. Listado de las variables que se obtienen utilizando la herramienta HAAT y una breve descripción de estas.

Abreviatura	Descripción
ID	Identificador numérico único
Título	Título colocado al texto para su reconocimiento (opcional)
Texto	Texto analizado por la HAAT
Longitud	Cantidad de palabras que componen el texto analizado
Oraciones	Numero de oraciones que contiene el texto
Palabras vacías	Cantidad de palabras poco relevantes que contiene el texto

Palabras únicas	Cantidad de palabras diferentes que contiene el texto
Riqueza léxica	Variedad del léxico utilizado en el texto
Facilidad de lectura	Evalúa la facilidad de lectura de un texto (prueba de Flesch)
Dificultad	Resultado de la prueba de Flesch
Polaridad negativa de sentimiento	Puntuación obtenida al extraer las palabras clasificadas como negativas en el léxico VADER
Polaridad neutral de sentimiento	Puntuación obtenida al extraer las palabras clasificadas como neutrales en el léxico VADER
Polaridad positiva de sentimiento	Puntuación obtenida al extraer las palabras clasificadas como positivas en el léxico VADER
Num de palabras negativas	Cantidad de palabras del texto analizado encontrados en el léxico Liu - Hu – Cheng de palabras negativas
Num de palabras neutrales	Cantidad de palabras del texto analizado no encontrados en los léxicos Liu - Hu – Cheng
Num de palabras positivas	Cantidad de palabras del texto analizado encontrados en el léxico Liu - Hu – Cheng de palabras positivas
Palabras positivas	Se muestran las palabras positivas detectadas en el texto
Palabras negativas	Se muestran las palabras negativas detectadas en el texto
Polaridad promedio de sentimiento	Puntuación compuesta ponderada normalizada del sentimiento del texto (VADER)
Polaridad emocional	Polaridad emocional del texto
Subjetividad emocional	Subjetividad emocional del texto
Emoción	Emoción detectada de acuerdo al modelo polaridad-subjetividad
Cantidad de verbos	Conteo de los verbos que aparecen en el texto
Cantidad de	Conteo de los sustantivos que aparecen en el

sustantivos	texto
Cantidad de adjetivos	Conteo de los adjetivos que aparecen en el texto
Cantidad de pronombres	Conteo de los pronombres que aparecen en el texto
Personas nombradas	Conteo de las entidades de tipo persona
Organizaciones nombradas	Conteo de las entidades de tipo organización
Lugares nombrados	Conteo de las entidades de tipo lugar/localización
Otras entidades nombradas	Conteo de otras entidades nombradas

A continuación se describe con mayor detalle cada característica obtenida:

- A) **ID:** Se asigna un identificador numérico único para cada texto procesado por la herramienta.
- B) **Título:** Es un dato opcional pero se recomienda colocarlo para distinguir rápida y fácilmente cada texto.
- C) **Texto:** Texto escrito utilizando un lenguaje natural (Español). Nota: La herramienta solo puede procesar texto en español, preferentemente de México.
- D) **Oraciones:** Una oración es una unidad de sentido que expresa una coherencia completa. Se delimita al inicio con una letra mayúscula y finaliza en un punto.
- E) **Palabras vacías (*Stop Words*):** Son las palabras que no contienen un significado (artículos, pronombres, preposiciones, etc.) pero son necesariamente incluidas en cualquier texto. Se realiza el filtrado utilizando la librería NLTK (Natural Language Toolkit, 2021).
- F) **Palabras únicas:** Se refiere a la cantidad de palabras diferentes que hay en el texto menos la cantidad de palabras vacías y menos otros símbolos y caracteres eliminados durante el procesamiento del texto.
- G) **Riqueza léxica:** Está definida por la siguiente ecuación:

$$\text{Riqueza léxica} = \frac{\text{Cantidad de caracteres únicos}}{\text{Cantidad total de caracteres}}$$

En donde el numerador es la cantidad de caracteres únicos utilizados en el texto y el denominador es la cantidad total de caracteres.

Por lo anterior se puede ver que la riqueza del léxico se obtiene con base en el conteo de caracteres del texto.

H) Facilidad de lectura: Prueba de facilidad de lectura de Flesch (Kincaid, Aagard, O'Hara & Cottrell, 1981), en donde las puntuaciones más altas indican un texto que es más fácil de leer y entender. En cambio los puntajes más bajos indican textos que son más difíciles de leer.

Matemáticamente se define de la siguiente forma:

$$\text{Prueba de Flesch} = 206.835 - 1.015 \left(\frac{\text{Cantidad de palabras}}{\text{Cantidad de oraciones}} \right) - 84.6 \left(\frac{\text{Cantidad de sílabas}}{\text{Cantidad de palabras}} \right)$$

En donde la cantidad de palabras es el conteo total de palabras en el texto, la cantidad de oraciones es el conteo total de oraciones en el texto y la cantidad de sílabas es el conteo de sílabas (fonemas pronunciados en una sola emisión o golpe de voz al decir una palabra).

I) Dificultad: Puntuación de la prueba de Flesch, interpretada de la siguiente manera: La puntuación puede interpretarse del siguiente modo:

Tabla 9. Descripción del puntaje de la prueba de Flesch.

Puntuación	Descripción
90-100	Texto muy fácil de comprender
80-90	Texto fácil de comprender (por ejemplo lenguaje conversacional)
70-80	Lectura relativamente fácil
60-70	Lenguaje llano (punto intermedio)
50-60	Texto relativamente difícil de entenderse
30-50	Texto de difícil entendimiento
0-30	Lectura muy difícil de ser comprendida

Es necesario considerar que el puntaje de legibilidad más alto posible de obtener, o de mayor facilidad de lectura, es 121.22 (Kincaid, Fishburne, Rogers & Chissom, 1975). En

cuanto al límite inferior no hay un puntaje mínimo teórico, por lo que es posible hacer que la puntuación sea tan baja como se determine de acuerdo a la composición del texto analizado (Textstat, 2021).

Análisis de sentimientos

Las características relacionadas con el análisis de sentimientos se extraen a partir de VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hu & Liu, 2004), una herramienta de análisis de sentimientos basada en reglas y en un léxico que está específicamente construido a partir de los sentimientos expresados y detectados en textos extraídos de redes sociales (Hutto & Gilbert, 2014).

J) Polaridad negativa de sentimiento: Se obtiene sumando las puntuaciones de valencia de cada palabra en el léxico VADER tomando aquellas palabras clasificadas como negativas y luego normalizando para estar entre -1 (extremo negativo máximo) y +1 (extremo positivo máximo) (VaderSentiment, 2021).

K) Polaridad neutral de sentimiento: Se obtiene sumando las puntuaciones de valencia de cada palabra en el léxico VADER tomando aquellas palabras clasificadas como neutrales y luego normalizando para estar entre -1 (extremo negativo máximo) y +1 (extremo positivo máximo).

L) Polaridad positiva de sentimiento: Se obtiene sumando las puntuaciones de valencia de cada palabra en el léxico VADER tomando aquellas palabras clasificadas como positivas y luego normalizando para estar entre -1 (extremo negativo máximo) y +1 (extremo positivo máximo).

M) Num de palabras negativas: Conteo de las palabras del texto analizado que se encuentran en el léxico Liu - Hu - Cheng (Liu, Bing & Hu, 2005) de palabras negativas. Para obtener las palabras se traduce el texto en Español a Inglés y posteriormente se busca cada palabra en el léxico.

N) Num de palabras neutrales: Corresponde a la cantidad de palabras que no se encuentran en el léxico positivo o negativo de Liu - Hu - Cheng.

O) Num de palabras positivas: Conteo de las palabras del texto analizado que se encuentran en el léxico Liu - Hu - Cheng de palabras positivas. Para obtener las palabras se traduce el texto en Español a Inglés y posteriormente se busca cada palabra en el léxico.

P) Palabras positivas: Listado con las palabras positivas encontradas en el léxico Liu - Hu - Cheng y el conteo de aparición de cada una de ellas.

Q) Palabras negativas: Listado con las palabras negativas encontradas en el léxico Liu - Hu – Cheng y el conteo de aparición de cada una de ellas.

R) Polaridad promedio de sentimiento (*compound*): Puntuación calculada sumando las puntuaciones de valencia de cada palabra en el léxico, realizando ajustes de acuerdo con las reglas de VADER y luego normalizando para estar entre -1 (extremo negativo máximo) y +1 (extremo positivo máximo). Esta es la métrica más útil si desea una única medida unidimensional de sentimiento para una oración o texto determinado. También se le puede denominar puntuación compuesta ponderada normalizada (Meiselman, 2021).

S) Polaridad emocional: Se determina al utilizar la librería textblob (TextBlob, 2021) para acceder al léxico de subjetividad para adjetivos y verbos en inglés (contenido dentro de la librería). El léxico está etiquetado con la polaridad emocional y un conjunto de cadenas de texto. Se buscan las cadenas etiquetadas de mayor similaridad al texto analizado, y de esa manera se determina la polaridad de la emoción. La puntuación de polaridad es un valor flotante dentro del rango [-1.0, 1.0] donde -1.0 es totalmente negativo y 1.0 es totalmente positivo.

T) Subjetividad emocional: Se determina al utilizar la librería textblob para acceder al léxico de subjetividad para adjetivos en inglés (contenido dentro de la librería). El léxico está etiquetado con la subjetividad emocional y un conjunto de cadenas de texto. Se buscan las cadenas etiquetadas de mayor similaridad al texto analizado, y de esa manera se determina la subjetividad de la emoción (Damerou & Indurkha, 2010). La puntuación de subjetividad es un valor flotante dentro del rango [0.0, 1.0] donde 0.0 es muy objetivo y 1.0 es muy subjetivo.

U) Emoción: Modelo de polaridad – subjetividad. Basado en el Modelo Descriptivo de los Estados Afectivos utilizado en UX (valencia y activación), las dos dimensiones psicofisiológicas son sustituidas por dimensiones lingüísticas: la polaridad y la subjetividad emocional. En cuanto a las emociones incorporadas son una adaptación de las 6 emociones básicas de Ekman (Ekman & Oster, 1981).

A continuación se muestra una comparativa entre ambos modelos:

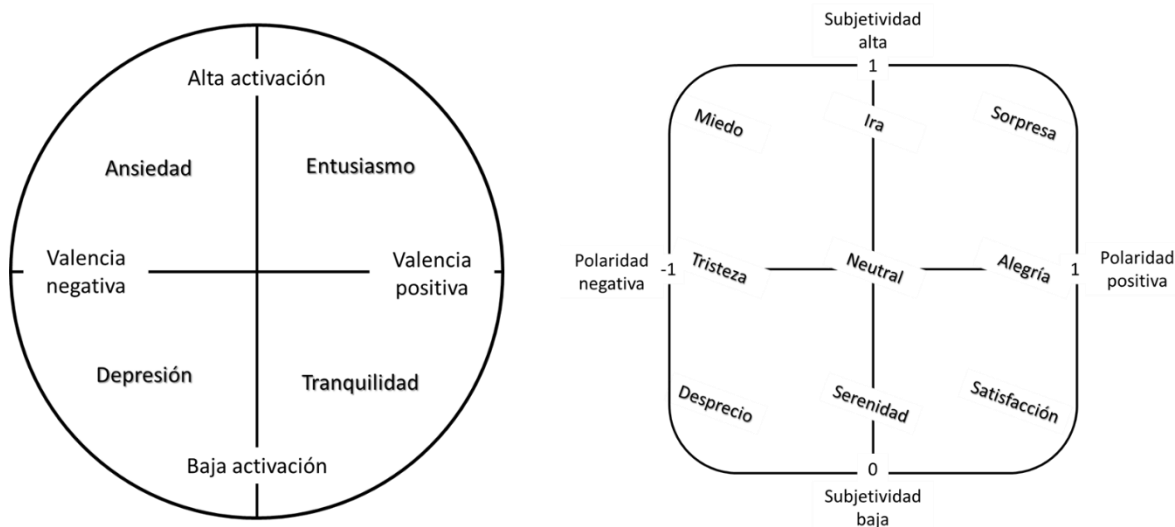


Figura 6. Modelo de valencia – activación (izquierda) y modelo de polaridad – subjetividad (derecha).

La figura izquierda muestra el Modelo Descriptivo de los Estados Afectivos mientras que a la derecha se muestra el Modelo de polaridad – subjetividad para el análisis de sentimientos en el texto. Utilizando un clasificador difuso de tipo 1 (basado en reglas) se obtiene la emoción, tomando como entrada las dos dimensiones mencionadas y obteniendo como salida alguna de las 9 emociones básicas posibles (ver Anexo 1).

Análisis de POS y NER

Las características relacionadas con la gramática del texto son aquellas referentes a la detección de la categoría gramatical de las palabras del texto (*Parts of speech* o POS) y al Reconocimiento de Entidades Nombradas (NER, por sus siglas en inglés). Para la extracción se utiliza la librería Spacy con el modelo pre entrenado es_core_news_sm (Spacy NLP, 2021).

V) Cantidad de verbos:

El

verbo es una categoría léxica que expresa una acción, movimiento, existencia, consecución, condición o estado del sujeto.

W) Cantidad de sustantivos: Los sustantivos son palabras cuyos referentes son clases de entidades fijas, no estados de hechos o relaciones gramaticales. Es la clase de palabras que se emplean para designar seres, entidades u objetos.

X) Cantidad de adjetivos: Categoría léxica que complementa un sustantivo para calificarlo. Expresa características o propiedades atribuidas a un sustantivo, ya sean concretas o abstractas. Los adjetivos son palabras que acompañan al sustantivo y que expresan cualidades y propiedades atribuidas a las entidades.

Y) Cantidad de pronombres: Palabras empleadas para referirse a personas, los animales o cosas sin nombrarlas, sustituyendo al sustantivo o variando el número, género y o persona en una oración.

Z) Personas nombradas: Conteo de la cantidad de palabras que hacen referencia a una persona dentro del texto.

AA) Organizaciones nombradas: Conteo de la cantidad de palabras que hacen referencia a una organización, empresa o institución dentro del texto.

AB) Lugares nombrados: Conteo de la cantidad de palabras que hacen referencia a un lugar o sitio geográfico concreto dentro del texto.

AC) Otras entidades nombradas: Conteo de la cantidad de palabras que hacen referencia a algún otro tipo de entidad dentro del texto.

Las características que se han mostrado fueron seleccionadas por considerarse representativas del lenguaje escrito. Por supuesto no se ha considerado la totalidad de variables gramaticales o de entidades nombradas, debido principalmente a que no son consideradas esenciales para describir una característica del lenguaje, sino que su uso principal es el de proporcionar un apoyo lingüístico para la construcción del lenguaje. Es absolutamente necesario tener en cuenta dos cuestiones relacionadas con el lenguaje: En primer lugar, el lenguaje es construido por cada individuo dependiendo de sus necesidades y con base en ciertas reglas gramaticales existentes. Por lo tanto, el lenguaje tiene una naturaleza variable que hace que incluso un mismo individuo, bajo las mismas condiciones, construya el lenguaje de formas diferentes. En segundo lugar, es importante tener en cuenta que el lenguaje escrito solamente puede proporcionar la información que se expresa dentro de su propia estructura, por lo tanto, el lenguaje solamente nos puede dar la información que representa y que contiene dentro de su representación, en este caso la escritura. Estas son las dos mayores limitantes que se han encontrado durante el desarrollo de este trabajo de tesis, y por consiguiente, la selección de las características extraídas se han visto influenciada por este par de limitantes que se han mencionado.

5.2.3 Pruebas para determinar precisión y óptimo funcionamiento

Un conjunto de pruebas iniciales se desarrolló con el objetivo de verificar el correcto funcionamiento de la herramienta, y para determinar la precisión de los módulos de análisis de sentimientos y análisis de emociones (pruebas 1 a 3). La prueba final (prueba 4) se desarrolló con datos recopilados de una prueba piloto llevada a cabo de forma electrónica (a través de la aplicación web desarrollada para tal fin) y en la que participaron voluntarias pertenecientes al grupo de estudio al cuál se dirigió el desarrollo de este proyecto de investigación, es decir, mujeres que han sido víctimas de algún tipo de violencia. En la Figura 7 se muestra la interfaz utilizada para el análisis de los conjuntos de datos de prueba.



Figura 7. Interfaz web de la herramienta HAAT.

Como ya se ha mencionado la interfaz, al ser web, requiere de un servidor web para su visualización, haciendo uso de un navegador web. Para el caso de las pruebas dicho servidor web se ha puesto en marcha en un dispositivo de cómputo. Así mismo, se han puesto en marcha las dos variantes de análisis del servidor (análisis individual y análisis múltiple). En la Figura 8 se muestra gráficamente el funcionamiento de las dos variantes del servidor de análisis, en donde se pueden notar los mensajes que arroja el servidor de acuerdo con su estado.

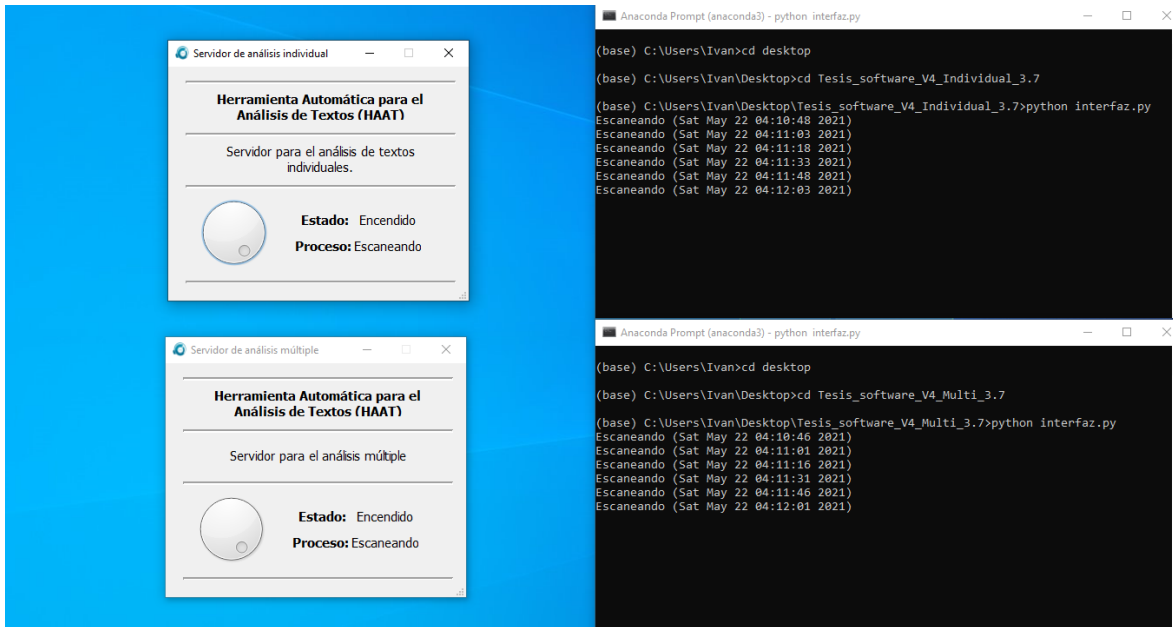


Figura 8. Servidores de análisis en ejecución.

5.3 Página web

En esta sección se muestra el proceso implementado para desarrollar las interfaces encargadas de la aplicación de la terapia de Escritura Emocional Autorreflexiva (EEA), así como la estructura de la base de datos utilizada para el almacenamiento de los datos ingresados por parte del usuario.

5.3.1 Página web

Al tratarse de un trabajo de investigación enfocado a una población de estudio con la necesidad de interactuar con un entorno virtual amigable e intuitivo, una de las primeras tareas fue el desarrollo de una página web que permitiera a las mujeres participantes en la terapia de EEA despejar sus principales dudas, así como tener un portal de inicio a través del cual se pueda proporcionar la información básica para que cada usuario individual pueda interactuar con la plataforma, así como conocer los objetivos y avances del proyecto.

En la Figura 9 se muestra el encabezado de la página web, en donde se incluye el logotipo desarrollado para este proyecto (esquina superior izquierda de la Figura), el menú de navegación (esquina superior derecha), así como un texto introductorio acompañado de una Figura representativa.



9. Encabezado de la página de inicio del portal web.

Figura

Partiendo de la página principal del portal se puede navegar por las diferentes interfaces que componen la página web, las cuales se describen a continuación y se muestran en la Figura 10:

- **Página de inicio:** Es el punto de partida para los y las usuarias que ingresan al portal. Se muestra información introductoria sobre esta parte del proyecto de tesis (aplicar una terapia de EEA), se incluye la respuesta a algunas preguntas frecuentes que podrían tener las usuarias, así mismo se describen brevemente los campos de investigación en los cuales de enfoca el proyecto y las instituciones involucradas, las cuales son el Tecnológico Nacional de México Campus CENIDET y el Instituto Nacional de Salud Pública (INSP).
- **Terapia:** Se proporciona información sobre la terapia de EEA, en qué consiste, cuál es el objetivo, así como los beneficios potenciales que tendrá para la persona participante el aplicar esta terapia alternativa. En la parte inferior de la ventana se proporciona el acceso al portal para realizar la terapia a través del navegador web (Ver sección 5.2).
- **Blog:** Se decidió incorporar un blog al portal web para proporcionar información actualizada sobre el estado del proyecto, cambios en la plataforma, actualizaciones, manuales y otros contenidos relevantes para facilitar a las usuarias la comprensión del proyecto y su evolución.
- **Contacto:** Como parte de la identidad creada para el proyecto, se habilitó una dirección de correo electrónico para la atención de dudas y sugerencias por parte de las personas usuarias.

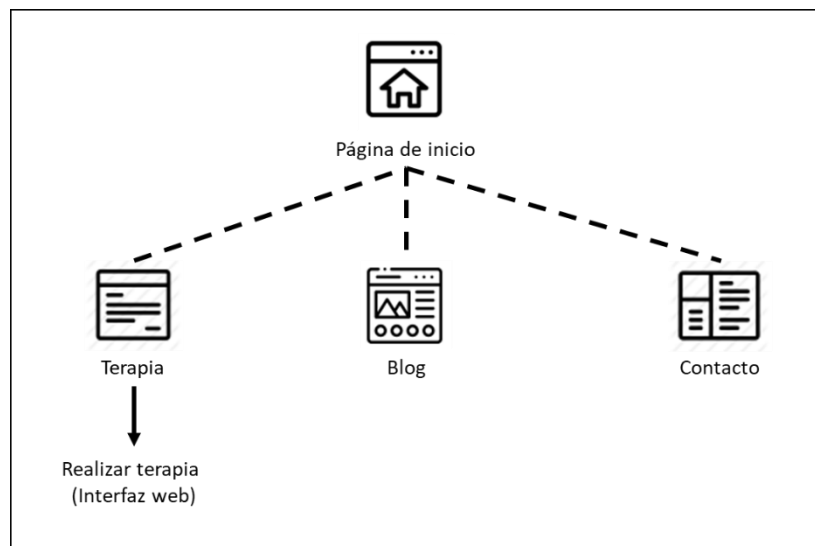


Figura 10. Esquema que muestra la relación entre las interfaces de la página web.

5.3.2 Terapia web

Si bien la página web mostrada en la Sección 5.1 se desarrolló con la finalidad de ser única y exclusivamente utilizada con fines informativos y de difusión, específicamente de la tarea consistente en aplicar una terapia de Escritura Emocional Autorreflexiva (EEA). Es el portal web para la aplicación de la misma terapia el que incorpora la interacción con las usuarias participantes.

A continuación, se describe cada una de las vistas incorporadas en el sistema para aplicar la terapia vía web:

- **Acceso:** Es la primera vista accesible al usuario de la herramienta. En esta ventana se solicita el ingreso del correo electrónico, contraseña y una clave de acceso. En caso de no contar con estos datos se debe proceder al registro (Nota: La clave de acceso es proporcionada por el administrador, y se implementó para evitar que cualquier persona pudiese realizar la terapia. De esta forma únicamente las participantes autorizadas pueden acceder.)
- **Registro:** Para crear un nuevo usuario en la plataforma de la terapia se debe proporcionar un nombre de usuario, una contraseña y una dirección de correo electrónico. Si los datos ingresados no existen en la base de datos y se encuentran en el formato correcto entonces se realiza el registro del nuevo usuario, notificando por medio de un mensaje informativo el estatus del proceso. Si el nuevo usuario se ha registrado correctamente deberá regresar a la ventana “Acceso” para proceder al inicio de sesión.
- **Inicio:** Al proporcionar las credenciales de acceso correctas la plataforma despliega la interfaz mostrada en la Figura 11. Se trata de la pantalla de navegación principal a través de la cual la persona participante puede desplazarse. En el encabezado se muestra el logotipo y título de la terapia (extremo superior izquierdo), mientras que en el extremo superior derecho se encuentra un menú de navegación secundario que permite al usuario cambiar su contraseña o cerrar su sesión.

En el extremo izquierdo de la vista se muestra un menú de navegación lateral, el cual permite el acceso a la ventana de instrucciones, a la ventana con el directorio de los centros de atención a mujeres víctimas de violencia, así como el acceso a los datos de contacto de la plataforma de aplicación de la terapia.

En cuanto a los elementos centrales estos representan cada una de las actividades que componen la terapia de escritura narrativa.

De color azul se muestra la actividad que debe realizar el usuario, en color gris las

tareas que bloqueadas que requieren que se complete la actividad anterior para ser habilitadas, y de color verde se encuentran las opciones completadas por la participante.

En la siguiente Figura se muestra la ventana de inicio:

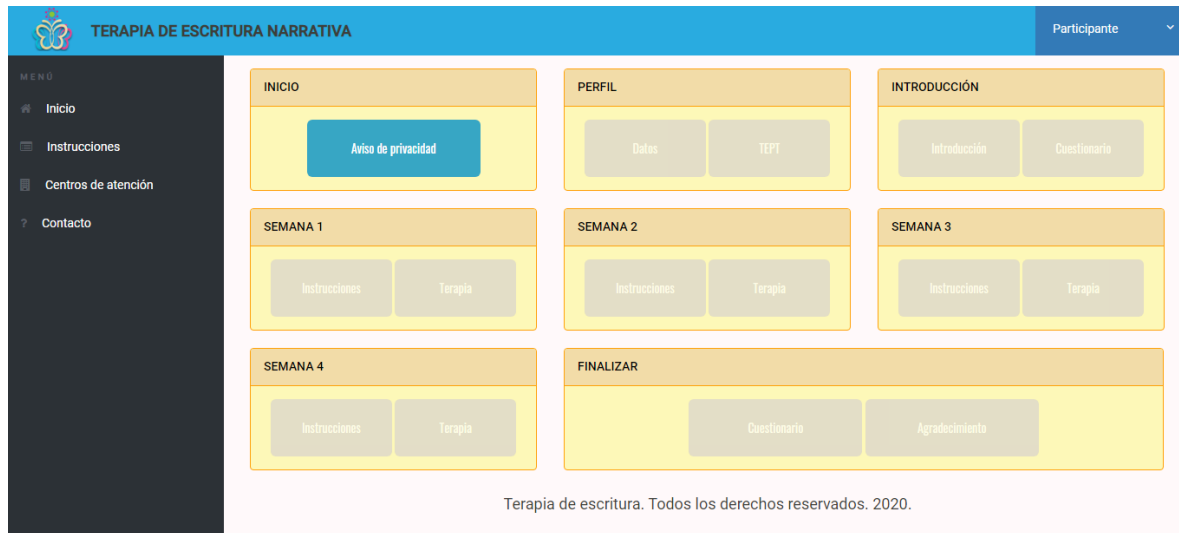


Figura 11. Menú de inicio de la ventana para realizar la terapia de EEA

- **Instrucciones:** Las instrucciones proporcionadas se componen por los pasos básicos necesarios para que cualquier persona con conocimientos básicos de computación pueda navegar y utilizar el sistema de aplicación de la terapia. Dichos pasos básicos se proporcionan con imágenes ilustrativas y textos cortos señalando las acciones a seguir para interactuar con los principales elementos.
- **Centros de atención:** En esta ventana se muestra la información correspondiente a los centros de atención para mujeres víctimas de violencia que están a disposición de las mujeres participantes. Debido a que se ha contemplado únicamente a la población de mujeres que habitan en el estado de Morelos, México, únicamente se incluyen datos informativos de esta entidad federativa. La información de contacto de los centros y dependencias dedicados a esta labor fueron extraídos del portal del gobierno federal, así como de datos publicados por el Instituto Mexicano del Seguro Social. Dicha información se encuentra disponible para su consulta de forma pública.
- **Contacto:** Se incluye el correo electrónico para brindar soporte a las personas usuarias de la terapia.

A continuación, se describen las vistas relacionadas directamente con la aplicación

de la terapia de escritura:

- **Aviso de privacidad:** Como aviso de privacidad se proporciona una carta de consentimiento informado a cada usuaria participante. En este apartado se especifica el uso de los datos, el propósito del estudio, así como orientar a la persona participante en caso de que su deseo sea causar baja del proyecto. Para completar la actividad se debe aceptar el aviso de privacidad presionando un campo de validación. Al realizar esta acción se puede continuar con el resto de los pasos, y se aceptan los términos y condiciones mencionados.
- **Datos:** En esta ventana se incorpora un cuestionario con dos secciones: En la primera de ellas se solicita a la participante que complete 9 preguntas relacionadas con su edad, el estado de la república en el que vive, su estado civil, entre otros. Es un cuestionario de naturaleza sociodemográfica que tiene como objetivo recopilar la información necesaria para perfilar a la participante. En la segunda sección se solicita que sean completadas 18 preguntas relacionadas con la violencia de pareja que se ha experimentado, antes y durante el confinamiento ocasionado por la pandemia de COVID19 a partir del segundo trimestre del año 2020.
- **TEPT:** Se realiza a cada participante una evaluación del trastorno de estrés postraumático consistente en la Lista Checable de Trastorno por Estrés Postraumático para DSM-5 (Durón Figueroa, 2019), y compuesta por 20 reactivos. El resultado no se le proporciona a la participante pero si se almacena en la base de datos para su posterior análisis.
- **Introducción:** Finalizando las actividades de perfilación y diagnóstico de la mujer participante se procede a mostrar la introducción a la terapia. Es un texto breve que describe de manera general la aplicación de la terapia de escritura narrativa.
- **Cuestionario:** Como vía para evaluar un posible cambio positivo o negativo en la participante al ser expuesta a la aplicación de la terapia de EEA se hace uso del inventario de ansiedad de Beck (BAI). Este instrumento se aplica antes de comenzar con la terapia y al finalizar la misma, sirviendo como base de comparación para determinar algunas características que señalen un cambio en la mujer participante. El resultado se almacena en la base de datos aunque no se muestra a la persona participante.
- **Instrucciones (Semana 1):** Ya que cada una de las sesiones que componen la terapia de escritura se aplican de manera semanal, una por semana durante cuatro semanas, en cada una de las cuatro actividades de escritura se proporcionan las instrucciones de escritura.

A continuación las instrucciones de escritura de la semana 1:

En esta sesión escribe un texto en el que cuentes tu experiencia de situación de violencia que has experimentado por parte de tu pareja. ¿Qué pasó? ¿Cómo te sentiste? ¿Cuáles fueron tus emociones?. Recuerda incluir también cómo te sentías, qué emociones, sentimientos y pensamientos experimentabas. Por favor incluye lo más de detalles posibles (el contexto, lugares, participantes, etc.). No hay límite de tiempo para realizar el ejercicio, y no te preocupes por la ortografía. Puedes escribir libremente, solo enfócate en lo que quieres contar.

- Terapia (Semana 1): Para realizar el ejercicio de escritura se proporciona a las participantes la interfaz mostrada en la Figura 12. Se compone de un campo para el ingreso de un título (el título para la narrativa de la participante es opcional), así como de un campo de texto para la escritura de la narrativa correspondiente a la terapia (campo obligatorio). Con la finalidad de evitar que la participante de por completada la actividad de forma accidental se debe validar un recuadro en la parte inferior de la vista.

Figura 12. Ventana para la aplicación del ejercicio de escritura.

- Instrucciones (Semana 2): Al habilitarse las actividades correspondientes a la semana dos se comienza por la sección de instrucciones.

A continuación, las instrucciones de escritura para la semana dos:

En esta segunda sesión queremos que profundices en tus emociones y sentimientos. Trata de describir con más detalle cómo te sientes a causa del episodio, o

episodios, de violencia que has sufrido. Escribe qué consecuencias y efectos ha tenido en ti y en las personas que te rodean.

Recuerda incluir también cómo te sentías, qué emociones, sentimientos y pensamientos experimentabas. Incluye información sobre el entorno, las personas que están a tú alrededor y todo detalle que consideres importante. No hay límite de tiempo para realizar el ejercicio, y no te preocupes por la ortografía. Puedes escribir libremente, solo enfócate en lo que quieres contar.

- **Terapia (Semana 2):** La actividad de escritura para la semana 2 cuenta con la misma estructura que el resto de ventanas de escritura. Un campo para ingresar un título a la narrativa (opcional), y un campo de texto para la escritura del ejercicio (obligatorio).
- **Instrucciones (Semana 3):** Al habilitarse las actividades correspondientes a la semana tres se comienza por la sección de instrucciones.

A continuación las instrucciones de escritura para la semana tres:

En esta tercera sesión debes abordar la escritura desde una perspectiva de tercera persona. Narra el evento como si tú fueras una persona que estuviera viendo la situación de violencia, en la cual no participas. Recuerda incluir también cómo te sentías, qué emociones, sentimientos y pensamientos experimentabas. Incluye información sobre el entorno, las personas que están a tú alrededor y todo detalle que consideres importante. No hay límite de tiempo para realizar el ejercicio, y no te preocupes por la ortografía. Puedes escribir libremente, solo enfócate en lo que quieres contar.

- **Terapia (Semana 3):** La actividad de escritura para la semana 3 cuenta con la misma estructura que el resto de ventanas de escritura. Un campo para ingresar un título a la narrativa (opcional), y un campo de texto para la escritura del ejercicio (obligatorio).
- **Instrucciones (Semana 4):** Al habilitarse las actividades correspondientes a la semana cuatro se comienza por la sección de instrucciones.

A continuación las instrucciones de escritura para la semana cuatro:

Para la última sesión, deberás crear una historia en la que te veas saliendo adelante de la situación que has vivido. ¿Cómo te vez en el futuro?. Por favor escribe cómo quieres que sea tu vida al superar la situación actual.

Recuerda incluir también cómo te sentías, qué emociones, sentimientos y pensamientos experimentabas

Incluye información sobre el entorno, las personas que están a tu alrededor y todo detalle que consideres importante. No hay límite de tiempo para realizar el ejercicio, y no te preocupes por la ortografía. Puedes escribir libremente, solo enfócate en lo que quieres contar.

- **Terapia (Semana 4):** La actividad de escritura para la semana 4 cuenta con la misma estructura que el resto de ventanas de escritura. Un campo para ingresar un título a la narrativa (opcional), y un campo de texto para la escritura del ejercicio (obligatorio).
- **Cuestionario:** Nuevamente se aplica el inventario de ansiedad de Beck (BAI). Este instrumento se aplica antes de comenzar con la terapia y al finalizar la misma, sirviendo como base de comparación para determinar algunas características que señalen un cambio en la mujer participante. El resultado se almacena en la base de datos aunque no se muestra a la persona participante.
- **Agradecimiento:** Finalmente se proporciona un mensaje de agradecimiento a las mujeres participantes que han terminado con la terapia de EEA. Esta es la última actividad de la terapia.

En la siguiente Figura se muestra de forma gráfica la relación entre cada una de las vistas incorporadas:

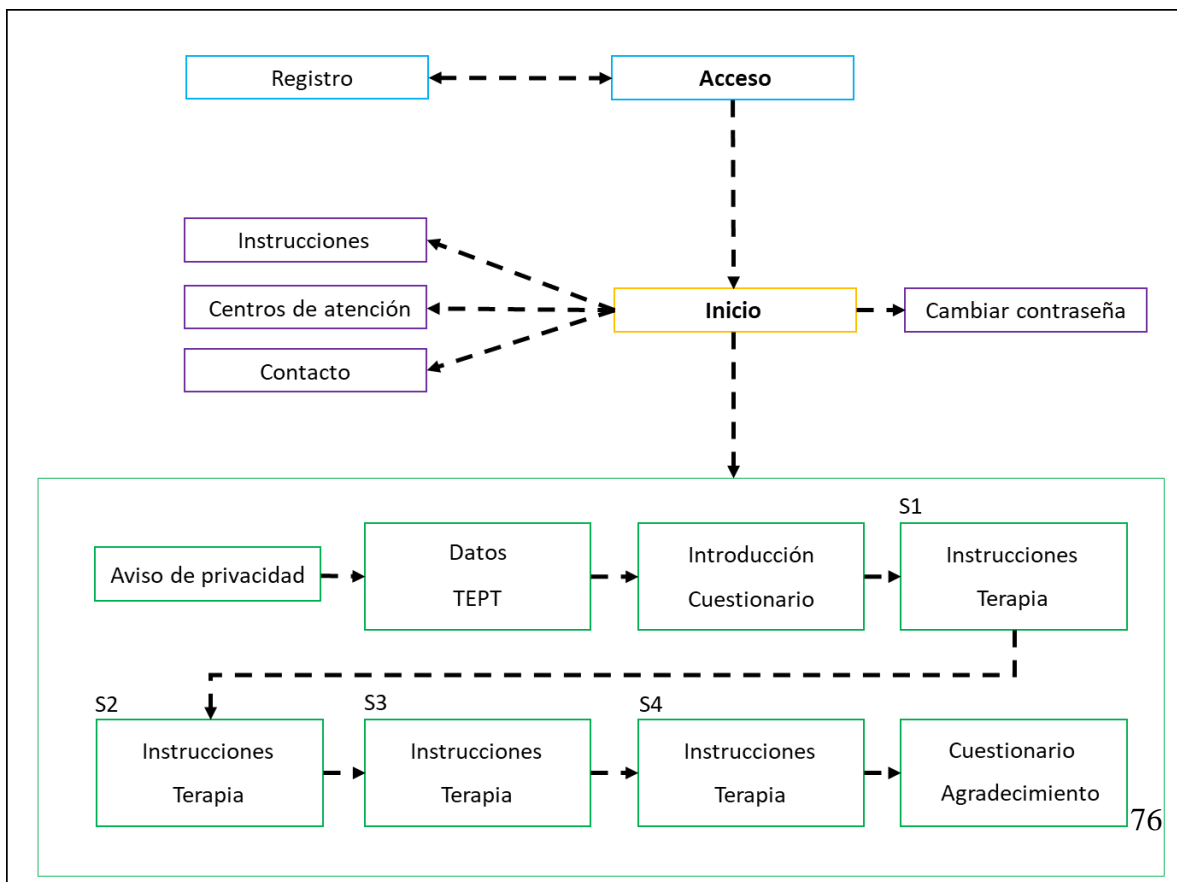


Figura 13. Relación de las interfaces incorporadas en el sistema web para aplicar la terapia.

Como ha sido posible observar en esta sección, la terapia utilizada en este trabajo de tesis se compone de 15 pasos. Los cuales se deben realizar de forma secuencial comenzando por aceptar los términos y condiciones de la investigación, siguiendo con la recopilación de datos sociodemográficos, de trastorno de estrés postraumático (TEPT) y de ansiedad, para posteriormente aplicar las actividades de escritura, una cada semana, durante cuatro semanas. Se finaliza con un cuestionario de ansiedad y un mensaje de agradecimiento. Este paradigma de aplicación de la terapia de escritura está basado en los trabajos que se han consultado como parte de la revisión de la literatura que ha incorporado este trabajo de tesis, tomando principalmente como base los trabajos de Pennebaker y su paradigma original de aplicación.

Un punto importante a tener en cuenta es que esta herramienta se pensó para su uso desde navegadores web ejecutados en equipos de cómputo, es decir, que los elementos se ajustan a un tamaño y dimensión propio de una pantalla de computadora y no se ajusta perfectamente en una pantalla de menores dimensiones, como lo son las pantallas de tabletas o teléfonos inteligentes. Para el caso que se prefiera utilizar un dispositivo móvil se desarrolló una aplicación móvil para el sistema operativo Android. Por lo tanto, la responsividad de los elementos gráficos de la interfaz de la terapia web es un trabajo a futuro que no se ha contemplado para este trabajo de tesis, precisamente porque existe una aplicación móvil que puede cubrir con ese aspecto.

Es necesario señalar que se trata de una aplicación dinámica, por lo tanto, existe una base de datos que permite el almacenamiento de algunos datos así como la consulta y recuperación de otros. Este aspecto se describe en la siguiente sección.

5.3.3 Base de datos

El objetivo de la interfaz para aplicar la terapia de escritura vía web es la recopilación de datos. Ese es su propósito y, por ello, ha sido necesario incorporar un Sistema Gestor de Base de Datos (SGBD) para el manejo del aspecto de almacenamiento de datos. De acuerdo a (Hueso Ibañez, 2016) una base de datos es un conjunto de datos relacionado y organizado con base en cierta estructura predefinida. Para realizar operaciones de inserción, eliminación, consulta o actualización se requiere de Sistemas Gestores de Base de Datos (SGBD), los cuales son herramientas que proporcionan un medio para la interacción entre los usuarios y las bases de datos.

Para los desarrollos involucrados en este trabajo de tesis se ha optado por incorporar el SGBD MySQL, el cual según (Christudas, 2019) es un SGBD de código

abierto y con gran popularidad desarrollado y respaldado por Oracle Corporation. Esta herramienta es del tipo relacional, almacenando los datos en tablas independientes. Entre sus principales ventajas está el de tener una estructura que permite optimizar la velocidad de las operaciones, así como ofrecer una estructura flexible al desarrollador. El software MySQL utiliza la licencia GPL (GNU General Public) y es un software de código abierto. Por supuesto hay que señalar una de sus principales desventajas, y es que al manejar grandes volúmenes de datos se pierde el funcionamiento óptimo de la herramienta, sin embargo, para el manejo de datos contemplado en este proyecto es más que suficiente la capacidad que proporciona esta herramienta, por lo que es el SGBD sobre el cual se almacenan los datos en todos los elementos que requieren de esta tarea desarrollados e incluidos en este trabajo de tesis.

La base de datos incorporada a la interfaz de la terapia web consta de 12 tablas, relacionadas todas ellas por medio de un identificador único (ID), y como usuario se utilizó el correo electrónico. A continuación se describe cada una de las tablas y sus respectivos campos:

- Tabla users: En esta tabla se almacenan los datos de los usuarios al momento de ser creados en el formulario de registro de la aplicación. Se genera un ID único para cada registro, se almacena el nombre, el correo electrónico y la contraseña en formato varchar con una longitud máxima para cada uno de ellos de 50 caracteres. Se incluye un campo int para señalar el status del usuario (1 significa que es un usuario activo).

También se incluye un campo para almacenar el código de acceso a la aplicación web, este código evita que cualquier persona externa al estudio pueda realizar la terapia.

- Tabla state: Esta tabla permite mantener actualizado el estado de avance del usuario con relación a la terapia. Los valores que abarca son del 1 al 15, correspondiendo estos a cada una de las actividades que componen la terapia. Esta tabla permite bloquear las actividades pendientes o completadas y dirigir al usuario a una actividad en específico. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. En formato int se almacena el estado (rango de valores del 1 al 15) y finalmente se incluye una columna para almacenar la fecha y hora de actualización del registro.
- Tabla deleteduser: Almacena un listado de todos los usuarios dados de alta. Aunque sea eliminado un usuario de la tabla de usuarios principal se almacena en esta tabla como respaldo. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Finalmente se incluye una columna para almacenar la fecha y hora de actualización del registro.

- Tabla perfil: Esta tabla de 12 columnas tiene como finalidad almacenar los datos obtenidos al responder el cuestionario de información sociodemográfica incorporado en la terapia de escritura. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Cada uno de los 9 reactivos del cuestionario sociodemográfico se almacena en una columna independiente, siendo los campos de entidad, sexo, escolaridad, etnia, ocupación, estado civil y víctima (¿La participante ha sido víctima de violencia?) almacenados en formato de texto, mientras que los campos edad e hijos (número de hijos) se almacenan en formato int con un máximo de 11 caracteres cada uno. También se incluye una columna para almacenar la fecha y hora de actualización del registro.
- Tabla violencia: En esta tabla se almacenan los datos recopilados al completar el cuestionario de violencia de pareja. Consta de 21 registros de los cuales 18 corresponden a los reactivos del cuestionario. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres.

Posteriormente se incluyen los 18 campos para cada uno de los reactivos del cuestionario, los cuales se almacenan en formato de texto y hacen referencia a si la persona participante ha sufrido humillación, agresión, golpes, insultos o ha sido forzada a realizar actos en contra de su voluntad. Cada reactivo incluye una respuesta para los eventos sufridos antes de la pandemia por COVID19 acontecida a partir del primer trimestre de 2020 así como su comparación con los eventos experimentados durante el periodo de tiempo del confinamiento. También se incluye una columna para almacenar la fecha y hora de actualización del registro.

- Tabla tept: En esta tabla se almacenan los resultados obtenidos al aplicar la Lista Checable de Trastorno por Estrés Postraumático para DSM-5 (Durón Figueroa, 2019). Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Después se incluyen 27 campos, de los cuales 2 son campos en formato textual varchar con un máximo de 2000 caracteres cada uno, seguido de las 23 columnas correspondientes a cada uno de los reactivos del instrumento, y los cuales se almacenan en formato numérico int con un posible rango de valores de 0 a 4. Posteriormente se incluye un campo para almacenar la sumatoria de todos los reactivos y, en otro campo se etiquetan los resultados de acuerdo al nivel de tept determinado por el inventario. La última columna corresponde a la fecha y hora de inserción de los datos.
- Tabla ansiedadpre: En esta tabla se almacenan los datos obtenidos al aplicar el inventario de ansiedad de Beck (Galindo Vázquez et al., 2015) antes de comenzar con la terapia de escritura. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. De las 23 columnas adicionales 21 corresponden a los ítems incorporados en el inventario de ansiedad de

Beck, los cuales se almacenan en formato numérico int. Posteriormente se incluye una columna en la que se almacena la sumatoria de todos los reactivos y en otro campo se guarda la etiqueta que determina el nivel de ansiedad de la persona participante. Finalmente se incluye una última columna corresponde a la fecha y hora de inserción de los datos.

- Tabla texto1: En esta tabla se almacenan los datos obtenidos al aplicar el ejercicio de escritura correspondiente a la semana 1 de la terapia. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Se incluye una columna para almacenar el título puesto por cada participante a su texto narrativo (es un dato opcional), seguido de una columna para el almacenamiento de los textos en formato de texto. La última columna corresponde a la fecha y hora de inserción de los datos.
- Tabla texto2: En esta tabla se almacenan los datos obtenidos al aplicar el ejercicio de escritura correspondiente a la semana 2 de la terapia. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Se incluye una columna para almacenar el título puesto por cada participante a su texto narrativo (es un dato opcional), seguido de una columna para el almacenamiento de los textos en formato de texto. La última columna corresponde a la fecha y hora de inserción de los datos.
- Tabla texto3: En esta tabla se almacenan los datos obtenidos al aplicar el ejercicio de escritura correspondiente a la semana 3 de la terapia. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Se incluye una columna para almacenar el título puesto por cada participante a su texto narrativo (es un dato opcional), seguido de una columna para el almacenamiento de los textos en formato de texto. La última columna corresponde a la fecha y hora de inserción de los datos.
- Tabla texto4: En esta tabla se almacenan los datos obtenidos al aplicar el ejercicio de escritura correspondiente a la semana 4 de la terapia. Se incluye un ID único para cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. Se incluye una columna para almacenar el título puesto por cada participante a su texto narrativo (es un dato opcional), seguido de una columna para el almacenamiento de los textos en formato de texto. La última columna corresponde a la fecha y hora de inserción de los datos.
- Tabla ansiedadpost: En esta tabla se almacenan los datos obtenidos al aplicar el inventario de ansiedad de Beck (Galindo Vázquez et al., 2015) después de realizar la terapia de escritura. Su estructura es igual que la tabla “ansiedadpre”. Se incluye un ID único para

cada registro, se almacena el correo electrónico en formato varchar con una longitud máxima de ellos de 50 caracteres. De las 23 columnas adicionales 21 corresponden a los ítems incorporados en el inventario de ansiedad de Beck, los cuales se almacenan en formato numérico int. Posteriormente se incluye una columna en la que se almacena la sumatoria de todos los reactivos y en otro campo se guarda la etiqueta que determina el nivel de ansiedad de la persona participante. Finalmente se incluye una última columna corresponde a la fecha y hora de inserción de los datos.

En la Figura 14 se muestra la relación de las tablas de la base de datos que incorpora la interfaz para la aplicación de la terapia vía web.

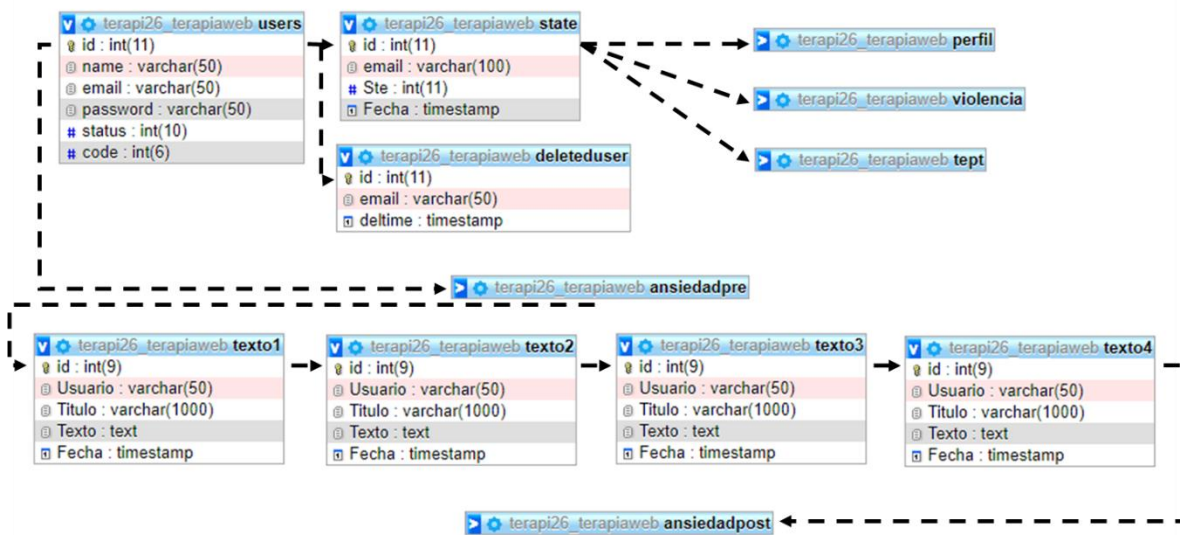


Figura 14. Estructura y relación de la base de datos de la terapia web

La estructura que se muestra en la Figura anterior se fue adaptando hasta dar como resultado un diseño de base de datos que cuenta con la capacidad de soportar las interacciones que requiere la propia interfaz web, la cual se desarrolló con las tecnologías web de uso estándar. Estas interacciones se dan entre el frontend y el backend de la siguiente manera:

- **HTML y CSS:** Etiquetado de los elementos de la interfaz y definición del estilo de los elementos gráficos. Se utilizó el framework bootstrap, en su versión v5.0.0-beta1 para realizar esta labor.
- **Javascript:** Lenguaje de programación dinámico e interpretado por el usuario. Se utilizó para mostrar mensajes de aviso, así como para controlar elementos dinámicos.

- PHP: Lenguaje interpretado por el servidor, el cual permitió la manipulación y consulta de la base de datos relacional (MySQL), así como el desarrollo de varios elementos de control dentro de la interfaz pero invisibles al usuario.

5.3.4 Aplicación móvil

Como plataforma complementaria al sistema web para la aplicación de la terapia de escritura se desarrolló una aplicación móvil para el sistema operativo Android. El concepto de esta aplicación es que sea utilizada por las usuarias que tengan un dispositivo móvil personal. Es importante que sea un dispositivo al que solamente tenga acceso la usuaria participante (un teléfono inteligente o una tablet) debido a que la aplicación móvil limita su funcionamiento a una persona por dispositivo, esto se debe a que la base de datos almacena como un usuario único a cada dispositivo, en vez de requerir que la participante complete un cuestionario de registro la aplicación genera un usuario de manera automática, por ello es que únicamente se puede manejar un usuario por dispositivo móvil.

El primer planteamiento del proceso para aplicar la terapia de escritura emocional autorreflexiva (EEA) contemplaba la distribución de la aplicación de forma pública, con lo que cualquier mujer víctima de violencia podría tener acceso a la aplicación de la terapia. Sin embargo, durante la maduración del proyecto el enfoque ha cambiado: en vez de ser una aplicación dirigida a la población de mujeres mexicanas en general ahora se ha segmentado para ser aplicada únicamente a una población de mujeres víctimas de violencia, específicamente para aquellas mujeres que acuden a los centros de apoyo y atención que se encuentran en el estado de Morelos, México.

A continuación se da una descripción de las vistas que componen la aplicación móvil (android), la aplicación se puede instalar en cualquier dispositivo con sistema operativo Android, API 6.0 o superior:

- Pantalla de carga: Al abrir la aplicación desde el menú de aplicaciones del dispositivo móvil se muestra una ventana de carga. Esta ventana muestra el logotipo desarrollado para servir como identidad de este trabajo de tesis, específicamente de los procedimientos en los que se interactúa directamente con el usuario.
- Pantalla principal: Después de la pantalla de inicio, la pantalla principal (Figura 15) contiene el logotipo de la aplicación, y cuatro botones de navegación: Botón “Comenzar”, en donde se redirige al usuario a una vista en la que se proporciona una descripción de la terapia. Botón “Privacidad”, en donde se muestran el aviso de consentimiento informado para las personas participantes. Botón “Realiza la terapia”, el cual dirige al usuario a la ventana de aplicación de la terapia. Botón “Directorio”, el cual envía al usuario al

directorio de contacto (muestra la dirección de los centros de atención para mujeres víctimas de violencia en el estado de Morelos, México).

También se muestran dos elementos comunes en todas las pantallas de la aplicación:

El Botón de cierre rápido, para salir de la aplicación rápidamente, y el menú lateral.

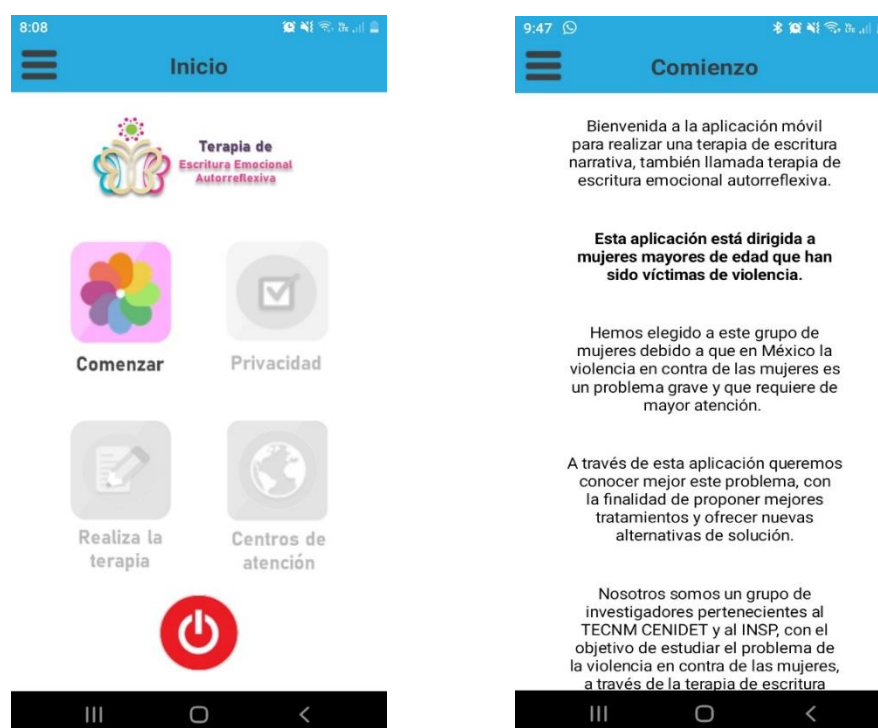


Figura 15. Aspecto de la pantalla de navegación principal de la aplicación móvil

- Menú lateral: El menú lateral es un elemento desplegable, se activa al presionar el correspondiente botón localizado en todas las vistas en la parte superior izquierda de la pantalla. Incorpora 5 acciones: a) Botón “Inicio”, el cuál dirige a la página principal. b) Botón “Instrucciones”, para ir a la ventana de instrucciones. c) Botón “Contacto”, para ir a la ventana que muestra los datos de contacto de los encargados de la app. d) Botón “Salir”, para cerrar la aplicación. e) Flecha para ocultar el menú lateral.
- Instrucciones: Vista en la que se muestran las instrucciones básicas para que cualquier usuario con conocimientos básicos de manejo de dispositivos móviles pueda interactuar con la aplicación y realizar la terapia de escritura. Se proporciona al usuario una combinación de imágenes y textos cortos para mostrar los principales elementos de la interface, así como las opciones de navegación.

- **Contacto:** En esta vista se incluye una dirección de correo electrónico habilitada específicamente para recibir comentarios, dudas o sugerencias por parte de las mujeres usuarias de la aplicación.
- **Privacidad:** Ya que la finalidad de la aplicación es la recopilación de datos para su posterior uso en investigaciones científicas, antes de comenzar la terapia se muestra un aviso de privacidad a los usuarios, el cual permite dar a conocer el uso y finalidad de la aplicación y de la recopilación de datos. Es necesario que el usuario confirme que está de acuerdo ya que, de no ser así, la aplicación no permitirá el avance a la ventana de aplicación de la terapia.
- **Realiza la terapia:** Es la vista en donde se aplica el procedimiento correspondiente a la terapia de escritura narrativa. La estructura y navegación se inspiró en la forma en la que se muestra el contenido en aplicaciones de e-learning o plataformas de cursos, guiando al usuario en todo el procedimiento, debiendo comenzar el proceso en la parte superior para ir completando cada una de las etapas hasta concluir el proceso terapéutico. El usuario navega haciendo uso de los botones azules, los cuales se muestran y desbloquean de acuerdo con el avance del usuario.
- **Centros de atención:** La finalidad de esta vista es mostrar a las mujeres usuarias los datos de contacto de los principales centros de atención para mujeres víctimas de violencia que hay en el estado de Morelos, México. Se proporciona un campo para desplegar una lista con los centros de atención, y al presionar sobre uno de ellos se despliega la información de contacto en forma de texto.

Las vistas descritas hasta ahora son puramente informativas. A continuación se describen las vistas que forman parte de la terapia de escritura emocional autorreflexiva (EEA), es necesario señalar que son los mismos pasos que la terapia de escritura accesible a través de la web, por lo que en este caso solamente se da una breve descripción de cada uno (para ver detalladamente cada etapa ver la Sección 5.2 de este reporte):

- **Datos:** Contiene un cuestionario de naturaleza sociodemográfica que tiene como objetivo recopilar la información necesaria para perfilar a la participante. En la segunda sección se incluye un cuestionario para explorar la violencia de pareja a la que se ha expuesto la participante.
- **TEPT:** Aplicación de la Lista Checable de Trastorno por Estrés Postraumático para DSM-5 (Durón Figueroa, 2019).
- **Introducción:** Es un texto breve que describe de manera general la aplicación de la terapia de escritura narrativa.

- Cuestionario: Aplicación del inventario de ansiedad de Beck (BAI) para que sirva como punto de comparación entre el antes y después de la aplicación de la terapia.
- Instrucciones (Semana 1): Instrucciones correspondientes a la tarea de escritura que se debe redactar durante la semana 1 de la terapia.
- Terapia (Semana 1): Se compone de un campo para el ingreso de un título (el título para la narrativa de la participante es opcional), así como de un campo de texto para la escritura de la narrativa correspondiente a la terapia (campo obligatorio). A través de esta vista se realiza el ejercicio de escritura.
- Instrucciones (Semana 2): Instrucciones correspondientes a la tarea de escritura que se debe redactar durante la semana 2 de la terapia.
- Terapia (Semana 2): Se compone de un campo para el ingreso de un título (el título para la narrativa de la participante es opcional), así como de un campo de texto para la escritura de la narrativa correspondiente a la terapia (campo obligatorio). A través de esta vista se realiza el ejercicio de escritura.
- Instrucciones (Semana 3): Instrucciones correspondientes a la tarea de escritura que se debe redactar durante la semana 3 de la terapia.
- Terapia (Semana 3): Se compone de un campo para el ingreso de un título (el título para la narrativa de la participante es opcional), así como de un campo de texto para la escritura de la narrativa correspondiente a la terapia (campo obligatorio). A través de esta vista se realiza el ejercicio de escritura.
- Instrucciones (Semana 4): Instrucciones correspondientes a la tarea de escritura que se debe redactar durante la semana 4 de la terapia.
- Terapia (Semana 4): Se compone de un campo para el ingreso de un título (el título para la narrativa de la participante es opcional), así como de un campo de texto para la escritura de la narrativa correspondiente a la terapia (campo obligatorio). A través de esta vista se realiza el ejercicio de escritura.
- Cuestionario: Nuevamente se aplica el inventario de ansiedad de Beck (BAI).
- Agradecimiento: Se proporciona un mensaje de agradecimiento a las mujeres participantes que han terminado con la terapia de EEA.

En la Figura 16 se muestra gráficamente la forma en la que se relacionan las diferentes vistas que incorpora la aplicación móvil desarrollada.

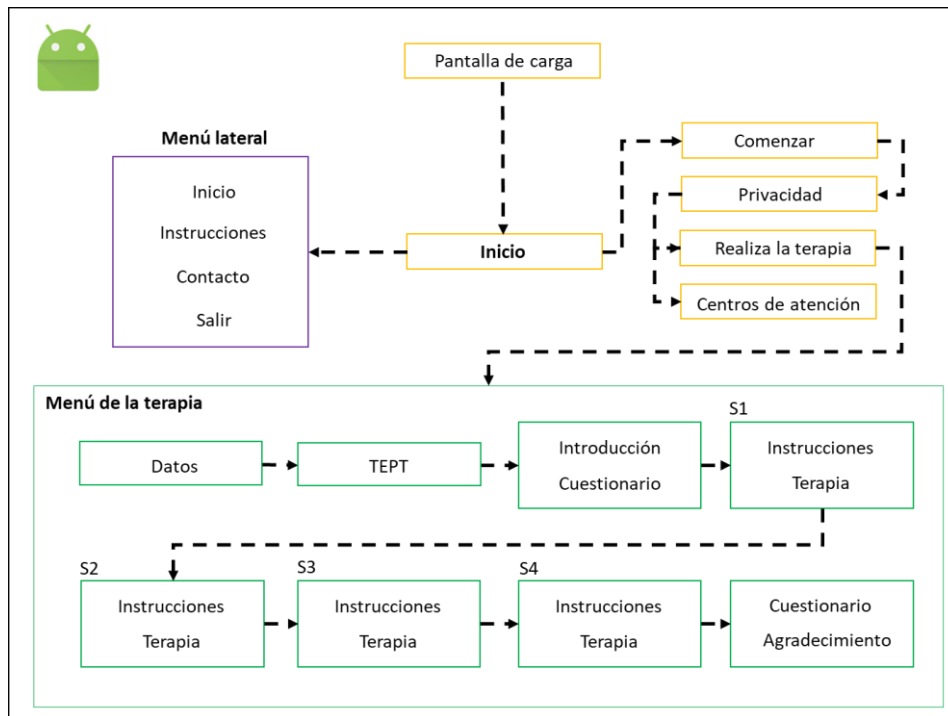


Figura 16. Diagrama de las interfaces que componen la aplicación móvil.

5.3.5 Base de datos

Con la finalidad de separar los datos insertados por medio de la aplicación móvil, de los que se ingresen por la interfaz web, se generó una base de datos relacional con el Sistema Gestor de Base de Datos MySQL para almacenar los datos procedentes de la aplicación móvil.

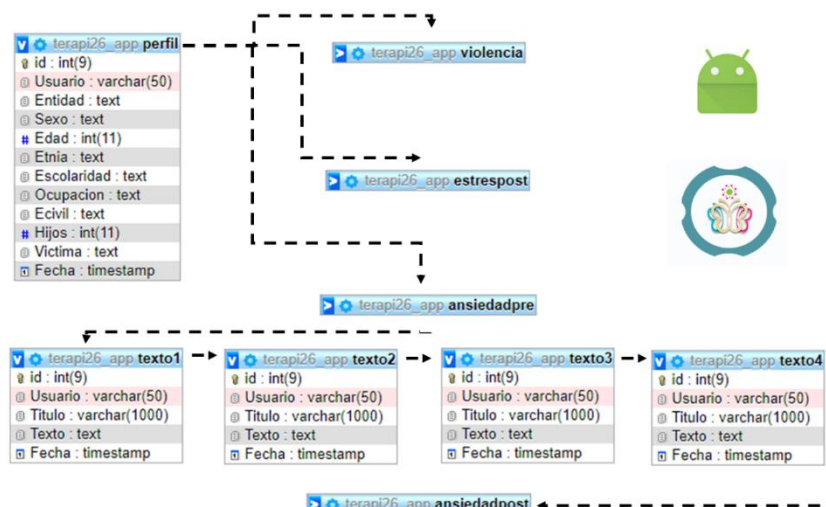


Figura 17. Estructura y relación de la base de datos de la aplicación móvil para la aplicación de la terapia.

A continuación se describen los componentes de cada una de las tablas incorporadas en la base de datos:

- Tabla perfil: Es la tabla en la que se almacenan los datos recopilados al aplicar un cuestionario sociodemográfico a las personas participantes. Incorpora un campo auto incrementable con el identificador único para cada registro, seguido por un campo para almacenar el usuario que se genera automáticamente. Después de los datos anteriores siguen los 8 espacios para almacenar las respuestas a las mismas 8 preguntas que incorpora el cuestionario de perfilación de la mujer participante. Finalmente hay un campo para insertar la fecha y hora de inserción del campo.
- Tabla violencia: Estructura en la que se almacenan las respuestas ingresadas por medio del formulario correspondiente al diagnóstico de violencia de pareja. Incorpora un campo con el identificador único para cada registro seguido por un campo para almacenar el usuario. También se encarga de almacenar las 18 respuestas a la misma cantidad de reactivos que se incluyen en la correspondiente vista gráfica de la aplicación.
- Tabla estrespost: En esta tabla se almacenan los resultados obtenidos al aplicar la Lista Checable de Trastorno por Estrés Postraumático para DSM-5 (Durón Figueroa, 2019). Se incluye un ID único para cada registro, el nombre de usuario, así como los campos correspondientes a cada reactivo del cuestionario de tept.
- Tabla ansiedadpre: En esta tabla se almacenan los datos obtenidos al aplicar el inventario de ansiedad de Beck (Galindo Vázquez et al., 2015). De la misma forma que en las columnas anteriores se incluye el identificador único, el nombre de usuario y un campo para registrar la fecha y hora de la inserción o actualización del campo.
- Tabla texto1: Almacenamiento para el primer ejercicio de escritura de la terapia, la estructura es similar al resto de los registros.
- Tabla texto2: Almacenamiento para el segundo ejercicio de escritura de la terapia, la estructura es similar al resto de los registros.
- Tabla texto3: Almacenamiento para el tercer ejercicio de escritura de la terapia, la estructura es similar al resto de los registros.
- Tabla texto4: Almacenamiento para el cuarto ejercicio de escritura de la terapia, la estructura es similar al resto de los registros.

- Tabla ansiedadpost: En esta tabla se almacenan los datos obtenidos al aplicar por segunda vez el inventario de ansiedad de Beck. De la misma forma que en las columnas anteriores se incluye el identificador único, el nombre de usuario y un campo para registrar la fecha y hora de la inserción o actualización del campo.

Nota: La comunicación entre la base de datos y la aplicación móvil se da mediante el método POST del protocolo HTTP, utilizando el servidor web como puente para realizar la inserción de datos.

Capítulo VI

Pruebas

En este capítulo se describe cada una de las pruebas realizadas para probar el funcionamiento de la herramienta de análisis de textos en la tarea de análisis de datos. En el Anexo 4 se detallan las métricas de evaluación utilizadas y los resultados obtenidos:

6.1 Pruebas con el corpus NVS2019

El corpus lingüístico denominado NVS2019 es una recopilación de textos escritos por mujeres víctimas de violencia, disponibles para consulta pública a través de redes sociales. Para su uso en las pruebas de la herramienta HAAT se construyó este corpus con 580 de los textos extraídos, todos ellos consultados, filtrados y recopilados de forma manual como parte de las actividades de este trabajo de tesis.

Para todas las pruebas que se realizaron con la herramienta HAAT las condiciones fueron las mismas: El mismo dispositivo de cómputo ejecutó la interfaz y el servidor de análisis, y dicho dispositivo se encontró conectado a internet por medio de una conexión inalámbrica vía WI-FI. Si bien esta configuración del equipo para las pruebas trajo consigo un problema: Un porcentaje de los textos no pudieron ser analizados por la herramienta. Otra posible causa de este porcentaje de error es que algunos textos contenían caracteres poco comunes, y ello pudo generar algún error dentro del proceso de análisis. Sin embargo, y de acuerdo con los resultados obtenidos en las pruebas, se ha concluido que el uso de una conexión inalámbrica en el dispositivo que aloja el servidor no es una configuración óptima. En la siguiente tabla se muestra la cantidad de textos que la herramienta pudo procesar para cada una de las tres pruebas que se ejecutaron con el corpus NVS2019, así como el porcentaje de error obtenido en cada prueba.

Tabla 10. Composición del corpus NVS2019 y porcentaje de error de procesamiento.

	Cantidad inicial	Textos procesados	% de éxito	% de error
Prueba 1				
Cantidad de textos	580	537	92.58%	7.42%
Negativos	580	537	92.58%	7.42%
Prueba 2				
Cantidad de textos	580	553	95.34%	4.66%
Negativos	580	553	95.34%	4.66%
Prueba 3				
Cantidad de textos	580	560	96.55%	3.45%
Negativos	580	560	96.55%	3.45%

En lo que se refiere a las características extraídas, el primer grupo de estas son descriptivas de la estructura del texto. En las tres pruebas se obtuvo el mismo resultado, mostrado en la Tabla 11:

Tabla 11: Características de la estructura del texto del corpus NVS2019

Estructura del texto				
	Total	Promedio	Menor	Mayor
Longitud (palabras)	112512	209.52	26	1366
Oraciones	3449	6.42	1	80
Palabras vacías	60777	113.18	13	719
Palabras únicas	34940	65.07	0	259
Riqueza léxica	457.355	0.85	0.381	1
Facilidad de lectura	-16493.6	-30.71	-1299.1	92.97

En cuanto a las características relacionadas con el análisis de sentimientos y de emociones, el análisis de resultados se realizó separando ambos conceptos. Para el análisis de sentimientos se obtuvo una precisión de 89.76% en la primera prueba, 90.24% en la segunda prueba y 89.64% en la tercera prueba. Los resultados obtenidos fueron muy satisfactorios, y eso se debe a que el corpus lingüístico utilizado es una herramienta de amplio uso en la comunidad de Procesamiento de Lenguaje Natural (PLN), por lo que es un conjunto de datos optimizado para su uso en el análisis de textos, y ello se ve reflejado en la precisión obtenida durante estas pruebas. En la Tabla 12 se muestran las características obtenidas en la segunda prueba, la que arrojó una precisión del 90.24%.

Tabla 12: Características de sentimiento extraídas del corpus NVS2019.

Análisis de sentimientos			
	Promedio	Menor	Mayor
Polaridad negativa de sentimiento	0.051	0.000	0.172
Polaridad neutral de sentimiento	0.943	0.828	1
Polaridad positiva de sentimiento	0.005	0.000	0.097
Núm. de palabras negativas	6.922	0.000	31
Núm. de palabras neutrales	103.098	19.000	387
Núm. de palabras positivas	3.286	0.000	21

Polaridad promedio de sentimiento	-0.625	-0.997	0.7074
--	--------	--------	--------

Positivos	23	Negativos	499
Neutrales	31	Precisión	90.24%

Los datos mostrados en la Tabla 12 se pueden resumir en la Figura 18 y en la visualización del comportamiento de la polaridad de sentimiento, es decir, de la composición del texto y la diferencia existente entre palabras positivas, neutrales y negativas. En este caso la herramienta realizó un análisis de cada una de las 34940 palabras únicas que forman el corpus y que forman los 553 textos que se analizaron en la segunda prueba y determinó, para cada texto, la polaridad con base en la cantidad de palabras positivas y negativas de cada texto.

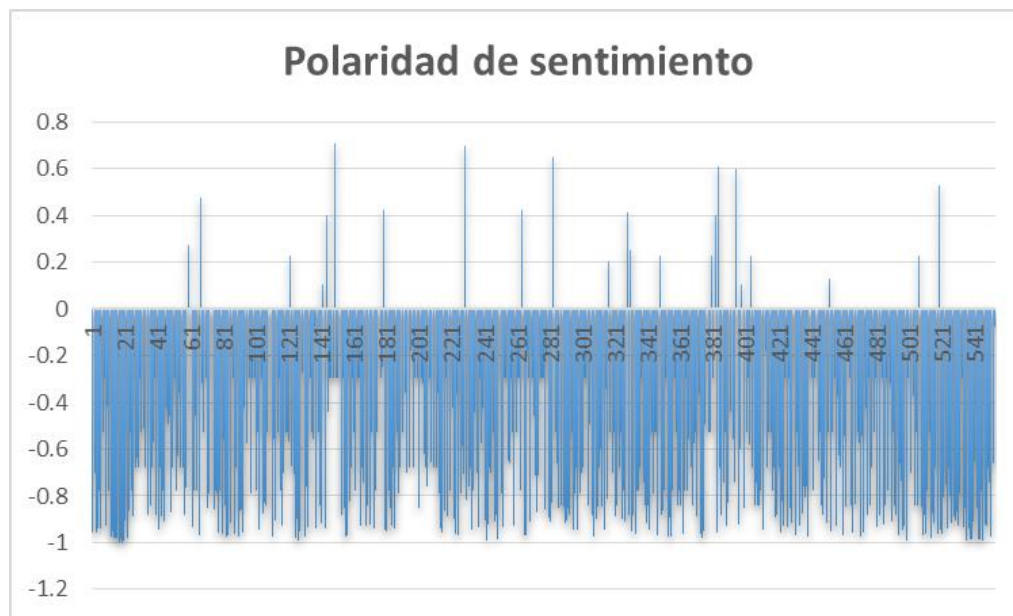


Figura 18. Puntaje de sentimiento de cada texto del corpus NVS2019

En lo referente a la tarea de análisis de emociones la precisión obtenida en la primera prueba fue del 9.87%, en la segunda prueba se obtuvo un 47.20% y en la tercera prueba 75%. Es evidente la gran diferencia de precisión entre cada una de las pruebas, esto se explica porque durante la primera y segunda prueba se estuvo utilizando la característica de polaridad emocional, una característica que es poco precisa, debido a que el corpus utilizado (TextBlob, 2021) está compuesto por textos cortos extraídos de redes sociales, principalmente de Twitter. Por lo tanto, su eficacia es más evidente en textos cortos, mientras que los corpus utilizados para el análisis de sentimientos fueron diseñados y

construidos para el análisis de textos de media y larga longitud. Sin embargo, la característica denominada subjetividad emocional es muy eficaz para determinar si un texto contiene un alto o un bajo porcentaje de palabras emocionales, independientemente de su longitud. Es justo esta característica de subjetividad emocional la que, junto con la polaridad de sentimiento, ocupan el rol de entradas en el módulo de clasificación emocional (ver Anexo 2).

En la Tabla 13 se muestra el resultado de la clasificación de emociones, así como el máximo puntaje obtenido de 75% en esta tarea. Si bien es posible mejorar la precisión, ya que a lo largo de las tres pruebas con este corpus se pudo mejorar, se debe considerar el tipo de texto que se analiza e, incluso, si esta característica es significativa en otros conjuntos de textos. Ya que el corpus NVS2019 está formado por textos escritos por mujeres víctimas de violencia, extraer esta característica es una buena forma de complementar el análisis de sentimientos, sin embargo, se sugiere que se contemple su utilidad para cada población de estudio.

Tabla 13: Resultados obtenidos del análisis de emociones en el corpus NVS2019

Emoción	Apariciones
Alegría	48
Desprecio	50
Ira	10
Miedo	16
Neutral	52
Satisfacción	3
Serenidad	9
Sorpresa	28
Tristeza	344
Positivos	88
Negativos	420
Neutrales	52
Precisión	75.00%

Viendo gráficamente las emociones que la herramienta encontró en los textos del corpus, y la clasificación que realizó, el resultado es el siguiente:

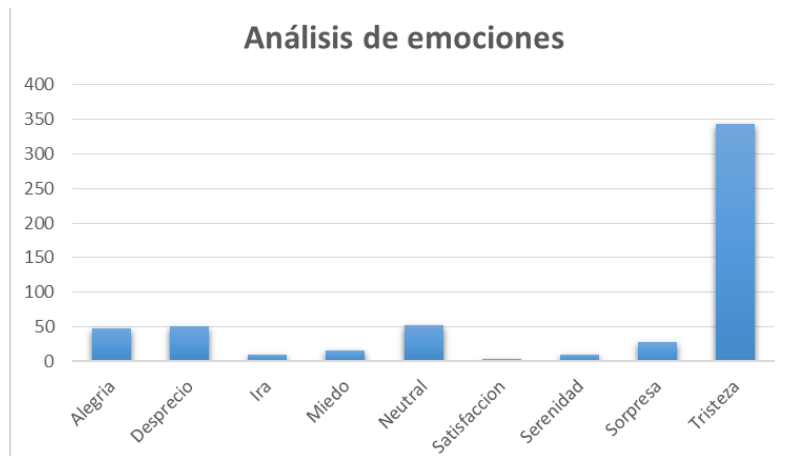


Figura 19. Etiquetado de emociones en el corpus NVS2019

En la Figura 19 se puede notar que la mayor parte de los textos fueron etiquetados como parte de alguna emoción negativa (Desprecio, Ira, Miedo y Tristeza), mientras que el resto se etiquetaron como neutral o positivo. Como se ha mencionado con anterioridad el lenguaje es construido por cada individuo a partir de las necesidades de comunicación que se tengan en un momento o contexto dado, y también el lenguaje solo puede transmitir aquella información contenida en su estructura (palabras, oraciones). El obtener una mayor precisión en la tarea de etiquetado emocional requeriría tener en cuenta estas problemáticas ya que, de lo contrario, se hace imposible obtener una precisión muy alta, y es que a pesar de que todos los individuos que escribieron en el corpus NVS2019 comparten la pertenencia a una población común, cada una de las personas participantes concibe el lenguaje desde su propia perspectiva y eso implica que le da un uso de acuerdo a sus necesidades, algo que inevitablemente impacta en el resultado obtenido por la herramienta HAAT, especialmente en lo que se refiere al análisis de sentimientos y emociones.

Comparativa de resultados Corpus NVS2019

En la siguiente tabla se muestran, a forma de comparación, los resultados en cuanto a la precisión obtenida en la tarea de análisis de sentimientos y emociones:

Tabla 14: Comparativa de precisión obtenida en las pruebas del corpus NVS2019

Prueba	Cantidad de textos	% de error	Precisión análisis de sentimientos	Precisión análisis de emociones
Prueba 1	580	7.42%	89.76%	9.87%
Prueba 2	580	4.66%	90.24%	30.38%
Prueba 3	580	3.45%	89.64%	75%

En promedio se obtuvo un porcentaje de error en el análisis de textos del 5.17%, mientras que la tarea de análisis de sentimientos tuvo el mejor resultado con un promedio de 89.88% y la precisión del análisis de emociones fue de 38.41% en promedio, aunque en este último caso no se tiene en cuenta que tras cada prueba se realizaron ajustes en el módulo encargado de esta tarea, lo que mejoró considerablemente el resultado hasta el máximo de 75% obtenido en la última prueba.

6.2 Pruebas con el Corpus (Keith, Fuentes & Meneses, 2017)

El conjunto de datos idóneo para las pruebas fue el corpus NVS2019, esto debido a que fue un conjunto de datos construido a partir de textos de una población concreta, mujeres víctimas de violencia. Sin embargo uno de los objetivos específicos de este trabajo de tesis hace referencia a la posibilidad de utilizar la herramienta HAAT para la extracción de características al realizar análisis de textos, no solo para poblaciones específicas, sino también para cualquier tipo de contenido textual a nivel de párrafo o superior. Para probar la capacidad de la herramienta al interactuar con textos con un origen o escritos bajo un contexto diferente se llevó a cabo una prueba con el corpus recopilado por (Keith, Fuentes & Meneses, 2017). Este conjunto de datos se encuentra compuesto por 192 textos positivos, 62 neutrales y 145 negativos.

Para tareas de clasificación se puede considerar un conjunto de datos desbalanceado, pero ya que la herramienta HAAT opera más bien dentro de la actividad de extracción de características y no propiamente de clasificación este desbalanceo no afectó el desempeño de las pruebas. En la Tabla 15 se muestra la cantidad de textos procesados con éxito para cada categoría de etiquetado, así como el porcentaje de error, recordando que este último aspecto se encuentra relacionado con el uso de una conexión inalámbrica de red para las pruebas lo que inevitablemente conllevó a breves y momentáneas interrupciones que afectaron la cantidad de textos que la herramienta pudo procesar, ya que al menos uno de los módulos requiere de una conexión a internet.

Tabla 15: Comparativa de precisión obtenida en las pruebas del corpus de revisión de artículos

	Cantidad inicial	Textos procesados	% de éxito	% de error
Positivos	192	187	97.39%	2.61%
Neutrales	62	61	98.38%	1.62%
Negativos	145	138	95.17%	4.83%

El corpus recopilado por (Keith, Fuentes & Meneses, 2017) es una recopilación de comentarios realizados por revisores académicos como parte del filtro aplicado para a cualquier trabajo de investigación que se desea publicar en una revista científica. Se trata

de textos en español y categorizados en textos positivos cuando el artículo fue aceptado para su publicación, neutrales cuando se dieron comentarios que no representaron una opinión negativa, y negativos cuando los comentarios realizaron implicaron el rechazo del trabajo o una crítica al mismo. Este corpus ya se encontraba etiquetado por lo que dichas etiquetas permitieron evaluar el desempeño de la herramienta HAAT.

En la siguiente tabla se muestran las características extraídas por la herramienta HAAT y referentes a la estructura sintáctica de los textos que componen el corpus. Se debe señalar que el análisis de cada categoría etiquetada (textos positivos, neutrales y negativos) se realizó de forma independiente, es decir, se realizó la separación de cada categoría etiquetada antes del análisis.

Tabla 16: Características relacionadas con la estructura sintáctica del corpus de revisión de artículos

Estructura del texto				
	Total	Promedio	Menor	Mayor
Textos positivos				
Longitud (palabras)	22192	122.61	7	686
Oraciones	1164	6.22	1	39
Palabras vacías	10373	55.47	1	321
Palabras únicas	8655	46.28	3	196
Riqueza léxica	157.559	0.84	0.426	1
Facilidad de lectura	3277.39	17.53	-256.04	109.04
Textos neutrales				
Longitud (palabras)	11190	189.66	11	654
Oraciones	584	9.57	1	37
Palabras vacías	5148	84.39	1	316
Palabras únicas	3688	60.46	6	158
Riqueza léxica	48.196	0.79	0.589	1
Facilidad de lectura	1032.9	16.93	-168.07	73.34
Textos negativos				
Longitud (palabras)	26655	200.41	16	1007
Oraciones	1380	10.00	1	40
Palabras vacías	12703	92.05	2	485
Palabras únicas	8673	62.85	8	200
Riqueza léxica	110.572	0.80	0.441	1
Facilidad de lectura	1180.69	8.56	-681.66	87.05

Con los resultados obtenidos de esta prueba se pudo constatar que la riqueza léxica aumento con respecto al resultado obtenido en la prueba con el corpus NVS2019.

Los textos del corpus de revisión de artículos fueron redactados por personas con un grado de estudios igual o superior al grado de licenciatura, además de que la redacción y el dominio de la lengua, principalmente en el manejo de un léxico más amplio, hacen que estos textos sean más complejos si se comparan con los textos extraídos de redes sociales, ya que estos últimos en su mayoría fueron redactados por personas con un dominio de la lengua promedio y en un formato de escritura libre, lo que por supuesto conllevó el desafío adicional de tratar con textos compuestos de caracteres poco comunes (emoticonos) o de faltas de ortografía o abreviaturas de uso general. Estas problemáticas se pudieron evitar en el caso de esta prueba, ya que por el propio nivel de dominio de los revisores de artículos científicos la mayor parte de los textos tenían una estructura gramatical impecable, aunque no por ello quiere decir que este corpus no contara con sus propios desafíos.

Si bien de las 30 variables que la herramienta HAAT puede extraer solo 12 de ellas están relacionadas con el análisis de sentimientos y etiquetado de emociones, son justo estas 12 variables las que mayor complejidad implican para poder representar correctamente la característica analizada. En la Tabla 17 se muestra una comparativa de la precisión obtenida al realizar análisis de sentimientos y el etiquetado de emociones.

Tabla 17: Precisión obtenida en el análisis de sentimientos y emociones con el corpus de revisión de artículos

Prueba	Cantidad de textos	% de error	Precisión análisis de sentimientos	Precisión análisis de emociones
Positivos	192	2.61%	11.22%	21.93%
Neutrales	62	1.62%	6.56%	18.03%
Negativos	145	4.83%	97.10%	76.81%

Como se puede observar en la Tabla 17 para la determinación de la polaridad emocional y el etiquetado de emociones la precisión fue de 97.10% y 76.81% respectivamente. Si se contrasta con los resultados obtenidos al analizar los textos pre etiquetados como neutrales la diferencia es enorme. Este resultado tan bajo en cuanto a la precisión obtenida se explica por dos cuestiones: En primer lugar se debe a que a pesar de que los textos fueron etiquetados originalmente como neutrales dentro del contenido de cada texto se puede encontrar un conjunto de palabras, elementos o patrones asociados ya sea a un contexto positivo o uno negativo. Por lo tanto, la herramienta HAAT no ha podido determinar la idea general del texto, una idea que tiende a lo neutral, y en vez de ello analizó los componentes del texto por lo que el resultado final fue una precisión prácticamente nula. En segundo lugar, la complejidad en la redacción de los textos afectó en este caso la precisión en las actividades de análisis de sentimientos y emociones, y ello se debe a que hay una correlación inversa entre la complejidad y la precisión.

En teoría, mientras menor sea la complejidad del texto mayor va a ser la precisión en el análisis de sentimientos y emociones, aunque en la práctica esto depende del perfil del individuo que realizó la redacción. Por lo tanto, las variables de sentimiento y emoción son las que sufren de una mayor dependencia a la propia redacción del texto. Un descubrimiento arrojado por esta prueba fue la identificación de una correlación inversa entre la característica “polaridad emocional” y la característica “polaridad de sentimiento”, algo que se puede apreciar al comparar la Figura 20 y la Figura 21.

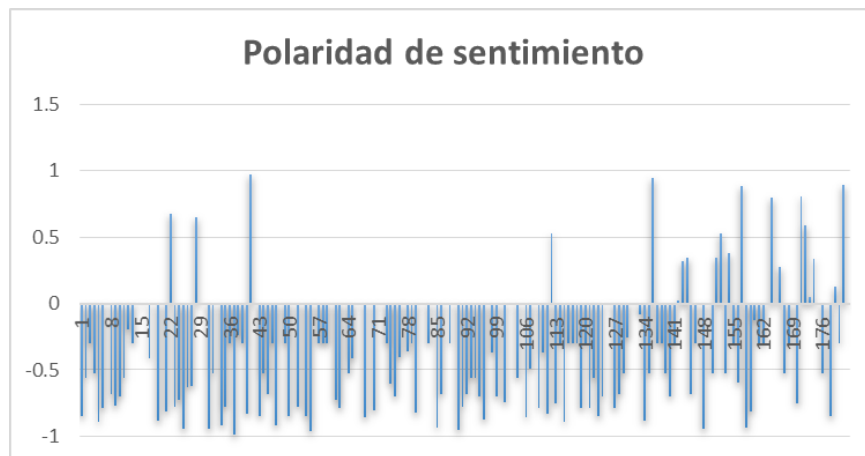


Figura 20. Polaridad de sentimiento con textos positivos corpus de revisión de artículos

En el caso del subconjunto de textos positivos se pudo apreciar un comportamiento con tendencia negativa en cuanto a la polaridad de sentimiento, mientras que un comportamiento con tendencia positiva de la polaridad emocional. Se debe tener en cuenta que las características de sentimiento se extrajeron a partir de un corpus construido con textos de mediana – larga extensión, y cuya estructura permitió obtener buenos resultados en la prueba con el corpus NVS2019.

Para la extracción de características emocionales el corpus utilizado fue diferente, recordando que se trató de un corpus de verbos y adjetivos etiquetados, este etiquetado se realizó utilizando textos cortos extraídos de redes sociales. Por lo tanto, la estructura de cada corpus demostró que la sensibilidad que proporciona cada conjunto de datos en la tarea de análisis de sentimientos y emociones varía de acuerdo con la estructura y redacción del texto.

Lo anterior explica porque, a pesar de tratarse del mismo sub conjunto de datos el resultado de las variables de sentimiento fue tan diferente del resultado de las variables emocionales, ya que simplemente cada corpus utilizado en dichas tareas de análisis es un

conjunto con sus propias características, etiquetado o incluso redacción, algo que impacta en el resultado final tal y como se muestra en la siguiente Figura, en donde se puede apreciar como la herramienta pudo detectar el contexto emocional positivo (un análisis que en el caso de la HAAT se hace con base en los verbos y adjetivos que aparecen en el texto), mientras que a nivel sentimiento el resultado no fue satisfactorio (debido a que para las tareas de sentimientos el corpus esta etiquetado a nivel de oración, y no gramaticalmente como en el caso emocional).

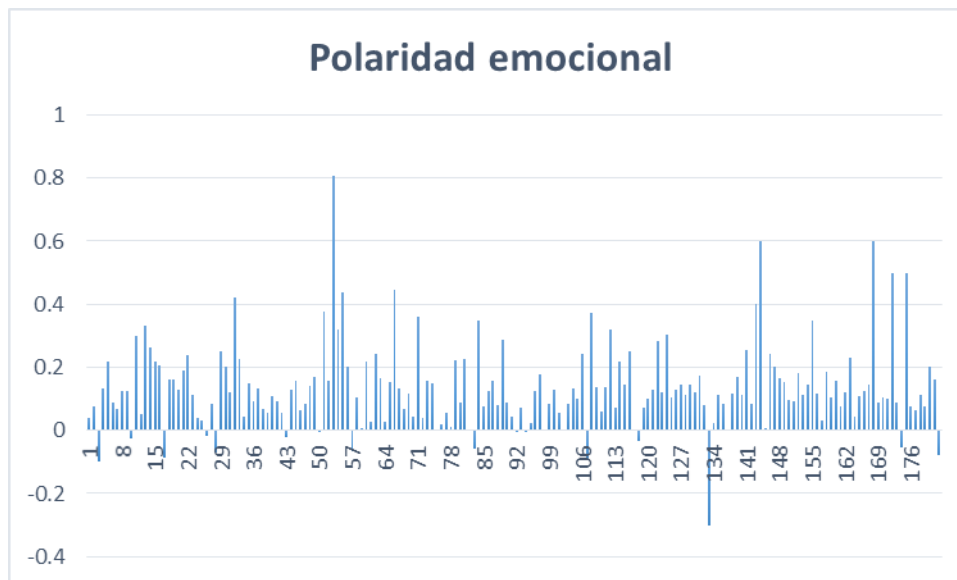


Figura 21. Polaridad emocional con textos positivos corpus de revisión de artículos

6.3 Pruebas con el corpus de la Prueba piloto de aplicación de la terapia EEA

La tercera prueba correspondió ya no al ajuste de la propia herramienta sino a su aplicación en el análisis de textos obtenidos a través de la aplicación de una terapia de escritura narrativa, también conocida como Terapia de Escritura Emocional Autorreflexiva (EEA). Para la aplicación de esta terapia se desarrollaron dos interfaces de aplicación, siendo la primera de ellas un sitio web a través del cual se sometió a una prueba piloto a 12 participantes voluntarias, todas ellas mujeres mayores de edad y víctimas de algún evento de violencia. De manera remota estas participantes realizaron las actividades que componen la terapia de escritura narrativa, y cuya duración fue definida en 5 semanas. De estas 5 semanas la primera correspondía al llenado de una serie de cuestionarios para recolectar datos sobre el perfil individual de cada participante, así mismo se aplicó un cuestionario para la detección de ansiedad así como un cuestionario para la detección de Trastorno de Estrés Postraumático (TEPT). En la Figura 22 se muestra el menú principal

de la interfaz web de la aplicación de la terapia de escritura, así como las 16 actividades comprendidas dentro de la aplicación de la terapia.

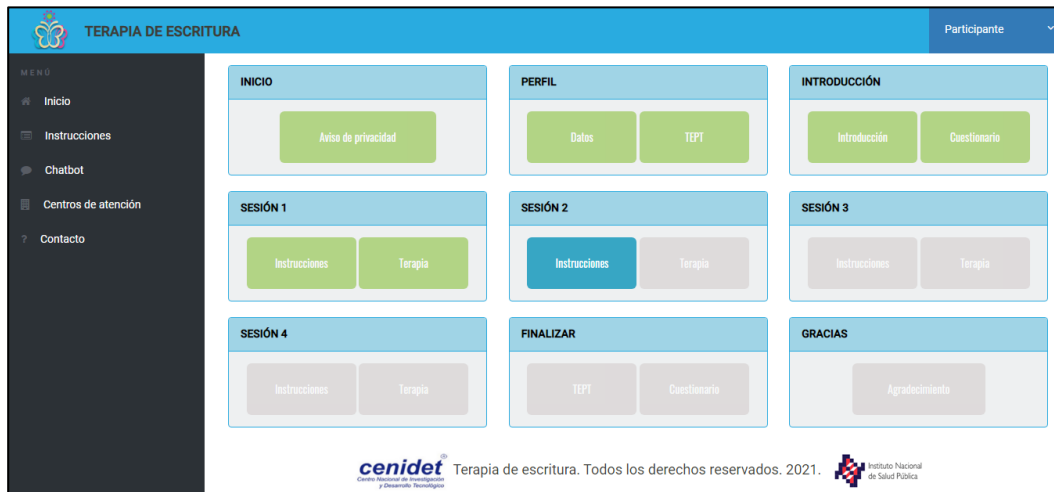


Figura 22. Interfaz web de aplicación de la terapia de escritura

De manera complementaria también se desarrolló una aplicación móvil para el sistema operativo Android con la finalidad de proporcionar otra opción para realizar la terapia, considerando que algunas de las mujeres participantes no contaban con un equipo de cómputo a su disposición y sí con un dispositivo móvil. A pesar de que el sitio web y la aplicación móvil fueron utilizados de manera inicial para la implementación de la prueba piloto, su diseño y desarrollo se pensó para su uso de forma masiva, lo que implicaría su liberación para cualquier mujer que cumpla con los requisitos de aplicación de la terapia, o incluso para otras poblaciones de estudio (ver Anexo 3).

En la Figura 22 se puede observar el estado final de algunas de las interfaces que componen la aplicación móvil. Haciendo uso de esta herramienta y de la interfaz web se pudieron obtener 32 textos de las 12 participantes, aclarando que algunas de ellas no terminaron todas las actividades de escritura, en la siguiente tabla se muestra el porcentaje de éxito en el procesamiento de los textos:

Tabla 18: Porcentaje de error en el análisis de los textos obtenidos de la prueba piloto de la terapia de escritura

	Cantidad inicial	Textos procesados	% de éxito	% de error
Cantidad de textos	32	32	100%	0%

En la tabla 18 se puede apreciar que los 32 textos obtenidos de la prueba piloto fueron procesados por la herramienta HAAT de forma satisfactoria. Por supuesto es un resultado que ha variado si se contrasta con el resultado obtenido en este aspecto en las pruebas anteriores, pero ha permitido comprobar que el porcentaje de error no es producto del funcionamiento de la herramienta sino más bien de la dependencia hacia recursos ajenos a la misma, como la conexión a internet o el propio dispositivo que aloja el servidor de análisis (la herramienta HAAT trabaja principalmente con la CPU del dispositivo).

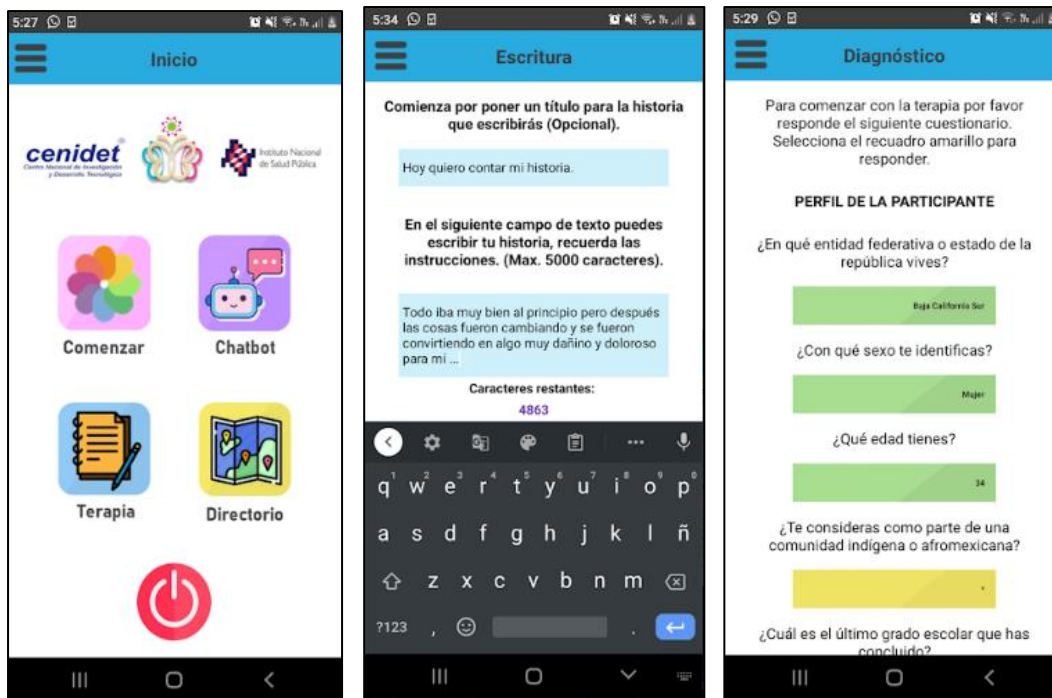


Figura 23. Aplicación móvil para la aplicación de la terapia de escritura

En cuanto a las características extraídas de esos textos se puede ver en la Tabla 19 la composición del conjunto de datos. En promedio la cantidad de palabras por texto fue de 135, una cantidad que si bien es mayor que un texto corto (tweet) se puede considerar como una longitud de texto baja.

Tabla 19: Estructura de los textos obtenidos de la prueba de la terapia de escritura

Estructura del texto				
	Total	Promedio	Menor	Mayor
Longitud (palabras)	4308	134.63	1	402
Oraciones	65	2.03	1	7
Palabras vacías	2214	69.19	0	219
Palabras únicas	1475	46.09	1	116

Riqueza léxica	27.603	0.86	0.699	1
Facilidad de lectura	-606.94	-18.97	-287.5	68.1

Al tratarse de pocos textos el grado de precisión de las actividades relacionadas con el análisis de sentimientos y de emociones no ha llegado al mismo nivel que en las pruebas anteriores. Para el caso de la precisión en el análisis de sentimientos se obtuvo un 66.67% y en la tarea de etiquetado emocional se obtuvo un 68.75%. A continuación, se muestran las características relacionadas con el análisis de sentimientos y el resultado obtenido:

Tabla 20: Características de sentimiento de los textos obtenidos de la terapia de escritura

Análisis de sentimientos			
	Promedio	Menor	Mayor
Polaridad negativa de sentimiento	0.058	0.000	0.174
Polaridad neutral de sentimiento	0.926	0.752	1
Polaridad positiva de sentimiento	0.016	0.000	0.074
Núm. de palabras negativas	4.750	0.000	14
Núm. de palabras neutrales	71.719	1.000	164
Núm. de palabras positivas	3.563	0.000	13
Polaridad promedio de sentimiento	-0.395	-0.978	0.7783

Positivos	4	Negativos	14
Neutrales	3	Precisión	66.67%
Errores	11		

La tabla anterior se puede resumir gráficamente en la Figura 24, en donde se nota cómo la mayor parte de los 32 textos obtenidos durante la prueba piloto de la terapia de escritura son calificados con un puntaje negativo, mientras que en menor medida hay algunos neutrales y otros positivos. Para determinar la precisión se realizó una revisión manual de cada texto y se etiquetó. De esta manera se pudo contrastar el resultado obtenido por medio del análisis de la herramienta con el etiquetado realizado de forma manual antes del inicio de las pruebas con la herramienta HAAT.

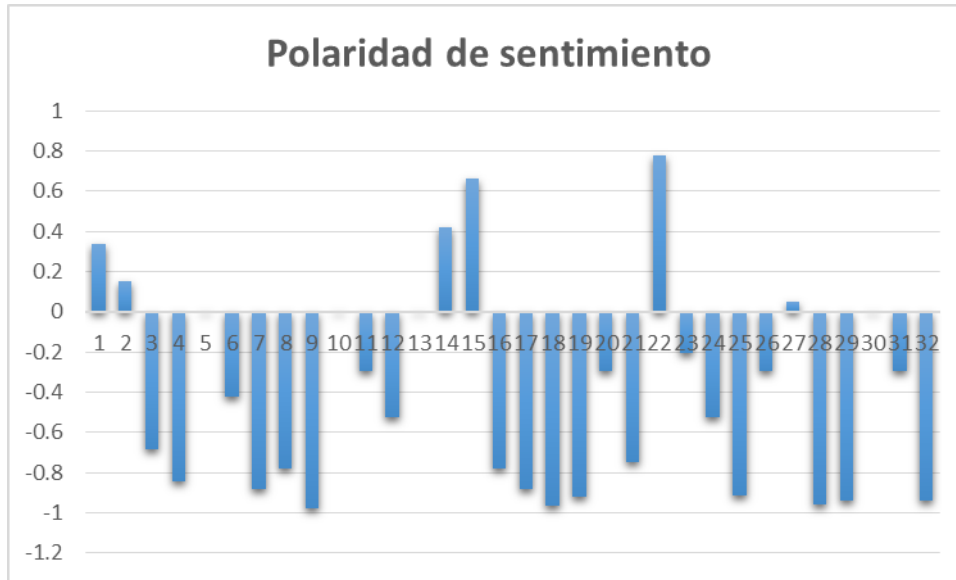


Figura 24. Polaridad de sentimiento de los textos obtenidos de la terapia de escritura

En la actividad de etiquetado de las emociones la precisión obtenida fue del 68.75%, resultando en el etiquetado de los diferentes textos de acuerdo con las categorías que se muestran en la tabla 21.

Tabla 21: Etiquetado de emociones de los textos obtenidos de la terapia de escritura

Emoción	Apariciones
Alegría	4
Desprecio	1
Ira	2
Miedo	2
Neutral	3
Satisfacción	0
Serenidad	1
Sorpresa	7
Tristeza	12

Positivos	12
Negativos	17
Neutrales	3
Errores	10
Precisión	68.75%

Visualizando los resultados mostrados en la Tabla 21 se puede observar en la siguiente Figura un etiquetado de los textos en donde 12 de los textos fueron reconocidos como textos con un contexto emocional de tristeza, una categoría que se considera como negativa en este trabajo de investigación. También se puede notar que en el conjunto de textos se encontraron textos etiquetados emocionalmente como neutrales o positivos y ello se debió a que dentro del conjunto de textos se incluyeron todos los textos generados durante la prueba piloto de la terapia de escritura, tanto los más negativos, los cuales se asocian a las primeras actividades de escritura terapéutica de la terapia, como los positivos que son asociados a las últimas actividades de la terapia de escritura.

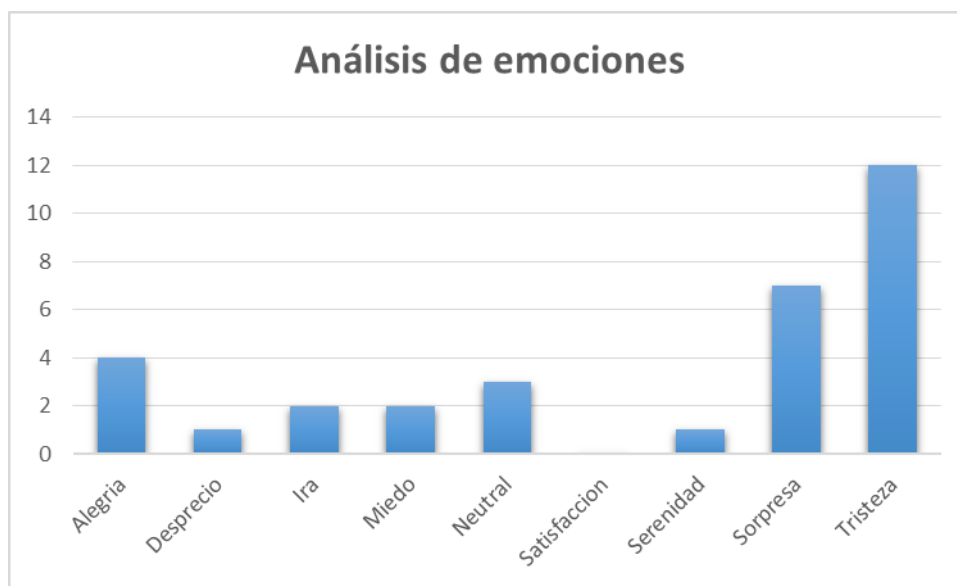


Figura 25. Etiquetado de emociones de los textos obtenidos de la terapia de escritura

Un dato diferenciador entre esta prueba piloto y el resto de las pruebas que se realizaron con la herramienta HAAT es que en este caso también se recopilaban cuestionarios con información adicional, como la evaluación del nivel de ansiedad y TEPT antes y después de la terapia, así como un cuestionario de violencia y uno más con la finalidad de obtener información sobre el perfil de cada persona participante, como su edad, entidad federativa de residencia, entre otras. Si bien dichos datos permiten recopilar hasta un total de poco más de 200 variables de análisis, ya que para esta prueba no se recopiló una cantidad considerable de datos, ya que la cantidad de participantes fue mínima, se espera que como trabajo futuro se puedan obtener más textos y aplicar un análisis multivalente a las más de 200 variables útiles, lo que en el futuro permitirá analizar a detalle el comportamiento de las variables y su impacto en el resultado de la terapia.

6.4 Precisión de los módulos de análisis de textos

Considerando que no todos los módulos que componen la herramienta HAAT, la cual se desarrolló con la finalidad de dar cumplimiento al objetivo general propuesto en este trabajo de tesis, tienen como finalidad el análisis de la información o la extracción de características, en la Tabla 22 se muestra un resumen con la precisión obtenida por cada uno de los módulos de análisis de la herramienta (ver Anexo 4).

Tabla 22: Resumen de la precisión obtenida por cada uno de los módulos de análisis de la herramienta HAAT

Módulo	Precisión	Librería	Observaciones
Estructura del texto	No aplica	Nltk, textstat	Se trata de medidas estadísticas
Análisis de sentimientos	89.64%	VaderSentiment + TextBlob	Puntaje obtenido por la HAAT
Etiquetado de emociones (Modelo de polaridad – subjetividad)	75%	TextBlob + Clasificador difuso	Puntaje obtenido por la HAAT
Clasificación de la categoría gramatical	97%	Spacy	Modelo es_core_news_sm pre entrenado
Reconocimiento de entidades nombradas	89%	Spacy	Modelo es_core_news_sm pre entrenado

El módulo de extracción de características estructurales del texto no cuenta con un valor de precisión debido a que las actividades que realizó se enfocaron en la minería de datos más que en clasificación o agrupamiento, por lo tanto, la precisión es un dato que no aplica en el caso de este módulo. Para los módulos de análisis de sentimientos y etiquetado de emociones la precisión se determinó promediando los resultados obtenidos en cada actividad durante las pruebas que se realizaron a la herramienta HAAT.

Este promedio arrojó una precisión del 89.64% en la actividad de análisis de sentimiento, específicamente en la asignación del puntaje de polaridad de sentimiento. En tanto que en el etiquetado de emociones se consiguió un promedio de 75% de precisión. Finalmente, el módulo de análisis gramatical obtuvo un 97% y el módulo de reconocimiento de entidades nombradas un 89%. Cabe señalar que estos últimos dos módulos utilizaron un modelo pre entrenado para realizar las correspondientes actividades.

Capítulo VI

Resultados y conclusiones

En el presente capítulo se muestran los resultados obtenidos al desarrollar el presente trabajo de investigación, así como proporcionar información relevante que podría ser utilizada como base para la realización de futuras investigaciones.

6.1 Resultados obtenidos

A continuación se hace una recapitulación de los resultados obtenidos durante las pruebas realizadas a la herramienta que se desarrolló con la finalidad de dar cumplimiento a los objetivos planteados en este trabajo de investigación.

1. El desarrollo de interfaces para llevar a cabo la aplicación de la terapia de escritura narrativa, tanto en su versión de página web como en su versión de aplicación móvil para sistema operativo Android, ha resultado en la obtención de una herramienta que sustituye el medio de aplicación tradicional de la terapia de escritura (con lápiz y papel) y que permiten amplificar el alcance de la aplicación de una intervención de este tipo.

2. De los experimentos llevados a cabo con la herramienta que se desarrolló se pueden concluir los siguientes supuestos:

a) Entre mayor sea la longitud del texto escrito se obtendrá una mayor precisión en los módulos de análisis de textos.

b) El módulo de análisis de sentimientos es más preciso cuando los textos que se analizan son de mediana y larga extensión en términos de palabras, en contraste, el módulo de etiquetado emocional depende de la cantidad de verbos y adjetivos que hay en el texto.

c) Los módulos de análisis de emociones y sentimientos tienden a tener un comportamiento diferente dependiendo de la forma en la que se encuentran redactados textos que se analizan, algo que no sucede en el resto de los módulos que extraen otras características.

d) El uso combinado de la polaridad de sentimiento con el grado de subjetividad emocional han demostrado poseer la capacidad de poder combinarse para poder etiquetar los textos en alguna de las posibles nueve emociones contempladas.

3. De la prueba piloto realizada aplicando la terapia de escritura a una muestra poblacional compuesta por mujeres víctimas de violencia, se puede concluir lo siguiente:

a) Aplicar la terapia en un período de tiempo de cinco semanas, tal y como fue en el caso de la prueba piloto, resulta en un considerable porcentaje de ausentismo por parte de las participantes.

b) El proceso de aplicación de la terapia resultó satisfactorio al realizarse utilizando las interfaces computacionales, teniendo el mismo desempeño que se obtendría al aplicar la terapia de forma convencional.

c) La herramienta de análisis HAAT cumple con el propósito de permitir el análisis de los textos obtenidos a partir de la aplicación de la terapia de escritura narrativa sin presentar complicaciones en el proceso.

4. La separación de las interfaces de aplicación de la terapia y de la herramienta de análisis de textos, permite que ambas funcionalidades se puedan utilizar ya sea de forma individual o de forma combinada. En el caso de las interfaces de aplicación de la terapia es posible adaptar las utilizar las para cualquier población de estudio que se requiera. En el caso de la herramienta HAAT uso no se limita al análisis de un tipo específico de textos, siendo posible implementar su uso en el análisis de cualquier otro contenido textual.

5. Se ha podido corroborar que esta herramienta aporta valor al realizar un proceso de extracción de características y etiquetado automático de elementos.

6.2 Conclusiones

Es difícil trabajar con textos porque lo que un individuo escribe refleja su propia realidad, una realidad dada por su propia perspectiva, por el entorno que lo rodea y por el propio contexto en el que se hace uso del lenguaje. Desde la perspectiva de las ciencias computacionales, y en específico del procesamiento de lenguaje natural, se han implementado soluciones para automatizar procesos de análisis e interpretación de textos aunque pocas veces este trabajo se ha dirigido hacia el estudio de un tratamiento psicológico alternativo como lo es la terapia de escritura emocional autoreflexiva, también conocida como terapia de escritura narrativa.

La terapia de escritura narrativa es una técnica que se ha estudiado desde la década de los 80 con resultados positivos para la mejora de la salud psicológica y física de los individuos involucrados en los estudios que se han realizado a lo largo de los años. Es una terapia cuya simplicidad y bajo costo permiten su uso para la exploración de diversas poblaciones de estudio de forma masiva. Aunque en el caso de México su estudio y el conocimiento de sus ventajas, así como de las posibilidades de aplicación se encuentra todavía en una etapa temprana de desarrollo.

En este trabajo de tesis no sólo se ha propuesto una herramienta que permitan la exploración de los datos obtenidos a partir de la aplicación será fiel escritura narrativa, sino que este diseño se ha visto materializado en forma de un programa de software cuya capacidad radica en extraer las características del lenguaje más importantes encontradas en los textos terapéuticos que se obtienen de la terapia, dicha herramienta se ha denominado Herramienta para el Análisis Automático de Textos” o HAAT.

Los desafíos abordados en este proyecto de tesis se enlistan a continuación:

Proponer un paradigma de aplicación de la terapia de escritura adaptado al estudio de una población compuesta por mujeres víctimas de violencia.

- Al comienzo del desarrollo del proyecto fue necesario realizar un análisis de las diferentes metodologías para la aplicación de la terapia de escritura. El resultado fue la generación de un proceso compuesto por 16 actividades repartidas en un período de cinco semanas. Sin embargo también se generó una configuración de este paradigma de aplicación en donde se sustituyen las cinco semanas por cinco días, rango de tiempo a través del cual se realizan las actividades correspondientes a la terapia.

- Las actividades de escritura de la terapia se incorporaron actividades instructiva es para brindar la información necesaria a las participantes, así como un cuestionario para obtener datos sobre el perfil de la participante, la evaluación del nivel de ansiedad y

trastornos de estrés postraumático en el momento previo al inicio de las actividades terapéuticas, así como medio de evaluación al finalizar las mismas.

Construir una interfaz para la aplicación remota de la terapia de escritura

- Se desarrolló un sitio web informativo a través del cual se pudo guiar a las participantes hacia las interfaces de aplicación de la terapia desarrolladas.

- Se desarrolló una plataforma web para el registro de las participantes voluntarias en la prueba piloto, así como para permitir la aplicación de la terapia desde cualquier lugar siempre y cuando se dispusiera de un equipo de cómputo con conexión a Internet.

- Como complemento se desarrolló una versión del sistema de aplicación de la terapia en forma de una aplicación móvil para el sistema operativo Android.

Desarrollo y prueba de una herramienta de análisis automático de textos para la extracción de características lingüísticas relevantes

- Se construyó un programa de software para automatizar la extracción de características lingüísticas de los textos. La versatilidad de dicho programa informático permite su uso en el análisis de textos escritos por diferentes poblaciones de estudio o con diferentes niveles de complejidad, uso del léxico, tipo de redacción, entre otros. Asimismo esta herramienta tiene la capacidad de analizar cuatro textos generados durante la terapia de escritura narrativa y extraer de forma simultánea las correspondientes características.

Este trabajo de investigación ha demostrado que es posible extraer un conjunto de características lingüísticas de un texto de forma automática y sin la intervención de un ser humano en el proceso de extracción. Adicionalmente esta herramienta puede aplicarse para el análisis de prácticamente cualquier tipo de texto o corpus lingüístico, por lo que permite aumentar el conocimiento que se puede extraer al aplicar técnicas de procesamiento de lenguaje natural para el análisis de textos.

La combinación de técnicas computacionales con tratamientos psicológicos ha demostrado ser acertada para poder desarrollar investigaciones relacionadas con problemáticas sociales en las cuales es importante interactuar con una gran cantidad de participantes, así como también permite la automatización de procesos que de otro modo tendrían que llevarse a cabo de forma manual lo que inevitablemente conlleva un mayor consumo de recursos, sobre todo en lo que se refiere al tiempo necesario para la aplicación de las investigaciones y el análisis de los resultados.

6.3 Trabajos futuros

En esta sección se describen las posibles actividades derivadas del trabajo de investigación.

Al analizar los resultados obtenidos a partir de las pruebas que se realizaron a la herramienta de análisis de textos HAAT, se considera que los siguientes trabajos futuros pueden complementar el desarrollo y la investigación que se llevaron a cabo en este proyecto de tesis:

- Sustituir el etiquetado de las emociones del clasificador difuso que se utiliza actualmente por otro tipo de técnicas que permitan obtener una mejora en la precisión del etiquetado de la emoción.

- Permitir el acceso a las interfaces de aplicación de la terapia a personas de la población en general que cumplan con los requisitos de inclusión para la participación en la terapia, es decir que se trate de mujeres mayores de edad que han sido víctimas de algún tipo de violencia y que quieren participar de forma voluntaria en la aplicación de una terapia de escritura narrativa.

- Agregar características adicionales para aumentar el número de variables de salida que se obtienen a partir del análisis, como el modelado de tópicos, análisis de N gramas, análisis de contexto, detección de falacias, entre otros.

- Migrar la herramienta de análisis de textos de un entorno local a un entorno en la nube con la finalidad de que se pueda utilizar a través de una conexión a Internet y en tiempo real.

- Realizar un análisis multivariante a los resultados obtenidos a través del análisis con la finalidad de detectar patrones de comportamiento y correlación entre las variables extraídas, así como para determinar cuáles de estas variables son más importantes y cuales son menos representativas del fenómeno que se estudia.

Construir un algoritmo de clasificación que tome como entradas las características que se extraen utilizando la herramienta HAAT y que permita clasificar a las personas participantes de acuerdo al cambio que se haya dado en las variables lingüísticas a través de cada una de las sesiones de la terapia de escritura narrativa.

Aportaciones y productos académicos

Derivado del trabajo de investigación realizado se obtuvieron los siguientes productos académicos:

Publicaciones

Título: Método para la extracción de tweets y su aplicación en comentarios en torno al coronavirus

Jornada de Ciencia y Tecnología Aplicada. Vol. 3, Núm. 1, Página. 117, Enero - Junio 2020.

URL: https://jcyta.cenidet.tecnm.mx/revistas/jcyta/04-Revista_JCyTA_Vol-3-Num-1_Ene-Jun_2020.pdf

Título: Lógica difusa para el análisis subjetivo de la cotización bursátil en la Bolsa Mexicana de Valores

Jornada de Ciencia y Tecnología Aplicada. Vol. 3, Núm. 2, Página 9, Julio - Diciembre 2020

URL: https://jcyta.cenidet.tecnm.mx/revistas/jcyta/05-Revista_JCyTA_Vol-3-Num-2_Jul-Dic_2020.pdf

Referencias

- Acosta, Olga. Aguilar, César & Araya, Fabiola. (2018). QuarryMeaning: Una aplicación para el modelado de tópicos enfocado a documentos en español. *Procesamiento del Lenguaje Natural*, 61, 197-200.
- Al-Mosaiwi, Mohammed & Johnstone, Tom. (2018). Linguistic markers of moderate and absolute natural language. *Personality and Individual Differences*, 134, 119-124.
- Almashraee, Mohammed & Monett, Dagmar & Paschke, Adrian. (2016). Emotion Level Sentiment Analysis: The Affective Opinion Evaluation. *Semantic Sentiment Analysis - ESWC2016*.
- Azim, M., & Bhuiyan, M. (2018). Text to Emotion Extraction Using Supervised Machine Learning Techniques. *TELKOMNIKA (Telecommunication Computing Electronics And Control)*, 16(3), 1394. doi: 10.12928/telkomnika.v16i3.8387
- B. Hirsh, Jacob. B. Peterson, Jordan. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, 43(3), 524-527.
- Baca Gómez, Yolanda Raquel. (2014). Desarrollo de un Servicio Web para Determinar la Polaridad de Textos de Redes Sociales en Español. Tesis de maestría en ciencias de la computación. Centro Nacional de Investigación y Desarrollo Tecnológico.
- Bahena, Leticia Aponte (2013). Identificación de características de depresión en usuarios de una red social. Tesis de maestría en ciencias de la computación. Centro Nacional de Investigación y Desarrollo Tecnológico.
- Baikie, K. A., & Wilhelm, K. (2005). Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment*, 11(5), 338–346. <http://doi.org/10.1192/apt.11.5.338>
- Bing, L. (2010). Libro de procesamiento de Lenguaje Natural (2nd ed., p. Capítulo Análisis de Sentimiento y Subjetividad). N. Indurkha and F. J. Damerau.
- CEAV, Comisión Ejecutiva de Atención a Víctimas. (2016). Diagnóstico cuantitativo sobre la atención de la violencia sexual en México. (1st ed.). México.

- Cortez, Mg. Augusto, Vega, Mg. Hugo, Pariona Jaime. (2009). Procesamiento de lenguaje natural. *Revista de ingeniería de sistemas e informática*. 6(2).
- Christudas, B. (2019). MySQL. *Practical Microservices Architectural Patterns*, 877-884. doi: 10.1007/978-1-4842-4501-9_27
- Damerau, F., & Indurkha, N. (2010). *Manual de procesamiento del lenguaje natural* (p. Capítulo Análisis de sentimientos y subjetividad.). Boca Raton, Estados Unidos: Taylor & Francis.
- Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A., Gelbukh, A., & Zhou, Q. (2016). Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. *Cognitive Computation*, 8(4), 757-771. doi: 10.1007/s12559-016-9415-7
- Dimsdale, E. Joel. Xin, Yu. Kleinman Arthur, Patel Vikram, E. Narrow, William, J. Sirovadka, Paul. A. Regier, Darrel. (2019). *Somatic presentations of mental disorders. Refining the Research Agenda for DSM-V*. (1st ed.). Elsevier Masson.
- Durón Figueroa, R. (2019). Adaptación de la Lista Checable de Trastorno por Estrés Postraumático para DSM-5 en Población Mexicana. *Acta De Investigación Psicológica*, 9(1), 26-36. <https://doi.org/10.22201/fpsi.20074719e.2019.1.03>
- Dworkin, E., & Schumacher, J. (2016). Preventing Posttraumatic Stress Related to Sexual Assault Through Early Intervention: A Systematic Review. *Trauma, Violence, & Abuse*, 19(4), 459-472. doi: 10.1177/1524838016669518
- Ekman, P., & Oster, H. (1981). Expresiones faciales de la emoción. *Estudios De Psicología*, 2(7), 115-144. doi: 10.1080/02109395.1981.10821273
- Ferrer, C., & R. Delgado, A. (2018). Revisión sistemática de las medidas del Trastorno por Estrés Postraumático. *Cuadernos De Neuropsicología (España)*, 12(1).
- Galindo Vázquez, O., Rojas Castillo, E., Meneses García, A., Aguilar Ponce, J., Álvarez Avitia, M., & Alvarado Aguilar, S. (2015). Propiedades psicométricas del inventario de ansiedad de beck (BAI) en pacientes con cáncer. *Psicooncología*, 12(1). https://doi.org/10.5209/rev_psic.2015.v12.n1.48903
- García, Felipe E., Vega Rojas, Nicole, Briones Araya, Francisco, & Bulnes Gallegos, Yasmín. (2018). Rumiación, crecimiento y sintomatología postraumática en

personas que han vivido experiencias altamente estresantes. *Avances en Psicología Latinoamericana*, 36(3), 443-457. <https://dx.doi.org/10.12804/revistas.urosario.edu.co/apl/a.4983>

Glass, Oliver. Dreusicke, Mark. Evans, John. Bechard, Elizabeth & Wolever, Ruth Q. (2019). Expressive writing to improve resilience to trauma: A clinical feasibility trial. *Complementary Therapies in Clinical Practice*, 34(1), 240-246.

Guerra, Cristóbal. Plaza, Hugo & Farkas, Chamarrita. (2017). Perfiles de estrés postraumático en adolescentes abusados sexualmente. *Psicoperspectivas*, 16(1), 67-79. <https://dx.doi.org/10.5027/psicoperspectivas-vol17-issue1-fulltext-879>

He, Q., Veldkamp, B., Glas, C., & de Vries, T. (2016). Automated Assessment of Patients' Self-Narratives for Posttraumatic Stress Disorder Screening Using Natural Language Processing and Text Mining. *Assessment*, 24(2), 157-172. doi: 10.1177/1073191115602551

Hu, M., & Liu, Bing. (2004). Minería y resumen de reseñas de clientes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. Seattle, Washington, USA.

Hueso Ibañez, L. (2016). *Administración de Sistemas gestores de Bases de Datos* (2nd ed., pp. 10-12). Ediciones de la U.

Hutto, C., & Gilbert, E. (2014). VADER: un modelo parsimonioso basado en reglas para el análisis de sentimientos del texto de las redes sociales. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI.

Isidoros Perikos, Ioannis Hatzilygeroudis. (2016). Recognizing emotions in text using ensemble of classifiers. *Engineering Applications of Artificial Intelligence*, 51, 191-201. <https://doi.org/10.1016/j.engappai.2016.01.012>.

Kanger, N., & Bathla, G. (2017). Recognizing Emotion in Text using Neural Network and Fuzzy Logic. *Indian Journal Of Science And Technology*, 10(12), 1-6. doi: 10.17485/ijst/2017/v10i12/100526

Keith, B., Fuentes, E., & Meneses, C. (2017). Un enfoque híbrido para el análisis de sentimientos aplicado a las revisiones de artículos. Consultado el 22/May/2021, de <https://sentic.net/wisdom2017fuentes.pdf>

Kincaid, J., Aagard, J., O'Hara, J., & Cottrell, L. (1981). *Sistema de edición de legibilidad*

por computadora. *IEEE Transactions On Professional Communication*, PC-24(1), 38 - 42. doi: 10.1109/TPC.1981.6447821

Kincaid, J., Fishburne, R., Rogers, R., & Chissom, B. (1975). Derivación de nuevas fórmulas de legibilidad (índice de legibilidad automatizado, conteo de niebla y fórmula de facilidad de lectura) para el personal alistado de la Marina. Estación Aeronaval de Memphis, TS: Jefe de Entrenamiento Técnico Naval.

Léxicos de sentimiento y emoción. (2021). Consultado el 22/May/2021, de <http://saifmohammad.com/WebPages/lexicons.html>

Llisterri, Joaquim & Torruella Casañas, Juan. (1999). Diseño de corpus textuales y orales. *Nuevas tecnologías en los estudios filológicos*, 45-81. (ISBN 84-89790-41-8).

Liu, Bing., Hu, M., & Cheng, J. (2005). *Observador de opiniones: análisis y comparación de opiniones en la Web*. Seattle, Chiba, Japan.

Lozano-Oyola, J., Gómez de Terreros Guardiola, M., Avilés Carvajal, I., & Sepúlveda García de la Torre, A. (2017). Sintomatología del trastorno de estrés postraumático en una muestra de mujeres víctimas de violencia sexual. *Cuadernos De Medicina Forense*, 23(3-4), 82-91.

McClure, G. (1987). Fórmulas de legibilidad: ¿útiles o inútiles?. *IEEE Transactions On Professional Communication*, PC-30(1), 12 - 15.

Meiselman, H. (2021). *Medición de la emoción* (2nd ed., p. Capítulo Análisis de sentimientos: detección automática de valencia, emociones y otros estados afectivos del texto). Woodhead Publishing.

Natural Language Toolkit. (2021). Consultado el 22/May/2021, de <https://www.nltk.org/>

Nasheeda, Aishath. Binti Abdullah, Haslinda. Krauss, Steven Eric & Ahmed, Nobaya Binti. (2019). Transforming Transcripts Into Stories: A Multimethod Approach to Narrative Analysis. *International Journal of Qualitative Methods*, 18, 1-9.

Parodi, Giovanni. (2008). Lingüística de corpus: Una introducción al ámbito. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 93-119. (<https://dx.doi.org/10.4067/S0718-48832008000100006>).

Pennebaker, J. (2004). Theories, therapies, and taxpayers. On the complexities of the expressive writing paradigm. *Clinical Psychology: Science and Practice*, 11,

138–142.

Pennebaker, J. (1997). Writing About Emotional Experiences as a Therapeutic Process. *Psychological Science*, 8(3), 162-166. doi: 10.1111/j.1467-9280.1997.tb00403.x

Pennebaker, J. (2004). Theories, therapies, and taxpayers. On the complexities of the expressive writing paradigm. *Clinical Psychology: Science and Practice*, 11, 138–142.

Ponce Cruz, P. (2010). *Inteligencia artificial con aplicaciones a la ingeniería* (1st ed.). AlfaOmega.

Proyecto Cuéntalo (2019). Recuperado de <http://proyectocuentalo.org/>. Consultado el 02/02/2020.

Reyes, José A., Montes, Azucena, González, Juan G., & Pinto, David E.. (2013). Clasificación de roles semánticos usando características sintácticas, semánticas y contextuales. *Computación y Sistemas*, 17(2), 263-272.

Rodriguez Hernandez, Gabriela, & Juárez Lugo, Carlos Saúl, & Ponce de León, María del Consuelo (2011). La culturalización de los afectos: Emociones y sentimientos que dan significado a los actos de protesta colectiva. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, 45(2),193-201.[fecha de Consulta 14 de Junio de 2021]. ISSN: 0034-9690. Disponible en: <https://www.redalyc.org/articulo.oa?id=28422741010>

Sidorov, G., Galicia, S., & Camacho, A. (2019). Construcción de un corpus marcado con emociones para el análisis de sentimientos en twitter en español. *Revista Escritos BUAP*, 1(1).

Silver, B. (2012). *BPMN method and style*. Aptos, CA: Cody-Cassidy Press.

Skovgaard Lyby, Marlene. Mehlsen, Mimi. Bonde Jensen, Anders. Howard Bovbjerg, Dana. S. Philipsen, Johanne & Wallot, Sebastian. (2019). Use of Recurrence Quantification Analysis to Examine Associations Between Changes in Text Structure Across an Expressive Writing Intervention and Reductions in Distress Symptoms in Women With Breast Cancer. *Frontiers in Applied Mathematics and Statistics*, 5.

Sobrinho Sande, José Carlos. (2018). *Análisis de sentimientos en Twitter*. Tesis de maestría

en ingeniería informática. Universidad Oberta de Catalunya.

Spacy NLP. (2021). Consultado el 22/May/2021, de <https://spacy.io/models/es>

Suster, Simon. Tulkens, Stéphan & Daelemans, Walter. (2017). QuarryMeaning: A Short Review of Ethical Challenges in Clinical Natural Language Processing. arXiv, 1703.10090.

Sutherland, Katie. (2016). Writing on Disability and Illness, and the Case for Self-reflexive Storytelling. *Global Media Journal*, 10.

TextBlob. (2021). Consultado el 22/May/2021, de <https://github.com/sloria/TextBlob>

Textstat. (2021). Consultado el 22/May/2021, de <https://github.com/shivam5992/textstat>

Trigueiro, Tatiane Ferreira, Silva, Marcelo Henrique da, Merighi, Miriam Aparecida Barbosa, Oliveira, Deíse Moura de, & Jesus, Maria Cristina Pinto de. (2017). Psychological suffering in the daily lives of women who have experienced sexual violence: a phenomenological study. *Escola Anna Nery*, 21(3), e20160282. Epub June 01, 2017. <https://dx.doi.org/10.1590/2177-9465-ean-2016-0282>

VaderSentiment. (2021). Consultado el 22/May/2021, de <https://github.com/cjhutto/vaderSentiment>

Vicente, Marta. Barros, Cristina. Peregrino, Fernando S., Agulló, Francisco & Lloret, Elena. (2015). La generación de lenguaje natural: análisis del estado actual. *Computación y Sistemas*, 19(4), 721-756. <https://dx.doi.org/10.13053/CyS-19-4-2196>

World Health Organization. (1999). Trastornos mentales y de comportamiento de la décima revisión de la clasificación internacional de las enfermedades (CIE-10). (1st ed.). Vol. 2.

World Health Organization (2018). Recuperado de https://www.who.int/topics/gender_based_violence/es/. Consultado el 05/11/2019.

World Health Organization. (2020). “Health care for women subjected to intimate partner violence or sexual violence”, Recuperado de https://apps.who.int/iris/bitstream/handle/10665/136101/WHO_RHR_14.26_eng.pdf;sequence=1, Consultado el 22/05/2020.

Anexos

Anexo 1: Clasificador difuso para el etiquetado de emociones

a) Planteamiento de la solución

Para el diseño del clasificador se tomaron como variables de entrada dos características.

A continuación se describe cada una de ellas:

- Polaridad de sentimiento: Se refiere al puntaje obtenido al realizar análisis de sentimientos de un texto utilizando el léxico etiquetado incluido en la librería VADER. El resultado es la suma de las puntuaciones de valencia obtenidas para cada palabra en el texto, normalizando el puntaje para encontrarse en un rango de entre -1 (extremo negativo máximo) y +1 (extremo positivo máximo)
- Subjetividad emocional: Valor obtenido al acceder al léxico de subjetividad para adjetivos y verbos en inglés contenido dentro de la librería textblob. El puntaje final se obtiene al buscar las cadenas de texto que tengan mayor similaridad al corpus etiquetado y sumando el conjunto de puntajes para obtener un resultado único. La puntuación de subjetividad es un valor flotante dentro del rango [0.0, 1.0] donde 0.0 es muy objetivo y 1.0 es muy subjetivo.

b) Estructura del clasificador difuso

Se consideran dos variables de entrada y una de salida: Polaridad de sentimiento (valor numérico en el intervalo [-1, 1]) y subjetividad emocional (valor numérico en el intervalo [0.0, 1.0]) como entradas. Emoción (miedo, ira, sorpresa, tristeza, neutral, alegría, desprecio, serenidad, satisfacción) como salida.

c) Definición de los conjuntos difusos

- Se definen 3 conjuntos difusos para polaridad de sentimiento (Entrada).
- Se definen 3 conjuntos difusos para subjetividad emocional (Entrada).
- Se definen 9 conjuntos difusos para emoción (Salida).

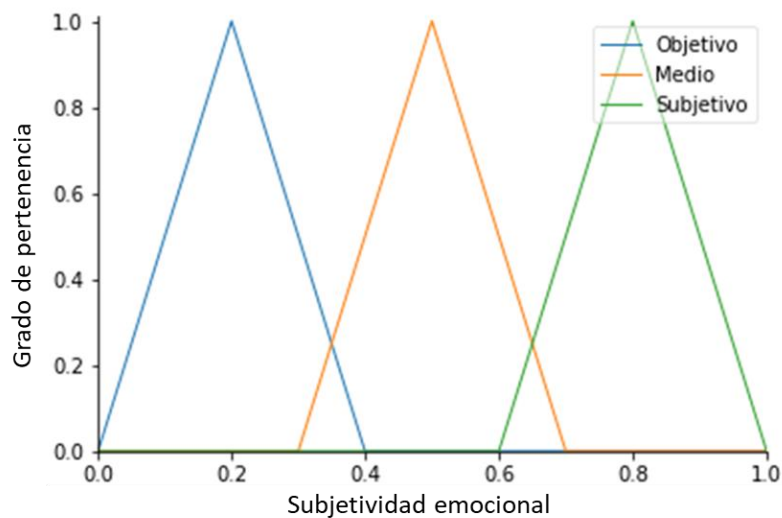
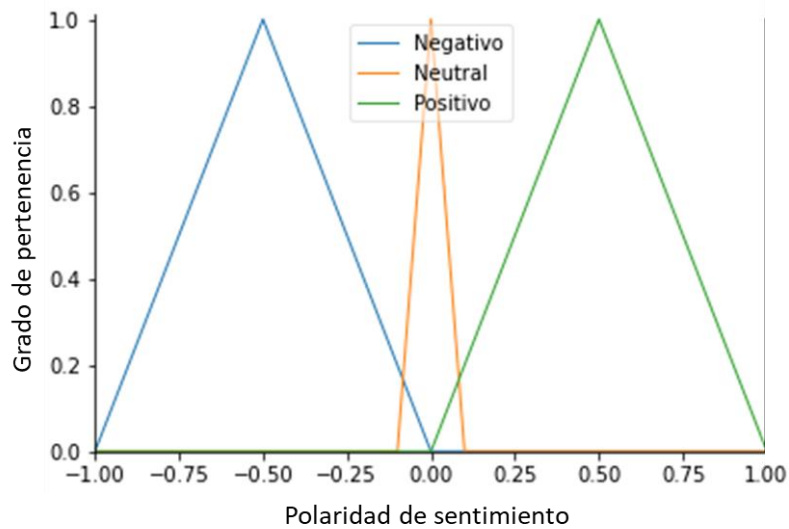
d) Definición de las funciones de pertenencia

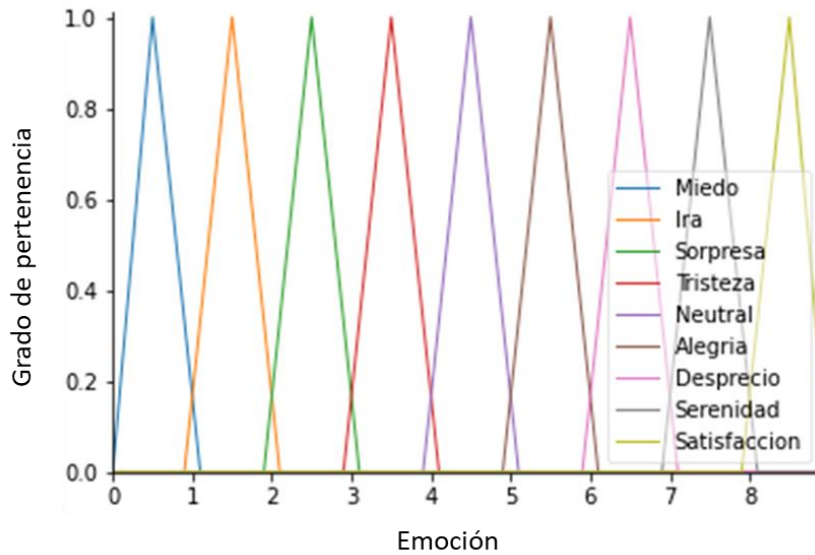
- Polaridad de sentimiento: Se consideran tres conjuntos difusos representados por funciones triangulares: Negativo (rango entre [-1, 0]), Neutral (rango entre [-0.1, 0.1]) y Positivo (rango entre [0, 1]).

- Subjetividad emocional: Se consideran tres conjuntos difusos representados por funciones triangulares: Objetivo (rango entre [0, 0.4]), Medio (rango entre [0.3, 0.7]) y Subjetivo (rango entre [0.6, 1]).

- Emoción: Se consideran 9 conjuntos difusos representados por funciones triangulares: Miedo (rango entre [0, 1.1]), Ira (rango entre [0.9, 2.1]), Sorpresa (rango entre [1.9, 3.1]), Tristeza (rango entre [2.9, 4.1]), Neutral (rango entre [3.9, 5.1]), Alegría (rango entre [4.9, 6.1]), Desprecio (rango entre [5.9, 7.1]), Serenidad (rango entre [6.9, 8.1]), Satisfacción (rango entre [7.9, 9]).

De manera gráfica se muestran las funciones de pertenencia a cada conjunto difuso a continuación:



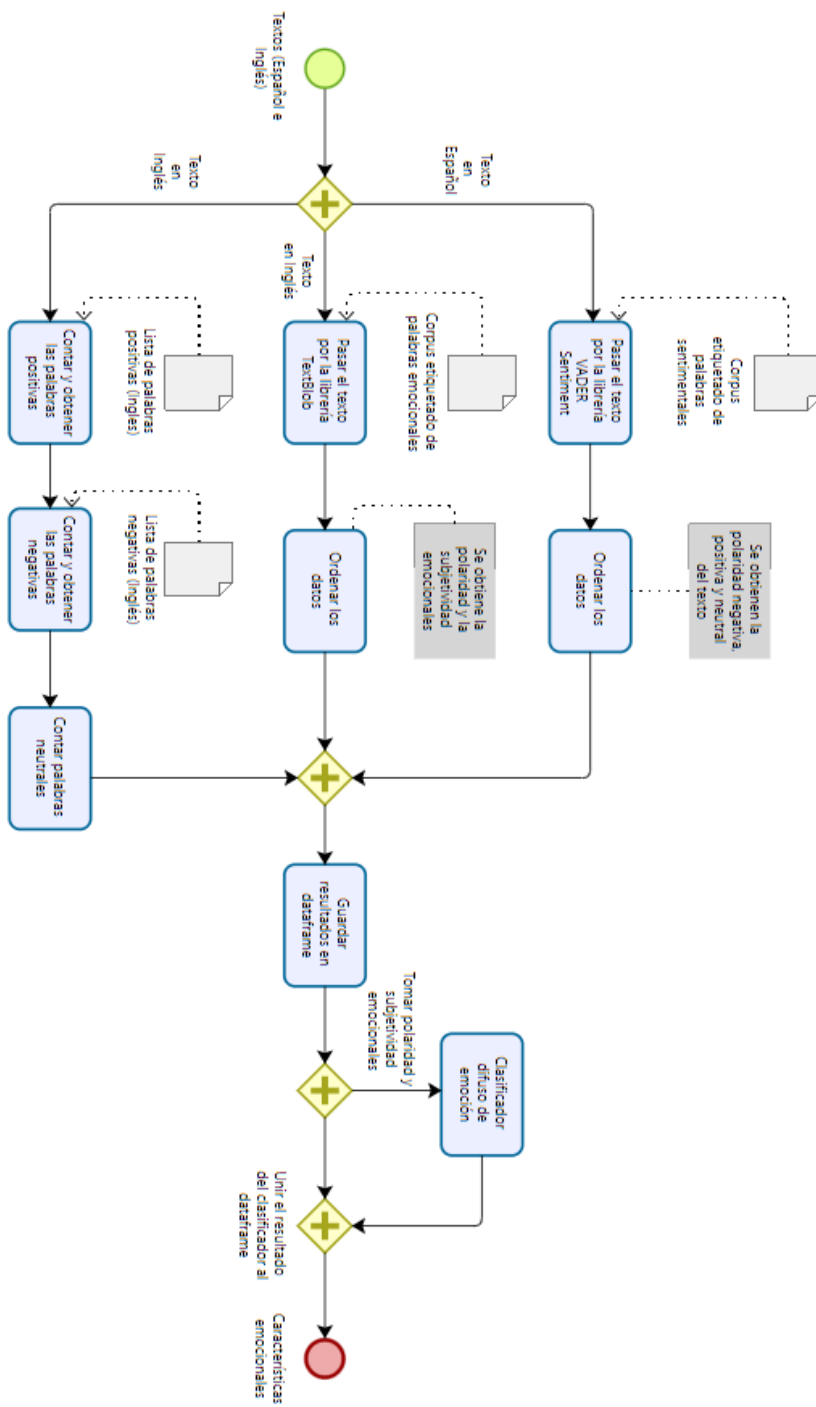


e) Diseño de reglas

Se trata de un sistema de inferencia difusa basado en reglas de tipo mamdani. Se incorporan 9 reglas, cada una de las cuales tiene como salida una de las posibles 9 emociones (miedo, ira, sorpresa, tristeza, neutral, alegría, desprecio, serenidad, satisfacción). A continuación se menciona el conjunto de reglas incorporadas en el clasificador:

- Regla 1: Nivel ['Negativo'] & Subjetividad ['Subjetivo'] = Emoción ['Miedo']
- Regla 2: Nivel ['Neutral'] & Subjetividad ['Subjetivo'] = Emoción ['Ira']
- Regla 3: Nivel ['Positivo'] & Subjetividad ['Subjetivo'] = Emoción ['Sorpresa']
- Regla 4: Nivel ['Negativo'] & Subjetividad ['Medio'] = Emoción ['Tristeza']
- Regla 5: Nivel ['Neutral'] & Subjetividad ['Medio'] = Emoción ['Neutral']
- Regla 6: Nivel ['Positivo'] & Subjetividad ['Medio'] = Emoción ['Alegría']
- Regla 7: Nivel ['Negativo'] & Subjetividad ['Objetivo'] = Emoción ['Desprecio']
- Regla 8: Nivel ['Neutral'] & Subjetividad ['Objetivo'] = Emoción ['Serenidad']
- Regla 9: Nivel ['Positivo'] & Subjetividad ['Objetivo'] = Emoción ['Satisfacción']

Anexo 2: Proceso de análisis de emociones y sentimientos

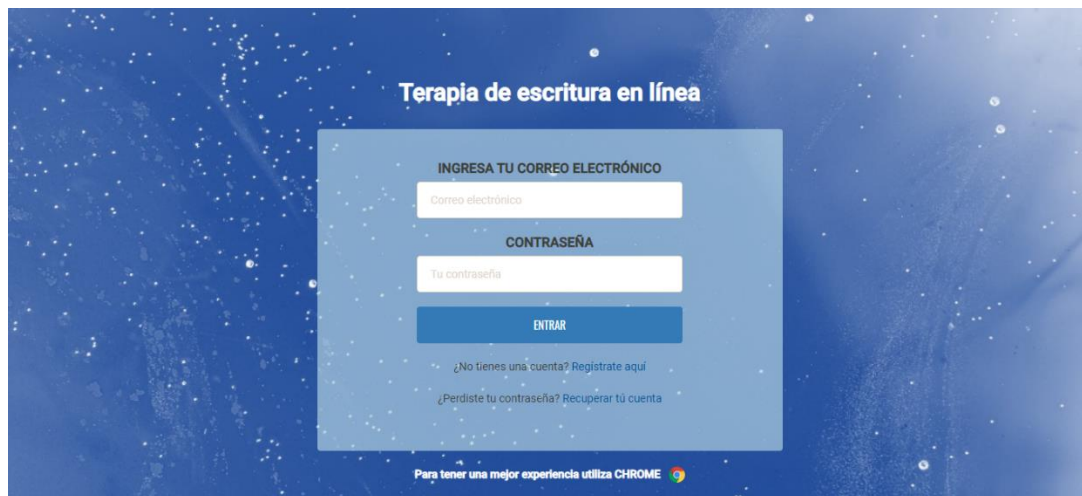


Anexo 3: Interfaces de aplicación de la terapia

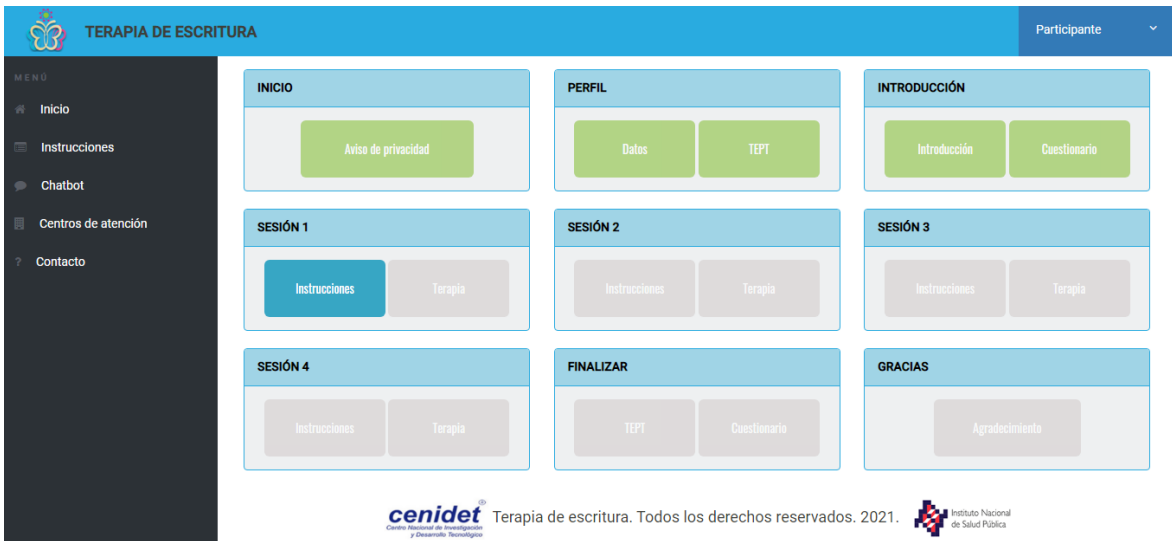
Como parte del desarrollo de este trabajo de tesis se desarrollaron interfaces para realizar la aplicación de la terapia de escritura narrativa, también llamada escritura Emocional Autorreflexiva (EEA). A continuación se detallan partes importante de las dos interfaces de aplicación:

a) Aplicación web

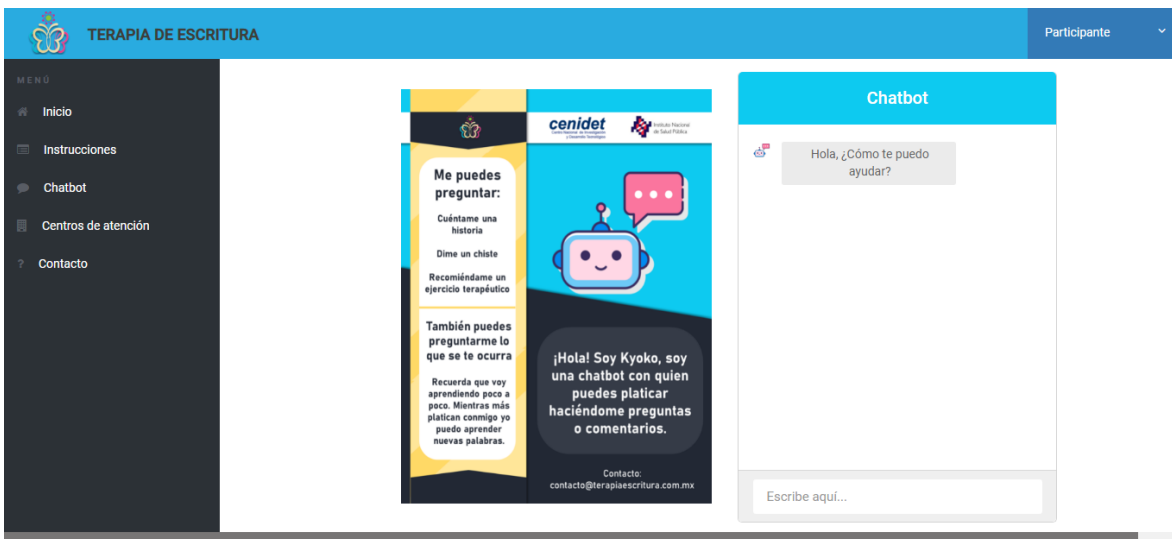
Con la finalidad de proveer un mecanismo de acceso seguro a la aplicación web para realizar la terapia de escritura se incorporó como interfaz inicial la pantalla de inicio de sesión en donde se solicita el correo electrónico y la contraseña del usuario para permitir su ingreso a la página de inicio de la terapia. De no contar con el registro correspondiente se proporciona al usuario la opción de registro, así como una opción para la recuperación de su cuenta en caso de perder la contraseña de acceso.



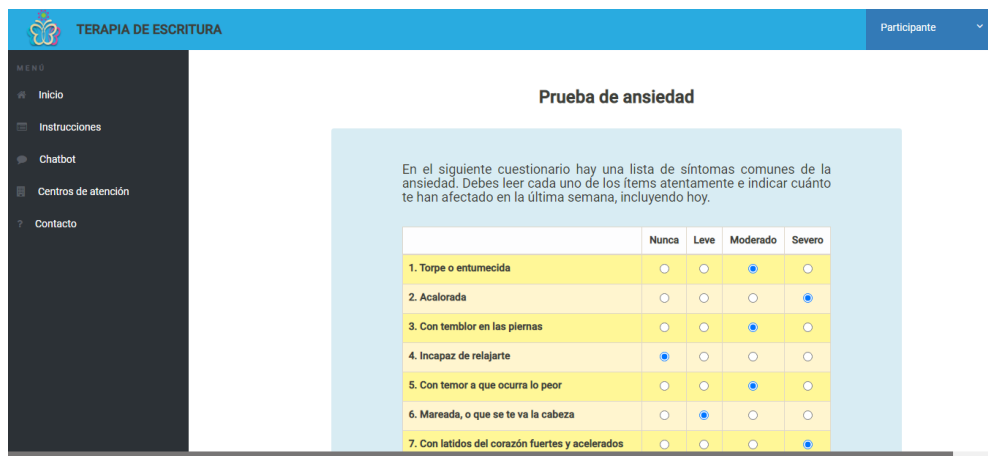
Al validarse en la información de inicio de sesión se muestra al usuario la pantalla de inicio a través de la cual se puede navegar ya sea en las distintas actividades que componen la terapia, y que se muestran en forma de rectángulos cuyo color varía dependiendo de la disponibilidad de la actividad, o también pudiendo seleccionar alguna de las cinco posibles opciones que se muestran en el menú lateral izquierdo, y que corresponden a la ventana de inicio, mostrar las instrucciones de uso de la plataforma de la terapia así como el procedimiento que el usuario participante debe seguir, interactuar con el chatbot incluido en la herramienta, visualizar en forma de mapa o a través de una lista el conjunto de centros de atención disponibles para brindar asesoramiento a mujeres víctimas de violencia, o visualizar los datos de contacto para poder proporcionar un comentario, sugerencia o para poder expresar sus dudas con relación a la plataforma de aplicación de la terapia.



El objetivo principal de esta interfaz es la aplicación de la terapia de escritura proporcionando un medio accesible a los usuarios interesados para realizar el procedimiento a través de Internet. Sin embargo, durante el período de realización de la prueba piloto en el cual se utilizó una pequeña muestra de la población objetivo se pudo constatar que la plataforma requería de un elemento adicional para mantener la atención de los usuarios participantes al menos durante el período de aplicación de la terapia, es decir, durante los cuatro o cinco días en los cuales se debe realizar la terapia. Para proporcionar este elemento de retención de los usuarios se incorporó un chatbot con opciones simples de conversación y con la única finalidad de proporcionar una funcionalidad adicional para mantener la atención de los participantes durante el tiempo necesario para que concluyan las actividades de la terapia.



En cuanto a las actividades que componen la terapia, se trata de 16 módulos que el usuario debe completar para concluir con este proceso. De estos módulos cinco de ellos corresponden a cuestionarios que debe de responder el usuario, lo que incluye un cuestionario para conocer el perfil del participante, un cuestionario para detectar la violencia de pareja, un par de cuestionarios para evaluar el nivel de ansiedad así como otro par de cuestionarios para la detección de síntomas correspondientes al trastorno de estrés postraumático.



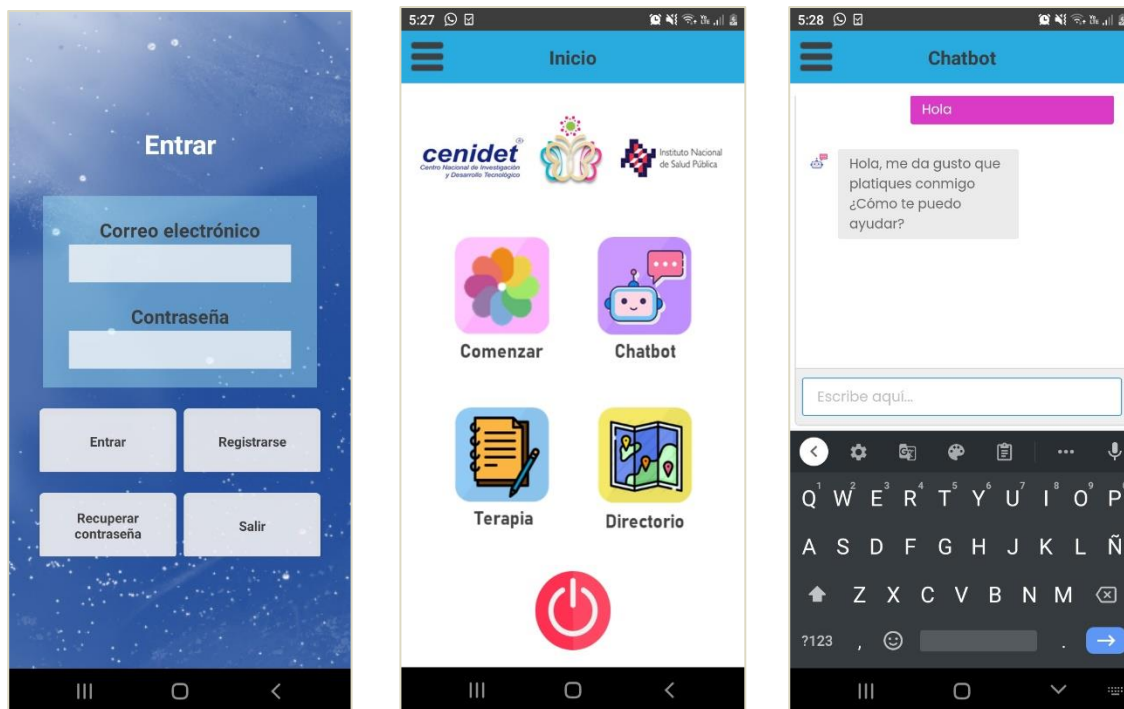
	Nunca	Leve	Moderado	Severo
1. Torpe o entumecida	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2. Acalorada	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
3. Con temblor en las piernas	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
4. Incapaz de relajarte	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Con temor a que ocurra lo peor	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
6. Mareada, o que se te va la cabeza	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Con latidos del corazón fuertes y acelerados	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Aparte de las actividades en forma de cuestionario, se incluyeron actividades de lectura a través de las cuales se brinda al usuario una serie de instrucciones o indicaciones acerca del proceso para llevar a cabo la terapia de escritura, siendo las restantes actividades correspondientes a los ejercicios terapéuticos de escritura (cuatro sesiones de escritura). Antes del inicio de cada sesión de escritura se le brinda al usuario la información necesaria para poder realizar el ejercicio, posteriormente se le muestra la interfaz a través de la cual realizará este ejercicio y en donde se le solicita que ingrese un título para el escrito que redactara, así como solicitar que escriban forma de texto libre la narrativa que quiere compartir y que hace referencia a la situación de violencia que ha experimentado.



b) Aplicación móvil

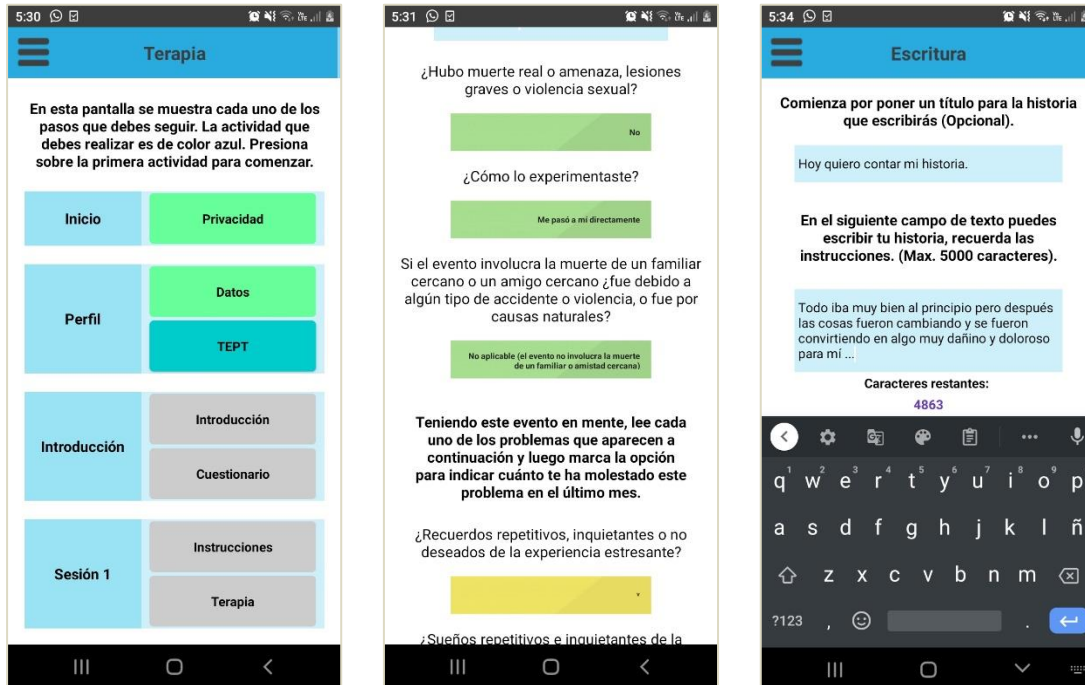
Como complemento a la interfaz en forma de páginas huérfanas se decidió construir una aplicación móvil nativa para el sistema operativo Android con la finalidad de poder abarcar un nicho mayor de posibles participantes. Esta decisión se basó en los resultados obtenidos durante la aplicación de la prueba piloto, en donde se pudo constatar que algunas de las mujeres participantes en el estudio tenían un mínimo dominio en el manejo de un equipo de cómputo, pero en la mayoría de los casos contaban con los conocimientos básicos sobre el uso de un dispositivo móvil.



A pesar de ser una aplicación móvil nativa el diseño de las interfaces debió de adaptarse de tal manera que fuese lo más parecida posible a la interfaz desarrollada para su uso a través de un navegador web. Por tal motivo el uso de colores, elementos gráficos así como funcionalidades es prácticamente en su totalidad el mismo que la versión web. Sin embargo uno de los cambios más portantes fue la inclusión de un menú principal simplificado, en el cual se incluyen las cuatro principales opciones para interactuar con la aplicación.

Un punto importante que se tomó en cuenta es que cualquier usuario puede realizar la terapia de escritura descargándose directamente la aplicación y sin tener que consultar la información proporcionada a través de la página web que se generó para el proyecto. Lo anterior significa que la aplicación móvil cuenta con toda la información y medios necesarios para poder explicar al usuario final el objetivo del proyecto la funcionalidad y

los pasos necesarios para llevar a cabo la realización de la terapia de escritura. Otra cuestión que se consideró es que a pesar de ser una aplicación que se instala en el dispositivo móvil esta depende de una conexión a Internet para su correcto funcionamiento, ya que se encuentra en constante interacción con una base de datos alojada en un servidor web conectado internet.



Al igual que en la interfaz web, los datos que se generan durante la interacción del usuario con las interfaces móviles se almacena en una base de datos remota. Este proceso de almacenar los datos es de vital importancia para el trabajo de investigación que se ha estado desarrollando, esto se debe a que estas interfaces permiten la recolección de datos útiles para su procesamiento con la herramienta de análisis de textos (HAAT) que se desarrolló en este trabajo de investigación.

Anexo 4: Cálculo de la precisión de la herramienta

A pesar de que el objetivo de la herramienta HAAT ha sido el de servir como plataforma para la extracción de características y no con fines de clasificación, dentro de la estructura del software se encuentran cinco módulos principales de extracción de características: Estructura del texto, Análisis de sentimientos, Etiquetado de emociones (Modelo de polaridad – subjetividad), Clasificación de la categoría gramatical y Reconocimiento de entidades nombradas. De estos cinco módulos principales el de Estructura del texto se encarga de obtener características mayormente estadísticas del texto, por lo tanto, no se puede evaluar su precisión, así como otras métricas.

En cuanto a los dos últimos módulos, el de análisis gramatical y de reconocimiento de entidades nombradas se utilizó en ambos casos el modelo pre entrenado `es_core_news_sm`, siendo implementado a través de la librería Spacy. Este modelo pre entrenado contó con el siguiente desempeño, esto de acuerdo con la documentación oficial de la librería:

Módulo	Precisión	Recall	F1-score
Clasificación de la categoría gramatical	0.97	0.97	0.97
Reconocimiento de entidades nombradas	0.89	0.89	0.89

Evaluación del modelo pre entrenado `es_core_news_sm`

Sin embargo, los módulos de Análisis de sentimientos y Etiquetado de emociones, a pesar de su naturaleza de etiquetadores automáticos sí que se pueden evaluar como clasificadores. A continuación se detalla el método para evaluar el rendimiento de estos dos módulos:

El primer paso fue generar la matriz de confusión obtenida a partir de los resultados obtenidos en el periodo de pruebas, recordando que esta matriz permite describir el desempeño del modelo teniendo como entrada los siguientes cuatro parámetros:

Por una parte se encuentran los verdaderos positivos y verdaderos negativos, que representan las observaciones que se predicen correctamente.

- Positivos verdaderos (TP, por sus siglas en inglés): Valores positivos clasificados correctamente, significando que el valor de la variable en el conjunto de datos de prueba es verdadero, al igual que el valor de la variable en el conjunto de datos obtenido por el modelo, en este caso por la herramienta HAAT.
- Negativos verdaderos (TN, por sus siglas en inglés): Valores negativos clasificados correctamente, significando que el valor de la variable en el conjunto de datos de prueba es falso, al igual que el valor de la variable en el conjunto de datos obtenido por el modelo.

Los dos valores adicionales de la matriz de confusión son los falsos positivos y falsos negativos, estos valores ocurren cuando hay una contradicción entre el valor de referencia y el valor obtenido por el modelo.

- Falsos positivos (FP, por sus siglas en inglés): Supuesto que ocurre cuando el valor de referencia es negativo y el valor obtenido por el modelo de predicción es positivo.
- Falsos negativos (FN, por sus siglas en inglés): Supuesto que ocurre cuando el valor de referencia es positivo y el valor obtenido por el modelo de predicción es negativo.

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	Positivo verdadero	Falso negativo
	0 (No)	Falso positivo	Negativo verdadero

Estructura de la matriz de confusión

Con base en los datos obtenidos durante las pruebas de la herramienta HAAT se procedió al llenado de la matriz de confusión para cada uno de los dos módulos evaluados y se obtuvieron las siguientes métricas:

- Precisión: Con esta métrica se midió la calidad del modelo en la tarea de clasificación. Se utilizó la siguiente fórmula para realizar el cálculo:

$$precision = \frac{TP}{TP + FP}$$

- Recall: Medida utilizada para determinar la cantidad de valores que el modelo es capaz de identificar. Se utilizó la siguiente fórmula para realizar el cálculo:

$$recall = \frac{TP}{TP + FN}$$

- F1-score: Métrica utilizada para combinar las medidas de precisión y recall en un sólo valor. Se utilizó la siguiente fórmula para realizar el cálculo:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Teniendo como base la matriz de confusión y las métricas de evaluación ya explicadas, se procedió a realizar la evaluación del módulo de Análisis de sentimientos y del módulo de Etiquetado de emociones, teniendo como entrada los datos obtenidos durante la etapa de pruebas.

a) Análisis de sentimientos

Pruebas con el corpus NVS2019

Prueba 1 Corpus NVS2019

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 482	FN = 43
	0 (No)	FP = 55	TN = 0

Matriz de confusión para el módulo de análisis de sentimientos (Prueba 1).

Precisión = $TP / TP+FP = 482 / 482+55 = 0.897579 = 89.76\%$

Recall = $TP / TP+FN = 482 / 482+43 = 0.918095 = 91.81\%$

F1 = $2 \cdot ((precision \cdot recall) / (precision+recall)) =$
 $2 \cdot ((0.8975 \cdot 0.9180) / (0.8975+0.9180)) = 0.907634 = 90.76\%$

Prueba 2 Corpus NVS2019

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 499	FN = 27
	0 (No)	FP = 54	TN = 0

Matriz de confusión para el módulo de análisis de sentimientos (Prueba 2).

$$\text{Precisión} = TP / TP+FP = 499 / 499+54 = 0.902350 = 90.24\%$$

$$\text{Recall} = TP / TP+FN = 499 / 499+27 = 0.948669 = 94.86\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.9023 \cdot 0.9180) / (0.94866 + 0.94866)) = 0.910082 = 91\%$$

Prueba 3 Corpus NVS2019

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 502	FN = 20
	0 (No)	FP = 58	TN = 0

Matriz de confusión para el módulo de análisis de sentimientos (Prueba 3).

$$\text{Precisión} = TP / TP+FP = 502 / 502+58 = 0.896428 = 89.64\%$$

$$\text{Recall} = TP / TP+FN = 502 / 502+20 = 0.961685 = 96.16\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.8964 \cdot 0.9616) / (0.8964 + 0.9616)) = 0.927856 = 92.78\%$$

En la siguiente tabla se resumen los resultados obtenidos de las tres pruebas con el corpus NVS2019:

Prueba	Precisión	Recall	F1-score
Prueba 1	89.76%	91.81%	90.76%
Prueba 2	90.24%	94.86%	91%
Prueba 3	89.64%	96.16%	92.78%
Promedio	89.88%	94.27%	91.51%

Comparativa de precisión, recall y F1-score obtenidos en las pruebas del corpus NVS2019 para el módulo de análisis de sentimientos.

Pruebas con el Corpus (Keith, Fuentes & Meneses, 2017)

Prueba con textos positivos

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 21	FN = 5
	0 (No)	FP = 166	TN = 0

Matriz de confusión para el módulo de análisis de sentimientos (Textos positivos).

$$\text{Precisión} = TP / TP+FP = 21 / 21+166 = 0.1122 = 11.22\%$$

$$\text{Recall} = TP / TP+FN = 21 / 21+5 = 0.8076 = 80.76\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.1122 \cdot 0.8076) / (0.1122 + 0.8076)) = 0.1972 = 19.72\%$$

Prueba con textos neutrales

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 4	FN = 1
	0 (No)	FP = 57	TN = 0

Matriz de confusión para el módulo de análisis de sentimientos (Textos neutrales).

$$\text{Precisión} = TP / TP+FP = 4 / 4+57 = 0.0656 = 6.56\%$$

$$\text{Recall} = TP / TP+FN = 4 / 4+1 = 0.8000 = 80\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.0656 \cdot 0.8000) / (0.0656 + 0.8000)) = 0.1212 = 12.12\%$$

Prueba con textos negativos

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 134	FN = 7
	0 (No)	FP = 4	TN = 0

Matriz de confusión para el módulo de análisis de sentimientos (Textos negativos).

$$\text{Precisión} = TP / TP+FP = 134 / 134+4 = 0.9710 = 97.10\%$$

$$\text{Recall} = TP / TP+FN = 134 / 134+7 = 0.9504 = 95.04\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.9710 \cdot 0.9504) / (0.9710 + 0.9504)) = 0.9606 = 96\%$$

En la siguiente tabla se resumen los resultados obtenidos de las tres pruebas con el corpus (Keith, Fuentes & Meneses, 2017):

Prueba	Precisión	Recall	F1-score
Textos positivos	11.22%	80.76%	19.72%
Textos neutrales	6.56%	80%	12.12%
Textos negativos	97.10%	95.04%	96%
Promedio	38.29%	85.26%	42.61%

Comparativa de precisión, recall y F1-score obtenidos en las pruebas del corpus (Keith, Fuentes & Meneses, 2017) para el módulo de análisis de sentimientos.

Prueba piloto de aplicación de la terapia EEA

Prueba EEA

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 22	FN = 0
	0	FP = 10	TN = 0

	(No)		
--	------	--	--

Matriz de confusión para el módulo de análisis de sentimientos (Prueba EEA).

$$\text{Precisión} = TP / TP+FP = 22 / 22+10 = 0.6667 = 66.67\%$$

$$\text{Recall} = TP / TP+FN = 22 / 22+0 = 1.0000 = 100\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.6875 \cdot 1) / (0.6875 + 1)) = 0.8000 = 80\%$$

b) Etiquetado de emociones

Pruebas con el corpus NVS2019

Prueba 1 Corpus NVS2019

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 53	FN = 43
	0 (No)	FP = 484	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Prueba 1).

$$\text{Precisión} = TP / TP+FP = 53 / 53+484 = 0.09869 = 9.87\%$$

$$\text{Recall} = TP / TP+FN = 53 / 53+43 = 0.552083 = 55.20\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.0986 \cdot 0.55208) / (0.0986 + 0.55208)) = 0.167317 = 16.73\%$$

Prueba 2 Corpus NVS2019

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 168	FN = 27
	0 (No)	FP = 385	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Prueba 2).

$$\text{Precisión} = TP / TP+FP = 168 / 168+385 = 0.303797 = 30.37\%$$

$$\text{Recall} = TP / TP+FN = 168 / 168+27 = 0.861538 = 86.15\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) =$$

$$2 \cdot ((0.3037 \cdot 0.8615) / (0.3037 + 0.8615)) = 0.449086 = 44.90\%$$

Prueba 3 Corpus NVS2019

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 420	FN = 20
	0 (No)	FP = 140	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Prueba 3).

$$\text{Precisión} = TP / TP + FP = 420 / 420 + 140 = 0.75 = 75\%$$

$$\text{Recall} = TP / TP + FN = 420 / 420 + 20 = 0.954 = 95.40\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.75 \cdot 0.954) / (0.75 + 0.954)) = 0.839788 = 84\%$$

En la siguiente tabla se resumen los resultados obtenidos de las tres pruebas con el corpus NVS2019:

Prueba	Precisión	Recall	F1-score
Prueba 1	9.87%	55.20%	16.73%
Prueba 2	30.37%	86.15%	44.90%
Prueba 3	75%	95.40%	84%
Promedio	38.41%	78.91%	48.54%

Comparativa de precisión, recall y F1-score obtenidos en las pruebas del corpus NVS2019 para el módulo de etiquetado de emociones.

Pruebas con el Corpus (Keith, Fuentes & Meneses, 2017)

Prueba con textos positivos

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 41	FN = 5
	0 (No)	FP = 146	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Textos positivos).

$$\text{Precisión} = TP / TP+FP = 41 / 41+146 = 0.2193 = 21.93\%$$

$$\text{Recall} = TP / TP+FN = 41 / 41+5 = 0.8913 = 89.13\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) =$$

$$2 \cdot ((0.2193 \cdot 0.8913) / (0.2193 + 0.8913)) = 0.3519 = 35.19\%$$

Prueba con textos neutrales

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 11	FN = 1
	0 (No)	FP = 50	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Textos neutrales).

$$\text{Precisión} = TP / TP+FP = 11 / 11+50 = 0.1803 = 18.03\%$$

$$\text{Recall} = TP / TP+FN = 11 / 11+1 = 0.9167 = 91.67\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) =$$

$$2 \cdot ((0.1803 \cdot 0.9167) / (0.1803 + 0.9167)) = 0.3014 = 30.14\%$$

Prueba con textos negativos

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 106	FN = 7
	0 (No)	FP = 32	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Textos negativos).

$$\text{Precisión} = TP / TP+FP = 106 / 106+32 = 0.7681 = 76.81\%$$

$$\text{Recall} = TP / TP+FN = 106 / 106+7 = 0.9381 = 93.81\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) =$$

$$2 \cdot ((0.7681 \cdot 0.9381) / (0.7681 + 0.9381)) = 0.8446 = 84.46\%$$

En la siguiente tabla se resumen los resultados obtenidos de las tres pruebas con el corpus (Keith, Fuentes & Meneses, 2017):

Prueba	Precisión	Recall	F1-score
Textos positivos	21.93%	89.13%	35.19%
Textos neutrales	18.03%	91.67%	30.14%
Textos negativos	76.81%	93.81%	84.46%
Promedio	38.92%	91.53%	49.93%

Comparativa de precisión, recall y F1-score obtenidos en las pruebas del corpus (Keith, Fuentes & Meneses, 2017) para el módulo de etiquetado de emociones.

Prueba piloto de aplicación de la terapia EEA

Prueba EEA

		Predicción del modelo	
		1 (Si)	0 (No)
Datos reales	1 (Si)	TP = 22	FN = 0
	0 (No)	FP = 10	TN = 0

Matriz de confusión para el módulo de etiquetado de emociones (Prueba EEA).

$$\text{Precisión} = TP / TP+FP = 22 / 22+10 = 0.6875 = 68.75\%$$

$$\text{Recall} = TP / TP+FN = 22 / 22+0 = 1 = 100\%$$

$$\text{F1} = 2 \cdot ((\text{precisión} \cdot \text{recall}) / (\text{precisión} + \text{recall})) = 2 \cdot ((0.6875 \cdot 1) / (0.6875 + 1)) = 0.8148 = 81.48\%$$