

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



Presentado por:

JOSÍAS BERNABÉ HUARI MAGINO

**“USO DEL SOFTWARE SISAM PARA LA DETECCIÓN
DE CONTENIDOS DENIGRANTES HACIA LA MUJER
EN PRODUCTOS AUDIO MUSICALES DEL GENERO
REGGAETÓN 2019”**

Asesor:

Mg. Richard Carrión Abollaneda

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO DE SISTEMAS (de acuerdo a Resolución N° 035-
2017-SUNEDU/CD)

ANDAHUAYLAS – APURÍMAC PERÚ

2020

APROBACION DEL ASESOR



Universidad Nacional José María Arguedas

Identidad y Esfuerzo para el Trabajo Productivo y el Desarrollo



APROBACION DEL ASESOR

Quién suscribe:

MSc. Richard Carrión Abollaneda, por la presente:

CERTIFICA,

Que, el Bachiller en Ingeniería de Sistemas, **JOSÍAS BERNABÉ HUARI MAGINO**, ha culminado satisfactoriamente el Informe Final intitulado: **"USO DEL SOFTWARE SISAM PARA LA DETECCIÓN DE CONTENIDOS DENIGRANTES HACIA LA MUJER EN PRODUCTOS AUDIO MUSICALES DEL GENERO REGGAETÓN 2019"** para optar el Título Profesional de Ingeniero de Sistemas.

San Jerónimo, 30 de junio del 2020

MSc Richard Carrión Abollaneda
Asesor

Josías Bernabé Huari Magino
Tesista

ACTA DE SUSTENTACION



Universidad Nacional José María Arguedas

Identidad y Excelencia para el Trabajo Productivo y el Desarrollo



FACULTAD DE INGENIERÍA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

ACTA DE SUSTENTACIÓN DE TESIS

En la Universidad Nacional José María Arguedas ubicado en el distrito de San Jerónimo de la Provincia de Andahuaylas, siendo las 08:00 horas del día 16 de julio del año 2020, se reunieron los docentes: Dra. Norma Lorena Catacora Flores, Ing. Roberto Quispe Quispe, MSc. Ivan Soria Solis en condición de integrantes del Jurado Evaluador del Informe Final de Tesis intitulado: "**USO DEL SOFTWARE SISAM PARA LA DETECCIÓN DE CONTENIDOS DENIGRANTES HACIA LA MUJER EN PRODUCTOS AUDIO MUSICALES DEL GENERO REGGAETÓN 2019**", cuyo autor es el Bachiller en Ingeniería de Sistemas **JOSÍAS BERNABÉ HUARI MAGINO**, el Asesor MSc. Richard Carrión Abollaneda, con el propósito de proceder a la sustentación y defensa de dicha tesis.

Luego de la sustentación y defensa de la tesis, el Jurado Evaluador **ACORDÓ: APROBAR** por **UNANIMIDAD** al Bachiller en Ingeniería de Sistemas **JOSÍAS BERNABÉ HUARI MAGINO**, obteniendo la siguiente calificación y mención:

Nota escala vigesimal		Mención
Números	Letras	
17	DIECISIETE	MUY BUENO

En señal de conformidad, se procedió a la firma de la presente acta.

.....
Dra. Norma Lorena Catacora Flores
Presidente del Jurado Evaluador

.....
Ing. Roberto Quispe Quispe
Primer Miembro del Jurado Evaluador

.....
MSc. Ivan Soria Solis
Segundo Miembro del Jurado Evaluador

DEDICATORIA

El presente trabajo va dedicado a Dios. Por su fortaleza y ayuda en el tiempo de crisis que sucumbió a mi país; parecía imposible, pero se logró. Creo en él.

Josías Huari Magino

AGRADECIMIENTO

Agradezco a Dios siempre como dice la biblia “Todo lo puedo en Cristo que me fortalece Filipenses 4: 13”. A mi asesor Mg. Richard Carrión Abollaneda por su apoyo en este trabajo. A mis padres por su amor y ejemplo, a mi tía Ruth por sus consejos, a mi abuelita Felicia Romero por su inmenso amor y cariño.

A la Universidad Nacional José María Arguedas donde pude culminar la carrera profesional de Ingeniería de Sistemas.

No sé qué me prepare el futuro, pero sé que con Cristo estaré bien.

Josías Huari Magino

INDICE

APROBACION DEL ASESOR	ii
ACTA DE SUSTENTACION.....	iii
DEDICATORIA.....	iv
AGRADECIMIENTO.....	v
INDICE	1
RESUMEN.....	5
ABSTRACT	6
CHUMASQ	6
INTRODUCCIÓN	7
CAPITULO 1: PROBLEMA DE INVESTIGACIÓN.....	8
1.1. Descripción del Problema.....	8
1.2. Formulación del Problema.....	10
1.3. Justificación	11
1.4. Objetivos	11
1.5. Delimitación de la Investigación.....	12
CAPITULO 2: ANTECEDENTES.....	12
2.1. Antecedentes.....	12
2.1.1. Antecedentes a Nivel Internacional.....	12
CAPITULO 3: MARCO TEORICO	14
3.1 Bases Teóricas Científicas	14
3.1.1 SISAM.....	14
3.1.2 Contenidos Denigrantes hacia la Mujer.....	16
3.2 Definición Conceptual	18
3.2.1. Procesamiento del Lenguaje Natural	18
3.2.2. Análisis de Sentimiento.....	18
3.2.3. Flujo de Trabajo del Análisis de Sentimiento.....	19
3.2.4. Tareas del Pre procesamiento de Texto	19
3.2.5. Técnicas de Clasificación de Sentimientos u Opinión	21

3.2.6. Niveles del Análisis de Sentimiento.....	21
3.2.7. Intensificadores en el análisis de sentimientos	22
3.2.8. Negación en el análisis de sentimientos.....	23
CAPITULO 4: METODOLOGÍA DE INVESTIGACIÓN.....	23
4.1. Hipótesis de investigación.....	23
4.1.1. Hipótesis General.....	23
4.1.2. Hipótesis Especificas.....	23
4.2. Operacionalización de Variables	24
4.3. Tipo de Investigación.....	25
4.4. Nivel de Investigación.....	25
4.5. Diseño de Investigación	25
4.6 Población y Muestra.....	25
4.6.1. Población.....	25
4.6.2. Muestra	25
4.7. Método de Investigación	26
4.8. Técnicas e Instrumentos de Acopio de Datos	26
4.9. Técnicas de Análisis de Datos	27
CAPITULO 5: RESULTADOS.....	27
5.1. Descripción de la solución desarrollada.....	27
5.1.1. Listas de Contenidos Denigrantes	28
5.1.2. Listas de Intensificadores	29
5.1.3. PRE PROCESAMIENTO	29
5.1.4. PROCESAMIENTO	31
5.2. Presentación de resultados	34
CAPITULO 6: DISCUSIÓN.....	37
6.1. Análisis y discusión de resultados	37
6.2. Comprobación de hipótesis.....	37
6.2.1. Comprobación de las hipótesis específicas.....	37
6.2.3. Comprobación de la hipótesis general.....	38
CONCLUSIONES	38
RECOMENDACIONES	39
REFERENCIAS BIBLIOGRAFICAS.....	40

ANEXOS.....	43
ANEXO 1: Programación Extrema XP	43
ANEXO 2: Implementación de la metodología ágil Programación Extrema(XP)	46
ANEXO 3: Matriz de Consistencia.....	77

RESUMEN DE FIGURAS

Figura 1. Diseño de Investigación	25
Figura 2. Etapas del Flujo de Trabajo del Análisis de Sentimiento	27
Figura 3. Grafica de estadística circular del resultado del análisis	34
Figura 4. Grafica de estadística circular del total de contenidos de cada tipo violencia	35
Figura 5. Grafica de barras de la cantidad de canciones de cada tipo de violencia	36
Figura 6. Primera Interacción	47
Figura 7. Segundo Interacción.....	48
Figura 8. Tercera Interacción.....	48
Figura 9. Cuarta Interacción	49
Figura 10. Diagrama de Flujo de datos Principal	51
Figura 11. Diagrama de Flujo de datos: Pre procesamiento	51
Figura 12. Diagrama de Flujo de datos: Procesamiento	52
Figura 13. Diagrama de Flujo de datos: Detección de Contenidos	53
Figura 14. Maquetación de la interfaz inicial del Software.....	53
Figura 15. Maquetación de la interfaz Pestaña Violencia Física	54
Figura 16. Maquetación de la interfaz Pestaña Violencia Psicológica	55

RESUMEN DE TABLAS

Tabla 1 Matriz de Operacionalización de Variables.....	24
Tabla 2. Cantidad de contenidos de cada tipo de violencia.....	28
Tabla 3. Valor de Mensaje y criterios.....	28
Tabla 4. Cantidad de Intensificadores por número palabra.....	29
Tabla 5. Resultados del Análisis de las canciones.....	34
Tabla 6. Cantidad de contenidos por tipo de violencia.....	35
Tabla 7. Cantidad de canciones por tipo de violencia	36
Tabla 8. Modelo de historia de usuario	46
Tabla 9. Tabla de Interacciones	49
Tabla 10. Cuadro de fechas de entrega	50
Tabla 11. Tarjeta CRC: Clase AlgoritmoLevenshtein.	56
Tabla 12. Tarjeta CRC: Clase Conteo_de_Repeticiones_y_Mensajes.....	56
Tabla 13. Tarjeta CRC: Clase Corrector Ortográfica.....	56
Tabla 14. Tarjeta CRC: Clase Intensificadores	57

Tabla 15.Tarjeta CRC: Clase Pre procesamiento.....	57
Tabla 16.Tarjeta CRC: Clase Procesos Negación.....	57
Tabla 17.Tarjeta CRC: Clase Procesos resalto palabras.....	58
Tabla 18.Tarjeta CRC: Clase Conectar.....	58
Tabla 19.Tarjeta CRC: Clase Extracción Mensj.....	58
Tabla 20.Tarjeta CRC: Clase Generar reporte.....	59
Tabla 21.Prueba de Aceptación: Corrección de Texto.....	73
Tabla 22.Prueba de Aceptación: Búsqueda, selección y extracción de contenidos.....	73
Tabla 23.Prueba de Aceptación: Búsqueda, selección y extracción de intensificadores.....	74
Tabla 24.Prueba de Aceptación: Búsqueda, selección y extracción de negación.....	74
Tabla 25.Prueba de Aceptación: Visualizar los contenidos por tipo de violencia.....	75
Tabla 26.Prueba de Aceptación: Visualizar los intensificadores y la negación por contenidos.....	75
Tabla 27.Prueba de Aceptación: Facilidad de Uso.....	76
Tabla 28.Prueba de Aceptación: Generar Reporte.....	76
Tabla 29.Matriz de Consistencia.....	77

RESUMEN DE IMÁGENES

Imagen 1.Captura de Imagen: Resultado de Corrección de texto.....	60
Imagen 2.Captura de Imagen: Resultado de Análisis de texto.....	61
Imagen 3.Captura de Imagen: JPanel – Contenidos Detectados.....	61
Imagen 4. Captura de Imagen: JPanel – Violencia Sexual.....	62
Imagen 5.Captura de Imagen: JTable – Lista - Intensificadores.....	63
Imagen 6.Captura de Imagen: JTable – Lista - Negación.....	63
Imagen 7.Captura de Imagen: JPanels.....	64
Imagen 8.Captura de Imagen: Tablas Mensajes.....	65
Imagen 9.Captura de Imagen: Tablas y campos del JPpanel Violencia Sexual.....	66
Imagen 10.Captura de Imagen: Botón Cargar Texto.....	67
Imagen 11.Captura de Imagen: Ventana para la selección de texto.....	67
Imagen 12.Captura de Imagen: Botones Pre proceso y Botón Análisis.....	68
Imagen 13.Captura de Imagen: JProgressBar.....	68
Imagen 14.Captura de Imagen: Menú de JPannels.....	69
Imagen 15.Captura de Imagen: Detalles de Mensaje.....	69
Imagen 16.Captura de Imagen: Ventana Ubicación de Mensaje.....	70
Imagen 17.Captura de Imagen: Ventana Detalles de Intensificador.....	71
Imagen 18.Captura de Imagen: Ventana Detalles de Negación.....	71
Imagen 19.Captura de Imagen: Botón Generar Reporte.....	72
Imagen 20.Captura de Imagen: Formato de Reporte.....	72

RESUMEN

El presente trabajo de tesis se realizó con el objetivo de determinar la influencia del uso del Software SISAM en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón; esto es en las letras de canciones del género reggaetón. Así poder demostrar también que el flujo de trabajo de análisis de sentimiento también puede ser usado para analizar letras de canciones. La investigación tuvo un diseño Pre experimental y su tipo de investigación fue Aplicada. La delimitación que tubo esta investigación fue temporal en ella especifico que se tomarían canciones del genero reggaetón de los años 2009 al 2019. Ya la población es ilimitada ya que existen muchas canciones del genero reggaetón , se selección una muestra de 400 canciones , se extrajo sus letras y se guardaron en archivos de textos .Para la recolección de las canciones se usó paginas donde muestren letras de canciones del genero reggaetón así como también lista de reproducciones de la plataforma de spotify. Se clasificó a los mensajes denigrantes en 4 tipos: Violencia Física, Violencia Psicológica, Violencia Sexual y Violencia Simbólica. El resultado del análisis presento que: el 60% del total de canciones presentaban 1 o más tipos de mensajes denigrantes. Por lo tanto, se concluye que el Software SISAM usando el Enfoque Basado en Diccionario del Análisis de Sentimiento si influye en la detección de contenidos denigrantes en letras de canciones del genero reggaetón.

Palabras clave: VF, Violencia Física. VPS, Violencia Psicológica. VSE, Violencia Sexual. VSIM, Violencia Simbólica. AS, Análisis de Sentimiento.

ABSTRACT

The present thesis work was carried out with the aim of determining the influence of the use of SISAM Software in the detection of content denigrating towards Women in audio-musical products of the Reggaetón Genre; This is in the lyrics of the reggaeton genre. So we can also demonstrate that the sentiment analysis workflow can also be used to analyze song lyrics. The research had a Research design that is Pre-experimental and its type of research was Applied. The delimitation that this investigation had was temporary, it specified that songs from the reggaeton genre would be taken from the years 2009 to 2019. Since the population is unlimited since there are many songs from the reggaeton genre, a sample of 400 songs was selected, their Lyrics and were stored in text files. For the collection of the songs, pages showing songs of the reggaeton genre were used, as well as a list of reproductions from the spotify platform. Denigrating messages were classified into 4 types: Physical Violence, Psychological Violence, Sexual Violence and Symbolic Violence. The result of the analysis showed that: 60% of the total songs presented 1 or more types of degrading messages. Therefore, it is concluded that the SISAM Software using the Dictionary-Based Approach to Sentiment Analysis does influence the detection of degrading content in reggaeton song lyrics.

Key words: VF, Physical Violence. VPS, Psychological Violence. VSE, Sexual Violence. VSIM, Symbolic Violence. AS, Sentiment Analysis.

CHUMASQ

Imaraykum Kay tesisqa rurakurqa imaynam Software Sisam nisqam riksichiwasun mana allin kaqkunata chay taki reggaeton ninsqampi.

Imatam rikurichin warmikunapi chay takikunata uyaritin, kay reggaeton taripa ukumpin kachkan mana allin. Chaynayataq riksichiwasun kay analisispay llankasqayqa chay takipa letrankunata analizananchpaq warmikunapa sonqompi sentesqanmanta puntataqa expermentasqam kay investigacionqa qinaspa ruraymanmi churasasqa. Kay investigacionqa manan wiñaypaqhu. Qinaspa riqsichini kay taki reggaeton akllasqa wata 2009 wak wat 2019. Runakunapas nisutam miraron qinaspa kay taki reggaetonpa achkam, chaymi akllani tawa pachak taki kunata , chaymantam orqoni letran kunata archivo waqaychasqa kanampaq. Chay taki reggaetonpa letrankuna qoñunaypaqmi usarqani paginakuna riksichikuqta chaymanta listape reproducesqanta plataforma spotify ninsqanwan. Tanapim rakinasqa chay mana allin mensajekuna: atipakuy cuerpoaq. atipakuy psicológica, atipakuy sexopaq chaymanta atipakuy simbolika nisqan.

Analisisqa resultadonmi : 60 % llapallan taki reggaetonpi kasqa hukmanta masyaq mana allin mensajekuna .

Chaynapim quispichini kay spftware sisam usaspay jinaspa sentimiento analisisqa dictionariopi tarinchik mana allinza taki reggaeton letrampi.

Importante Ramaykuna :

VF, Atipakuy cuerpo. VPS , atipakuy psicológica nisqan, VSE, atipakuy sexo. VSIM, atipakuy simbolika ninsqan. AS, sentimentokuna riksiy.

INTRODUCCIÓN

En la actualidad existen muchas canciones publicadas en diferentes plataformas, todos tenemos acceso a ellas desde los más pequeños y en distintos formatos (.mp3, mp4., etc.). A muchos nos atrae ciertos tipos de música ya sea por su letra, ritmo o mensaje. La música también es un medio por donde podemos expresar

nuestras emociones, contar historias, contar hechos reales a través de la letra que acompañan a la melodía. Pero existen también canciones donde las letras también pueden afectarnos ya que tienen palabras, frases que son mensajes que ofenden y denigran. Estos mensajes son los que se guardan en nuestra memoria a causa de haberlos escuchado repetidamente. Su efecto más terrible es cuando lo usamos al momento de comunicarnos o expresarnos trayendo como consecuencia ofender o denigrar a alguien. Por otra parte, existen métodos de análisis de textos que buscan determinar su polaridad respecto algún tema como es el Análisis de Sentimiento (AS). Haciendo uso de ciertos procesos y putas para lograrlo. Es así como surge la idea de crear un software que detecte mensajes denigrantes mediante el AS. La programación en java es un lenguaje siguen siendo utilizado en la actualidad. Existiendo programas donde uno pueden crear desde pequeños a grandes proyectos ya sea desde aplicativos de tareas específicas hasta un sistema complejo. La metodología de Desarrollo de Software utilizada para esta investigación es la Programación Extrema (XP), la cual permitió cubrir las necesidad y requerimientos importantes para el software. Por lo consiguiente, el presente informe está constituido de la siguiente manera:

Problema de Investigación: Aquí se encuentra la descripción del problema, la formulación del problema, la justificación, los objetivos y la delimitación de la investigación.

Antecedentes: Aquí encontraremos los Antecedentes para la presente investigación.

Marco Teórico: Son las bases teóricas científicas y las definiciones conceptuales.

Metodología de Investigación: Aquí se encuentra las hipótesis, operacionalización de variables, tipo de investigación, nivel de investigación, diseño de investigación, población - muestra, método de investigación, técnicas e instrumentos de acopio de datos y técnicas de análisis de datos.

Resultados: Aquí está la descripción de la solución y presentación de los resultados.

Discusión: Aquí se encuentra el análisis y discusión de los resultados.

Y por último están las conclusiones, recomendaciones y anexos

CAPITULO 1: PROBLEMA DE INVESTIGACIÓN

1.1. Descripción del Problema

La violencia ejercida contra la mujer es un problema que en el Perú se vive a diario y que obedece a estructuras jerárquicas patriarcales que reproducen

una cultura donde las mujeres son vistas como objetos desechables y maltratadas (Melendez, 2005) En el año 2004, la cifra a nivel de la VII Región–Lima de la PNP de denuncias por maltrato llego a 41.567. Existe tipos de violencia contra la mujer uno de ellos es violencia psicológica o emocional, consistente en las amenazas, los comentarios degradantes, el lenguaje sexista y el comportamiento humillante. Estos son componentes frecuentes de la conducta violenta hacia las mujeres que pueden tener consecuencias para su bienestar psíquico o emocional. El considerar a la mujer como un objeto y no como ser humano también se ha visto mucho en la música. Uno de los géneros musicales que ha marcado mucho en esto es el reggaetón. Según una investigación hecha en la universidad de Antioquia (Ramírez Noreña, 2012) “El reggaetón desde el principio tuvo un grado de aceptación muy alto y se convirtió en la música preferida de un número considerable de personas, especialmente adolescentes y jóvenes. Tuvo gran influencia en la industria musical y aún sigue vigente. La gran mayoría de estas canciones tienen contenidos fuertes y utilizan un léxico erótico donde la mujer se presenta como un objeto sexual”. Estas canciones suelen tener mensajes vulgares y ofensivos que denigran a la mujer. (Noa, 2018) La figura femenina es expuesta a la violencia frecuentemente a través de las letras textuales del reggaetón y su libre difusión sobre todo en los medios de comunicación, ya que en él se transmiten mensajes que incentiva la violencia, el machismo y crea estereotipos, creencias que predisponen un pensamiento erróneo sobre la imagen femenina en donde principalmente se observa que la mujer, es quien viste prendas muy ajustadas dejando ver sus atributos, bonita, voluptuosa de muy poca moral y ética, dispuesta a satisfacer las exigencias masculinas siendo constantemente insultada con sobre nombres como “gata” “zorra” , que además acepta indicaciones que la ponen en una situación de vulnerabilidad como “ponte contra la pared” “ te voy a dar duro” entre otros que se pueden escuchar en cualquier canción de reggaetón. Aunque para las industrias discográficas más les interesa la tendencia que despierta en los oyentes la canción que en pocas ocasiones suelen percatarse del contenido de los temas musicales ya que el ritmo, la matización, la armonía, etc. que son características de la música suelen hacer que pase desapercibido la letra de la canción siendo así su mayor problema al final ya que algún oyente puede hacer uso de esa canción para alguna circunstancia y percatarse de que contiene palabras o frases que están incentivando temas como el racismo, violación ,maltrato y aborto .Esto se agrava aún más ya que en la actualidad las redes sociales son la forma más rápida de publicar o hacer llegar las noticias haciendo que el ranking de la canción con respecto a sus descargar o reproducciones baje. Ya que para algunos puede representar un pasatiempo escuchar música, pero para otros

representa algo más. (Pérez-Aldeguer, 2014) Es una herramienta para desarrollar la competencia intercultural. Permite reconocer, comprender y valorar adecuadamente comportamientos, actitudes y expectativas laborales de personas pertenecientes a otros entornos culturales haciendo conocer las características culturales que tiene el país o cultura del cual procede la persona. Estas canciones pueden causar no solo pérdidas sino también problemas sociales que incluyan al compositor y/o autor de dichos temas musicales haciéndose creer tener relación con dichos temas ofensivos provocando una caída no solo en su audiencia o seguidores sino también pérdidas en aquellas empresas que usaron dichos temas para fomentar campañas o auspicios de productos. Para las plataformas de streaming music (Spotify, Google Play Music, YoutubeMusic y Deezer) el descubrir estos temas musicales es un esfuerzo arduo siendo así que en muchas ocasiones son los usuarios quienes reportan que tales canciones no son aptas o son insultantes. Estos servicios no cuentan con un software capaz de encargarse de estar analizando las letras de cada canción antes de ser publicadas. Ya que toda detección de algún producto musical proviene de los consumidores. El contar con un software que pueda estar al pendiente de estos problemas permitiría el alertar a la empresa antes publicarlo para sus clientes. Algunos avances se han elaborado con respecto a análisis de frases o pequeños trozos de textos que algunos los llaman comentarios en la búsqueda de conocer o entender las actitudes, opiniones y emociones expresadas en estos. Este proceso es conocido como Análisis de sentimiento es uno de los campos que incluye la Inteligencia Artificial. Se conoce que ha sido utilizado para procesar comentario que los usuarios dejan en los foros o chats respecto a un tema en particular y que mediante ciertos procesos internos se ha llegado a conocer la connotación positiva o negativa del lenguaje. Existen investigaciones que se han desarrollado con respecto al Análisis de Sentimientos en universidades. Además, este proceso no solo se enfoca para situaciones que demande grandes cantidades de datos sino también investigaciones de situaciones específicas como la investigación desarrollada en la Universidad de Guayaquil en el 2016 donde se usó este proceso en el análisis de los comentarios de Twitter de los socios de un club deportivo. Se propuso encontrar el nivel de satisfacción por los servicios que brindaba dicho club (VITERI ALVARADO & JAVIER, 2016).

1.2. Formulación del Problema

1.2.1. Problema General

¿Cuál es la influencia del uso del Software SISAM en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón 2018?

1.2.2. Problemas Específicos

- ¿Cómo influye el nivel de análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón?
- ¿Cómo influye el flujo de trabajo del análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón?
- ¿Cómo influye el diccionario léxico en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón?

1.3. Justificación

Para las plataformas de Streaming Music el tener un software que se encargue de detectar aquellos temas musicales que contengan mensajes inapropiados les permitiría preocuparse de otros asuntos. Ya que el estar en constante supervisión de las canciones que están en su repositorio para detectar dichos problemas o esperar el comentario de un usuario que reportando el inconveniente hace que haya una constante preocupación de ser una plataforma confiable para todo público. Se entiende que en la actualidad respecto a estas plataformas no hay límite de edad o restricción para el acceso a estos servicios ya que hasta un menor de edad puede hacer uso de estos ya que la música no puede llegar a ser clasificada para que tipo de edad ya que cada ser humano tiende a tener diferentes gustos o preferencias. Además, esta investigación permitirá a otros investigadores poderlo aplicar no solo en el contenido de canciones también en videos y poder incluso implementar ciertas normas que restrinjan el acceso a ciertas personas a dichos videos musicales y canciones insultantes median la identificación de su edad u otras normas.

1.4. Objetivos

1.4.1. Objetivo General

Determinar la influencia del uso del Software SISAM en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.

1.4.2. Objetivos Específicos

- Determinar la influencia del nivel de análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.
- Identificar la influencia del flujo de trabajo del análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.
- Determinar la influencia del diccionario léxico en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.

1.5. Delimitación de la Investigación

1.5.1. Temporal

Las canciones del género reggaetón que se analizará serán desde el año 2009 al 2019.

CAPITULO 2: ANTECEDENTES

2.1. Antecedentes

Para fortalecer la presente investigación fue necesaria revisar estudios anteriores o relacionados con el tema, a fin de buscar un aporte al mismo, a continuación:

2.1.1. Antecedentes a Nivel Internacional

Lina Torres Samboni (Torres Samboni, 2015), en su proyecto de tesis presentado en la Escuela Colombiana de Ingeniería Julio Garavito titulada, “Análisis de Sentimientos sobre el Posconflicto Colombiano utilizando herramientas de Minería de Texto”, en su investigación plantea conocer cuál es la opinión de la población colombiana sobre el Posconflicto que aqueja a la población. Lina plantea como objetivo de su proyecto perfilar la opinión positiva, negativa o neutra del posconflicto a partir de información no estructurada proveniente de sectores políticos, medios de comunicación, guerrilleros y generadores de opinión colombianos. Lina en esta investigación para el análisis de sentimiento hizo uso de un diccionario de palabras llamado Spanish Emotion Lexicon (SEL) que clasifica las opiniones en

base a emociones básicas como es: alegría, enojo, miedo, tristeza, sorpresa y disgusto. Además, hizo uso del software KNIME como plataforma de minería de datos quien a su vez permite poder cargar el diccionario y realizar el análisis. Al final Lina concluyo que el uso del análisis de sentimiento y la minería de datos permitió obtener información más representativa sobre la opinión de cada sector que fueron parte del análisis.

Montesinos García Lucas (Montesinos García, 2014), en su proyecto de tesis para la Facultad de Ciencias Físicas y Matemáticas del Departamento de Ingeniería Eléctrica de la Universidad de Chile titulada , “Análisis de Sentimientos y Predicción de Eventos en Twitter” ,en su estudio plantea el uso del análisis de sentimiento para predecir resultados futuros respecto a algún evento o suceso en este caso los resultados de encuestas de las elecciones primarias chilenas del año 2013; donde los candidatos fueron Andrés Allamand y Pablo Longueira. El objetivo de su trabajo fue desarrollar un método que sea capaz de extraer la opinión de los usuarios chilenos en Twitter, a través de un análisis de sentimientos y comprobar si estas predicen los resultados de las encuestas. En su proyecto Lucas hizo un estudio de las palabras positivas y negativas que aparecen con mayor frecuencia en los comentarios llegando a clasificar 700 palabras. A este diccionario se le agrego otros diccionarios léxicos como es Spanish Emotion Lexicon (SEL). Lucas obtuvo como resultado que, al analizar usuarios con polaridad positiva, no supera en cuanto a los usuarios negativos, ya que ellos obtuvieron un resultado que llega a casi un 100% de acierto. Lucas concluyo que los resultados marcan la misma tendencia al dar mayor rechazo a Longueira que a Allamand de la misma forma que la encuesta CEP lo hace.

Rutilio Rodolfo López Barbosa (López Barbosa, 2015) , en su proyecto de tesis para el Departamento de Ciencias de la Computación de la Universidad de Alcalá titulada “Aplicación del Análisis de Sentimientos a la Evaluación de Datos Generados en Medios Sociales” en su estudio plantea que en el Twitter los usuarios han desarrollado la capacidad de disminuir el tamaño de las palabras mediante el uso de abreviaturas, contracciones, acrónimos, síncopas, apócope y el truncado de palabras, todo esto combinado con el uso de la jerga de Internet. Estas características pueden complicar el análisis de los tweets y afectar por lo tanto su utilidad. El objetivo de su trabajo fue estudiar el grado de confiabilidad de las herramientas en línea de

análisis de sentimiento que trabajan con Twitter. Para esto Rutilio selecciono 6 herramientas de análisis de sentimientos y lo probó con 20 objetos diferentes (productos y servicios). Los resultados que obtuvo con cada herramienta eran comparados con los resultados que obtenía por el análisis de expertos. Estos expertos fueron personas especializadas en el análisis del lenguaje humano y utilizaron sus propios mecanismos y herramientas de análisis. El investigador obtuvo como resultado que las herramientas utilizadas no lograron ser confiables en todos los objetos al cien por ciento logrando solo ciertos porcentajes. Como es el ejemplo de la herramienta SentUAH que fue aceptable en la mayoría de los casos con porcentajes mayores al 50% en todos los casos. El investigador concluyo, que el análisis de sentimientos es una disciplina en progreso y con las tecnologías y las herramientas de software disponibles para esta tarea, no es aún evidente su aportación como artefactos cien por ciento confiables.

CAPITULO 3: MARCO TEORICO

3.1 Bases Teóricas Científicas

3.1.1 SISAM

Sistema de Análisis Musical (SISAM). Es un software de aplicación que comprende un conjunto de componentes lógicos necesarios que hacen posible la realización de tareas específicas. Está diseñado para encontrar y verificar existencia de contenidos denigrantes hacia la mujer en las letras de las canciones. Las dimensiones de la variable independiente (SISAM) son: NIVEL DE ANÁLISIS DE SENTIMIENTO, DICCIONARIO LÉXICO y FLUJO DE TRABAJO DEL ANÁLISIS DE SENTIMIENTO. El software hace uso del Enfoque Basado en Diccionario del Análisis de Sentimiento. Para mayor entendimiento se añade la definición de Enfoque Basado en Diccionario y Análisis de Sentimiento

- **Análisis de Sentimiento**

Según Miguel, Darnos, David P., Mireya T. y Beatriz (Rosales Quiroga, Vilariño Ayala, Pinto, Tovar, & Beltrán, 2016) es una tarea de clasificación de textos dentro del área del Procesamiento del Lenguaje Natural, su objetivo es dado una opinión de usuario poder detectar la polaridad de esta, ya sea positiva, negativa o neutra.

- **Enfoque Basado en Diccionario**

Según Medhat, Hassan y Korashy (Medhat, Hassan, & Korashy, 2014a) es conocido en inglés como Dictionary-Based Approach y es un tipo de técnica de clasificación de sentimientos que pertenece al Enfoque basado en Léxico (Lexicon-based Approach). Según Fernández (Fernández, 2016) este tipo de enfoque se caracteriza por el uso de diccionarios de términos (palabras o frases) con orientación semántica.

3.1.1.1 **Características del Software**

- ✓ Portable; pueden ser ejecutado desde una memoria como desde el escritorio
- ✓ Utiliza una base de datos portable; sus datos y tablas se encuentran comprimidos en archivos en una carpeta.
- ✓ Contiene paneles ordenados por tipo de violencia; donde se muestran información y ubicación de los mensajes detectados.
- ✓ Admite archivos con extensión .txt
- ✓ Analiza un texto a la vez.
- ✓ Genera un reporte al detectar coincidencias con algún tipo de violencia en el texto importado.

3.1.1.2 **Nivel de Análisis de Sentimiento**

Es el nivel que define la profundidad del análisis de sentimiento. Se considera tres niveles de profundidad: Nivel de documento, Nivel de Oración y Nivel de aspecto - entidad (Liu, 2012b) ; es decir es una forma de clasificar e identificar la opinión. Por ello el nivel que se seleccione determinará si se clasificará la opinión general del documento (texto) o la opinión de cada oración o la opinión de cada aspecto y entidad. Para este proyecto se eligió el Nivel de Documento.

3.1.1.3 Diccionario Léxico

Consiste en un listado de términos (pueden ser palabras o multi palabras) que tienen asociados una determinada orientación de sentimiento. Esta orientación puede ser de polaridad (positiva / negativa), de intensidad de sentimiento o de tipo de emoción (Baviera Puig, 2017). Estos diccionarios presentan principalmente adjetivos, que son los que aportan más información a la hora de analizar los sentimientos, aunque también incluye verbos, adverbios y sustantivos (Montesinos García, 2014). Con lo mencionado diremos que esta dimensión nos mostrara cuan útil es mediante el número aciertos al realizar la comparación.

3.1.1.4 Flujo de Trabajo del Análisis de Sentimiento

Según Dannyll, Darío y María (Vélez Román, Zambrano Zambrano, & Zambrano Zambrano, 2017) señala que es una serie de procedimientos y que el flujo o esquema de trabajo clásico dentro del análisis de sentimiento contiene una serie de tareas de pre-procesamiento de texto; que se encargan de obtener las características de interés o relevantes del mismo. Seguido de una etapa de procesamiento; esta etapa puede comprender una serie de tareas de análisis, clasificación y la aplicación de diversas metodologías para evaluar.

Esta dimensión identificará y analizará los textos (opinión) para así detectar los contenidos denigrantes que contiene el texto.

3.1.2 Contenidos Denigrantes hacia la Mujer

Antes de definir lo que son Contenidos Denigrantes a la Mujer se debe entender lo que es Denigración a la Mujer.

Según Yolanda Ruiz (Ruiz, 2008) Denigración a la mujer es el acto de atacar su entorno; es decir es el acto de generar circunstancias o situaciones incómodas para la mujer ya sea en lo social, laboral, cultural y demás. Creando un ambiente desagradable y peligroso para la mujer. Con lo expuesto anteriormente se define que los contenidos denigrantes a la mujer son aquellas palabras, oraciones y frases del texto que a través de su significado o mensaje generan situaciones que afectan el entorno de la mujer.

Para identificar las dimensiones de la variable dependiente se verifica primero si existe alguna relación de causa-efecto respecto a los mensajes o ideas plasmados en los textos y el entorno de la mujer. Segundo se define el tipo de mensajes o ideas que debe mostrar las palabras, oraciones y frases del texto para ser considerados como contenidos denigrantes.

Según Lissette (Lozano, 2016) señala que los mensajes de las canciones del género reggaetón si afectan a la mujer en su entorno social. Además, señala que las canciones con mensajes de violencia a la mujer son una de las causas. Ya que los mensajes de violencia a la mujer si tiene un efecto en el entorno de la mujer serán considerados como contenidos denigrantes a la mujer.

Según Karina, Emilia, Isabel, Adonai, Simone (Arévalo, Chellew, Figueroa-Cofré, Arancibia, & Schmied, 2018) en las letras de las canciones del género reggaetón se puede encontrar mensajes relacionados con ciertos niveles de violencia al género femenino. Además, señalan que los niveles de violencia encontrados con mayor frecuencia son:

- **La Violencia Física**
Frecuencia en la cual se menciona al interior de cada canción el uso de golpes contra una mujer.
- **La Violencia Psicológica**
Cantidad de insultos o juicios despectivos contra la mujer (calificativos hacia la mujer haciendo referencia a animales perra, etc.), frecuencia con la que se amenaza a la mujer.
- **La Violencia Sexual**
Cantidad de veces en que se menciona que se lleva a cabo un acto sexual por la fuerza, haber tocado a una mujer sin su consentimiento y frecuencia con la que se relata un coqueteo no recíproco.
- **La Violencia Simbólica.**
Se la califica a la mujer como fácil o difícil en función de su disposición o no a tener relaciones sexuales, se manifiesta interés por las características físicas de la mujer por sobre las

características personales o psicológicas, se señala a la mujer como un objeto de deseo sexual.

Cabe señalar que estos niveles afectan directamente a la mujer en su entorno a través de mensajes subjetivos, violentos y ofensivos. Por ello son considerados para la variable dependiente como referencia para sus dimensiones las cuales son:

- Palabras y/o frases que transmitan mensajes de violencia física a la Mujer.
- Palabras y/o frases que transmitan mensajes de violencia sexual a la Mujer.
- Palabras y/o frases que transmitan mensajes de violencia psicológica a la Mujer.
- Palabras y/o frases que transmitan mensajes de violencia simbólica a la Mujer.

3.2 Definición Conceptual

3.2.1. Procesamiento del Lenguaje Natural

Consiste en la utilización de un lenguaje natural para comunicarnos con la computadora, debiendo ésta entender las oraciones que le sean proporcionadas, el uso de estos lenguajes naturales, facilita el desarrollo de programas que realicen tareas relacionadas con el lenguaje o bien, desarrollar modelos que ayuden a comprender los mecanismos humanos relacionados con el lenguaje (Cortez, Vega, & Pariona, 2009)

3.2.2. Análisis de Sentimiento

El análisis de opinión, sentimiento y contenido (Sentiment Analysis) es un área de investigación del Procesamiento del Lenguaje Natural que cuenta con más de 15 años de desarrollo (Moreno-Sandoval, 2014). Es el campo de estudio que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, eventos, temas y sus atributos (Liu, 2012b)

Según Medhat, Hassan y Korashy (Medhat et al., 2014a) señalan que el análisis de sentimiento hace uso de técnicas de clasificación de sentimiento de las cuales las principales son las técnicas enfoque de aprendizaje automático y las técnicas de enfoque basado en léxico.

3.2.3. Flujo de Trabajo del Análisis de Sentimiento

Son procedimientos que dependen en gran medida de varios aspectos, como puede ser el tipo de texto de entrada, el tipo de resultados que se desean obtener o qué información sobre los sentimientos queremos que nos provean y finalmente qué métodos de clasificación se desean utilizar. El esquema de trabajo clásico dentro del análisis de sentimiento inicia con el pre procesamiento de texto seguido de la etapa de procesamiento (Vélez Román, Zambrano Zambrano, & Zambrano Zambrano, 2017) .

3.2.3.1. Pre procesamiento

Según Luciana Dubiau (Dubiau & Arte, 2013), consiste en una secuencia de transformaciones en los textos de entrada antes de su procesamiento. Sus transformaciones dependen de las tareas del procesamiento de texto que se esté realizando. Además, señala que el objetivo del pre procesamiento es mejorar los resultados del clasificador.

3.2.3.2. Procesamiento de texto

Dannyll, Darío y María (Vélez Román, Zambrano Zambrano, & Zambrano Zambrano, 2017), esta etapa puede comprender una serie de tareas, técnicas de clasificación y de diversas metodologías para evaluar la opinión del texto.

3.2.4. Tareas del Pre procesamiento de Texto

Para Roberto Hernández Petlachi y Xiaoou Li (Hernández Petlachi & Li, 2014) definieron realizar ciertas correcciones al texto en la fase de pre procesamiento , algunas de ellas son las siguientes :

3.2.4.1. Corrección de signos de puntuación

Es importante hacer resaltar que los signos de puntuación en el análisis de sentimiento son muy importantes debido a que si se colocan de manera incorrecta, pueden afectar negativamente al resto del proceso (Hernández Petlachi & Li, 2014).

3.2.4.2. Corrección de abreviaturas

Se sustituyen algunas de las palabras abreviadas más habituales por su forma gramatical reconocida (por ejemplo: “q” → “que”, “xq” → “porque”) (Hernández Petlachi & Li, 2014).

3.2.4.3. Corrección de palabras

Para la corrección ortográfica utilizamos el algoritmo de Levenshtein con su noción de distancia (Hernández Petlachi & Li, 2014).

Para Hugo Librado Jacobo (HUGO, 2016) considera que debe tener la fase pre procesamiento ciertas tareas a cumplir algunas de ellas son:

3.2.4.4. Normalización y acentuación de términos

La normalización de términos se aplica a aquellos que han sido modificados por la repetición de caracteres y signos de puntuación. Por ejemplo, las palabras “largooooo...” y “Doloooo”, de las cuales es fácil deducir que equivalen al adjetivo “largo” y al nombre “dolor”. La normalización hace necesaria una función que reduzca el número de caracteres repetidos y elimine los signos de puntuación, a excepción de algunas letras, las cuales se usan de forma doble y son comunes en muchas palabras del idioma español, estas letras son “c”, “e”, “l” y “r”. Debido a que en el idioma español se usa una acentuación gráfica, las vocales acentuadas cambian su valor numérico y por lo tanto cambian al término, de ahí que una computadora considere diferente la palabra “reelección” de la palabra “reeleccion”, en consecuencia, aunque “reelección” se encuentre dentro del vocabulario de la computadora, la computadora establece que “reeleccion” no se encuentra, haciendo necesario acentuar las palabras (HUGO, 2016).

3.2.4.5. Eliminación de texto no útil

Eliminación de texto que no aporta información relevantes (HUGO, 2016).

3.2.4.6. Algoritmo de Levenshtein

La metodología que desarrolló Vladimir Levenshtein [30] en los años 60 es ampliamente conocida y usada en la corrección de textos. Básicamente consiste en realizar operaciones de sustitución, inserción y extracción de caracteres encontrar el número mínimo cambios o la distancia mínima entre 2 palabras (HUGO, 2016).

3.2.5. Técnicas de Clasificación de Sentimientos u Opinión

Según Medhat, Hassan y Korashy (Medhat et al., 2014a) las técnicas de clasificación de sentimiento se pueden dividir en enfoque de aprendizaje automático, enfoque basado en léxico y enfoque híbrido. Además, señalan que el Enfoque Basado en Lexicón se divide en dos las cuales son: Enfoque basado en diccionario y Enfoque basado en corpus.

Según Natalia Paola Hernández Muñoz (MUÑOZ, 2016) el enfoque no supervisado basado en un lexicón (en inglés, unsupervised lexicon-based approaches) es también llamado enfoque semántico, o en inglés semantic-based approaches.

Javier Mateo Fernández (Fernández, 2016) señala que los enfoques semánticos, se caracterizan por el uso de diccionarios de términos (lexicons) con orientación semántica de polaridad u opinión. Típicamente los sistemas pre procesan el texto y lo dividen en palabras, con la apropiada eliminación de las palabras de parada y una normalización lingüística por stemming o lematización, y luego comprueban la aparición de los términos del lexicon para asignar el valor de polaridad del texto mediante la suma de valores de polaridad de los términos.

3.2.5.1. Enfoque Basado en Diccionario

Medhat, Hassan y Korashy (Medhat et al., 2014a) describe como, un pequeño conjunto de palabras de opinión se recopila manualmente con orientaciones conocidas. Luego, este conjunto crece al buscar en los conocidos corpus WordNet o tesoro sus sinónimos y antónimos. Las palabras recién encontradas se agregan a la lista de semillas y luego comienza la siguiente iteración. El proceso iterativo se detiene cuando no se encuentran nuevas palabras. Una vez completado el proceso, se puede realizar una inspección manual para eliminar o corregir errores.

3.2.6. Niveles del Análisis de Sentimiento

Según Lui (Liu, 2012b), señala que el análisis del sentimiento se ha investigado principalmente en tres niveles: Nivel de Documento (Document Level), Nivel de frase u oración (Sentence Level) y Nivel de Aspecto y Entidad (Entity and Aspect Level).

3.2.6.1. Nivel de Documento

Se asume que cada documento expresa opiniones sobre una única entidad u objeto, así como que cada opinión es emitida por un solo emisor o autor (Rincón García, 2016)

3.2.6.2. Nivel de Oración

La tarea en este nivel va a las oraciones y determina si cada oración expresó una opinión positiva, negativa o neutral. Neutral por lo general significa que no hay opinión. Este nivel de análisis está estrechamente relacionado con la clasificación de la subjetividad que distingue oraciones que expresan información objetiva de las que expresan información subjetiva (Liu, 2012b).

3.2.6.3. Nivel de Aspecto y Entidad

Este es el nivel de análisis de grano más fino de las líneas de investigación actuales, ya que es el que consigue extraer más información de las opiniones. En vez de atender a las construcciones de lenguaje (documentos, párrafos, frases, etc.) se centra en la opinión directamente, bajo la idea de que una opinión está compuesta de un sentimiento (positivo o negativo) y un objetivo (Rincón García, 2016) .

3.2.7. Intensificadores en el análisis de sentimientos

En el análisis de sentimientos existen términos que hacen reducir o aumentar el sentimiento total del mensaje a analizar. Estos, no siempre son fáciles de identificar y pueden variar dependiendo del contexto (Montesinos García, 2014).

Los intensificadores son un grupo de modificadores emocionales a tener en cuenta cuando se plantea desarrollar un sistema de análisis de sentimientos. Los intensificadores se definen como términos capaces de ampliar o disminuir la intensidad emocional del texto al que afectan. La intensificación basada en porcentajes propone que cada intensificador tenga un porcentaje asociado, positivo si es amplificador y negativo si es decrementador. Esto permite que las oraciones semánticas se modifiquen de forma que se tenga en cuenta las polaridades de los elementos involucrados. El mecanismo seguido en el sistema para la detección de intensificadores en los textos consiste en la comparación de los términos de cada oración con todas las entradas de la lista de intensificadores proporcionada por el diccionario SODictionariesV1.11Spa. Cuando el sistema detecta una

coincidencia, comprueba a su vez que el rol gramatical del término identificado es el adecuado, y si es así, este es marcado como intensificador. Posteriormente se le suma el porcentaje ya sea para incrementar o decrementar la palabra (Hernández Petlachi & Li, 2014).

3.2.8. Negación en el análisis de sentimientos

La detección de la negación es un proceso complejo que no solo abarca la identificación de las palabras o términos que expresan la negación, sino también su ámbito de acción; es decir, las palabras afectadas por dicha negación. La negación es una de las construcciones lingüísticas que más influye en el análisis de polaridad de los textos. La negación es una construcción habitual a la hora de expresar una opinión. Una forma de negar una idea en el español es mediante el término “no”, aunque existen otros negadores como “tampoco”, “sin” y “nunca” que son también utilizados con frecuencia (Hernández Petlachi & Li, 2014).

CAPITULO 4: METODOLOGÍA DE INVESTIGACIÓN

4.1. Hipótesis de investigación

4.1.1. Hipótesis General

El uso del Software SISAM influye significativamente en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON.

4.1.2. Hipótesis Específicas

- El nivel de análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.
- El flujo de trabajo del análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.
- El diccionario léxico influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.

4.2. Operacionalización de Variables

Tabla 1 Matriz de Operacionalización de Variables

MATRIZ DE OPERACIONALIZACIÓN DE LAS VARIABLES			
VARIABLE	DEFINICION	DIMENSION	INDICADORES
Software (SISAM) Variable Independiente	<p>Es un software de aplicación que comprende un conjunto de componentes lógicos necesarios que hacen posible la realización de tareas específicas. Está diseñado para encontrar y verificar existencia de contenidos denigrantes hacia la mujer en las letras de las canciones. El software hace uso del Enfoque Basado en Diccionario del Análisis de Sentimiento.</p> <p>Análisis de Sentimiento Es el campo de estudio que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, temas, eventos, temas y sus atributos (Liu, 2012a).</p> <p>Enfoque Basado en Diccionario Según Medhat, Hassan y Korashy (Medhat, Hassan, & Korashy, 2014b) es conocido en inglés como Dictionary-Based Approach y es un tipo de técnica de clasificación de sentimientos usada en el análisis de sentimiento que pertenece al Enfoque basado en Léxico (Lexicon-based Approach). Según Fernández (Fernández, 2016) este tipo de enfoque se caracteriza por el uso de diccionarios de términos (palabras o frases) con orientación semántica.</p>	Flujo de Trabajo del Análisis de Sentimiento	Numero de Contenidos denigrantes
		Diccionario Léxico	Numero de Aciertos
		Nivel de Análisis de Sentimiento	Documento (Texto)
Contenidos de Denigrantes hacia la Mujer (Variable Dependiente)	<p>Antes de definir lo que son Contenidos Denigrantes a la Mujer se debe entender que</p> <p>Según Ruiz (Ruiz, 2008) Denigración a la mujer es el acto de atacar su entorno; es decir es el acto de generar circunstancias o situaciones incómodas para la mujer ya sea en lo social, laboral, cultural y demás. Creando un ambiente desagradable y peligroso para la mujer.</p> <p>Por lo cual se define que los contenidos denigrantes a la mujer son aquellas palabras, oraciones y frases del texto que a través de su significado o mensaje generan situaciones que afectan el entorno de la mujer.</p> <p>Según Lissette (Lozano Salas, 2016) señala que los mensajes del género reggaetón si afectan a la mujer en su entorno social .Además señala que las canciones con mensajes de violencia a la mujer son una de las causas.</p>	Palabras y/o frases que transmitan mensajes de violencia física a la mujer	Numero de Palabras y/o frases
		Palabras y/o frases que transmitan mensajes de violencia sexual a la mujer	Numero de Palabras y/o frases
		Palabras y/o frases que transmitan mensajes de violencia psicológica a la mujer.	Numero Palabras y/o frases
		Palabras y/o frases que transmitan mensajes de violencia simbólica a la mujer.	Numero Palabras y/o frases

Fuente: Elaboración propia.

4.3. Tipo de Investigación

El tipo de investigación de este proyecto es Aplicada.

4.4. Nivel de Investigación

El nivel de investigación de este proyecto es Experimental

4.5. Diseño de Investigación

El diseño de la Investigación es Pre experimental; diseño que realiza la aplicación de un tratamiento para después observar los resultados obtenidos.



Figura 1. Diseño de Investigación

Como se puede observar en la figura 1; se manipulará a la variable independiente X con ellos se obtendrá las O observaciones

G1: El grupo experimental

4.6. Población y Muestra

4.6.1. Población

En la actualidad existen tantas canciones del género reggaetón que se hace difícil de contar, pero no imposible por ello la población será Infinita.

4.6.2. Muestra

La muestra será probabilística, ya que se incluirá una cantidad de productos que serán escogidos aleatoriamente. La muestra resulta ser 384 canciones del género reggaetón. Pero se consideró hacer usar una muestra mayor de 404 canciones

Ecuación:

$$n = \frac{Z^2 * p * q}{e^2} ; \frac{1.96^2 * 0.5 * 0.5}{0.05^2} = 384 \text{ canciones}$$

Z : Nivel de Confianza; en este caso sera 1.96 que es por defecto

p : Porcentaje de la poblacion que tiene el atributo deseado por defecto es 0.5

q : Porcentaje de la poblacion que no tiene lo deseado por defecto es 0.5

e : Error e estimacion ; maximo aceptado por defecto es 5%

n : Tamaño de Muestra

4.7. Método de Investigación

La investigación usa un método cuantitativo ya que se quiere probar y medir que el software permite la detección de contenidos denigrantes a la mujer en las letras musicales de canciones del género reggaetón.

Para el desarrollo del software se hace uso de una de las metodologías ágiles de desarrollo de software que es Programación Extrema (XP) ya que el software tiene un uso específico y se realizará ajustes constantes en su desarrollo. Para mayor detalle sobre la metodología de desarrollo ágil Programación Extrema(XP) se encuentra en el Anexo 1.

4.8. Técnicas e Instrumentos de Acopio de Datos

La recolección de datos que en este caso son letras de canciones del género reggaetón las técnicas son registros.

- a. Para la dimensión **Palabras y/o frases que transmitan mensajes de violencia física** se llevará a cabo un registro de canciones del género reggaetón que muestre actos de violencia física a la mujer en las letras de las canciones; esto se realizará a través de búsqueda de noticias, repositorios, páginas webs y blogs.
- b. Para la dimensión **Palabras y/o frases que transmitan mensajes de violencia sexual** se llevará a cabo un registro de canciones del género reggaetón que muestre actos de violencia sexual a la mujer a la mujer en las letras de las canciones; esto se realizará a través de búsqueda de noticias, repositorios, páginas web con letras de canciones y blogs.
- c. Para la dimensión de **Palabras y/o frases que transmitan mensajes de violencia psicológica** se llevará a cabo un registro de canciones del género reggaetón que muestre actos de violencia psicológica a la mujer en las letras de las canciones; esto se realizará a través de búsqueda de noticias, repositorios, páginas web con letras de canciones y blogs.

- d. Para la dimensión de **Palabras y/o frases que transmitan mensajes de violencia simbólica** se llevará a cabo un registro de canciones del género reggaetón que muestre actos de violencia simbólica a la mujer en las letras de las canciones; esto se realizará a través de búsqueda de noticias, repositorios, páginas web con letras de canciones y blogs.
- e. Para aumentar y completar la cantidad de la muestra definida en el proyecto se realizará la búsqueda de las canciones más escuchadas del género reguetón; esto a través de lista que publica cada plataforma de streaming music cada año y que lo realiza a través de encuestas.

4.9. Técnicas de Análisis de Datos

- a. Análisis de Sentimiento
- b. Se aplicará el software al grupo experimental
- c. Se comparará y analizará la información obtenida de las pruebas para sacar las conclusiones de la investigación.
- d. Por último, se elaborará un informe final de todos los datos, información, que se ha obtenido con la investigación.

CAPITULO 5: RESULTADOS

5.1. Descripción de la solución desarrollada

Para saber cómo el uso del Software SISAM (basado en el análisis de sentimiento) influye en la detección de contenidos denigrantes hacia la Mujer; nos amparamos en el Flujo de Trabajo del Análisis de Sentimiento que esta implementado en el software SISAM para detectar dichos contenidos. Este flujo de trabajo consta de las siguientes etapas:



Figura 2. Etapas del Flujo de Trabajo del Análisis de Sentimiento

Fuente: Elaboración propia.

Anteriormente a la ejecución de este flujo se define crear ciertas listas necesarias y que están implementadas en el software las cuales son: lista de contenidos denigrantes y lista de Intensificadores.

5.1.1. Listas de Contenidos Denigrantes

Se crea 4 listas de contenidos denigrantes cada lista está relacionada con un tipo de violencia ya especificada anteriormente en este informe los cuales son: Violencia Física (VF), Violencia Psicológica (VPS), Violencia Sexual (VSE) y Violencia Simbólica (VSIM).

Cada lista contiene una cantidad de contenidos que están relaciones con el tipo de violencia como se muestra en la siguiente tabla:

Tabla 2. Cantidad de contenidos de cada tipo de violencia

VF	VPS	VSE	VSIM
12	16	107	167

Fuente: Elaboración propia.

También se puso un valor a cada contenido; esto valores se midieron desacuerdo a criterios definidos que se muestra en la siguiente tabla:

Tabla 3. Valor de Mensaje y criterios

Valor	Criterios
1	Poco Denigrante
2	Denigrante
3	Muy Denigrantes

Fuente: Elaboración propia.

A continuación, se definen estos criterios:

- **Poco Denigrante**
Es cuando el contenido es muy subjetivo o menos insultante; es decir puede ser denigrante dependiendo al uso o punto de vista, como también puede que no representa un insulto muy desagradable para uno, sino uno común que suelen usar.
- **Denigrante**
Es cuando el mensaje es subjetivo; es decir se le suelen decir a la mujer, pero depende del contexto y el ambiente en que se le dice.

- **Muy Denigrante**

Es cuando el contenido o palabra no usualmente se escucha decirle a una mujer y contiene un mensaje altamente denigrante.

5.1.2. Listas de Intensificadores

Para crear estas listas se hace uso del diccionario SODictionariesV1.11Spa. que contiene una lista de intensificadores relacionados con un valor decimal y puede ser negativo o positivo. Estos son agrupados por número de palabras; es decir intensificadores de 3 palabras, intensificadores de 2 palabras e intensificadores de 1 palabra, dando como resultado la creación de tres listas. A continuación, se muestra en una tabla la cantidad de intensificadores en cada lista.

Tabla 4. Cantidad de Intensificadores por número palabra

Lista de Intensificadores 1 Palabra	Lista de Intensificadores 2 Palabras	Lista de Intensificadores 3 Palabras
122 Intensificadores.	28 Intensificadores.	12 Intensificadores.

Fuente: Elaboración propia.

5.1.3. PRE PROCESAMIENTO

En esta etapa se hace una limpieza y corrección del texto. Ya que se detectó que las letras de canciones del genero reggaetón encontradas en páginas que las publican, suelen tener errores ortográficos, dobles o más espacios en blanco, triple o más saltos de línea, caracteres innecesarios para el análisis y abreviaciones de palabras. Por ello se definió realizar los siguientes procesos en esta etapa los cuales son: eliminación de caracteres innecesarios, eliminación de espacios y saltos de línea innecesarios, normalización y corrección ortográfica.

5.1.3.1. Eliminación de caracteres innecesarios

La lista de caracteres innecesarios se creó de acuerdo a los caracteres que se pudo observar que aparecen en los textos. Busca detectar los caracteres para luego reemplazarlos por espacios en blanco.

Se muestra un ejemplo:

Texto antes de ser eliminado los caracteres:

ella se dejo llevar (se dejo llevar),

Texto luego de ser eliminado los caracteres:

ella se deajo llevar se deajo llevar

5.1.3.2. Eliminación de espacios y saltos de línea innecesarios

Busca detectar en el texto dos o más espacios en blanco seguidos; siendo luego reemplazados un espacio en blanco. También busca detectar tres o más saltos de línea seguidos para luego reemplazarlos por dos saltos.

5.1.3.3. Normalización

Se creó una lista de abreviaturas y sus palabras de acuerdo a lo que se pudo observar en los textos y las que se suelen escuchar en las canciones. Busca estas abreviaciones y las reemplaza por las palabras correspondientes. Además, realiza otra tarea también como es la de corregir palabras con repeticiones de caracteres (vocales y consonantes) a excepción de las que se usan usualmente en las palabras como son: “c”, “e”, “l” y “r”.

Se muestra un ejemplo:

Texto antes de ser normalizado:

Y **terminamo'** haciéndolo.

Texto luego de ser normalizado:

y **terminamos** haciéndolo

5.1.3.4. Corrección ortográfica

Hace uso del algoritmo de Levenshtein para la corrección ortográfica del texto y del Diccionario Corpus de 5000 palabras Referencia del Español Actual (CREA), además de otras 165 palabras más, haciendo un total de 5165 palabras. La forma como función es muy sencilla cada palabra es analizada en el algoritmo con cada palabra del diccionario si durante el proceso el algoritmo devuelve como resultado 0; quiere decir que la palabra buscada está bien escrita ya que lo encontró en el diccionario así que ya no sigue buscando. Otro caso es cuando devuelve 1; quiere decir que ha encontrado una palabra parecida a la palabra buscada y si al finalizar de

compararlo con todas las palabras no encuentra alguna palabra parecida la reemplaza por la primera palabra que devolvió como número el uno.

Se muestra un ejemplo:

Texto antes de ser corregido:

la **musica** que nos entona.

Texto luego de ser corregido:

la **música** que nos entona

Aunque el algoritmo presenta buenos resultados en su corrección también tiene ciertos problemas al momento de analizar una palabra que debería llevar tilde, ya que en la mayoría de los casos devuelve un valor mayor a uno al compararlo con la palabra correcta. Por ello en algunos casos no corrige de manera correcta la filmación.

5.1.4. PROCESAMIENTO

En esta etapa se realiza el análisis. Esta Etapa inicia en el momento que el usuario le da en el botón análisis del software. Esta etapa del Análisis se divide en tres procesos que trabajan de manera coordinada:

5.1.4.1. Detección de contenidos

Este proceso se divide en dos subprocesos las cuales son:

- **Conteo**

Antes de hacer el conteo de contenidos, este subproceso divide el texto en líneas de texto para hacer la búsqueda de los contenidos de cada tipo por línea de texto. Es así que se realiza 4 conteos y búsquedas; esto es por cada tipo de violencia. Este subproceso pasa como dato al subproceso siguiente que es la extracción de contenidos una lista de las cantidades de contenidos que presenta el texto de cada tipo de violencia.

- **Extracción de Contenido**

También este subproceso divide el texto en líneas de texto para hacer la extracción de los contenidos por línea de texto. Este subproceso se ejecuta ni bien el subproceso conteo termina de ejecutarse, este subproceso recibe de su antecesor una lista de conteo en donde se ve cuantos contenidos de cada tipo de violencia tiene el texto ingresado. Luego de tener este dato seguidamente empieza con la extracción de esos contenidos de cada tipo y además extrae el número de línea donde se encontró el contenido y la cantidad de veces que éste aparece en la misma línea de texto. Durante la extracción de cada contenido detectado, este subproceso envía cada contenido y la línea de texto donde se ubica dicho contenido, a los procesos de detección de negación y de intensificadores; esto es para que verifiquen que si existe alguna negación e intensificador que pueda afectar o anular al contenido. A final termina guardando los contenidos que no hayan sido anulados dependiendo al valor que devuelvan los resultados de los procesos de detección de negación y guarda el intensificador que se detectó que le afecta, además también se inserta su número de línea de texto donde aparece, su valor y las veces que se repite en la línea de texto.

5.1.4.2. Detección de Negación

Para la detección de negación se definió también 4 tipos de negación dependiendo a 4 casos en los cuales pueden aparecer.

- **Negación de Tipo 1**

Es cuando se encuentra una palabra negativa antes del contenido denigrante; en este caso se considera estas palabras negativas: no, nunca, jamás y tampoco.

- **Negación de Tipo 2**

Es cuando se encuentra la conjugación de NO + VERBO SER antes del contenido denigrante; en este caso se

consideró estas conjugaciones además de otras tres adicionales: "No es", "No eres", "No son", "No serás", "No quiero", "No deseo", "No debo", "No debería".

- **Negación de Tipo 3**

Es cuando se encuentra la conjugación de NO – NUNCA - JAMAS + NI antes del contenido denigrante. A continuación, un ejemplo:

No te voy a pegar ni voy a violarte

- **Negación de Tipo 4**

Es cuando se encuentra la palabra SIN antes del contenido denigrante.

Estos tipos de negaciones están diseñados en base a los distintos casos en que se podría negar los contenidos denigrantes. Ya que se dan casos en que existen la repetición del contenido en la misma línea de texto se crea una regla para este caso la cual es: Para que sea anulado por completo el contenido y sus repeticiones; todas ellas deben ser negados caso contrario se considerará aun el contenido.

5.1.4.3. Detección de Intensificadores

Se define dos tipos de intensificadores dependiendo a dos casos en los cuales pueden aparecer.

- **Intensificador de Tipo 1**

Es cuando el intensificador se encuentra dentro del contenido denigrante.

- **Intensificador de Tipo 2**

Es cuando el intensificador se encuentra antes del contenido denigrante.

Los Intensificadores cumplen la función de aumentar el peso emocional del contenido denigrante o también disminuirlo. Por ello en el caso de que algún intensificador afectara a algún contenido o alguna repetición del contenido; esto hará que al valor total del contenido se le sume el valor del intensificador.

La implementación de la metodología del de la metodología ágil Programación Extrema(XP) en el desarrollo del software SISAM se encuentra en el Anexo 2.

5.2. Presentación de resultados

A continuación, se presenta los resultados habiendo usado como muestra 404 canciones del genero reggaetón.

Tabla 5.Resultados del Análisis de las canciones

Nº Canciones con Contenidos	Nº Canciones sin Contenidos	Total de Canciones
245 canciones	159 canciones	404 canciones

Fuente: Elaboración propia

Seguidamente se muestra una gráfica de la representa porcentual de los resultados del análisis.

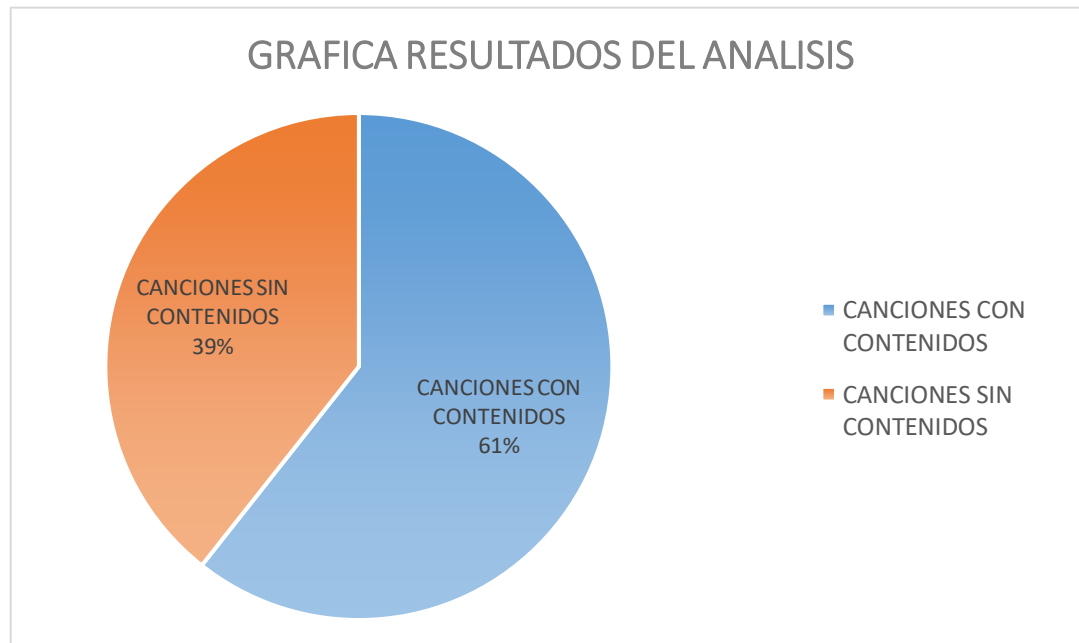


Figura 3.Grafica de estadística circular del resultado del análisis

Fuente: Elaboración propia

A continuación, se muestra cantidad de contenidos detectados de cada tipo de violencia.

Tabla 6. Cantidad de contenidos por tipo de violencia

V. FISICA	V. PSICOLOGICA	V. SEXUAL	V. SMBOLICA	TOTAL
18 Cont.	165 Cont.	329 Cont.	533 Cont.	1045 Cont.

Fuente: Elaboración propia

A continuación, se muestra una gráfica estadística hecha con los datos de la Tabla 6 respecto a la cantidad de cada tipo de violencia.

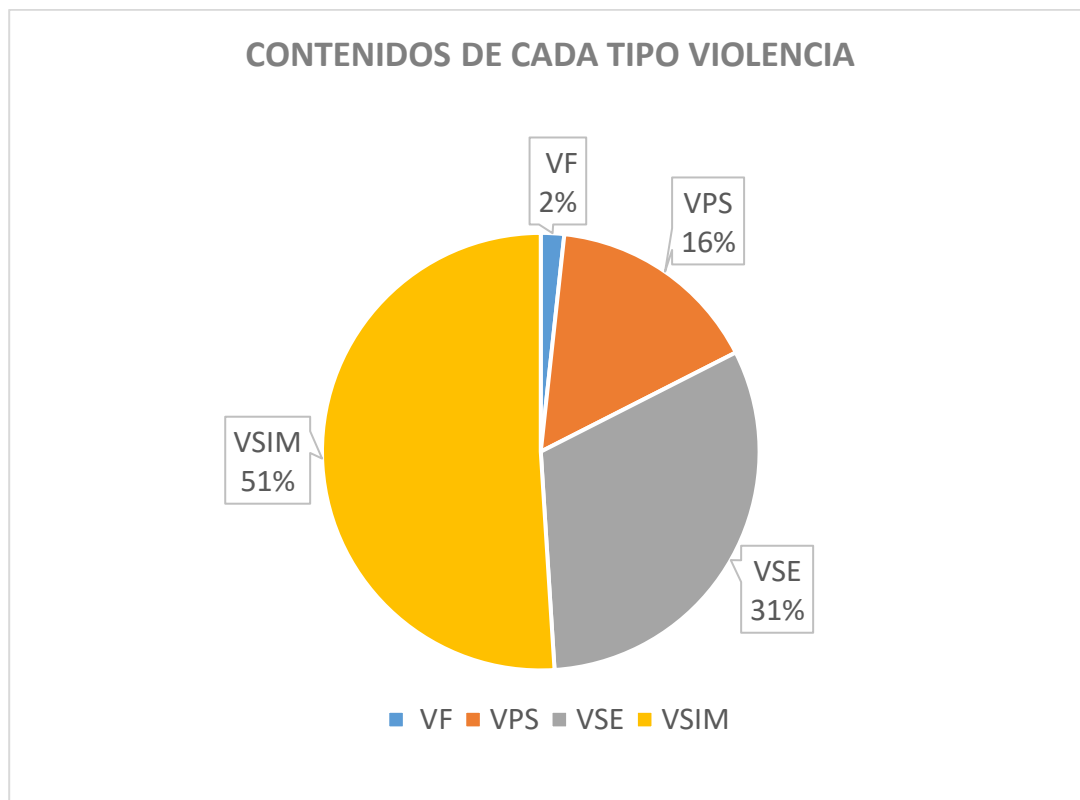


Figura 4. Grafica de estadística circular del total de contenidos de cada tipo violencia

Fuente: Elaboración propia.

A continuación, se muestra la cantidad de canciones de cada tipo de violencia.

Tabla 7. Cantidad de canciones por tipo de violencia

V. FISICA	V. PSICOLOGICA	V. SEXUAL	V. SMBOLICA
10 Canc.	51 Canc.	120 Canc.	167 Canc.

Fuente: Elaboración propia

A continuación, se muestra una gráfica estadística hecha con los datos de la Tabla 7 respecto a la cantidad de cada tipo de violencia.

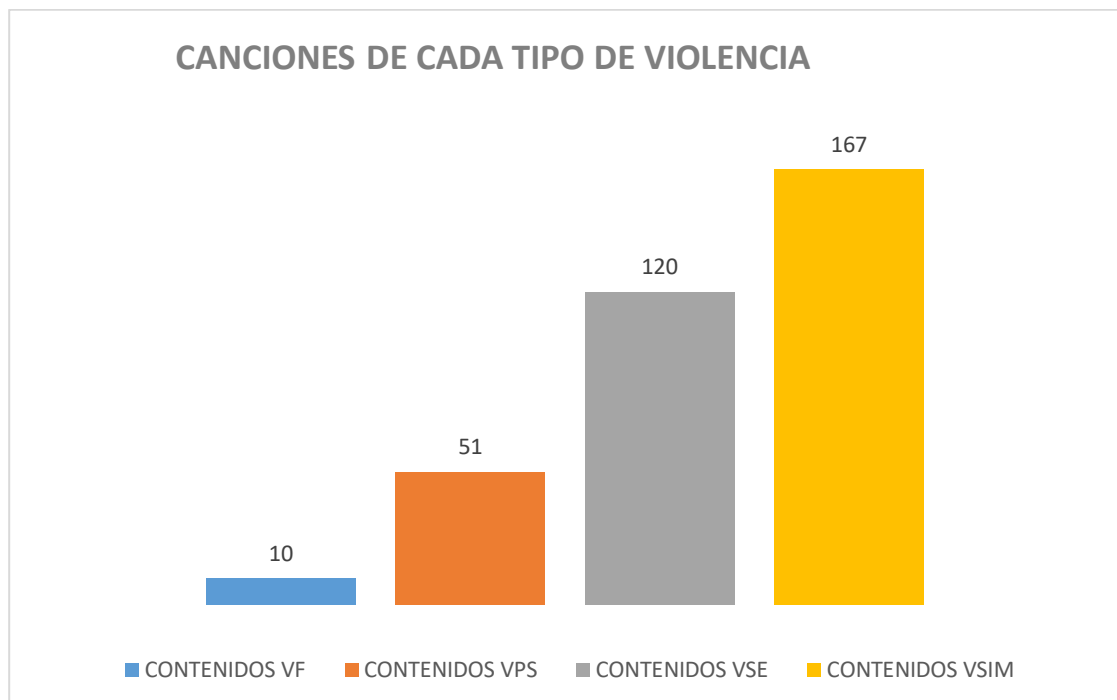


Figura 5. Gráfica de barras de la cantidad de canciones de cada tipo de violencia

Fuente: Elaboración propia

CAPITULO 6: DISCUSIÓN

6.1. Análisis y discusión de resultados

El análisis se realizó con una muestra de 404 canciones. Los resultados obtenidos del análisis mostraron que el software detectó que más de la mitad del total de canciones tienen contenidos denigrantes, esto resultados se muestran en la Tabla 5. Es así que del 100% de canciones ingresadas al software, el 61% de estas presentan contenidos denigrantes es así como se muestra en la Gráfico 1.

6.2. Comprobación de hipótesis

6.2.1. Comprobación de las hipótesis específicas

Hipótesis específica 1

H1: El nivel de análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.

De acuerdo a los resultados que muestra la tabla 5, el nivel de análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer ya que el análisis a nivel de documento realizado a cada canción en el software dio como resultado que del total de canciones el 61% presentaba contenidos denigrantes. Por lo tanto, la hipótesis H1 es verdadera

Hipótesis específica 2

H2: El flujo de trabajo del análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.

De acuerdo a los resultados obtenidos, el flujo de trabajo del análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón. Ya que el flujo de trabajo del análisis de sentimiento que consta de las fases de pre procesamiento y procesamiento, como también de sus técnicas y procesos dieron como guía al desarrollo y elaboración de los procedimientos y tareas del software. Por lo cual apoyaron de manera significativa en la detección de los contenidos que como se muestran en la tabla 6 ha dado como resultado un total de 1045 contenidos denigrantes detectados del total de canciones analizadas. Por lo tanto, la hipótesis H2 es verdadera

Hipótesis específica 3

H3: El diccionario léxico influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón.

De acuerdo a los resultados obtenidos, el diccionario léxico influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género Reggaetón, ya que con el diccionario léxico elaborado donde este está dividido en 4 listas, una por cada tipo de violencia. Es éste diccionario quien también apoyo en la detección, dando como resultado tener un total de 1045 aciertos como se muestra en la tabla 6. Por lo tanto, la hipótesis H3 es verdadera

6.2.3. Comprobación de la hipótesis general

H0: El uso del Software SISAM influye significativamente en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON.

Con los resultados obtenidos que se muestran en las Tablas 5,6 y 7, vemos que el uso del software si influye de manera significativa en la búsqueda, selección y análisis para la detección contenidos denigrantes hacia la Mujer. Por lo tanto, la hipótesis H0 es verdadera

CONCLUSIONES

1. Con los resultados obtenidos que se muestran en las Tablas 5,6 y 7, vemos que el uso del software SISAM influye de manera significativa tanto en la búsqueda, selección y análisis para detectar los contenidos denigrantes hacia la mujer.
2. El nivel de análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer. El análisis a nivel de documento realizado a cada canción en el software dio como resultado poder verificar que del total de canciones analizadas el 61% presente contenidos denigrantes.

3. El flujo de trabajo del análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer; el diseñar un software en base a las fases del análisis de sentimiento, así como algunas de sus técnicas y procedimientos, dio como resultado un total de 1045 contenidos denigrantes detectados del total de canciones analizadas
4. El diccionario léxico influye en la detección de contenidos denigrantes hacia la Mujer; la creación de un diccionario léxico que contenga los distintos contenidos denigrantes y luego agruparlos por tipo de violencia; dio como resultados que se obtengan un total 1045 aciertos en la comparación con el diccionario durante la búsqueda de los contenidos de todas las canciones.

RECOMENDACIONES

Se plantean algunas recomendaciones con el fin de mejorar los resultados obtenidos en la presente investigación. Con el objetivo de mejorar la detección de los contenidos denigrantes hacia la mujer. Estas recomendaciones son:

- ✓ Incrementar el diccionario léxico con más contenidos; cuan mayor sea la cantidad de contenidos mucho mejor serán los resultados.
- ✓ Incluir para análisis los caracteres que permitan añadan más información respecto a los contenidos denigrantes.
- ✓ Incluir un traductor de inglés a español ya que en el reggaetón suelen usar palabras en inglés.
- ✓ Implementar el enfoque de aprendizaje automático del análisis de sentimiento para detección de los contenidos.

REFERENCIAS BIBLIOGRAFICAS

- Arévalo, K., Chellew, E., Figueroa-Cofré, I., Arancibia, A., & Schmied, S. (2018). Ni pobre diablo ni candy: Violencia de género en el reggaetón. *Revista de Sociología*, 33(1), 7–23. <https://doi.org/10.535/0719-529X.2018.51797>
- Baviera Puig, T. (2017). Técnicas para el Análisis de Sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Dígitos: Revista de Comunicación Digital*, 1(3), 33–50. <https://doi.org/10.7203/rd.v1i3.74>
- Cortez, A., Vega, H., & Pariona, J. (2009). Procesamiento de lenguaje natural robusto. *Primer Encuentro de Grupos de Investigación Sobre Procesamiento Del Lenguaje*.
- Dubiau, L., & Arte, E. (2013). *Procesamiento de Lenguaje Natural en Sistemas de Análisis de Sentimientos*.
- Fernández, J. M. (2016). *Análisis de contenidos en Social Media: clasificación de mensajes e identificación de influyentes en el Banco Central Europeo (BCE)*. 70. Retrieved from http://eprints.ucm.es/38544/1/TFM_javmateo_DEF_v3.pdf
- Fojtik, R. (2011). Extreme programming in development of specific software. *Procedia Computer Science*, 3, 1464–1468. <https://doi.org/10.1016/j.procs.2011.01.032>
- Gregorio, R. M., & Jorge, F. Z. (2002). *Programación eXtrema y Software Libre*. 1–23. Retrieved from <http://es.tldp.org/Presentaciones/200211hispalinux/ferrer/robles-ferrer-ponencia-hispalinux-2002.pdf>
- Hernández Petlachi, R., & Li, X. (2014). Análisis de sentimiento sobre textos en Español basado en aproximaciones semánticas con reglas lingüísticas. *Tass 2014*.
- HUGO, L. J. (2016). *Análisis automático de opiniones de productos en redes sociales*.
- José, J. (2008). Reglas y prácticas en eXtreme Programming. *Universidad de Vigo. España*, 1–22. Retrieved from <http://iie.fing.edu.uy/~josej/docs/XP - Jose Joskowicz.pdf>
- Liu, B. (2012a). *Sentiment Analysis and Opinion Mining*. (May).

- Liu, B. (2012b). SentimentAnalysis-and-OpinionMining. *Opinion Analysis for Online Reviews*, (May), 79–94. https://doi.org/10.1142/9789813100459_0007
- López Barbosa, R. R. (2015). *Aplicación del Análisis de Sentimientos a la Evaluación de Datos Generados en Medios Sociales*. 226. Retrieved from <http://dspace.uah.es/dspace/bitstream/handle/10017/23329/Tesis Rutilio Rodolfo Lopez Barbosa.pdf?sequence=1&isAllowed=y>
- Lozano, L. (2016). *Análisis de los mensajes que transmiten los cantantes reguetoneros y su influencia en los jóvenes de 15 y 19 años en la ciudadela “San Enrique” en el Cantón Duran, provincia del Guayas*. Retrieved from <http://repositorio.ug.edu.ec/bitstream/redug/13999/1/tesis Lissette Lozano finall.pdf>
- Lozano Salas, L. G. (2016). *UNIVERSIDAD DE GUAYAQUIL TEMA : ANÁLISIS DE LOS MENSAJES QUE TRANSMITEN LOS CANTANTES REGUETONEROS Y SU INFLUENCIA EN LOS JÓVENES ENTRE 15 Y 19 AÑOS DE LA CIUDADELA SAN ENRIQUE DEL CANTÓN DURÁN , PROVINCIA DEL GUAYAS EN EL 2016 . Trabajo de titulación que.*
- Marycarmen, D. L., & Antonio, C. G. (2013). *La Programación Extrema*. (November 2013). <https://doi.org/10.13140/RG.2.2.29359.43687>
- Medhat, W., Hassan, A., & Korashy, H. (2014a). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Medhat, W., Hassan, A., & Korashy, H. (2014b). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Melendez, L. (2005). La Violencia contra la mujer: Femicidio en el Perú. *CMP Flora Tristán*, 40. [https://doi.org/10.1016/S0191-8141\(00\)00092-4](https://doi.org/10.1016/S0191-8141(00)00092-4)
- Montesinos García, L. (2014). *Análisis de sentimientos y predicción de eventos en twitter*. Retrieved from <http://repositorio.uchile.cl/handle/2250/130479>
- Moreno-Sandoval, A. (2014). Análisis de opinión y contenido en los medios sociales. *Instituto de Ingeniería Del Conocimiento*, (October), 11. <https://doi.org/10.13140/2.1.3585.7928>
- MUÑOZ, N. P. H. (2016). *METODOLOGÍA PARA EL DISEÑO Y CONSTRUCCIÓN DE UN LEXICÓN DE OPINIÓN, BASADO EN COMENTARIOS DE TWITTER APLICADO AL PROYECTO “OPINIONZOOM.”*
- Noa, K. D. (2018). *El Reggaetón y violencia contra la mujer en la Asociación Juventud Emprendedora para el desarrollo e innovación del Distrito De Cabanillas*. 1–84. Retrieved from http://repositorio.unap.edu.pe/bitstream/handle/UNAP/7448/Noa_Calla_Kather

in_Duanis.pdf?sequence=1&isAllowed=y

- Pérez-Aldeguer, S. (2014). La música como herramienta para desarrollar la competencia intercultural en el aula. *Perfiles Educativos*, 36(145), 175–187. [https://doi.org/10.1016/S0185-2698\(14\)70644-0](https://doi.org/10.1016/S0185-2698(14)70644-0)
- Ramírez Noreña, V. K. (2012). El concepto de mujer en el reggaeton: análisis lingüístico. The concept of women in the reggaeton: linguistic analysis. *Lingüística y Literatura*, 33(62), 227–243. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=4236110&info=resumen&idioma=ENG>
- Rincón García, S. (2016). *Minería de textos y análisis de sentimientos en sanidadysalud.com*. 2015–2016. Retrieved from http://eprints.ucm.es/39524/1/memoriaTFM_sergio_rincon_garcia.pdf
- Rosales Quiroga, M. A., Vilariño Ayala, D., Pinto, D., Tovar, M., & Beltrán, B. (2016). Análisis de sentimientos basado en aspectos: un modelo para identificar la polaridad de críticas de usuario. *Research in Computing Science*, 115(1), 171–180. <https://doi.org/10.13053/racs-115-1-14>
- Ruiz, Y. (2008). La violencia contra la mujer en la sociedad actual: análisis y propuestas de prevención. *Fòrum de Recerca*, 13, 188–200.
- Torres Samboni, L. A. (2015). Análisis de sentimientos sobre el pos-conflicto colombiano utilizando herramientas de minería de texto. *Universidad de Los Andes*, 1–54. Retrieved from <https://repositorio.escuelaing.edu.co/handle/001/403>
- Vélez Román, D. J., Zambrano Zambrano, M. I., & Zambrano Zambrano, D. M. (2017). Innovación para el Análisis de Sentimientos en Texto, una revisión de la técnica actual aplicando metodologías de crowdsourcing. *IX Congreso Internacional de La Red de Investigación y Docencia En Innovación Tecnológica*, 158(July 2017), 138–146.
- VITERI ALVARADO, R. J., & JAVIER, R. (2016). *Análisis De Sentimiento En Twitter De Los Socios De Un Club De Futbol a Través De La Evaluación De Herramienta Que Manejan Gran Volumen De Información*.

ANEXOS

ANEXO 1: Programación Extrema XP

Se basa en una serie de reglas y principios que se han ido gestando a lo largo de toda la historia de la ingeniería del software. Usadas conjuntamente proporcionan una nueva metodología de desarrollo software que se puede englobar dentro de las metodologías ligeras, que son aquéllas en la que se da prioridad a las tareas que dan resultados directos y que reducen la burocracia que hay alrededor tanto como sea posible (Gregorio & Jorge, 2002).

Es una metodología ágil y flexible que hace hincapié en la interconexión de la propuesta y las etapas de implementación. Las actividades básicas son: Planificación y gestión, Diseñando, Codificación y Prueba (Fojtik, 2011).

A. Planificación

El proyecto comienza recopilando “Historias de usuarios”, las que sustituyen a los tradicionales “casos de uso”. Una vez obtenidas las “historias de usuarios”, los programadores evalúan rápidamente el tiempo de desarrollo de cada una. Si alguna de ellas tiene “riesgos” que no permiten establecer con certeza la complejidad del desarrollo, se realizan pequeños programas de prueba (“spikes”), para reducir estos riesgos (José, 2008).

B. Diseño

Según Marycarmen y Antonio (Marycarmen & Antonio, 2013) la metodología XP hace especial énfasis en los diseños simples y claros. Los conceptos más importantes de diseño en esta metodología son los siguientes:

- **Simplicidad**

Un diseño simple se implementa más rápidamente que uno complejo. Por ello XP propone implementar el diseño más simple posible que funcione. Se sugiere nunca adelantar la implementación de funcionalidades que no correspondan a la iteración en la que se esté trabajando.

Características fundamentales del código: Testeable, legible, comprensible y explicable.

- **Metáforas**

Una “metáfora” es algo que todos entienden, sin necesidad de mayores explicaciones. La metodología XP sugiere utilizar este

concepto como una manera sencilla de explicar el propósito del proyecto, y guiar la estructura y arquitectura del mismo.

- **Solución “spike”**

Una solución “spike”, es una solución muy simple para plantear posibles soluciones, de manera, que solamente se aborda el problema en concreto y se aísla de otro tipo de preocupaciones.

- **Refactorización**

La recodificación consiste en escribir nuevamente parte del código de un programa, sin cambiar su funcionalidad, a los efectos de hacerlo más simple, conciso y/o entendible.

C. Desarrollo o Implementación

Según Marycarmen y Antonio (Marycarmen & Antonio, 2013) en la etapa de desarrollo cuenta con:

- **Cliente disponible**

Uno de los requisitos de XP es tener al cliente disponible. No solo para ayudar al equipo de desarrollo, sino para ser parte del mismo. Todas las fases de XP requieren comunicación con el cliente.

Las historias de usuario son escritas por el cliente con la ayuda de los desarrolladores, además de establecer la prioridad de las mismas. Su presencia asegura que los desarrollos cubren toda la funcionalidad descrita.

- **Estándares de codificación**

Todos los programadores deben escribir y documentar el código en la misma manera. El código debe seguir los estándares de codificación. Las normas de codificación ayudan a mantener el código legible y fácil de mantener y re factorizar.

- **Implementación dirigida por las Pruebas Unitarias**

En las metodologías tradicionales, la fase de pruebas, incluyendo la definición de los tests, se realizaba al final del proyecto, o al final del desarrollo de cada módulo. La metodología XP propone un modelo inverso, en el que, lo primero que se escriben son los test que el

sistema debe pasar. Luego, el desarrollo debe ser el mínimo necesario para pasar las pruebas previamente definidas. Las pruebas a las que se refiere esta práctica, son las pruebas unitarias, realizados por los desarrolladores. La definición de estos test al comienzo, condiciona y dirige el desarrollo.

- **Programación en parejas**

XP promueve que todo el código sea escrito en parejas trabajando en el mismo ordenador. La programación en parejas incrementa la calidad del código sin impactar en la fecha de entrega. En contra de lo que parece, dos personas que trabajan en un mismo equipo añadirán la misma funcionalidad que dos personas trabajando por separado, excepto que el código será de mucha mayor calidad.

D. Pruebas

Según José (José, 2008) en la fase de pruebas considera tres actividades a realizar las que son :

- **Pruebas unitarias**

Los módulos deben de pasar las pruebas unitarias antes de ser liberados o publicados.

- **Detección y corrección de errores**

Cuando se encuentra un error, éste debe ser corregido inmediatamente, y se deben tener precauciones para que errores similares no vuelvan a ocurrir.

- **Pruebas de aceptación**

Las pruebas de aceptación son creadas en base a las historias de usuarios, en cada ciclo de la iteración del desarrollo.

ANEXO 2: Implementación de la metodología ágil Programación Extrema(XP)

A continuación, se describe la experiencia obtenida en el desarrollo del software SISAM utilizando la metodología XP.

1. Planificación

Iniciaremos con esta fase de la metodología XP. Dentro de esta fase comenzamos con los Historias de Usuario.

1.1. Historias de Usuario

El cliente es quien a través de esta herramienta puede expresar los requerimientos que son necesarios para la funcionalidad del proyecto. Esta herramienta se enfocan en definir lo que el usuario necesita realizar. En este caso las funcionalidades serán puestas por el desarrollador a continuación se muestra el modelo de historia de usuario. Se consideró 8 historias.

Tabla 8. Modelo de historia de usuario

HISTORIA DE USUARIO		
ID :	Título de Historia :	
Usuario :		
Quiero :		
Para :		
Estimación:	Prioridad :	Valor :

Fuente: Elaboración propia

1.2. Plan de entregas (Release Plan)

Ya que el software tiene un fin específico, se pasó a ordenar las historias de usuarios de acuerdo a sus prioridades y evaluar el tiempo de desarrollo de cada una de ellas. Se consideró todas las historias de usuario; su clasificación fue según el orden de importancia. Se consideró cuatro interacciones y para determinar el tiempo se propuso como medida 10 semanas, cada semana estuvo compuesta de 4 días que son: lunes, martes, miércoles y jueves. Cada día se consideró

trabajar 6 horas. Todas estas se aprobaron por las circunstancias que empezaron a suceder.

1.2.1. Primera Interacción

Para esta primera interacción se eligió 2 historias de usuario las cuales son:

- Corrección de Texto (Hist. Usus. 1)
- Búsqueda, selección y extracción de contenidos (Hist. Usus. 2)

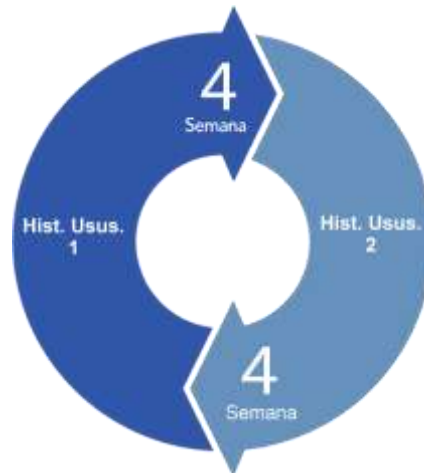


Figura 6. Primera Interacción

Fuente: Elaboración propia

1.2.2. Segunda Interacción

Para esta primera interacción se eligió 2 historias de usuario las cuales son:

- Búsqueda, selección y extracción de intensificadores (Hist. Usus. 3)

- Búsqueda, selección y extracción de negación (Hist. Usus. 4)

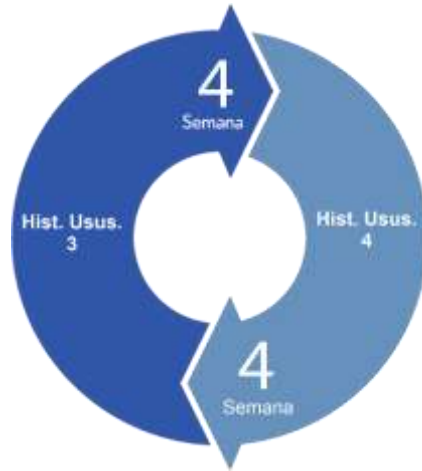


Figura 7. Segundo Interacción

Fuente: Elaboración propia

1.2.3. Tercera Interacción

Para esta primera interacción se eligió 2 historias de usuario las cuales son:

- Visualizar los contenidos por tipo de violencia (Hist. Usus. 5)
- Visualizar los intensificadores y la negación por contenidos (Hist. Usus. 6)

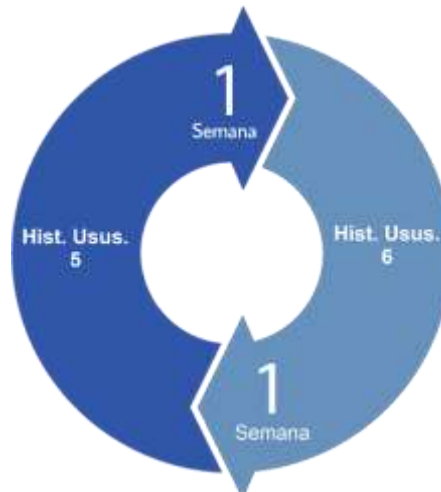


Figura 8. Tercera Interacción

Fuente: Elaboración propia

1.2.4. Cuarta Interacción

Para esta primera interacción se eligió 2 historias de usuario las cuales son:

- Fácil Uso (Hist. Usus. 7)
- Generar Reporte (Hist. Usus. 8)

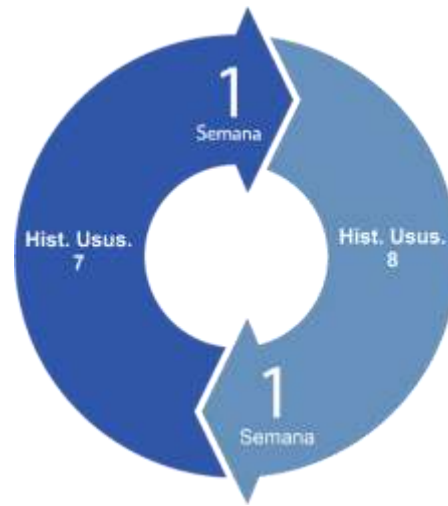


Figura 9. Cuarta Interacción

Fuente: Elaboración propia

1.3. Tiempo de Proyecto

Se midió el tiempo que requiere para la ejecución de cada interacción en función de las historias de usuario, su prioridad y su complejidad de desarrollo.

Tabla 9. Tabla de Interacciones

	Interacción 1	Interacción 2	Interacción 3	Interacción 4
Cant. Horas	96	96	24	24
Cant. Semanas	4	4	1	1
Horas por Semana	24	24	24	24
Historias de Usuario	2	2	2	2

Fuente: Elaboración propia

1.4. Plan de Entregas

Aquí se fijaron las fechas de entregas según el desarrollo de las historias de usuarios para su verificación. Las fechas son las siguientes:

Tabla 10. Cuadro de fechas de entrega

Fecha de Entrega	
1º Interacción	30 de abril
2º Interacción	31 de mayo
3º Interacción	8 de junio
4º Interacción	15 de junio

Fuente: Elaboración propia

2. Diseño

La metodología programación extrema siempre propone implementar el diseño más simple posible.

2.1. Diseños Simples

En esta parte se elaboró el flujo de datos y la interfaz gráfica. Estas facilitan la comprensión del trabajo que realiza el software como su forma de usar.

A continuación, se muestra los diagramas elaborados y las interfaces elaboradas. Empezamos con lo diagramas de flujo de datos.

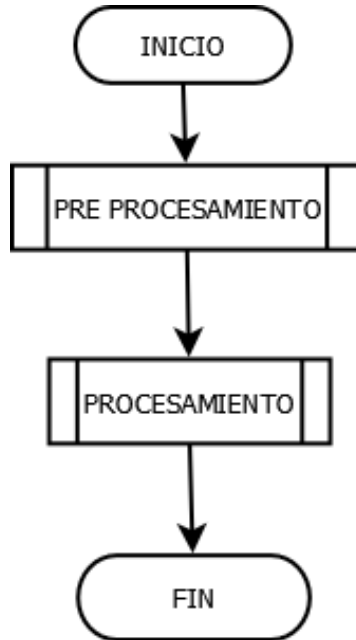


Figura 10. Diagrama de Flujo de datos Principal

Fuente: Elaboración propia

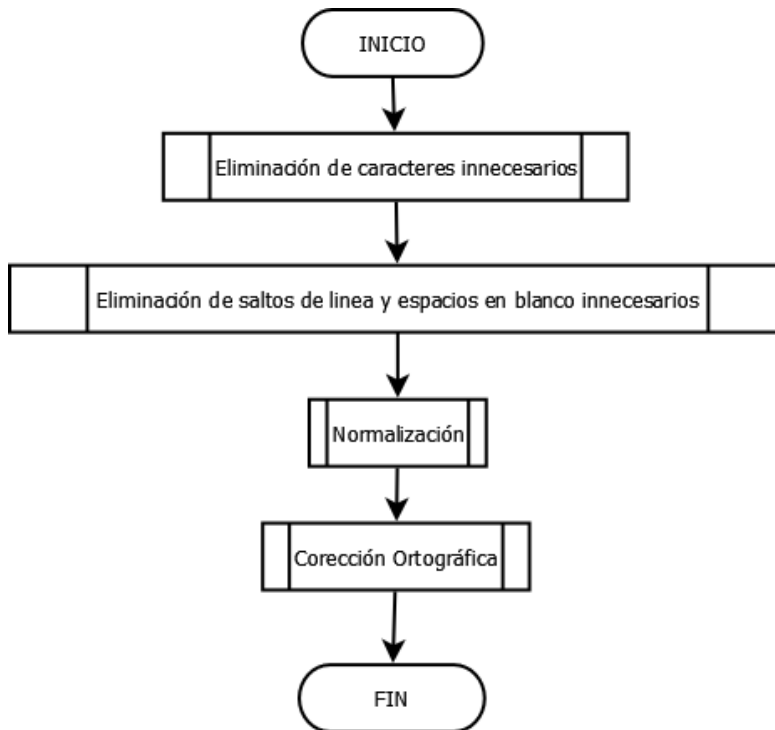


Figura 11. Diagrama de Flujo de datos: Pre procesamiento

Fuente: Elaboración propia

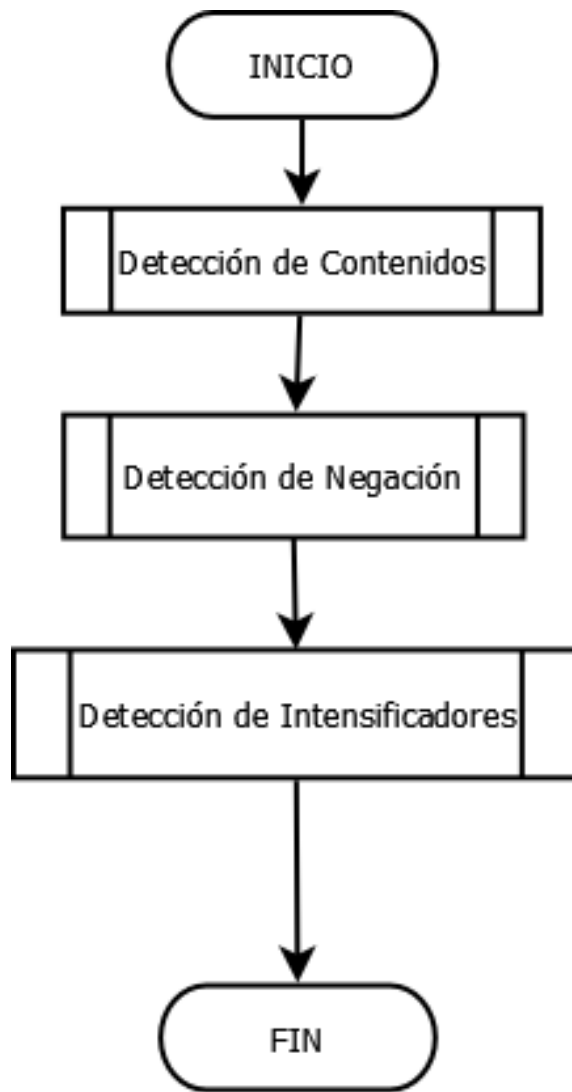


Figura 12. Diagrama de Flujo de datos: Procesamiento

Fuente: Elaboración propia

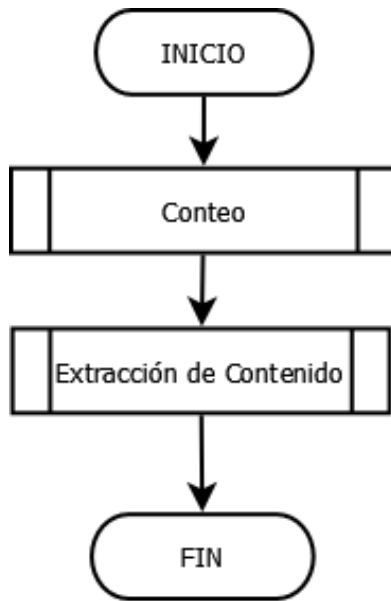


Figura 13. Diagrama de Flujo de datos: Detección de Contenidos

Fuente: Elaboración propia

Ahora se mostrará la maquetación del software para su próxima implementación. En primera instancia se tiene la interfaz de inicial del software. Esto es lo que aparecerá al ejecutar el software.

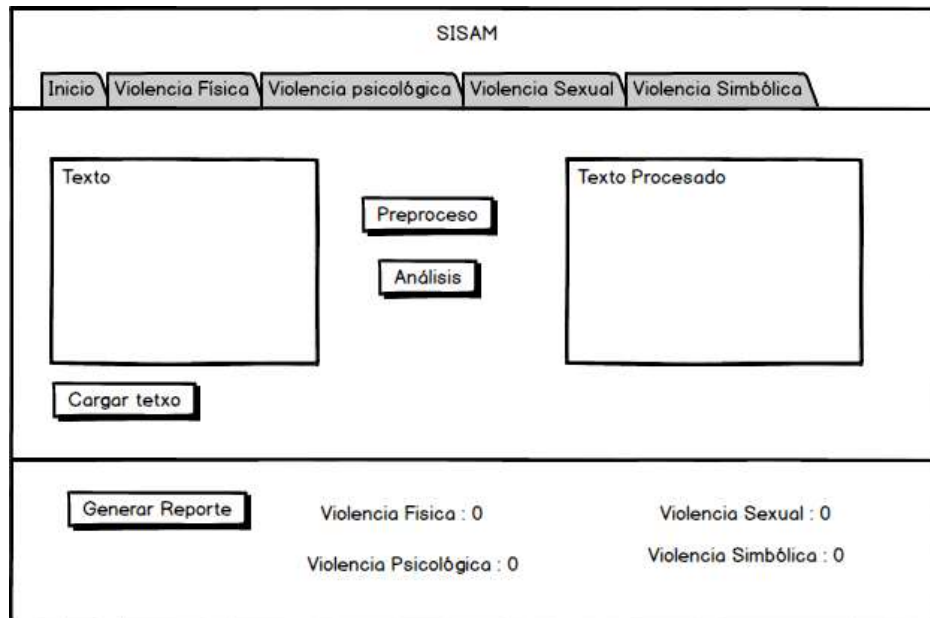


Figura 14. Maquetación de la interfaz inicial del Software

Fuente: Elaboración propia

Como se puede ver en la Figura 6, en la interfaz que se tiene 5 JPanels esto es para poder distribuir los contenidos de cada tipo de violencia física así cada uno puede tener su propio espacio.

Ahora se muestra la maqueta de la interfaz de las Pestaña titulada Violencia Física. Cabe mencionar que las 4 pestañas que llevan el título de los 4 tipos de violencia tienen el mismo diseño.

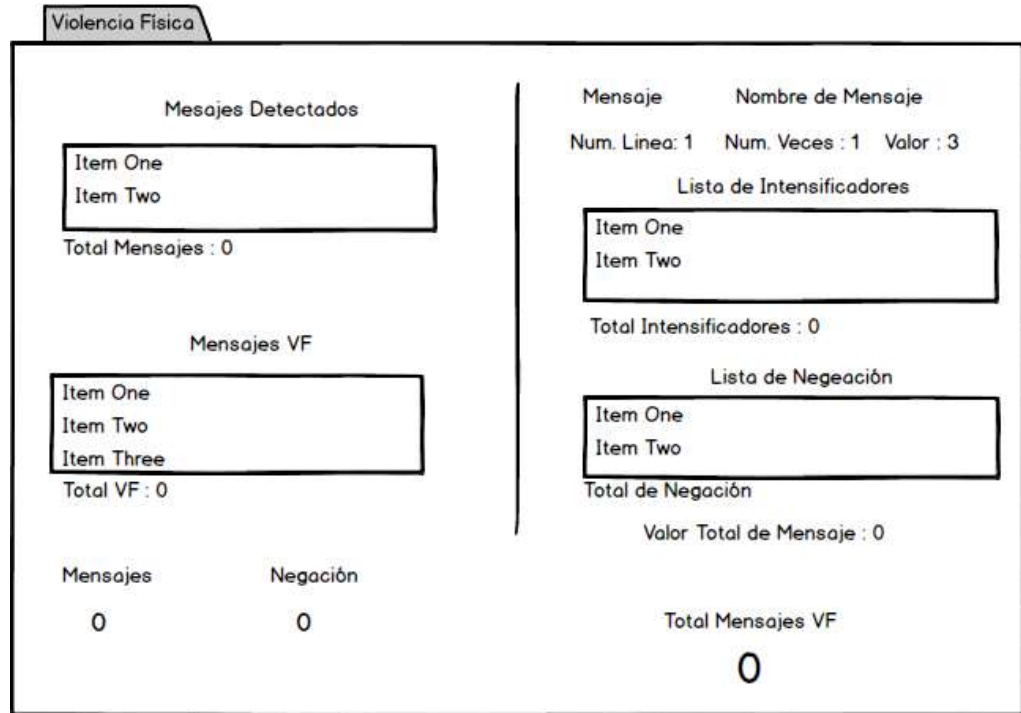


Figura 15. Maquetación de la interfaz Pestaña Violencia Física

Fuente: Elaboración propia

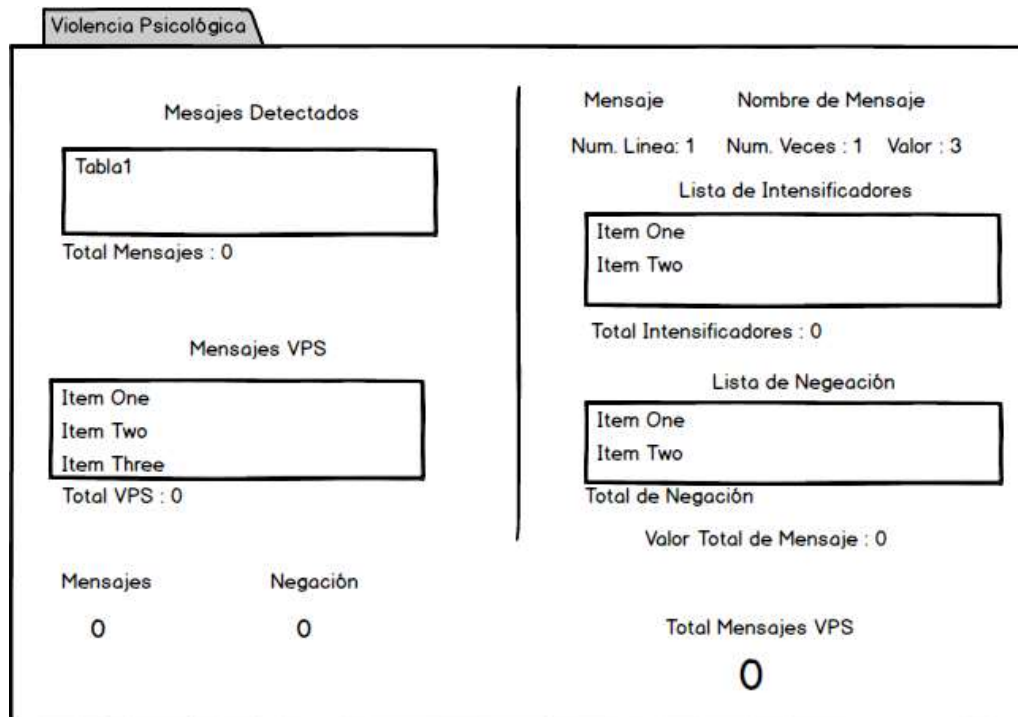


Figura 16. Maquetación de la interfaz Pestaña Violencia Psicológica

Fuente: Elaboración propia

2.2. Glosario de términos

No fue necesario ya que se encuentra ordenado el código y en este caso la variable se encuentra respectivamente nombradas, así como también se hace uso de clases para una distribución mejor del código.

2.3. Riesgos

La programación extrema aconseja considerar los posibles riesgos que puede suceder en el transcurso del desarrollo y darle atención. Pero el software no presenta algún riesgo en su desarrollo.

2.4. Funcionalidad extra

No fue necesario agregar alguna funcionalidad adicional al software.

2.5. Re factorizar

No fue necesario el volver a nuevamente escribir parte de código del programa.

2.6. Tarjetas CRC

A continuación, se le presenta las distintas tarjetas CRC que se utilizó.

Tabla 11. Tarjeta CRC: Clase AlgoritmoLevenshtein.

AlgoritmoLevenshtein	
Compara las palabras que recibe y busca su similitud	Ninguna

Fuente: Elaboración propia

Tabla 12. Tarjeta CRC: Clase Conteo_de_Repeticiones_y_Mensajes

Conteo_de_Repeticiones_y_Mensajes	
Cuenta por tipo de violencia cuantos contenidos denigrantes existen	BD Sisam Conectar

Fuente: Elaboración propia

Tabla 13. Tarjeta CRC: Clase Corrector Ortográfica

Corrector Ortográfico	
Divide el texto en palabras para luego buscar alguna palabra mal escrita	AlgoritmoLevenshtein Conectar BD Sisam

Fuente: Elaboración propia

Tabla 14. Tarjeta CRC: Clase Intensificadores

Intensificadores	
Busca intensificadores que afecte al contenido detectado	Conectar BD Sisam

Fuente: Elaboración propia

Tabla 15. Tarjeta CRC: Clase Pre procesamiento

Pre procesamiento	
<ul style="list-style-type: none"> • Elimina caracteres innecesarios • Elimina espacios en blanco y saltos de línea innecesarios • Reemplaza abreviaturas por palabras • Normaliza • Corrige algunos casos de faltas ortográficas 	Ninguno

Fuente: Elaboración propia

Tabla 16. Tarjeta CRC: Clase Procesos Negación

Procesos Negación	
Busca negación que afecte al contenido detectado	Ninguno

Fuente: Elaboración propia

Tabla 17.Tarjeta CRC: Clase Procesos resalto palabras

Procesos resalto palabras	
Se encargan de posición de palabra(s) en el texto para que luego lo resalten; dependiendo al caso puede ser el contenido denigrante o palabra negativa o palabra intensificadora	Ninguno

Fuente: Elaboración propia

Tabla 18.Tarjeta CRC: Clase Conectar

Conectar	
Se encargan de conectarse con la BD Derby	Derby.jar Derbyclient.jar

Fuente: Elaboración propia

Tabla 19.Tarjeta CRC: Clase Extracción Mensj

Extracción Mensj	
Se encargan de extraer los contenidos denigrantes, las negaciones y los intensificadores e insertar en la Base de Datos.	Conteo de Repeticiones y Mensajes
Verifica si al contenido detectado le afecta algún intensificador	BD Sisam
Verifica si al contenido detectado le afecta alguna negación	Procesos Negación
	Conectar
	Intensificadores

Fuente: Elaboración propia

Tabla 20. Tarjeta CRC: Clase Generar reporte

Generar reporte	
Se encargará de generar el reporte cuando se ha detectado algún contenido denigrante en el texto el cual no ha sido afectado por alguna negación	Conectar BD Sisam

Fuente: Elaboración propia

3. Desarrollo

Como se mencionó en la característica del software sisam que este es portable; tanto la base de datos, el ejecutable, las librerías y las imágenes están en una carpeta y estas pueden ser transportadas y ejecutadas desde una unidad extraíble como desde una pc.

Tal como se mencionó en la fase de planificación existen 4 interacciones a continuación se muestra el desarrollo de la primera interacción que consta de 2 historias de usuario las que son:

- Corrección de Texto.
- Búsqueda, selección y extracción de contenidos.

Para la implementación de toda la interfaz se usó el editor de código fuente Netbeans 8.2, el cual nos permitió poder desarrollar con facilidad los componentes del software. Cabe mencionar que se usaron las librerías derby.jar y derbyclient.jar para conectarse a la base de datos portable.

Ahora con respecto a las funcionalidades del software, en la historia del usuario el cliente menciona que el software debe realizar la corrección de texto. Para ver esta corrección el software tiene un botón que se encarga de esta funcionalidad mostrando inclusive el resultado de la corrección en un área de texto que lleva el título de PREPROCESADO.



Imagen 1. Captura de Imagen: Resultado de Corrección de texto

Fuente: Elaboración propia

La siguiente historia de usuario menciona que el software debe realizar la búsqueda, selección y extracción de los contenidos denigrantes. Esta funcionalidad fue incluida en un solo botón llamado ANALISIS, al darle clic en el botón realiza dicha funcionalidad tanto la búsqueda, selección y extracción.



Imagen 2.Captura de Imagen: Resultado de Análisis de texto

Fuente: Elaboración propia

Si el análisis texto encuentra contenidos esto se podrá verificar al irse al panel titulado CONTENIDOS DETECTADOS, allí se puede ver si encontró algún contenido de algún tipo de violencia.

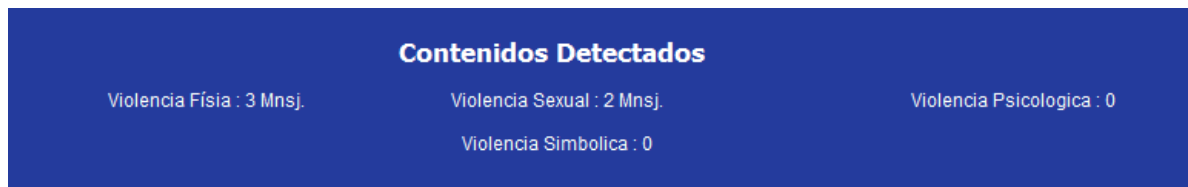


Imagen 3.Captura de Imagen: JPanel – Contenidos Detectados

Fuente: Elaboración propia

La segunda interacción contiene las siguientes historias de usuario:

- Búsqueda, selección y extracción de intensificadores.
- Búsqueda, selección y extracción de negación.

La funcionalidad que solicita el usuario sobre la búsqueda, selección y extracción de intensificadores se ejecuta al presionar el botón análisis; se implementó de la siguiente manera:

Ya que después de que el software verifique que existe algún contenido denigrante y verifica que existe o no alguna negación que afecte al contenido, este envía un pedazo de texto con el contenido al proceso Intensificadores para que verifique si en el texto se encuentra algún intensificador que afecte al contenido.

Mensajes Detectados			
Indice	Mensaje	Num. Linea	Valor
1	voy a tocarlo	Linea 27	1
2	múltiples orgasmos causarte	Linea 28	3

Total : 2 Mensaje(s)

Mensajes con Violencia Sexual (Mensaje - Negación)				
Indice	Mensaje	Num. Linea	Interf.	Val. Tot.
1	voy a tocarlo	Linea 27	Tipo 2	1.5
2	múltiples orgasmos causarte	Linea 28	Ninguna	3.0

Total : 2 Mensaje(s) VSE

MENSAJES VSE		MENSAJES VSE NEGADOS	
2		0	

Lista - Intensificadores				
Indice	Intensificador	Tipo	Num. Mensj.	Valor
1	más	2	1º Mensj.	0.5

Cant. de Intenf. : 1

Lista - Negación			
Indice	Negación	Tipo	Num. Mensj.

Cant. de Neg. : 0

Val. Tot. Mensj. (Val. Mensj. + Val. Tot. Intenf. - Tot. Neg.) : 1.5

TOTAL MENSAJES VSE

2

Imagen 4. Captura de Imagen: JPanel – Violencia Sexual

Fuente: Elaboración propia

En la Imagen 4, podemos ver que en el JPanel titulado Violencia Sexual existe una tabla llamada Lista-Intensificadores es donde se carga los

intensificadores que afecten al contenido como sus repeticiones en la misma línea de texto.

Lista - Intensificadores				
Indice	Intensificador	Tipo	Num. Mensj.	Valor
1	más	2	1º Mensj.	0.5
Cant. de Intnf. : 1				

Imagen 5.Captura de Imagen: JTable – Lista - Intensificadores

Fuente: Elaboración propia

La siguiente historia de usuario es la búsqueda, selección y extracción de negación; esta funcionalidad se implementó en el software haciendo que se ejecute seguido de la búsqueda y extracción del contenido, estos son los procesos de Detección de negación. Todo esto se pone en ejecución al presionar el botón Análisis.

Lista - Negación			
Indice	Negacion	Tipo	Num. Mensj.
Cant. de Neg. : 0			

Imagen 6.Captura de Imagen: JTable – Lista - Negación

Fuente: Elaboración propia

Cada JPanel que lleva el nombre de algún tipo de violencia contiene una tabla titulada LISTA – NEGACION donde se suele mostrar las negaciones encontradas.

La tercera interacción contiene las siguientes historias de usuario:

- Visualizar los contenidos por tipo de violencia.
- Visualizar los intensificadores y la negación por contenidos.

La funcionalidad de visualizar los contenidos por tipo de violencia, se consideró al momento de diseñar las interfaces del software. Así es que dependiendo de los tipos contenidos detectados, estos se mostraran al usuario por tipo de violencia que esta organizados por JPanels.

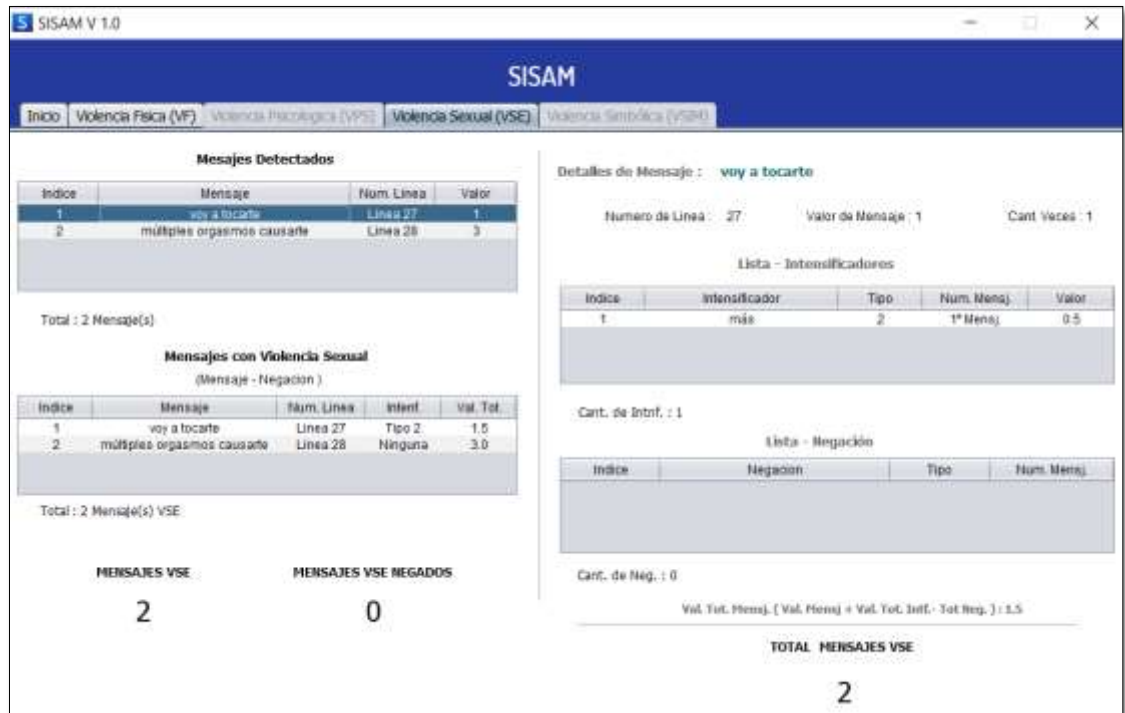


Imagen 7.Captura de Imagen: JPanels

Fuente: Elaboración propia

Mensajes Detectados				
Indice	Mensaje	Num. Linea	Valor	
1	voy a tocarte	Linea 27	1	
2	múltiples orgasmos causarte	Linea 28	3	
Total : 2 Mensaje(s)				
Mensajes con Violencia Sexual (Mensaje - Negacion)				
Indice	Mensaje	Num. Linea	Intenf.	Val. Tot.
1	voy a tocarte	Linea 27	Tipo 2	1.5
2	múltiples orgasmos causarte	Linea 28	Ninguna	3.0
Total : 2 Mensaje(s) VSE				

Imagen 8.Captura de Imagen: Tablas Mensajes

Fuente: Elaboración propia

Como se muestra en la Imagen 8, estos contenidos se muestran en la tabla titulada Mensajes Detectados, pero también algunos de ellos se muestran en la tabla titulada Mensajes con + (nombre tipo de violencia); esto es porque algunos de los contenidos puede que hayan sido afectados por alguna negación es por ello esos contenidos no se mostrarían en esta tabla.

La siguiente historia de usuario es visualizar los intensificadores y la negación por contenidos; esto consideró al momento de diseñar las interfaces del software. Se implementó de la siguiente manera:

Cuando se termina de realizar el análisis, si se detectó algún contenido el software lo mostrará en el JPanel al cual pertenezca por su tipo de violencia. Como se puede ver en la imagen 7. Se activó los JPannels Violencia Sexual y Física; esto es porque se encontró contenidos de tipo violencia física y sexual. Ahora una vez que se habilitan estos JPannels los contenidos que corresponden son cargados en los JPannels como se muestra en la Imagen 4.

Es así que para poder ver si el contenido ha sido afectado por alguna negación e intensificador, solo basta con hacer clic sobre el contenido que se muestra en la tabla titulada Mensajes Detectados o sobre la tabla titulada Mensajes con (Tipo de Violencia), para que se cargue en los campos de su lado izquierdo como se muestra a continuación:

The screenshot displays a software interface with two main sections. The left section contains two tables: 'Mensajes Detectados' and 'Mensajes con Violencia Sexual (Mensaje - Negacion)'. The right section shows 'Detalles de Mensaje' for the message 'voy a tocarle' and two lists: 'Lista - Intensificadores' and 'Lista - Negación'.

Mensajes Detectados

Indice	Mensaje	Num. Linea	Valor
1	voy a tocarle	Linea 27	1
2	múltiples orgasmos causarte	Linea 28	3

Total : 2 Mensaje(s)

Mensajes con Violencia Sexual (Mensaje - Negacion)

Indice	Mensaje	Num. Linea	Intenf.	Val. Tot.
1	voy a tocarle	Linea 27	Tipo 2	1.5
2	múltiples orgasmos causarte	Linea 28	Ninguna	3.0

Total : 2 Mensaje(s) VSE

Detalles de Mensaje : voy a tocarle

Numero de Linea : 27 Valor de Mensaje : 1 Cant. Veces : 1

Lista - Intensificadores

Indice	Intensificador	Tipo	Num. Mensj.	Valor
1	más	2	1* Mensj.	0.5

Cant. de Intenf. : 1

Lista - Negación

Indice	Negacion	Tipo	Num. Mensj.
--------	----------	------	-------------

Imagen 9. Captura de Imagen: Tablas y campos del JPanel Violencia Sexual

Fuente: Elaboración propia

La cuarta interacción contiene las siguientes historias de usuario:

- Fácil Uso.
- Generar Reporte.

La historia de usuario fácil uso se consideró en el diseño de la interfaz del software. Ya que el funcionamiento del software se redujo en lo más simple posible.

Primero se debe cargar el texto que se desea analizar para esto en primer JPanel está el botón Cargar texto.

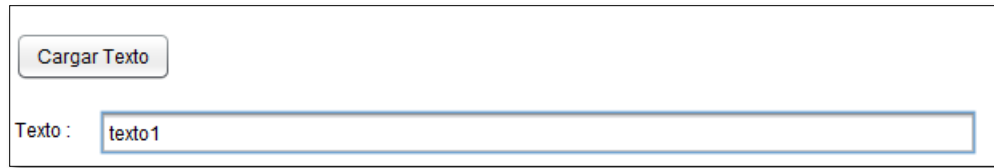


Imagen 10.Captura de Imagen: Botón Cargar Texto

Fuente: Elaboración propia

Una vez presionado el botón se nos abrirá una ventana donde buscaremos y seleccionaremos el texto, es importante que el texto este con extensión .txt de lo contrario no lo cargara.

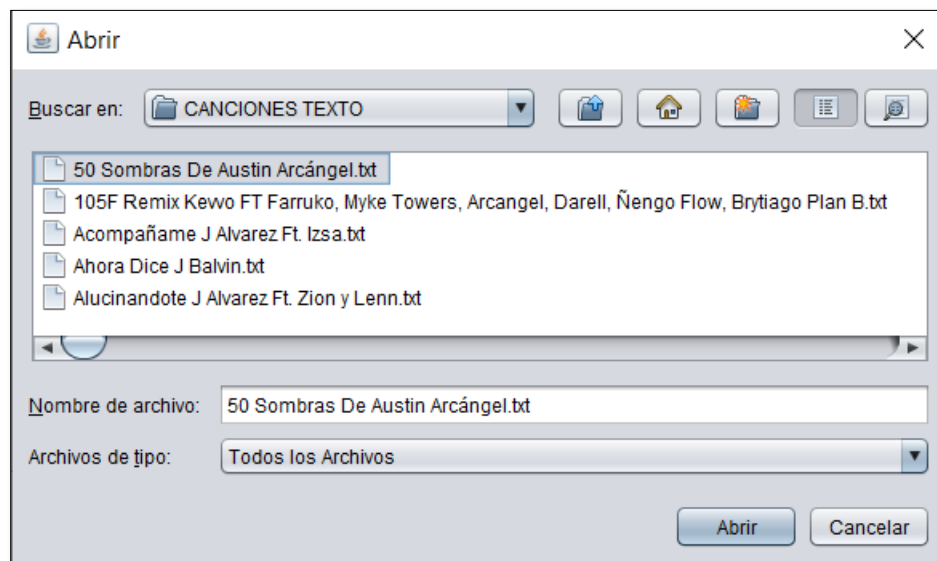


Imagen 11.Captura de Imagen: Ventana para la selección de texto

Fuente: Elaboración propia

Se incluyó la ejecución de todos los subprocessos del pre procesamiento en un solo botón, así como también la ejecución de todos los subprocessos del procesamiento se incluyeron en un solo botón.

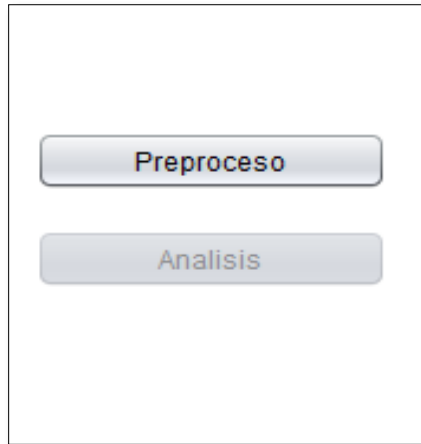


Imagen 12. Captura de Imagen: Botones Pre proceso y Botón Análisis

Fuente: Elaboración propia

Seguido de la importación del texto se presiona el botón Pre procesado para su ejecución. Durante el pre procesamiento el software ocultará el JFrame y se mostrará una JProgressBar mostrando que se está realizando el pre procesamiento.



Imagen 13. Captura de Imagen: JProgressBar

Fuente: Elaboración propia

Una vez terminada la ejecución del JProgressBar aparece el JFrame principal con el botón Análisis activado y el botón Pre procesamiento desactivado. El siguiente paso será presionar el botón Análisis para iniciar el análisis del texto. Este proceso es sumamente rápido. Al finalizar se desactivará el botón Análisis y se activará el botón Preproceso.

La interfaz en general cuenta con 5 JPanel ordenados de la siguiente manera: Inicio, Violencia Física (VF), Violencia Psicológica (VPS), Violencia Sexual (VSE), Violencia Simbólica (VSIM). Para que así el usuario pueda acceder a los tipos contenidos encontrados y cambiar de un tipo a otro, Por eso que, seguido después de haberse hecho el análisis, los JPanel que tiene

el nombre de los tipos de violencia; cada uno de ellos se activara dependiendo al tipo violencia que pertenezca los contenidos que se detectaron además de que podrá conocer y ver a los contenidos detectados en los JPanels. Así también el usuario conocerá que solo esos tipos de violencia presenta el texto

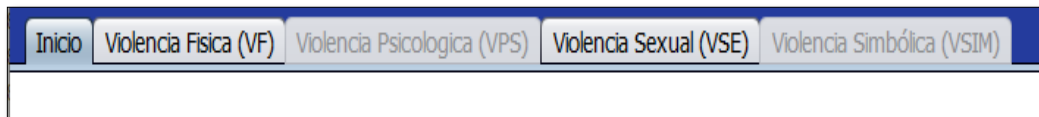


Imagen 14.Captura de Imagen: Menú de JPanels

Fuente: Elaboración propia

También el software te permite ver a detalle el mensaje en una ventana para ello al seleccionar el contenido de alguna de esas tablas los datos del contenido se cargan en los campos y tablas que se encuentra al lado izquierdo como se ve en la figura 7. Al seleccionar una los contenidos nos dirigiremos al lado izquierdo del JPanel Allí encontraremos más información del contenido como se muestra en la Imagen 15.

Detalles de Mensaje : voy a tocarte

Numero de Linea : 27 Valor de Mensaje : 1 Cant. Veces : 1

Lista - Intensificadores

Indice	Intensificador	Tipo	Num. Mensj.	Valor
1	más	2	1º Mensj.	0.5

Cant. de Intnf. : 1

Lista - Negación

Indice	Negacion	Tipo	Num. Mensj.
--------	----------	------	-------------

Cant. de Neg. : 0

Val. Tot. Mensj. (Val. Mensj + Val. Tot. Intf.- Tot Neg.) : 1.5

Imagen 15.Captura de Imagen: Detalles de Mensaje

Fuente: Elaboración propia

Ahora para ver en una ventana más información del contenido nos dirigiremos al nombre del contenido que se encuentra resaltado al lado del JLabel Detalles de Mensaje, le daremos click y se abrirá un JDialog con detalles más del mensaje como es, su ubicación en el texto y un pedazo de texto donde se encuentra, además se muestra al contenido resaltado en párrafo.

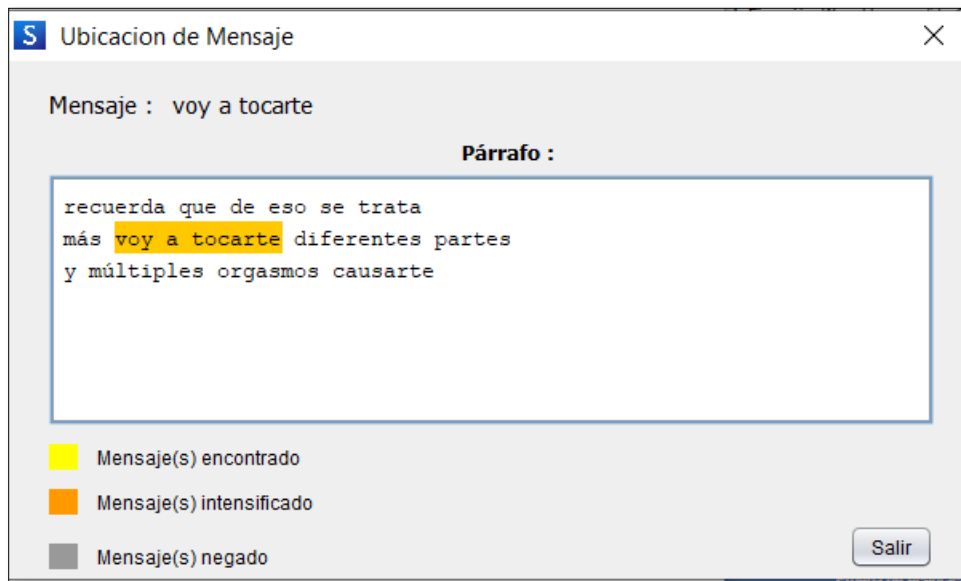


Imagen 16. Captura de Imagen: Ventana Ubicación de Mensaje

Fuente: Elaboración propia

De la imagen 16, Dependiendo al caso en que se encuentre el mensaje será resaltado en el párrafo si es de naranjado es que es un mensaje que esta afectado por un intensificador, si fuera plomo es que esta anulado por una alguna palabra negativa y si es amarillo no está afectado ningún intensificador o negación.

Las siguientes imágenes son capturas de las ventanas que muestran detalles adicionales de los intensificadores y negación que se detecten, esto se muestra cuando le damos clic sobre el intensificador o negación detectado.

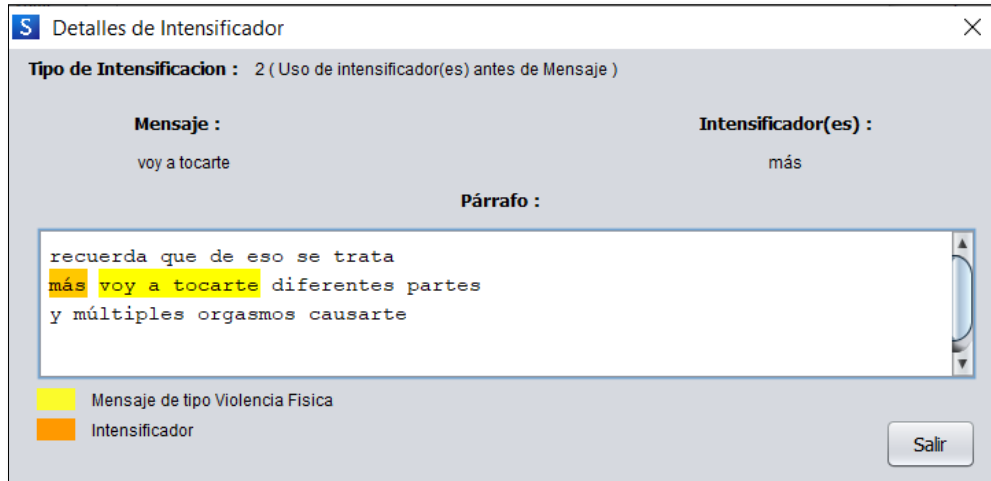


Imagen 17. Captura de Imagen: Ventana Detalles de Intensificador

Fuente: Elaboración propia

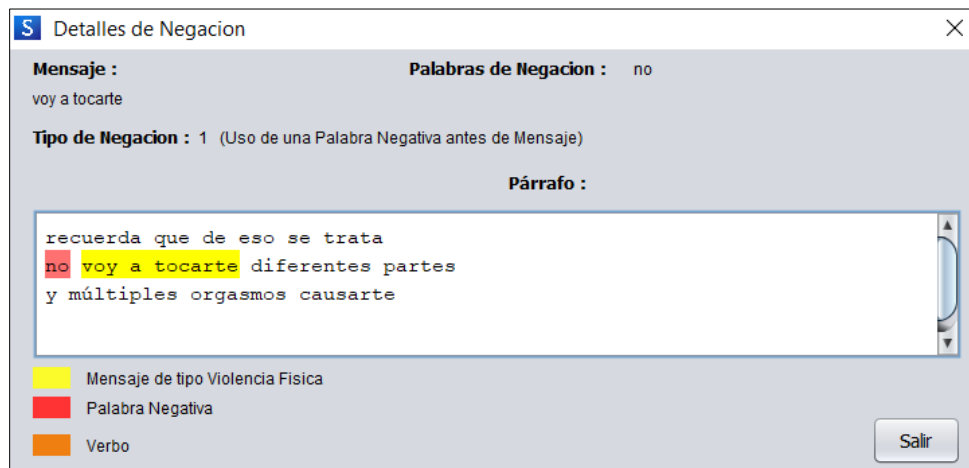


Imagen 18. Captura de Imagen: Ventana Detalles de Negación

Fuente: Elaboración propia

La siguiente historia de usuario es Generar Reporte; se programó para que el software permita poder generar un reporte solo cuando se halla detectado contenidos denigrantes que no hayan sido afectados por alguna negación. El botón se activará solo en este caso de lo contrario estará inactivo. Al darle clic en el botón reporte se generará un reporte con extensión .pdf

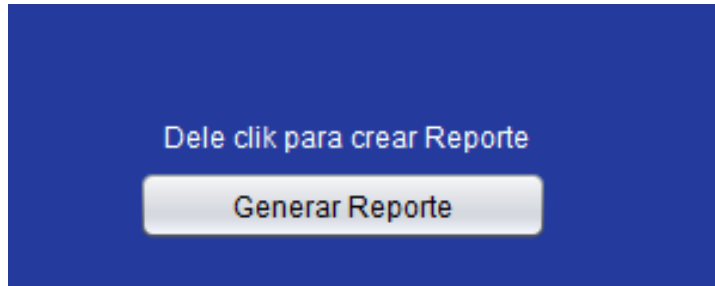


Imagen 19.Captura de Imagen: Botón Generar Reporte

Fuente: Elaboración propia

SISAM
Sistema de Análisis Musical

Reporte de Análisis del Texto Ingresado 'Vacía Sin Mí Ozuna feat. Darell.txt'

DETALLES GENERALES (INDICADORES)

Texto				
Vacía Sin Mí Ozuna feat. Darell.txt				
Dicc. V. Física	Dicc. V. Psicológica	Dicc. V. Sexual	Dicc. V. Simbólica	Num. Total Contenidos
0 Aciertos	1 Aciertos	0 Aciertos	1 Aciertos	2 Msj.

MENSAJES DEL TIPO VIOLENCIA PSICOLÓGICA (VPS)

Detalles de Mensajes (VPS) Detectados

INDICE	MENSAJE V_PS	TEXTO
1	puta	te acuestas con este flor de pelotero y tú eres v_ps y me sobran los cuerpos

Lista de Mensajes VPS

INDICE	MENSAJE	NUM. VECES	INTENSIFICADORES	VALOR
1	puta	1	Ninguna	3.0

MENSAJES DEL TIPO VIOLENCIA SIMBÓLICA (VSIM)

Detalles de Mensajes (VSIM) Detectados

INDICE	MENSAJE V_SIM	TEXTO
1	el culo	mami dios bendiga a los doctores por v_sim que te hicieron para que sepa bendecidos los que te metieron

Lista de Mensajes VSIM

INDICE	MENSAJE	NUM. VECES	INTENSIFICADORES	VALOR
1	el culo	1	Ninguna	2.0

Imagen 20.Captura de Imagen: Formato de Reporte

Fuente: Elaboración propia

4. Pruebas

A continuación, se procederá a presentar las pruebas de aceptación desarrolladas en función a las historias de usuario:

Tabla 21. Prueba de Aceptación: Corrección de Texto

Prueba de Aceptación	
Numero de Caso: 1	Numero de Historia de Usuario : 1
Nombre de Caso de prueba: Corrección de Texto	
Descripción: Realizar la corrección ortográfica, reemplazo abreviaturas, eliminación de caracteres innecesarios y como salto de línea y espacios en blanco innecesarios en el texto	
Condiciones de ejecución: Software ejecutando , Texto cargado al software, conexión a Base de Datos	
Entradas / Pasos: El usuario procederá a presionar el botón Pre procesado para iniciar la corrección	
Resultado Esperado: Texto Pre procesado	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 22. Prueba de Aceptación: Búsqueda, selección y extracción de contenidos

Prueba de Aceptación	
Numero de Caso: 2	Numero de Historia de Usuario : 2
Nombre de Caso de prueba: Búsqueda, selección y extracción de contenidos	
Descripción: Realizar la búsqueda , el conteo y extracción de los contenidos denigrantes	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto	
Entradas / Pasos: El usuario procederá a presionar el botón Análisis para iniciar búsqueda, selección y extracción	
Resultado Esperado: Contenidos detectados	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 23. Prueba de Aceptación: Búsqueda, selección y extracción de intensificadores

Prueba de Aceptación	
Numero de Caso: 3	Numero de Historia de Usuario : 3
Nombre de Caso de prueba: Búsqueda, selección y extracción de intensificadores	
Descripción: Realizar la búsqueda , el conteo y extracción de intensificadores que puedan afectar a los contenidos denigrantes	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto, extracción de los contenidos denigrantes , búsqueda de negación.	
Entradas / Pasos: Seguido del proceso de la detección de contenidos y la detección de negación sigue la detección de intensificadores.	
Resultado Esperado: Intensificadores detectados	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 24. Prueba de Aceptación: Búsqueda, selección y extracción de negación

Prueba de Aceptación	
Numero de Caso: 4	Numero de Historia de Usuario : 4
Nombre de Caso de prueba: Búsqueda, selección y extracción de negación	
Descripción: Realizar la búsqueda , el conteo y extracción de los de negación que anulen a los contenidos denigrantes	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto y extracción de los contenidos denigrantes.	
Entradas / Pasos: Seguido del proceso de la detección de contenidos sigue la detección de negación.	
Resultado Esperado: Negación detectada	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 25. Prueba de Aceptación: Visualizar los contenidos por tipo de violencia

Prueba de Aceptación	
Numero de Caso: 5	Numero de Historia de Usuario : 5
Nombre de Caso de prueba: Visualizar los contenidos por tipo de violencia	
Descripción: El software debe mostrar los contenidos detectados por tipo de violencia	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto y su Procesamiento por completo. Y Haber detectado contenidos denigrantes	
Entradas / Pasos: El usuario procederá a por ver los contenidos que se encuentran agrupados por tipo de violencia en 4 JPanels	
Resultado Esperado: Visualización de los contenidos por tipo de violencia generada	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 26. Prueba de Aceptación: Visualizar los intensificadores y la negación por contenidos

Prueba de Aceptación	
Numero de Caso: 6	Numero de Historia de Usuario : 6
Nombre de Caso de prueba: Visualizar los intensificadores y la negación por contenidos	
Descripción: El software debe mostrar los intensificadores y negación que se encontraron de los contenidos	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto y su Procesamiento por completo. Y Haber detectado contenidos denigrantes	
Entradas / Pasos: El usuario procederá a por ver los intensificadores y negación al seleccionar el contenido del que desea ver	
Resultado Esperado: Visualización los intensificadores y la negación por contenidos generada	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 27. Prueba de Aceptación: Facilidad de Uso

Prueba de Aceptación	
Numero de Caso: 7	Numero de Historia de Usuario : 7
Nombre de Caso de prueba: Facilidad de Uso	
Descripción: Visualización y comprobación de la funcionalidad y manejo fácil	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto	
Entradas / Pasos: El usuario procederá a hacer uso del software, primero importará el texto , seguido presionara el botón de Preproceso y luego espera que se active el botón análisis para presionarlo , luego si se detectó algún contenido podrá verlos en los JPanels que se hallan activado, y podrá generar el reporte del análisis	
Resultado Esperado: Satisfacción del usuario al manejo del software portable	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

Tabla 28. Prueba de Aceptación: Generar Reporte

Prueba de Aceptación	
Numero de Caso: 8	Numero de Historia de Usuario : 8
Nombre de Caso de prueba: Generar Reporte	
Descripción: El software debe permitir el poder generar un reporte de los contenidos detectados	
Condiciones de ejecución: Haber Realizado el Pre procesamiento del texto y su Procesamiento por completo. Y Haber detectado contenidos denigrantes	
Entradas / Pasos: El usuario procederá a presionar el botón Generar Reporte	
Resultado Esperado: Reporte Generado	
Evaluación: Prueba Satisfactoria	

Fuente: Elaboración propia

ANEXO 3: Matriz de Consistencia

Tabla 29. Matriz de Consistencia

PROBLEMA	OBJETIVO	HIPOTESIS	VARIABLE	METODOLOGIA	POBLACION
<p>Problema General: ¿Cuál es la influencia del uso del Software SISAM en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018?</p>	<p>Objetivo General: Determinar la influencia del uso del Software SISAM en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018</p>	<p>Hipótesis General: El uso del Software SISAM influye significativamente en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018?</p>	<p>Variable Independiente: Software SISAM</p> <p>Indicadores:</p> <ul style="list-style-type: none"> • Numero de palabras. • Caracteres Extraídos. • Numero de Caracteres irrelevantes. • Numero de Coincidencias en la comparación. • Contenidos detectados. 	<p>Tipo de investigación: La investigación aplicada.</p> <p>Nivel de investigación: El nivel de investigación es Experimental</p> <p>Diseño de investigación: El diseño de la Investigación es Pre experimental</p> <p>Dónde: G : Grupo de Muestra</p> <p>X: Es la variable independiente. O: Observaciones</p> <p>G->X->O</p> <p>Método de investigación: El método de investigación a utilizar es el método cuantitativo.</p>	<p>Población de estudio: Canciones del género REGGAETON. Ya que la cantidad no es exacta se seleccionara de las distintas plataformas de música (Spotify ,paginas web).</p> <p>Muestra de estudio: Conformado por 404 Canciones.</p>
<p>Problemas Específicos:</p> <ul style="list-style-type: none"> • ¿Cómo influye el nivel de análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018? • ¿Cómo influye las fases de análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018? • ¿Cómo influye el tipo de datos analizado en la detección de 	<p>Objetivos Específicos:</p> <p>Determinar que el nivel de análisis de sentimiento influye en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON – 2018.</p> <p>Identificar la influencia de las fases de análisis de sentimiento en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018</p> <p>Determinar que el tipo de datos analizado influye en la detección de contenidos denigrantes hacia la Mujer</p>	<p>Hipótesis Específicos:</p> <p>El nivel de análisis de sentimiento apoyará en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON – 2018.</p> <p>Las fases de análisis de sentimiento apoyarán en la detección de contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018</p> <p>El tipo de datos analizado apoyará en la detección de contenidos denigrantes hacia la Mujer en</p>	<p>Variable Dependiente: Contenidos de Denigración a la Mujer</p> <p>Indicadores:</p> <ul style="list-style-type: none"> • Documento (Texto digital) • Texto • Etiquetación de las oraciones • Detección del uso de Adjetivos calificativos denigrantes (Diccionario de adjetivos calificativos) • Detección de veces que se menciona a la mujer de manera directa o indirecta (Sustantivo o sujeto) 	<p>G : Grupo de Muestra</p> <p>X: Es la variable independiente. O: Observaciones</p> <p>G->X->O</p> <p>Método de investigación: El método de investigación a utilizar es el método cuantitativo.</p>	<p>Muestra de estudio: Conformado por 404 Canciones.</p>

<p>contenidos denigrantes hacia la Mujer en productos audio musicales del Género REGGAETON - 2018?</p>	<p>en productos audio musicales del Género REGGAETON - 2018</p>	<p>productos audio musicales del Género REGGAETON - 2018</p>	<ul style="list-style-type: none"> • Detección de contenidos o frases denigrantes (mensajes ofensivos) • Resultados obtenidos por los procesos realizados 		
---	---	--	---	--	--