

# Universidad Siglo 21



Licenciatura en Informática

Trabajo Final de Graduación

Sistema clasificador de emociones para las comunicaciones digitales privadas por medio de técnicas de inteligencia artificial

Leandro M. Boisselier

DNI 24.134.688

Legajo VINF07560

Año 2020

## **Resumen**

El análisis de las comunicaciones digitales privadas entre participantes plantea múltiples desafíos a los investigadores que buscan extraer información de utilidad para cumplir objetivos dentro de proyectos de investigación. La gran cantidad de procesamiento manual que deben realizar para dividir y clasificar los mensajes dificulta el abordaje de proyectos con grandes cantidades de interacciones.

Mediante el uso de técnicas de inteligencia artificial, aplicadas al procesamiento de lenguaje natural, se construyó un sistema que recibe un archivo exportado de un sistema de mensajería. Este documento contiene los mensajes que se intercambiaron en un período de tiempo entre dos o más participantes y lo descompone en unidades atómicas que son procesadas a partir de diferentes algoritmos.

La información que se obtiene como producto final de este procesamiento es de gran valor para un investigador de las comunicaciones digitales que le permite descubrir, mediante una representación gráfica y analítica, cómo afectan las palabras y los recursos multimodales en las emociones de los participantes.

El resultado obtenido permite caracterizar la interacción digital en términos estadísticos de personas intervinientes, palabras utilizadas, emociones detectadas y puntos de inflexión dentro de una línea de tiempo.

Palabras Claves: humanidades digitales, comunicaciones, emociones, emoticones

## **Abstract**

Analyzing private digital communications between participants poses multiple challenges for researchers seeking to extract useful information to meet objectives within research projects. The large amount of manual processing that you must perform to split and classify messages makes it difficult to approach projects with large amounts of interactions.

Using artificial intelligence techniques, applied to natural language processing, a system was built that receives an exported file from a messaging system. This document contains messages that were exchanged in a period of time between two or more participants and breaks it down into atomic units that are processed by different algorithms.

The information obtained as the final product of this processing is of great value to a digital communications researcher that allows him to discover, through a graphical and analytical representation, how words and multimodal resources affect the emotions of the participants.

The result allows to characterize digital interaction in statistical terms of interveners, words used, detected emotions and turning points within a timeline.

**Keywords:** digital humanities, communications, emotions, emojis

## Índice

Título .....	1
Introducción.....	1
Descripción del área problemática.....	2
Justificación.....	3
Objetivo general.....	4
Objetivos específicos .....	4
Marco teórico referencial .....	5
Antecedentes.....	5
Dominio del problema.....	5
Actividad del cliente.....	6
TICs.....	7
Competencias.....	14
Diseño metodológico .....	14
Relevamiento .....	16
Relevamiento estructural.....	17
Relevamiento funcional.....	18
Relevamiento de documentación .....	19
Proceso de negocio .....	20
Diagnóstico y Propuesta.....	20
Diagnóstico .....	20
Propuesta .....	22
Objetivos, Límites y Alcances del Prototipo.....	23
Objetivos del Prototipo.....	23
Límites.....	23
Alcance .....	23
Descripción del Sistema .....	23
Requerimientos Funcionales.....	23
Requerimientos No Funcionales .....	24
Diagrama de Casos de uso.....	24
Descripción de Casos de uso .....	24
Diagrama de Colaboración de análisis.....	31
Diagramas de Clase.....	31
Prototipos de interfaces de pantallas .....	32
Diagrama de despliegue .....	33
Diagrama de componentes.....	34
Seguridad.....	34

Análisis de Costos.....	40
Análisis de Riesgos .....	44
Conclusiones.....	47
Demo .....	48
Referencias .....	49
Anexos.....	54
Anexo 1 Formulario recepción de archivos.....	54
Anexo 2 - Fragmento de un archivo anonimizado.....	55

## Índice de Ilustraciones

Ilustración 1. Diagrama de Gantt.....	16
Ilustración 2. Organigrama.....	17
Ilustración 3. Proceso de negocio .....	20
Ilustración 4. Diagramas de casos de uso.....	24
Ilustración 5. Diagrama de secuencia .....	31
Ilustración 6. Diagrama de Clases .....	32
Ilustración 7. Pantalla carga archivo .....	32
Ilustración 8. Procesar Archivo.....	32
Ilustración 9. Ver Reporte.....	33
Ilustración 10. Diagrama de despliegue .....	33
Ilustración 11. Diagrama de componentes .....	34
Ilustración 12. Probabilidad de ocurrencia de eventos.....	46

## Índice de Tablas

Tabla 1. Owasp TOP 10 2017 .....	35
Tabla 2. Análisis de Costos de Recursos Humanos .....	41
Tabla 3. Costos de Hardware.....	41
Tabla 4. Análisis de Riesgos .....	45

## **Título**

Sistema clasificador de emociones para las comunicaciones digitales privadas por medio de técnicas de inteligencia artificial

## **Introducción**

Las comunicaciones digitales son una trama de intercambios entre usuarios, sistemas y dispositivos que basan parte de su éxito, es decir, el alcance del propósito comunicativo, en la adecuada comprensión de las emociones de los mensajes. En la conversación cara-a-cara, los hablantes procesan mucha información de manera conjunta (gestos, entonaciones, silencios) que, en el intercambio de mensajes digitales, no aparecen. La ajustada interpretación de la emoción (positiva, negativa o neutra) del mensaje es necesaria para poder comunicarnos. Es por ello que, en los entornos digitales, los interlocutores despliegan un amplio abanico de estrategias compensatorias destinadas a suplir estas carencias que incluyen recursos lingüísticos y multimodales.

En este punto, emerge la siguiente pregunta: ¿es posible llevar a cabo una identificación sistemática de los procedimientos lingüísticos que intervienen en la transmisión de las emociones? Diversos enfoques lingüísticos se han ocupado de explorar la manera en que los enunciados codifican la evaluación emocional. De entre todas ellas, probablemente, la Teoría de la Valoración (Martin; White, 2003) constituye la propuesta más interesante y productiva para los recursos lingüísticos empleados por los interlocutores para expresar y negociar determinadas posiciones intersubjetivas en los discursos (Kaplan, 2004).

Ahora bien, cabe plantearse si es posible automatizar este proceso. La Inteligencia Artificial<sup>1</sup> (IA) lleva años aplicándose en tal tarea. Autores como Vela, Cantamutto y Nuñez (2021) sostienen que la Teoría de la Valoración se torna especialmente relevante al ofrecer un correlato cualitativo para explicar los procedimientos que permiten explicitar las valoraciones intersubjetivas aplicable a proyectos de minería de datos. Estas sinergias, entre modelos cualitativos y cuantitativos, se tornan imprescindibles en proyectos de desarrollo como el aquí desarrollado.

Identificado este nicho, el proyecto ha cubierto necesidades detectadas por algunos investigadores que trabajan en el ámbito del análisis de las interacciones digitales privadas para el análisis de los mensajes que intercambian los interlocutores. En Cantamutto y Vela (2015) se

---

<sup>1</sup> Inteligencia Artificial: “Estudio de los cálculos que hacen posible percibir, razonar y actuar” (Winston, 1992)

identifican algunos de los problemas con que se topan los estudios de corpus sobre interacciones digitales escritas, entre ellos, la recuperación de los datos, la sistematización de estos y su análisis automático. Estas dificultades se complican si, además, se necesita identificar la polaridad emocional de los mensajes y la manera en que influyen entre sí los distintos mensajes que constituyen un intercambio comunicativo.

En este sentido, el sistema propuesto ha resuelto un problema real con el que se encuentran cotidianamente los investigadores de las comunicaciones digitales<sup>2</sup> al momento de determinar cuál es el flujo de emociones dentro de un grupo de mensajes intercambiados entre múltiples participantes a lo largo del tiempo. Los mensajes clasificados a través de su indexación emocional permiten detectar cuál(es) intervención(es) resultan más conflictivas y cuáles son los elementos que propician el conflicto. Además, se pueden visualizar, a partir de una línea del tiempo, cuándo se acrecientan o disminuyen las intervenciones según sea la emoción predominante.

Es por ello que consideramos que la propuesta no solo trabaja en la línea de otros modelos de análisis de sentimientos, sino que permite ir un paso más allá al atender a un factor que raramente (no se han ubicado trabajos en este sentido) ha sido abordado: la interacción emocional entre los distintos mensajes. Resulta importante recordar en este punto que, tal y como sostiene el análisis de las conversaciones (Briz, 2003), las intervenciones no están aisladas unas de otras, sino que, en un intercambio de carácter dialógico, como es el caso de las interacciones por WhatsApp, estos se articulan en intercambios o pares de adyacencia que se determinan mutuamente, también desde la perspectiva emocional.

### *Descripción del área problemática*

El uso creciente de los medios de comunicación por parte de las personas para intercambiar ideas y consensuar objetivos de diversas índoles ha despertado el interés de organizaciones comerciales y civiles para comprender mejor cómo se desarrollan las interacciones y cuáles son factores que generan cambios en la percepción de los participantes. En el contexto pandemia/pospandemia, los intercambios digitales formaron parte de las diferentes esferas de la actividad humana. Debido a la necesidad de mantener distancia social, la educación, el trabajo, las relaciones interpersonales, fueron canalizadas por medios digitales. Los intercambios en contextos no presenciales y el aumento exponencial de los intercambios virtuales requiere de proyectos que aborden diferentes aspectos

---

<sup>2</sup> Investigadores de las comunicaciones digitales: Dentro del movimiento de las Humanidades Digitales, estudian comunicaciones en los términos más amplios y las interacciones personales en distintos canales de comunicación modernos

de estos intercambios: negociación de (des)acuerdo, estrategias de cortesía y también comunicación de contenidos emocionales.

Con el objetivo de poder analizar los factores que inciden en el éxito comunicativo, se promueven cada vez más proyectos de investigación dedicados a estudiar las comunicaciones digitales. Los estudios de esta envergadura enfrentan a los investigadores a diferentes retos metodológicos y, entre ellos, es la gestión del conjunto de mensajes que deben analizar para dar respuestas a esos interrogantes. Las interacciones de carácter público han sido mayormente estudiadas pero los intercambios privados (como las interacciones por WhatsApp) aún no. Facilitar y automatizar la comprensión de esos factores beneficiará tanto a organizaciones como a sus clientes ya que permitirá mayores posibilidades de recibir una atención de calidad, gestionar los conflictos y permitir la detección de los factores que inciden en el éxito o fracaso de las comunicaciones.

### **Justificación**

Tal como se explicó recién, el sistema automatizado de análisis de mensajes permite trabajar con una cantidad de mensajes más significativa que la que podría estudiar una persona en forma manual. La identificación de la emoción de cada mensaje, a partir de las palabras que lo componen, lleva un tiempo de procesamiento mucho menor cuando es realizado por un sistema ya que no se expone a un individuo al desgaste mental que podría provocar clasificaciones tan exhaustivas y de este modo se evitan algunas clasificaciones potencialmente incorrectas. Si, además, el resultado de la clasificación debe procesarse bajo distintos criterios para obtener resultados, el trabajo que logra el sistema es mucho más eficiente ya que solo se ve limitado por las capacidades de procesamiento del sistema que se seleccione para ejecutar la tarea. De este modo, con este sistema se obtienen resultados en forma rápida y eficiente y se aumenta la cantidad de mensajes que se pueden procesar dentro de un proyecto de investigación.

El análisis manual requiere que los investigadores utilicen algún criterio de trabajo que podría no coincidir cuando se colabora en equipos ya que, al ser tareas rutinarias, cada integrante de un equipo podría establecer sus propias variaciones en cuanto al procesamiento de los fenómenos, la catalogación de los mensajes y registración de resultados. Estas variaciones generarían dificultades en el momento de conciliar los datos procesados.

En este sentido, el uso de un sistema aporta, además de la velocidad, una homogeneidad en la catalogación de mensajes y los objetos resultante que facilita la incorporación de esos resultados en conjuntos más grandes que luego pueden ser procesados como unidades mayores. Por lo que un

impacto logrado con la utilización del sistema es la estandarización de los resultados facilitando el intercambio de información entre investigadores.

Entre los objetos resultantes de la ejecución del sistema que se estandarizan están las planillas de cálculos, gráficos y archivos de texto en formato plano. Este último es plausible de ser analizado a partir de muchas herramientas de análisis lingüístico como, por ejemplo, el paquete AntConc (Antonhy, 2004).

Una innovación, con respecto a la clasificación de los mensajes que solo procesan los elementos lingüísticos, se produce al considerar los emojis como entidades que aportan información relevante para la valoración positiva, negativa o neutra de los mensajes. El repertorio de emojis disponibles que ofrece el consorcio UNICODE además de ser muy extenso está en constante crecimiento: en un contexto de análisis humano, el tiempo requerido es aún mayor que para el análisis del texto convencional debido a la necesidad de recorrer listas de emojis que, por su característica gráfica, no tienen un orden determinado y no se recorren en forma secuencial. En tal sentido, al disponer de tablas de asociación de emojis con su sentimiento asociado el sistema accede en forma inmediata al significado que se propone.

### **Objetivo general**

Catalogar los mensajes de texto intercambiados por los participantes de una conversación de mensajería instantánea aplicando técnicas de procesamiento de lenguaje natural con el fin de detectar las participaciones que provocan cambios en el estado emocional de las intervenciones siguientes.

### **Objetivos específicos**

- Catalogar mensajes utilizando técnicas de procesamiento de lenguaje natural asignando un valor emocional negativo, positivo o neutro a cada uno con el fin de procesar luego el conjunto de resultados.
- Generar una línea de tiempo de los mensajes catalogados usando algoritmos de ordenación usando como criterios el participante, el momento y la emoción para poder comparar con los mensajes adyacentes.
- Generar una visualización de las intervenciones donde se observan cambios de estado emocional en la línea de tiempo.

## **Marco teórico referencial**

### *Antecedentes*

A continuación, se presentan los principales antecedentes organizados en dos ejes. Estudios sobre emociones en mensajería instantánea, en primer lugar, y, en segundo lugar, sobre sistemas de detección automática de sentimientos. Esta pequeña reseña ha puesto de manifiesto que, hasta donde llega la búsqueda realizada, no existen antecedentes de sistemas similares a los aquí propuestos.

En torno al estudio de la mensajería instantánea, como WhatsApp, la mayoría de las investigaciones que abordan las emociones en estas plataformas privadas analizan o bien pequeños corpus (dada la dificultad para acceder a este tipo de datos) o bien sobre aspectos puntuales como los emojis y las funciones que estos tienen en la manifestación de sentimientos. Ejemplo de ello es el artículo sobre la estructura del intercambio de Alcantará-Plá (2014), quien estudia las unidades discursivas de WhatsApp a partir del análisis automático de un corpus de 176.000 palabras. Por su parte, Vázquez Cano, Mengual Andrés y Roig Vila (2015) presentan una investigación lexicométrica de los aspectos lingüísticos y paralingüísticos de la escritura digital sincrónica de adolescentes españoles también en esa plataforma.

En relación con los sistemas de detección automática de sentimientos, múltiples estudios han dedicado su atención, con propósitos diferentes, al análisis de los sentimientos. Por ejemplo, Baviera (2017) expone la importancia de este tipo de análisis automático para la detección de la opinión. La red social Twitter suele ser la plataforma predilecta para medir el termómetro social.

### *Dominio del problema*

En la actualidad, el trabajo con corpus, a saber, grandes colecciones de datos lingüísticos estructurados, se impone en la investigación lingüística. Es por ello que se reclaman herramientas que permitan trabajar de forma automatizadas esos datos. En palabras de Bolaño Cuéllar (2015: 15), “la lingüística de corpus se inscribe en el mismo paradigma de las disciplinas lingüísticas que se dedican a estudiar las diferentes manifestaciones del uso del lenguaje en contextos reales de interacción comunicativa, lo que, en otras palabras, corresponde a un enfoque lingüístico funcional”.

Como en cualquier intercambio comunicativo, los interlocutores que participan de una interacción digital llevan a cabo un conjunto de elecciones comunicativas destinadas a transmitir de forma eficaz sus propósitos comunicativos (Verschueren, 2002). Algunas de estas elecciones tienen una naturaleza verbal, las palabras, y las diversas formas que adquieren, pero otras tienen naturaleza

multimodal, tal es el caso, por ejemplo, de los emojis (Danesi, 2017) o de otros recursos gráficos como stickers o gifs animados (Herring y Dainas, 2017). Recuperar la intención comunicativa de un mensaje resulta un proceso muy complejo y el efecto que produce en el interlocutor (es decir, la emociones que emergen). Es decir, no solo implica el procesamiento de una gran cantidad de información, sino es necesario clasificarla para determinar de qué manera interactúan los diferentes componentes de los mensajes en el desarrollo del intercambio.

El campo de estudio del análisis de sentimientos constituye una línea fructífera de investigación a través de la cual se ha intentado explicar formalmente los diversos modos en los que las emociones se evidencian en los enunciados digitales.

La mayoría de estas investigaciones se han centrado en el estudio de redes públicas como por ejemplo Twitter, en detrimento de otras redes sociales privadas como WhatsApp. En particular, el análisis de las redes sociales públicas, como Twitter, presenta grandes ventajas: muchos datos accesibles, disponibles tanto para el investigador como para las herramientas de análisis automático de grandes cantidades de datos, disponibilidad pública de los datos, fácil recuperación y ordenación de la información.

Estas razones justifican que exista una importante tradición de estudios sobre las comunicaciones digitales a través de redes sociales públicas (Mayer-Schönberger & Cukier, 2013), mientras que la dificultad en el acceso a los datos producidos en plataformas privadas como Telegram, WhatsApp o similares ha provocado que desde el campo computacional no se haya atendido en profundidad la inteligencia artificial aplicada a la interpretación de estos intercambios.

Nuestra perspectiva teórica incluye aportaciones de la lingüística de corpus a partir de la recolección de datos y su posterior análisis automático (Balbachán, 2014) y del análisis de la comunicación digital (Thurlow y Mroczek, 2011; Cantamutto y Vela Delfa, 2016). En tal sentido, son de utilidad las definiciones de los mensajes como intervenciones menores dentro de un intercambio mayor, la conversación por WhatsApp (Alcantará-Plá, 2014).

#### *Actividad del cliente*

Lingüistas, filólogos, psicólogos y sociólogos han tratado de abordar, desde sus marcos teóricos y metodológicos de trabajo, esta labor. Sin embargo, sus acercamientos eminentemente cualitativos se topan con la dificultad de manejar grandes masas de datos. Algo necesario cuando se busca hacer generalizaciones consistentes

El estudio automatizado de ciertos fenómenos lingüísticos, como las frecuencias de ciertas unidades o las colocaciones, es decir, las combinaciones de unidades, ha avanzado mucho en las últimas décadas. Pero la correcta interpretación de un mensaje es algo más complejo que identificar recurrencias lingüísticas. En tal sentido, es necesario identificar la orientación emocional de los mensajes.

### *TICs*

Para el desarrollo del apartado tecnológico es conveniente realizar una clasificación del contexto en que es aplicada cada tecnología o concepto. Se reconocen los siguientes aspectos relevantes: Conceptos generales y transversales al proyecto; diseño, desarrollo y codificación; software de base para servidor y estación de trabajo; software de orquestación de servicios; lenguajes de programación; tecnologías aplicadas al funcionamiento del sistema en general; tecnologías y conceptos aplicados a resolver los requerimientos funcionales del proyecto; persistencia y gestión de datos; tecnologías aplicadas a la usabilidad de la aplicación, interfaces de usuario, visualización de resultados; tecnologías aplicadas a la conectividad física y lógica.

Se presentan a continuación los conceptos en detalle:

Conceptos generales y transversales al proyecto:

- **Software Libre:** “«Software libre» es el software que respeta la libertad de los usuarios y la comunidad. A grandes rasgos, significa que los usuarios tienen la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar el software. Es decir, el «software libre» es una cuestión de libertad, no de precio.” (GNU, s.f.)
- **Open Source:**

Originalmente, la expresión open source (o código abierto) hacía referencia al software open source (OSS). El software open source es un código diseñado de manera que sea accesible al público: todos pueden ver, modificar y distribuir el código de la forma que consideren conveniente. El software open source se desarrolla de manera descentralizada y colaborativa, así que depende de la revisión entre compañeros y la producción de la comunidad. Además, suele ser más económico, flexible y duradero que sus alternativas propietarias, ya que las encargadas de su desarrollo son las comunidades y no un solo autor o una sola empresa. (RedHat, S.f.)

- Hash: “En esencia, una función resumen (hash, en inglés), proporciona una secuencia de bits de pequeña longitud, que va asociada al mensaje, aunque contiene menos información que éste, y que debe resultar muy difícil de falsificar.” (Lucena López, 2010)
- WhatsApp: es un sistema de mensajería instantánea (Alcantará-Plá, 2014) que permite el intercambio de textos breves, audio, video e imágenes a partir de una lista de contactos predeterminada.
- Criptografía asimétrica:
 

Los algoritmos asimétricos o de clave pública han demostrado su interés para ser empleados en redes de comunicación inseguras (Internet). Introducidos por Whitfield Diffie y Martin Hellman a mediados de los años 70, su novedad fundamental con respecto a la criptografía simétrica es que las claves no son únicas, sino que forman pares. (Lucena López, 2010)

Sobre los aspectos de Diseño, Desarrollo y Codificación se definen las siguientes tecnologías:

- NetBeans IDE: Utilizado para escribir el código fuente
 

es un entorno de desarrollo - una herramienta para que los programadores puedan escribir, compilar, depurar y ejecutar programas. Está escrito en Java - pero puede servir para cualquier otro lenguaje de programación. Existe además un número importante de módulos para extender el NetBeans IDE. NetBeans IDE es un producto libre y gratuito sin restricciones de uso. (Netbeans, s.f)
- Git: Utilizado para el control de versiones del código fuente “es un sistema de control de versiones distribuido gratuito y de código abierto diseñado para manejar todo, desde proyectos pequeños a muy grandes, con velocidad y eficiencia.” (Git, s.f.)

Procesador de textos y Planilla de cálculos:

- LibreOffice: Conjunto de herramientas de oficina de licencia libre, utilizado para el procesamiento de texto y elaboración de planillas de cálculo. (LibreOffice, s.f)
- Google Suite: Conjunto de servicios de oficina ofrecidos por la empresa Google. Entre sus servicios se encuentra la provisión de formularios dinámicos.

Captura y edición de imágenes:

- GIMP: Software de edición de imágenes de tipo mapa de bits. Utilizado para la edición y ajuste fino de imágenes y capturas de pantalla. Según la página oficial “es un editor de imágenes multiplataforma disponible para GNU / Linux, OS X, Windows y más sistemas operativos” (Gimp, s.f.)

#### Diseño de diagramas:

- Umbrello: Utilizado para la edición de diagramas UML. Según el sitio oficial “Umbrello UML Modeller es un programa de diagrama de lenguaje de modelado unificado (UML) basado en la tecnología KDE. UML le permite crear diagramas de software y otros sistemas en un formato estándar para documentar o diseñar la estructura de sus programas.” (Umbrello, s.f.)  
Pencil: “está diseñado con el propósito de proporcionar una herramienta de creación de prototipos GUI gratuita y de código abierto que las personas pueden instalar y usar fácilmente para crear maquetas en plataformas de escritorio populares.” (Pencil, s.f.)
- Diagrams: Utilizado para la elaboración de diagramas UML. Según la página oficial “diagrams.net es un software de diagramación desplegable de código abierto, en línea, de escritorio y de contenedor” (Diagrams, s.f.)

#### Software de base para servidor y estación de trabajo:

- Ubuntu Server: Utilizado como sistema operativo de base en el servidor y como sistema operativo de base en las imágenes de Docker.
- Ubuntu Desktop: Utilizado como sistema operativo en la estación de trabajo de codificación.
- Windows Home: Utilizado como sistema operativo en las estaciones de trabajo de los usuarios del sistema.

#### Software de orquestación de servicios:

- Contenedor:

Un contenedor es una unidad estándar de software que empaqueta el código y todas sus dependencias para que la aplicación se ejecute de forma rápida y confiable de un entorno informático a otro. Una imagen de contenedor de Docker es un paquete de software ligero, independiente y ejecutable que incluye todo lo necesario para ejecutar una aplicación: código, tiempo de ejecución, herramientas del sistema, bibliotecas del sistema y configuraciones. (Docker, s.f)

- Docker Engine: Es el proceso que ejecuta y controla contenedores en los hosts. (Smith, R, 2017)
- docker-compose: Herramienta utilizada para orquestar la composición de servicios (Smith, R, 2017) Se utilizó para preparar los entornos de Desarrollo y de Producción

#### Lenguajes de programación:

- PHP: “PHP (acrónimo recursivo de PHP: Hypertext Preprocessor) es un lenguaje de código abierto muy popular especialmente adecuado para el desarrollo web y que puede ser incrustado en HTML.” (PHP, s.f)
- Python: “Python es un lenguaje de programación de tipo interpretado, funcional, orientado a objetos e interactivo. Comúnmente es comparado (dando la preferencia :) ) a Lisp, Tcl, Ruby, Perl, C#, Visual Basic, Visual Fox Pro, Scheme, Java, MatLab e incluso Julia. Posee una mayor madurez que muchos de ellos, de creciente potencia y multifuncionalidad.” (Python, s.f)

#### Tecnologías aplicadas al funcionamiento del sistema en general:

- Apache: Servidor Web.  
El proyecto del servidor HTTP Apache es un esfuerzo por desarrollar y mantener un servidor HTTP de código abierto para sistemas operativos modernos, incluidos UNIX y Windows. El objetivo de este proyecto es proporcionar un servidor seguro, eficiente y extensible que proporcione servicios HTTP en sincronía con los estándares HTTP actuales. (Apache HTTP, s.f.)
- Yii2 Framework:  
Yii es un marco de programación web genérico, lo que significa que se puede utilizar para desarrollar todo tipo de aplicaciones web utilizando PHP. Debido a su arquitectura basada en componentes y al soporte sofisticado de almacenamiento en caché, es especialmente adecuado para desarrollar aplicaciones a gran escala como portales, foros, sistemas de administración de contenido (CMS), proyectos de comercio electrónico, servicios web RESTful, etc. (Yii2,Introduction: About Yii, 2020)

#### Tecnologías y conceptos aplicados a resolver los requerimientos funcionales del proyecto:

- Procesamiento Natural del Lenguaje (PNL):

Las personas que están profundamente involucradas en el estudio del lenguaje son lingüistas, mientras que el término "lingüista computacional" se aplica al estudio de los lenguajes de procesamiento con la aplicación de la computación. Básicamente, un lingüista computacional será un científico de la computación que tenga suficiente comprensión de idiomas y pueda aplicar sus habilidades computacionales para modelar diferentes aspectos del lenguaje. Mientras que los lingüistas computacionales abordan el aspecto teórico del lenguaje, la PNL no es más que la aplicación de la lingüística computacional. (Hardeniya, Perkins, Chopra, Joshi, Mathur, 2016)

- Librería NLTK Python: Librería que incorpora funcionalidades de procesamiento de texto natural basada en Python.

La mayoría de las herramientas están escritas en Java y tienen funcionalidades similares. Algunos de ellos son robustos y tienen una variedad diferente de herramientas de PNL disponibles. Sin embargo, cuando se trata de la facilidad de uso y la explicación de los conceptos, NLTK puntúa muy alto. NLTK también es un kit de aprendizaje muy bueno porque la curva de aprendizaje de Python (en la que está escrito NLTK) es muy rápida. NLTK ha incorporado la mayoría de las tareas de PNL, es muy elegante y fácil de trabajar. Por todas estas razones, NLTK se ha convertido en una de las bibliotecas más populares de la comunidad de PNL. (Hardeniya, Perkins, Chopra, Joshi, Mathur, 2016)

- Análisis de Sentimientos: El análisis de sentimientos o la generación de sentimientos es una de las tareas de la PNL. Se define como el proceso de determinar los sentimientos detrás de una secuencia de caracteres. Puede usarse para determinar si el hablante o la persona que expresa los pensamientos textuales está de humor feliz o triste, o si representa una expresión neutral. (Hardeniya, Perkins, Chopra, Joshi, Mathur, 2016)

- UNICODE:

es un set de caracteres universal, es decir, un estándar en el que se definen todos los caracteres necesarios para la escritura de la mayoría de los idiomas hablados en la actualidad que se usan en la computadora. Su objetivo es ser, y, en gran medida, ya lo ha logrado, un superconjunto de todos los sets de caracteres que se hayan codificado (w3, s.f)

- Emoji:

Los emoji son pictografías (símbolos pictóricos) que normalmente se presentan en forma colorida y se utilizan en línea en el texto. Representan cosas como caras, clima, vehículos y edificios, comida y bebida, animales y plantas, o íconos que representan emociones, sentimientos o actividades. Para la computadora, son simplemente otro personaje, pero las personas se envían miles de millones de emoji todos los días para expresar amor, agradecimiento, felicitaciones o cualquier número de un conjunto de ideas en crecimiento. (Unicode.org, s.f.)

#### Persistencia y gestión de datos

- MySQL: Sistema Gestor de Bases de Datos Relacionales (SGDBR)

Con más de 10 millones de instalaciones, MySQL es probablemente el sistema de administración de bases de datos más popular para servidores web. Desarrollada a mediados de la década de 1990, ahora es una tecnología madura que impulsa muchos de los destinos de Internet más visitados de la actualidad. (Nixon, 2018)

- PHPMyadmin: Cliente web basado en PHP que permite conectarse a un SGDB MySQL y realizar tareas de administración y ejecución de consultas.

es una herramienta de software gratuita escrita en PHP, destinada a manejar la administración de MySQL a través de la Web. phpMyAdmin admite una amplia gama de operaciones en MySQL y MariaDB. Las operaciones de uso frecuente (administración de bases de datos, tablas, columnas, relaciones, índices, usuarios, permisos, etc.) se pueden realizar a través de la interfaz de usuario, mientras aún tiene la capacidad de ejecutar directamente cualquier declaración SQL. (Phpmyadmin, s.f.)

Tecnologías aplicadas a la usabilidad de la aplicación, interfaces de usuario, visualización de resultados:

- HTML5: Lenguaje de marcas para la elaboración de páginas web.

representa un avance sustancial en el diseño, el diseño y la usabilidad web. Proporciona una forma sencilla de manipular gráficos en un navegador web sin recurrir a complementos como Flash, ofrece métodos para insertar audio y video en páginas web (nuevamente sin complementos) y elimina varias inconsistencias molestas que se infiltraron en HTML durante su evolución. (Nixon, 2018)

- CSS3: Hojas de estilo que permiten dar formato a las páginas web. “es el complemento fundamental de HTML, ya que garantiza que el texto HTML y las imágenes incrustadas se presenten de forma coherente y adecuada para la pantalla del usuario.” (Nixon, 2018)
- JavaScript: Lenguaje de programación intepretado que se ejecuta en el navegador web utilizado para dar dinamismo a las páginas.
 

se creó para permitir el acceso de secuencias de comandos a todos los elementos de un documento HTML. En otras palabras, proporciona un medio para la interacción dinámica del usuario, como verificar la validez de la dirección de correo electrónico en los formularios de entrada y mostrar mensajes como "¿Realmente quiso decir eso?" (Nixon, 2018)
- Bootstrap: Librería que permite maquetar una página web utilizando CSS y JavaScript. Según el sitio oficial es “el marco de trabajo más popular del mundo para crear sitios con capacidad de respuesta para dispositivos móviles, con BootstrapCDN y una página de inicio de plantilla” (Bootstrap, s.f.)
- JQuery:
 

Potente y flexible como es JavaScript, con su gran cantidad de funciones integradas, aún necesita capas adicionales de código para cosas simples que no se pueden lograr de forma nativa o con CSS, como animaciones, manejo de eventos y comunicación asincrónica (Nixon, 2018)
- Openmoji: Librería de imágenes que tienen una representación en formato de imagen de emojis. “el primer sistema de emoji de código abierto e independiente hasta la fecha.” (Openmoji, s.f.)
- D3.js: Librería que utiliza JavaScript y SVG para la generación de visualizaciones dinámicas.
 

D3.js es una biblioteca de JavaScript para manipular documentos basados en datos. D3 le ayuda a dar vida a los datos mediante HTML, SVG y CSS. El énfasis de D3 en los estándares web le brinda todas las capacidades de los navegadores modernos sin atarse a un marco propietario, combinando poderosos componentes de visualización y un enfoque basado en datos para la manipulación DOM. (D3, s.f)

Tecnologías aplicadas a la conectividad física y lógica:

- TCP/IP: Conjunto de protocolos de comunicaciones que permite la comunicación entre dos sistemas. (Tanenbaum, Wetherall, 2012)

- HTTP:

es un estándar de comunicación que rige las solicitudes y respuestas que se envían entre el navegador que se ejecuta en la computadora del usuario final y el servidor web. El trabajo del servidor es aceptar una solicitud del cliente e intentar responder a ella de una manera significativa, generalmente ofreciendo una página web solicitada; por eso se usa el término servidor. La contraparte natural de un servidor es un cliente, por lo que ese término se aplica tanto al navegador web como a la computadora en la que se ejecuta. (Nixon, 2018)

- SSL: Utilizado para encriptar de extremo a extremo las comunicaciones entre dos dispositivos “es una Capa de Sockets Seguros que construye una conexión segura entre dos sockets, incluyendo: Negociación de parámetros entre el cliente y el servidor; Autenticación del servidor por el cliente; Comunicación secreta; Protección de la integridad de los datos.” (Tanenbaum, Wetherall, 2012)
- Switch: Dispositivo concentrador de redes de datos que actúa hasta el nivel de capa de enlace en el protocolo TCP/IP.

Un switch tiene varios puertos, cada uno de los cuales se puede conectar a una computadora. El trabajo del switch es transmitir paquetes entre las computadoras conectadas a él, y utiliza la dirección en cada paquete para determinar a qué computadora se lo debe enviar. (Tanenbaum, Wetherall, 2012)

- Access point: Punto de acceso inalámbrico para conectar una estación de trabajo con una infraestructura de red. Utilizado para que las estaciones de trabajo puedan conectarse sin cables a la red local. Definido como “enrutador inalámbrico o estación base; transmite paquetes entre las computadoras inalámbricas y también entre éstas e Internet”. (Tanenbaum, Wetherall, 2012)

### *Competencias*

No se han encontrado soluciones similares en publicaciones, papers o disponibles en Internet.

Durante las búsquedas se encuentran soluciones parciales que responden a alguno de los puntos del proceso, pero de utilizarlas aun requieren de una intervención manual por parte del investigador.

### **Diseño metodológico**

Como metodología de desarrollo se utilizó UML, en particular el marco de trabajo proporcionado por el Proceso Unificado de Desarrollo (PUD).

La materialización del proyecto requirió del uso de métodos de modelado y procesamiento de estructuras de datos, para procesar los mensajes y dividirlos en unidades de análisis. Luego de su análisis, se ordenaron los resultados.

Las unidades requirieron de un análisis utilizando técnicas de procesamiento de lenguaje natural pertenecientes a la rama de la Inteligencia artificial con el objetivo de obtener una clasificación emocional.

Las técnicas de recolección de datos utilizadas han sido la observación y la entrevista personal debido a la posibilidad de poder contar con acceso directo a investigadores de las comunicaciones digitales que permitieron observar la metodología de trabajo que utilizan en la actualidad. Mediante la indagación se pudo consensuar los requerimientos funcionales que permiten resolver las problemáticas más relevantes.

Para poder obtener una cantidad de ejemplos reales sobre los cuales realizar los análisis se habilitó un espacio de recepción de archivos donde voluntarios cargaron exportaciones de sus sistemas de mensajería en forma manual. El formulario de recepción informa que los datos serán procesados con fines de investigación. Los nombres de las personas intervinientes son sustituidos mediante la técnica de anonimización empleada por Cantamutto (2018) que consiste en el reemplazo de nombres por otros similares a través de un sistema de equivalencia ya puesto a prueba. De este modo, se garantiza la no trazabilidad de los datos digitales (de-Matteis, 2014).

En los anexos 1 y 2 se encuentran ejemplos del formulario de recepción de archivos y de un archivo anonimizado.

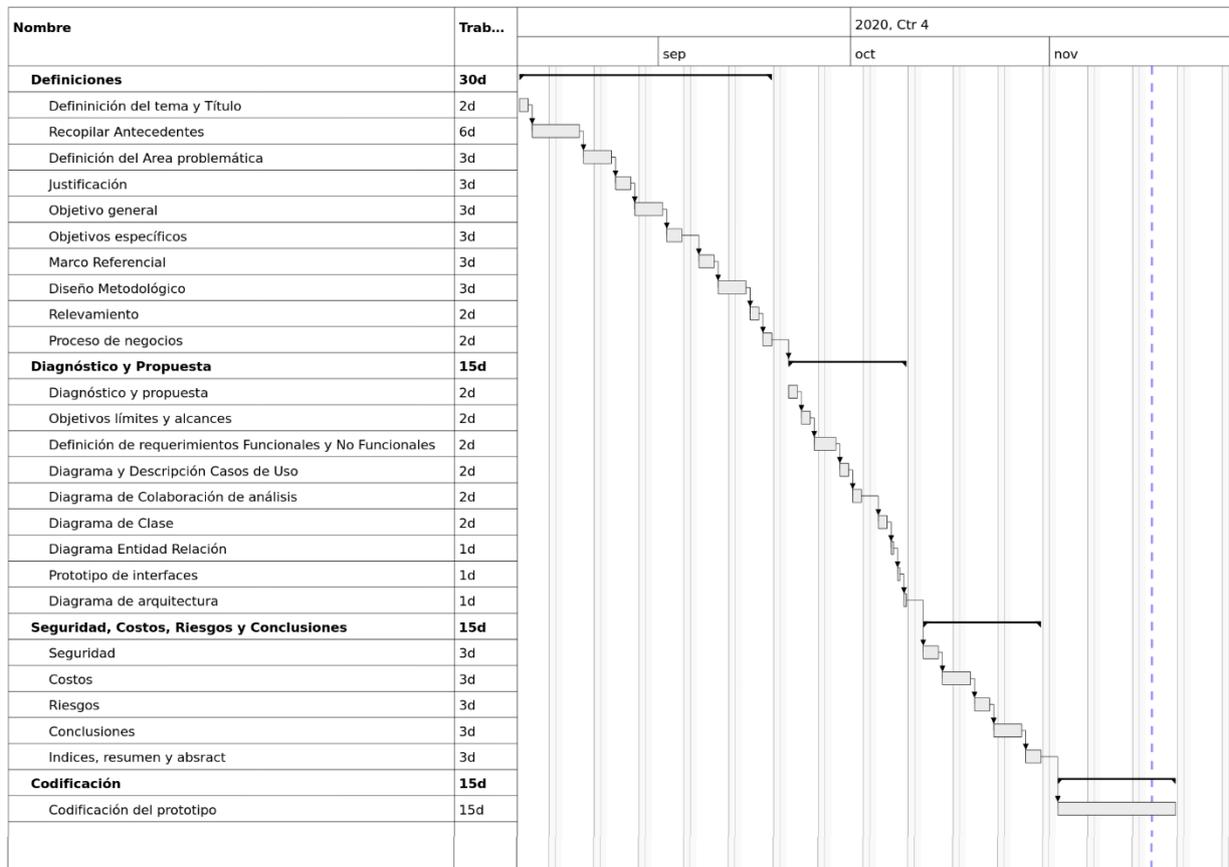
El sistema utiliza como gestor de bases de datos relacionales Mysql con el propósito de facilitar la carga de las entradas a procesar y guardar los estados intermedios de procesamiento y objetos de salida resultantes. La administración de la base se realizó con PHPMysql por la flexibilidad de poder realizar la gestión desde un navegador web.

Para interactuar con el usuario se utilizó una interfaz web con lenguaje Html5, CSS, Bootstrap, JavaScript y JQuery, lo que permitió lograr interfaces de usuario cómodas, flexibles y modernas.

El lenguaje de programación utilizado para resolver las tareas asociadas al análisis del mensaje es Python debido a que proporciona una mayor madurez en librerías de procesamiento de texto, seleccionando como principal NLTK. Para el resto de las tareas se utiliza el lenguaje PHP apoyado en el framework de desarrollo Yii2 que refuerza la construcción de aplicaciones de calidad y seguridad.

El siguiente diagrama de Gantt contiene la descripción y organización de los hitos cumplidos para lograr el objetivo de completar el presente Trabajo Final de Graduación.

*Ilustración 1. Diagrama de Gantt*



Fuente: elaboración propia

## Relevamiento

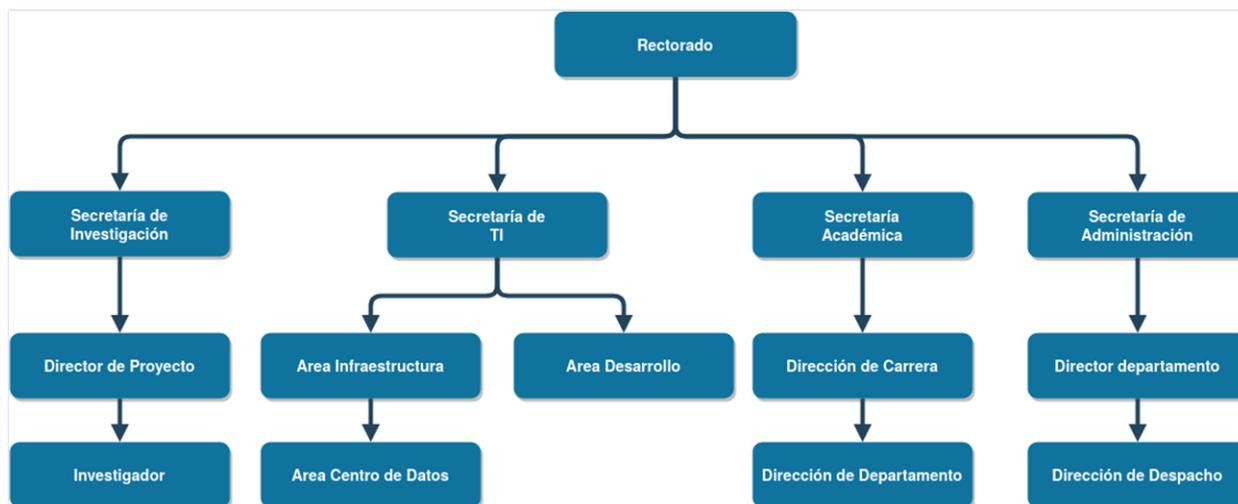
El proyecto tiene como destinatarios principales a los investigadores de las comunicaciones digitales que, en la actualidad, no se encuentran todos dentro de una única organización, sino que se hallan insertos en una variedad de organizaciones públicas y privadas en las que se desarrollan proyectos de investigación. Por ese motivo se hace necesario modelar una organización genérica que contemple una estructura organizativa clásica que podría estar presente dentro de la diversidad de organizaciones donde se utilizará el sistema.

Se ha tomado como modelo genérico... una organización educativa en la que los investigadores desarrollan proyectos de investigación relacionados con la comunicación digital.

### Relevamiento estructural

Un modelo representativo de las unidades organizativas de la organización genérica en la que se desenvuelven los investigadores de las comunicaciones digitales es el siguiente. Al ser un proyecto de tecnología requiere del apoyo de una unidad organizativa especializada en tecnología capaz de poner en marcha los servicios necesarios para el funcionamiento.

Ilustración 2. Organigrama



Fuente: elaboración propia

Las unidades organizativas de interés son:

- **Secretaría de Investigación:** Se encarga de gestionar los proyectos de investigación que se desarrollan en la universidad. Evalúa propuestas de nuevos proyectos, controla la ejecución administrativa y presupuestaria durante la ejecución y verifica que se finalicen cumpliendo todos los requisitos formales. Incentiva a estudiantes, profesores y graduados a involucrarse en la investigación.
- **Secretaría de Administración:** Se encarga de gestionar los recursos económicos de la universidad y de cumplir la ejecución del presupuesto anual. Verifica que los movimientos económicos que se producen dentro de los proyectos de investigación se encuentren debidamente fundamentados y se mantengan dentro del presupuesto destinado a cada proyecto en particular.
- **Secretaría de TI:** Gestiona los recursos humanos y técnicos para asegurar el funcionamiento de los dispositivos de computación y telecomunicaciones.

- Director de proyecto: Responsable de la ejecución de un proyecto de investigación, coordina las actividades entre los investigadores con el fin de lograr los objetivos y metas necesarios para cumplir con el objetivo general del proyecto.
- Investigador: Realiza tareas dentro del proyecto de investigación con el propósito de cumplir los distintos pasos de la planificación del proyecto con el propósito de obtener información de valor para determinar la validez de las hipótesis.

### *Relevamiento funcional*

Relevadas las funciones de las estructuras organizativas se concluye que en la ejecución de las distintas tareas participarán en mayor medida investigadores, destacando los siguientes procesos cuando se trata de estudiar las comunicaciones digitales:

1. Seleccionar archivo con mensajes
2. Dividir en unidades procesables separados por interviniente
3. Clasificar la emoción de los mensajes
4. Generar una Línea de tiempo
5. Detectar puntos de inflexión
6. Generar archivos y reportes

El siguiente es un análisis de cada proceso:

- Proceso: Seleccionar archivos con mensajes
  - Roles: Usuario
  - Pasos: El usuario selecciona un archivo de texto plano que contiene los mensajes que desea procesar. El archivo debe ser relevante al proyecto en el que está trabajando.
- Proceso: Dividir en unidades procesables separados por interviniente
  - Roles: Usuario, Procesador de Textos, Base de Datos
  - Pasos: Se debe recorrer el archivo de texto plano separando los mensajes por emisor y separando la fecha y hora en que fue generado, texto del mensaje y emojis encontrados.
- Proceso: Clasificar la emoción de los mensajes
  - Roles: Usuario, Base de Datos
  - Pasos: Se debe tomar un mensaje individual y se observan las palabras en el cuerpo del texto y emojis asociados. Al finalizar la interpretación se lo valora como positivo, neutro o negativo y se guarda su ponderación asociada.

- Proceso: Generar una Línea de tiempo
  - Roles: Usuario, Base de Datos, Planilla de Cálculos
  - Pasos: Se arma una línea temporal que inicia en la fecha y hora del mensaje que tenga la fecha más antigua y finaliza en la fecha y hora del mensaje más nuevo. Se establece un intervalo en minutos y se completa el intervalo con fechas y horas consecutivas separadas por el intervalo establecido. Se genera una columna por cada participante de la conversación y se ubican los mensajes en una fila según la fecha y hora del mensaje y el intervalo en el que quede enmarcada.
- Proceso: Detectar puntos de inflexión
  - Roles: Usuario, Planilla de Cálculos
  - Pasos: Se recorre la línea de tiempo y se analiza el contexto cercano a cada mensaje buscando un cambio de valoración entre el mensaje actual y los que se encuentren adyacentes en el intervalo. Se realiza una anotación sobre el mensaje para destacarlo.
- Proceso: Generar archivos y reportes
  - Roles: Usuario, Planilla de Cálculos
  - Pasos: Se utiliza la línea de tiempo con la valoración para generar un reporte que contenga una lista de las personas intervinientes en el mensaje, estadística asociada a cada una incluyendo cantidad de mensajes, palabras utilizadas, cantidad de emociones detectadas, uso de emojis. Se generan gráficos de línea de tiempo destacando los mensajes donde se registraron puntos de inflexión. Los resultados se incorporan al proyecto de investigación.

### *Relevamiento de documentación*

Como documentación respaldatoria de los procesos se destacan los siguientes elementos:

Archivo de exportación de mensajes: Es un documento de entrada que contiene los mensajes en formato plano en secuencia

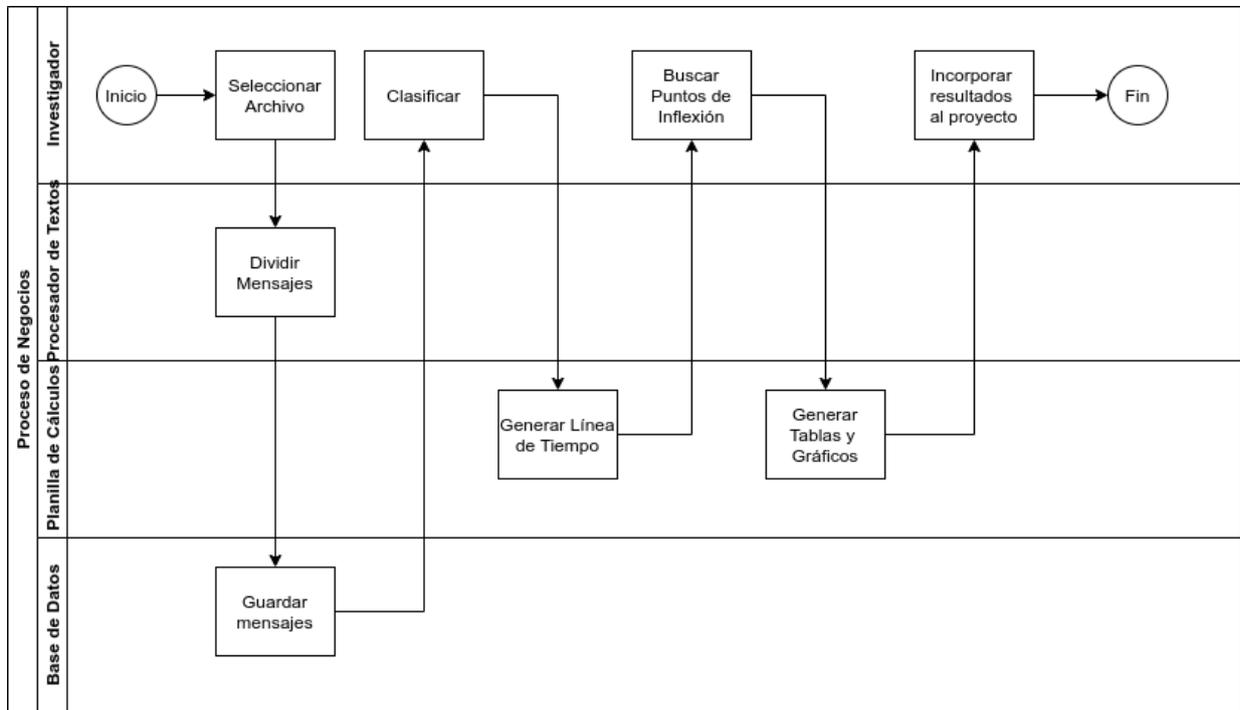
Listado de mensajes en línea de tiempo: Es un documento de salida que contiene el resultado de aplicar métodos de procesamiento de datos.

Listado de clasificación de emociones: Es un documento de salida que contiene la línea de tiempo donde se destacan los mensajes que producen una variación en las emociones siguientes.

## Proceso de negocio

Analizados los pasos necesarios para completar el objetivo general del sistema se puede visualizar en el diagrama de la figura el proceso de negocios general.

Ilustración 3. Proceso de negocio



Fuente: elaboración propia

## Diagnóstico y Propuesta

### Diagnóstico

Durante el relevamiento de procesos se han encontrado los siguientes problemas:

- Proceso: Seleccionar archivo con mensajes
  - Problemas:
    - Dificultad para encontrar los archivos relacionados con el proyecto
    - Dificultad para medir el avance cuando se procesan varios archivos
  - Causas:
    - No se etiquetan correctamente los archivos correspondientes a un proyecto y quedan todos en un directorio mezclados con archivos de otros proyectos.
    - Los procedimientos manuales requieren de un registro adicional para llevar un control de cuántos archivos se han procesado para tener claridad de cuántos archivos y tiempo restan para completar la tarea.

- Proceso: Dividir en unidades procesables separados por interviniente
  - Problemas:
    - Dificultad para la separación de los mensajes que forman parte del archivo de texto.
    - Dificultad para separar los emojis.
  - Causas:
    - Los mensajes no están siempre separados por saltos de línea y se debe buscar en el texto indicadores de inicio y fin.
    - Los emojis se visualizan como caracteres Unicode que en el procesador de textos se ven como dos o más caracteres contiguos que ante un error pueden ser rotos por no ser reconocidos como una entidad única.
- Proceso: Clasificar la emoción de los mensajes
  - Problemas:
    - Mucho tiempo de procesamiento para lograr la valoración del mensaje
  - Causas:
    - El procesamiento manual de los mensajes requiere de una lectura y comprensión del texto y un análisis de los emojis involucrados para que el investigador concluya en una valoración. El trabajo repetitivo genera agotamiento y se reduce la cantidad de mensajes que pueden ser procesados en un período de tiempo.
- Proceso: Generar una Línea de tiempo
  - Problemas:
    - Se debe reconstruir frecuentemente la línea de tiempo.
  - Causas:
    - La tarea manual de construir la tabla de tiempo implica encontrar parámetros de fechas y horas máximos y mínimos, cantidades de participantes y la asignación manual de cada mensaje dentro de su marco de tiempo correspondiente. Un parámetro mal detectado requiere que se deban descartar muchos pasos ya realizados desperdiciando tiempo y esfuerzo.
- Proceso: Detectar puntos de inflexión
  - Problemas:
    - Dificultad para detectar los cambios de estados entre mensajes adyacentes

- Causas:
  - El proceso es manual y requiere de la capacidad de observación del operador que se ve afectada por el procesamiento rutinario de gran cantidad de mensajes.
- Proceso: Generar archivos y reportes
  - Problemas:
    - Se deben regenerar manualmente todos los archivos y reportes cuando se producen cambios en las etapas anteriores.
  - Causas:
    - La generación manual de archivos con resultados y gráficos se basa en la línea de tiempo y las valoraciones asociadas a los mensajes. Cuando se realizan cambios que impactan en la línea de tiempo se debe descartar todo lo realizado y generar nuevos archivos y reportes desperdiciando tiempo y energías del investigador.

### *Propuesta*

Como solución a los problemas que se mencionaron en el diagnóstico se propuso la realización de un sistema web en el que los investigadores puedan gestionar múltiples proyectos y asociar a cada uno los recursos que serán objeto de investigación. Los recursos atraviesan por estados de procesamiento que permite diferenciar los que están pendientes de procesar.

El investigador selecciona un recurso para procesar de los que están pendientes y obtiene como resultado los archivos y reportes que le aportan la información necesaria para cumplir con los objetivos de su proyecto.

El tratamiento que se le da a cada archivo con mensajes durante el procesamiento resuelve las dificultades de separar desde un archivo de texto los mensajes, asociarlos a una fecha y un emisor, valorarlos emocionalmente según el contenido de las palabras del texto y los emojis encontrados, generar una línea de tiempo con un marco de ventana ajustable, detectar la variación de emociones entre los mensajes adyacentes y finalmente generar una planilla de cálculos con los informes y tablas que el investigador debe adjuntar al proyecto de investigación.

Este procesamiento automatizado resuelve el problema de tener que regenerar archivos, reportes y líneas de tiempo si ocurren cambios en los objetos de entrada o se desea ajustar el

intervalo de la línea de tiempo, permitiendo al investigador elaborar nuevos escenarios mediante la selección y ajuste de parámetros del procesamiento.

## **Objetivos, Límites y Alcances del Prototipo**

### *Objetivos del Prototipo*

Generar un reporte en formato de línea de tiempo con la valoración emocional de mensajes obtenidos de un archivo exportado desde un sistema de mensajería privada donde se destaquen los mensajes que producen cambios de estado en las interacciones siguientes.

### *Límites*

El sistema contempla desde la recepción de un archivo conteniendo las interacciones de comunicaciones digitales hasta la generación y descarga de archivos y reportes con los resultados del procesamiento.

### *Alcance*

En el prototipo se desarrollan las siguientes funcionalidades:

- Gestión de Proyectos
- Asociación de archivos a proyectos
- Procesamiento de archivos del proyecto con parametrización
- División del texto del archivo en mensajes
- Clasificación de emoción en cada mensaje
- Generación de la línea de tiempo
- Detección de puntos de inflexión
- Generación de reportes y archivos descargables

## **Descripción del Sistema**

### *Requerimientos Funcionales*

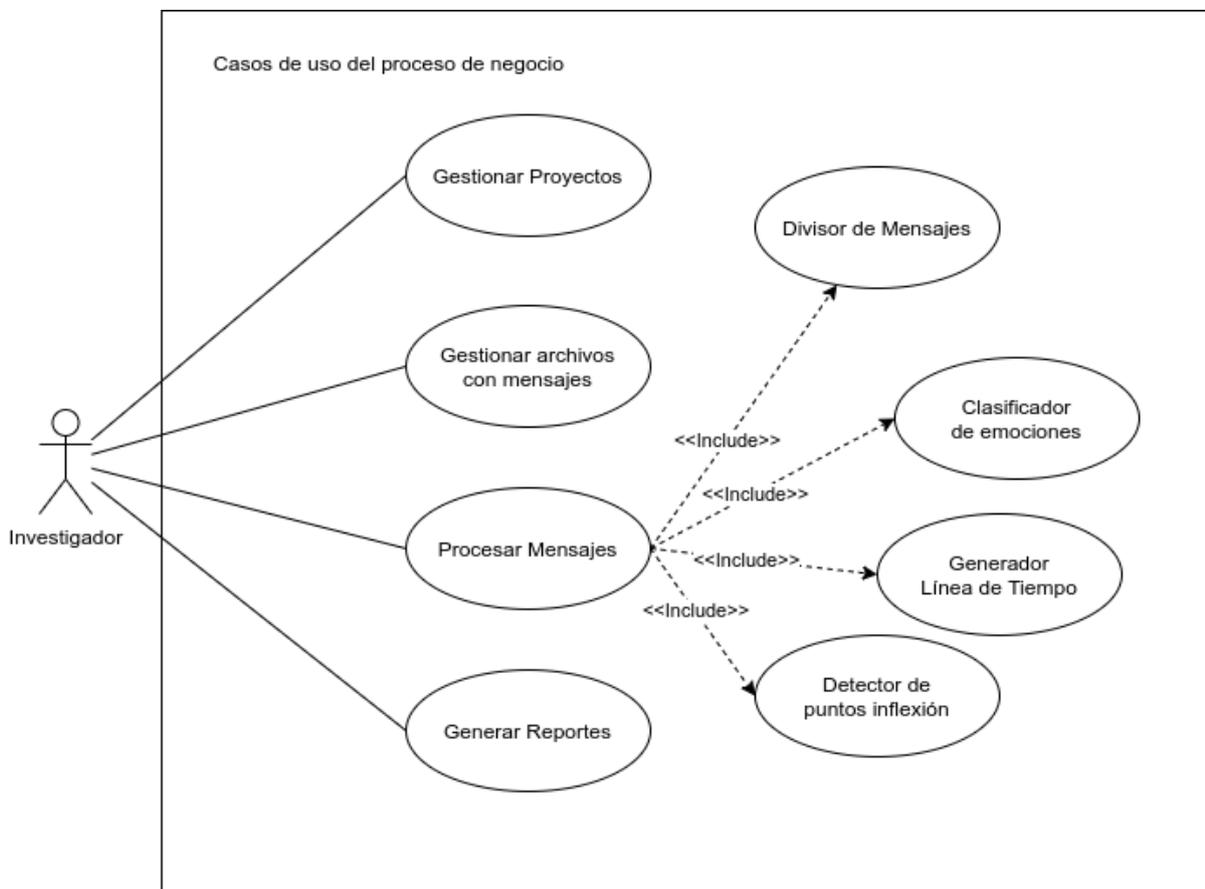
- RF01 - Gestionar proyectos
- RF02 - Gestionar archivos con mensajes
- RF03 - Procesar archivos de mensajes
- RF04 – Generar de reportes por proyectos

*Requerimientos No Funcionales*

- RNF01 El sistema puede utilizarse desde cualquier sistema operativo que disponga de un navegador web.
- RNF02 El sistema debe adaptarse a cualquier tamaño de dispositivo cliente.
- RNF03 El sistema deberá estar accesible para un acceso local y remoto.
- RNF04 El sistema debe utilizar protocolos de seguridad para proteger las comunicaciones entre los clientes y el servidor.
- RNF05 El sistema debe almacenar la contraseña de usuario utilizando una función de resumen que genere un resultado de al menos 184 bits.

*Diagrama de Casos de uso*

*Ilustración 4. Diagramas de casos de uso*



Fuente: Elaboración Propia

*Descripción de Casos de uso*

Se describen cuatro casos de uso significativos

<b>RF01</b>	Gestionar Proyectos
<b>Versión</b>	1.0 (11/04/2020)

<b>Actor</b>	Investigador		
<b>Descripción</b>	Como investigador quiero dar de alta un nuevo proyecto		
<b>Precondición</b>	No hay condiciones previas		
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>	
	1	El investigador solicita ingresar un nuevo proyecto	
	2	El sistema solicita que ingrese un número de proyecto	
	3	El sistema solicita que ingrese el nombre del proyecto	
	4	El sistema solicita que ingres una descripción del proyecto	
	5	El sistema informa que el proyecto fue guardado exitosamente	
<b>Postcondición</b>	Se guarda el proyecto en la base de datos		
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>	
	2	Si el número de proyecto ya existe	
		2.1	Informar error de número duplicado
		2.2	Se cancela el caso de uso
<b>Importancia</b>	Importante		
<b>Comentarios</b>	El alta de un proyecto es requerida para poder asociar archivos con mensajes		

<b>RF02</b>	Gestionar Archivos con Mensajes	
<b>Versión</b>	1.0 (11/04/2020)	
<b>Actor</b>	Investigador	
<b>Descripción</b>	Como investigador quiero cargar un nuevo archivo con mensajes asociado a un proyecto	
<b>Precondición</b>	El proyecto debe estar cargado	
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>
	1	El investigador solicita ingresar un nuevo archivo
	2	El sistema solicita que seleccione un número de proyecto existente
	3	El sistema solicita que adjunte el archivo
	4	El sistema solicita que ingres una descripción del archivo
	5	El sistema informa que el archivo fue guardado exitosamente
<b>Postcondición</b>	Se guarda el archivo asociado al proyecto en la base de datos	
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>
	2	Si el número de proyecto no existe
		2.1

	2.2	Se cancela el caso de uso
<b>Importancia</b>	Importante	
<b>Comentarios</b>	El alta de un proyecto es requerida para poder asociar archivos con mensajes	

<b>RF03</b>	Procesar Mensajes	
<b>Versión</b>	1.0 (11/04/2020)	
<b>Actor</b>	Investigador	
<b>Descripción</b>	Como investigador quiero realizar el procesamiento de un mensaje	
<b>Precondición</b>	El archivo debe estar cargado previamente	
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>
	1	El investigador solicita ingresar el procesamiento de un mensaje
	2	El sistema solicita que ingrese un número de proyecto
	3	El sistema solicita que seleccione el archivo a procesar
	4	El sistema descompone archivo en unidades de mensajes
	5	Para cada mensaje único
	5.1	Generar valoración de mensaje
	6	El sistema genera una línea de tiempo
	7	Para cada mensaje único
	7.1	Insertar mensaje en línea de tiempo
8	El sistema recorre la línea de tiempo marcando los mensajes que tienen una variación emocional con los adyacentes	
9	El sistema informa que el proceso ha finalizado exitosamente	
<b>Postcondición</b>	Se guarda la clasificación del archivo en la base de datos	
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>
	2	Si el número de proyecto no existe
	2.1	Informar error de número duplicado
	2.2	Se cancela el caso de uso
<b>Importancia</b>	Importante	
<b>Comentarios</b>	Si hubiera procesamientos anteriores se reemplazan por el de la última ejecución	

<b>RF03A</b>	Divisor de Mensajes	
<b>Versión</b>	1.0 (11/04/2020)	
<b>Actor</b>	RF03	

<b>Descripción</b>	Se debe dividir un archivo de texto plano en mensajes individuales		
<b>Precondición</b>	Paso 03 de RF03		
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>	
	1	Se recibe un archivo de texto	
	2	Se divide el archivo en bloques de mensajes usando una expresión regular que detecta los márgenes del mensaje	
	3	Para cada mensaje único	
		3.1	Usando expresiones regulares separar fecha, remitente y mensaje
3.2	Generar un nuevo mensaje con los datos de fecha, remitente y mensaje		
4	Finalizar		
<b>Postcondición</b>	Los mensajes quedan en la tabla de mensajes		
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>	
	2	Si no se pueden encontrar los límites	
		2.1	Informar posible error de formato
		2.2	Retornar un error
<b>Importancia</b>	Importante		
<b>Comentarios</b>			

<b>RF03B</b>	Generar Valoración del de un mensaje	
<b>Versión</b>	1.0 (11/04/2020)	
<b>Actor</b>	RF03	
<b>Descripción</b>	Generar una valoración de positivo, negativo o neutro sobre un mensaje	
<b>Precondición</b>	Paso 04 de RF03 Se debe contar con una tabla de asociación de emociones a palabras: -1 negativa, 0 neutra, 1 positiva Se debe contar con una tabla de asociación de emociones a emojis: -1 negativa, 0 neutra, 1 positiva Se debe contar con una lista de palabras no relevantes (Stopwords)	
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>
	1	Se recibe un mensaje que contiene un texto
	2	Se separan las palabras y emojis del texto del mensaje y se descartan las palabras Stopwords
	3	Inicializar variables para ir acumulando resultados de valoración
	4	Para cada palabra
4.1		Comparar la palabra con una lista de asociación y emoción y obtener una valoración entre -1, 0 y 1

	4.2	Sumar la valoración obtenida en el acumulador
	5	Para cada Emoji
	5.1	Comparar el Emoji con una lista de asociación y emoción y obtener una valoración entre -1, 0 y 1
	5.2	Sumar la valoración obtenida en el acumulador
	6	Determinar la valoración total entre -1, 0 y 1 según el acumulador sea negativo, cero o positivo
<b>Postcondición</b>	Se devuelve una valoración entre -1, 0 y 1	
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>
	2	No hay palabras ni emojis procesables
	2.1	Informar posible error de formato
	2.2	Retornar un error
<b>Importancia</b>	Importante	
<b>Comentarios</b>	Se tienen en cuenta solo palabras y emojis, no se consideran los signos de puntuación	

<b>RF03C</b>	Generador de línea de tiempo		
<b>Versión</b>	1.0 (11/04/2020)		
<b>Actor</b>	RF03		
<b>Descripción</b>	Retornar una estructura de datos que contenga los participantes con sus mensajes agrupados por grupos según una ventana de tiempo		
<b>Precondición</b>	Paso 05 de RF03 Se debe contar con la lista de mensajes con su fecha y hora, emisor y valoración		
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>	
	1	Se recibe como parámetro el tamaño de ventana	
	2	Recorrer la lista de mensajes para separar la lista de emisores	
	3	Inicializar estructura de datos con los emisores, ventana y mensajes	
	4	Para cada mensaje	
		4.1	Calcular la ventana de tiempo que le corresponde.
		4.2	Insertar en la estructura de datos para el participante y ventana que le corresponden.
6	Retornar la estructura		
<b>Postcondición</b>	Se devuelve una estructura de datos con los participantes y los mensajes agrupados en una ventana de tiempo.		

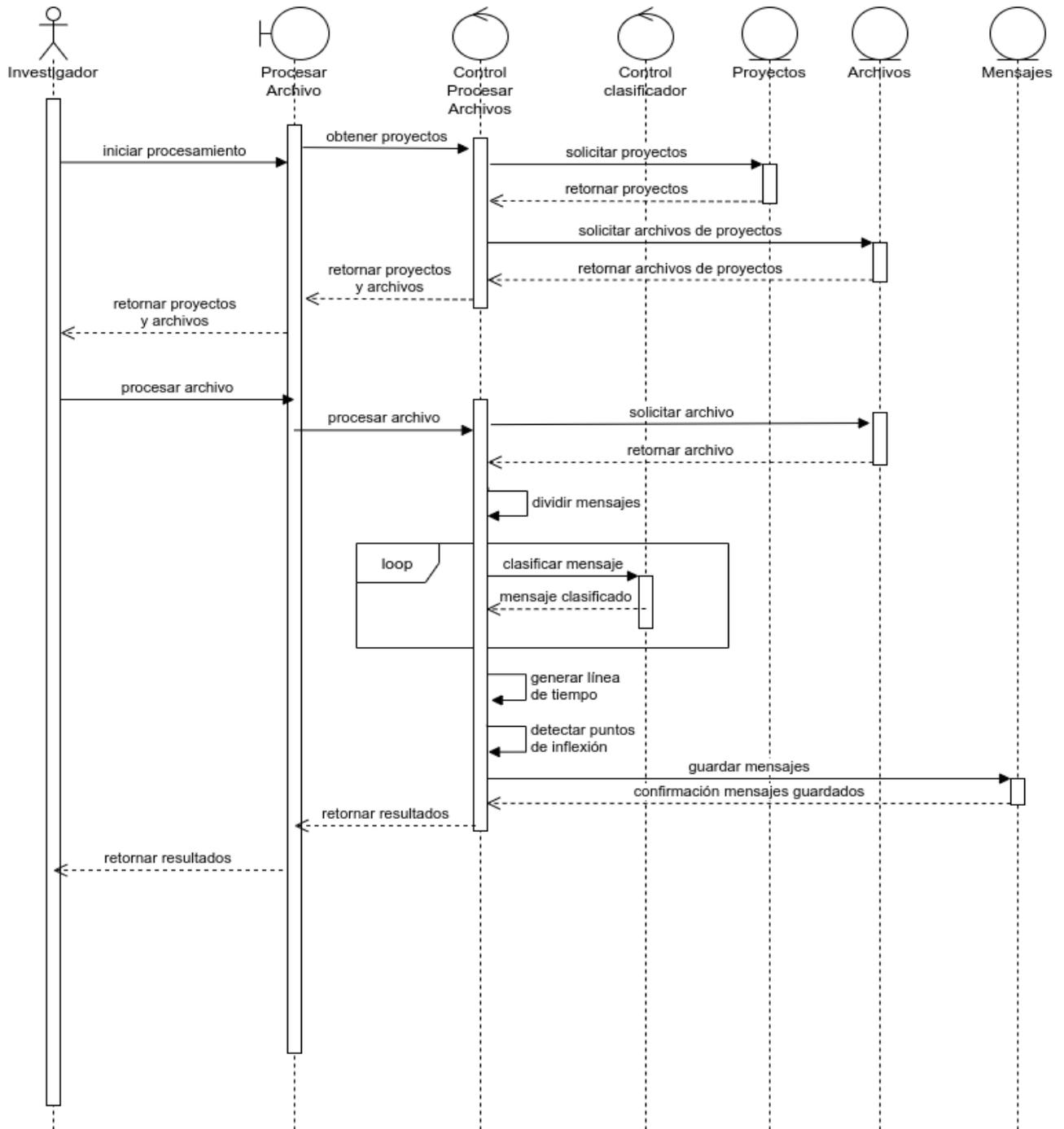
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>
	2	No hay mensajes
	2.1	Informar que no se encuentran mensajes
	2.2	Retornar un error
<b>Importancia</b>	Importante	
<b>Comentarios</b>		

<b>RF03D</b>	Detector de Puntos de Inflexión		
<b>Versión</b>	1.0 (11/04/2020)		
<b>Actor</b>	RF03		
<b>Descripción</b>	Recorrer la lista de mensajes y comparar la valoración emocional de los mensajes de otros participantes pero que están situados en una ventana temporal adyacente con el fin de determinar un posible impacto de ese mensaje.		
<b>Precondición</b>	Paso 07 de RF03 Se debe contar con la lista de mensajes con su fecha y hora, emisor y valoración		
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>	
	1	Se recibe la estructura de datos de línea de tiempo	
	2	Para cada ventana de tiempo	
	2.1	Para cada mensaje M en la ventana	
		2.1.1	Para cada uno de los otros participantes de la conversación
			2.1.1.1
6	Retornar		
<b>Postcondición</b>	Los mensajes considerados de impacto quedan marcados como puntos de inflexión		
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>	
	2	Si no hay ventanas de tiempo	
	2.1	Informar que la estructura no tiene mensajes	
	2.2	Retornar un error	
<b>Importancia</b>	Importante		
<b>Comentarios</b>			

<b>RF04</b>	Generar Reportes		
<b>Versión</b>	1.0 (11/04/2020)		
<b>Actor</b>	Investigador		
<b>Descripción</b>	Como investigador quiero imprimir un reporte del resultado de procesar un archivo		
<b>Precondición</b>	El archivo debe haber sido procesado previamente		
<b>Secuencia Normal</b>	<b>Paso</b>	<b>Acción</b>	
	1	El actor solicita generar un reporte	
	2	El sistema solicita que ingrese un número de proyecto	
	3	El sistema solicita que seleccione el archivo a procesar	
	4	El sistema pregunta si desea descargar	
	4.1	El sistema envía el reporte para descarga	
	5	El sistema muestra el reporte	
<b>Postcondición</b>	No hay		
<b>Excepciones</b>	<b>Paso</b>	<b>Acción</b>	
	2	Si el número de proyecto no	
		2.1	Informar error de proyecto inexistente
		2.2	Se cancela el caso de uso
<b>Importancia</b>	Importante		
<b>Comentarios</b>	El archivo para descargar es un archivo que contiene las tablas y gráficos que se visualizan en el reporte		

Diagrama de Colaboración de análisis

Ilustración 5. Diagrama de secuencia

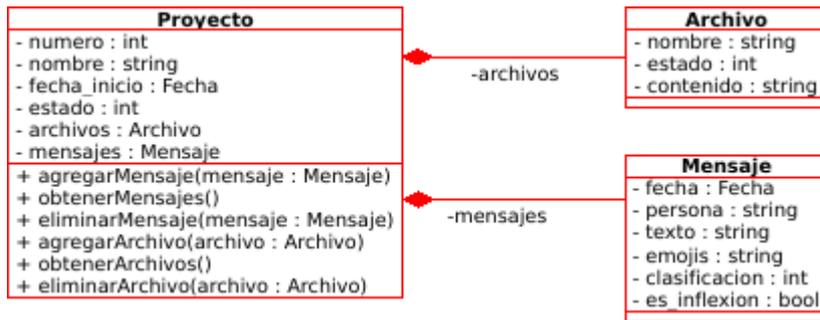


Fuente: elaboración propia

Diagramas de Clase

El diagrama representa las clases que son relevantes al Core del proyecto. Se omiten los setters y getters y se omiten las relacionadas con el funcionamiento genérico: Usuarios, Roles, Perfiles y Estados en general.

Ilustración 6. Diagrama de Clases



Fuente: elaboración propia

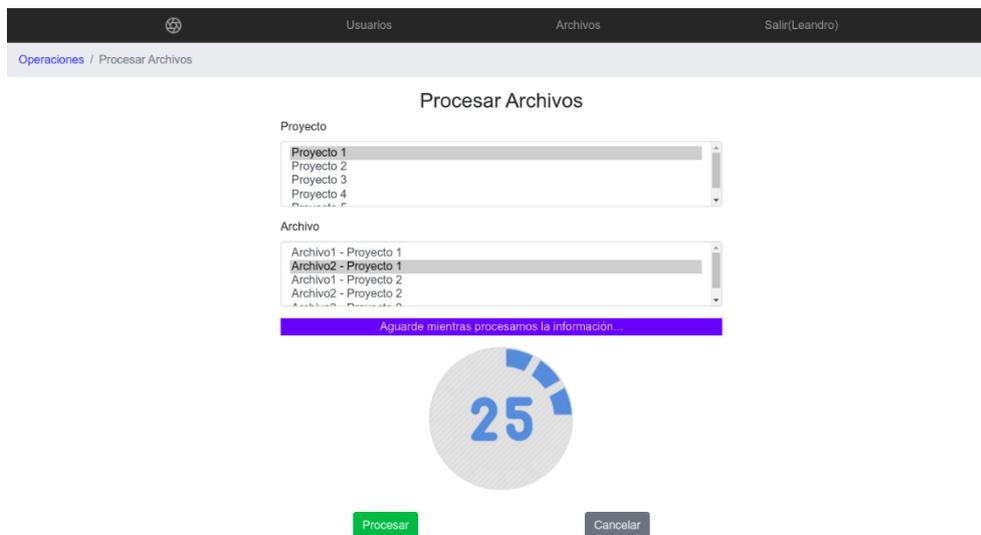
### Prototipos de interfaces de pantallas

Ilustración 7. Pantalla carga archivo



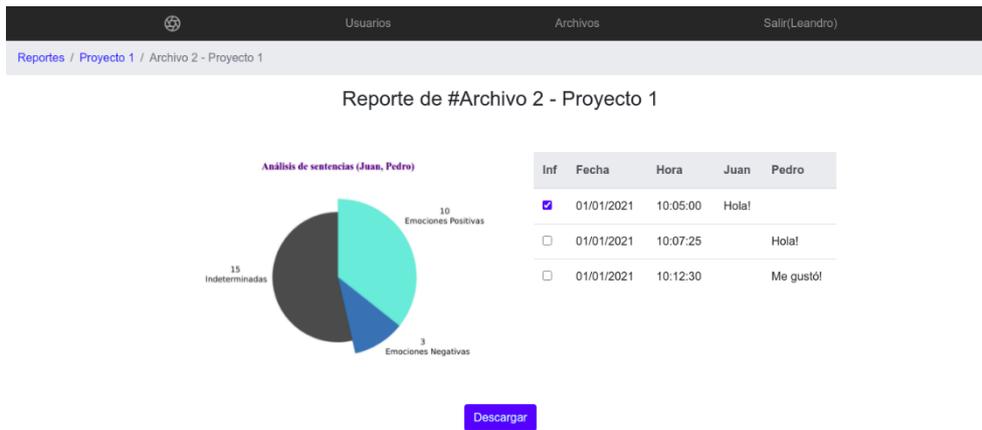
Fuente: elaboración propia

Ilustración 8. Procesar Archivo



Fuente: elaboración propia

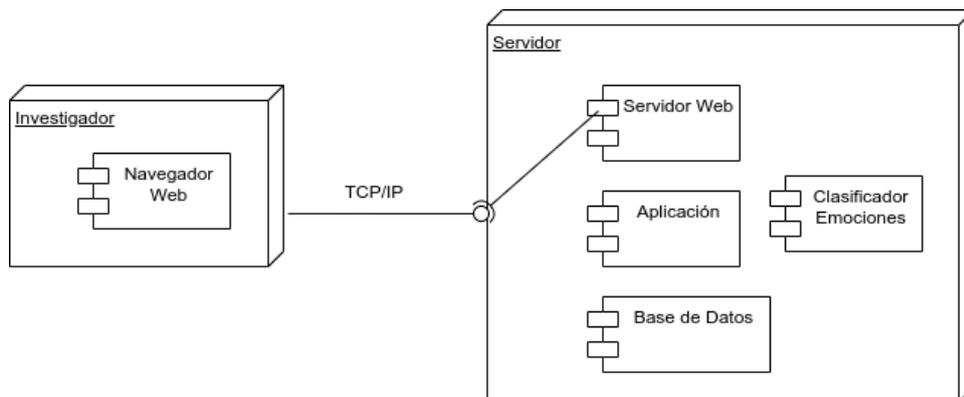
Ilustración 9. Ver Reporte



Fuente: elaboración propia

Diagrama de despliegue

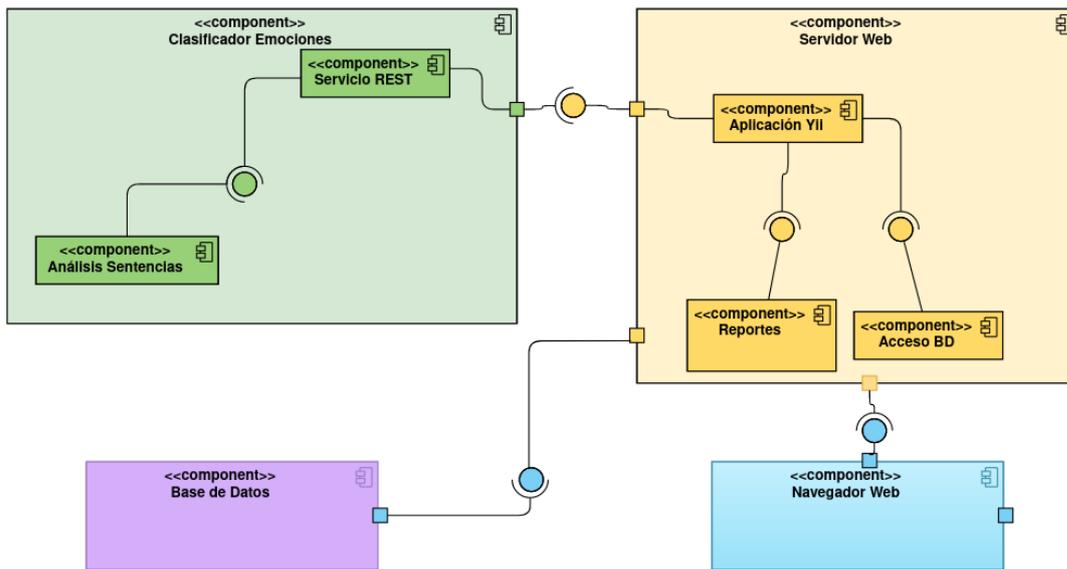
Ilustración 10. Diagrama de despliegue



Fuente: elaboración propia

## Diagrama de componentes

Ilustración 11. Diagrama de componentes



Fuente: elaboración propia

## Seguridad

Para desarrollar la cuestión de seguridad, debido a la diversidad de elementos que intervienen, se consideran aspectos relacionados con la seguridad de la aplicación, seguridad del servidor y de la infraestructura, seguridad de los datos y seguridad de las comunicaciones.

En aspecto de la seguridad de la aplicación se listan los componentes que pueden ser objeto de ataque y se explica cómo la aplicación gestiona el riesgo tomando como referencia las diez vulnerabilidades más importantes del Proyecto Abierto de Seguridad en Aplicaciones Web (OWASP) según la última publicación de 2017. (Owasp, 2020)

El software inseguro está debilitando las finanzas, salud, defensa, energía, y otras infraestructuras críticas. A medida que el software se convierte en algo crítico, complejo e interconectado, la dificultad de lograr seguridad en las aplicaciones aumenta exponencialmente. El ritmo vertiginoso de los procesos de desarrollo de software actuales incrementa aún más el riesgo de no descubrir vulnerabilidades de forma rápida y precisa. (Owasp, 2020)

La elección de las diez vulnerabilidades se realiza mediante un proceso que se nutre de diversas fuentes:

se basa principalmente en el envío de datos de más de 40 empresas que se especializan en seguridad de aplicaciones y una encuesta de la industria que fue completada por más de 500 personas. Esta información abarca vulnerabilidades recopiladas de cientos de organizaciones y más de 100.000 aplicaciones y APIs del mundo real. Las 10 principales categorías fueron seleccionadas y priorizadas de acuerdo con estos datos de prevalencia, en combinación con estimaciones consensuadas de explotabilidad, detectabilidad e impacto. (Owasp, 2020)

Tiene como objetivo mitigar el riesgo en la seguridad de las aplicaciones y para ello responde a la pregunta “¿Cuáles son los riesgos en seguridad de aplicaciones? Los atacantes pueden, potencialmente, utilizar diferentes rutas a través de su aplicación para perjudicar su negocio u organización. Cada uno de estos caminos representa un riesgo que puede o no ser suficientemente grave como para merecer atención.” (Owasp, 2020)

La siguiente tabla contiene las definiciones conceptuales de cada vulnerabilidad:

*Tabla 1. Owasp TOP 10 2017*

<b>Vulnerabilidad</b>	<b>Definición</b>
A1:2017 Inyección	Las fallas de inyección, como SQL, NoSQL, OS o LDAP ocurren cuando se envían datos no confiables a un intérprete, como parte de un comando o consulta. Los datos dañinos del atacante pueden engañar al intérprete para que ejecute comandos involuntarios o acceda a los datos sin la debida autorización.
A2:2017 Pérdida de Autenticación	Las funciones de la aplicación relacionadas a autenticación y gestión de sesiones son implementadas incorrectamente, permitiendo a los atacantes comprometer usuarios y contraseñas, token de sesiones, o explotar otras fallas de implementación para asumir la identidad de otros usuarios (temporal o permanentemente).
A3:2017 Exposición de datos sensibles	Muchas aplicaciones web y APIs no protegen adecuadamente datos sensibles, tales como información financiera, de salud o Información Personalmente Identificable (PII). Los atacantes pueden robar o modificar estos datos protegidos inadecuadamente para llevar a cabo fraudes con tarjetas de crédito, robos de identidad u otros delitos. Los datos sensibles requieren métodos de protección adicionales, como el cifrado en almacenamiento y tránsito.
A4:2017 Entidades Externas XML (XXE)	Muchos procesadores XML antiguos o mal configurados evalúan referencias a entidades externas en documentos XML. Las entidades externas pueden utilizarse para revelar archivos internos mediante la URI o archivos internos en servidores no actualizados, escanear puertos de la LAN, ejecutar código de forma remota y realizar ataques de denegación de servicio (DoS).

<p>A5:2017 Pérdida de Control de Acceso</p>	<p>Las restricciones sobre lo que los usuarios autenticados pueden hacer no se aplican correctamente. Los atacantes pueden explotar estos defectos para acceder, de forma no autorizada, a funcionalidades y/o datos, cuentas de otros usuarios, ver archivos sensibles, modificar datos, cambiar derechos de acceso y permisos, etc.</p>
<p>A6:2017 Configuración de Seguridad Incorrecta</p>	<p>La configuración de seguridad incorrecta es un problema muy común y se debe en parte a establecer la configuración de forma manual, ad hoc o por omisión (o directamente por la falta de configuración). Son ejemplos: S3 buckets abiertos, cabeceras HTTP mal configuradas, mensajes de error con contenido sensible, falta de parches y actualizaciones, frameworks, dependencias y componentes desactualizados, etc.</p>
<p>A7:2017 Secuencia de Comandos en Sitios Cruzados (XSS)</p>	<p>Los XSS ocurren cuando una aplicación toma datos no confiables y los envía al navegador web sin una validación y codificación apropiada; o actualiza una página web existente con datos suministrados por el usuario utilizando una API que ejecuta JavaScript en el navegador. Permiten ejecutar comandos en el navegador de la víctima y el atacante puede secuestrar una sesión, modificar (defacement) los sitios web, o redireccionar al usuario hacia un sitio malicioso.</p>
<p>A8:2017 Deserialización Insegura</p>	<p>Estos defectos ocurren cuando una aplicación recibe objetos serializados dañinos y estos objetos pueden ser manipulados o borrados por el atacante para realizar ataques de repetición, inyecciones o elevar sus privilegios de ejecución. En el peor de los casos, la deserialización insegura puede conducir a la ejecución remota de código en el servidor.</p>
<p>A9:2017 Componentes con vulnerabilidades conocidas</p>	<p>Los componentes como bibliotecas, frameworks y otros módulos se ejecutan con los mismos privilegios que la aplicación. Si se explota un componente vulnerable, el ataque puede provocar una pérdida de datos o tomar el control del servidor. Las aplicaciones y API que utilizan componentes con vulnerabilidades conocidas pueden debilitar las defensas de las aplicaciones y permitir diversos ataques e impactos.</p>
<p>A10:2017 Registro y Monitoreo Insuficientes</p>	<p>El registro y monitoreo insuficiente, junto a la falta de respuesta ante incidentes permiten a los atacantes mantener el ataque en el tiempo, pivotar a otros sistemas y manipular, extraer o destruir datos. Los estudios muestran que el tiempo de detección de una brecha de seguridad es mayor a 200 días, siendo típicamente detectado por terceros en lugar de por procesos internos</p>

Fuente: Adaptación de T10 (Owasp, 2020, p7)

Luego de haber presentado las definiciones de las vulnerabilidades se expone la forma en la que la aplicación está preparada para mitigarla o si acaso pudiera ser afectada por la misma:

- A1:2017-Inyección: El riesgo de Inyección (es decir, cuando el sistema ejecuta acciones indeseadas) se mitiga utilizando el Framework de desarrollo Yii2 que filtra, sin intervención del desarrollador, las entradas. De este modo, se refuerza el acceso a la base de datos mediante llamadas parametrizadas, con un preprocesamiento de la consulta, que impiden la ejecución de código inyectado en la base de datos.
- A2:2017-Pérdida de Autenticación: Se establece una política de contraseñas para el acceso a la base de datos y una desconexión automática después de un tiempo de inactividad.
- A3:2017-Exposición de datos sensibles: No se proveen Apis para consumo de aplicaciones externas y todo el tráfico entre los clientes y el servidor viaja protegido mediante una capa de sockets seguros (SSL) que protege todo el que se envía entre dos sistemas. (Websecurity, 2020)
- A4:2017-Entidades Externas XML (XXE): No se admite la importación de archivos XML, solo archivos de texto (en formato .txt) y se gestionan sin ser interpretados.
- A5:2017-Pérdida de Control de Acceso: El acceso a las acciones se encuentra restringido a los roles correspondientes y los objetos digitales, archivos y registros son asociados a cada usuario, controlándose en las acciones de descarga o visualización que se trate del usuario correcto.
- A6:2017-Configuración de Seguridad Incorrecta: El sistema cuenta con una configuración, que separa los ambientes de trabajo: Un entorno de desarrollo en la máquina local, y un entorno de producción, es decir público, que se ubica en el servidor. De esta manera no se mezclan configuraciones y se minimiza la posibilidad de que lleguen a desplegarse en producción servicios configurados en forma incorrecta. Asimismo, los componentes de la aplicación se actualizan periódicamente recibiendo actualizaciones de seguridad en el momento en que están disponibles.
- A7:2017-Secuencia de Comandos en Sitios Cruzados (XSS): El framework de desarrollo Yii2 satiniza, es decir transforma a una forma segura, todas las salidas antes de ser enviadas al usuario, por lo que cualquier intento de XSS queda invalidado.
- A8:2017-Deserialización Insegura: No se reciben objetos serializados, es decir transformados a un formato que luego pueda ser utilizado dentro de un sistema como un elemento nativo, bajo ningún mecanismo por lo que no es un riesgo.

- A9:2017- Componentes con vulnerabilidades conocidas: Las librerías y el framework Yii2 se mantienen actualizados con un sistema de gestión de paquetes que mantiene las versiones y dependencias y estos son instalados cuando se publica una nueva actualización. En el contexto de infraestructura, al utilizar un despliegue de contenedores que funcionan en forma orquestada permite que las imágenes que integran un servicio se actualicen como unidades intercambiables que se reemplazan en forma íntegra al ser actualizadas. (Fernandez Gines X., 2018)

El sistema operativo anfitrión tiene su propio mecanismo de actualización que está automatizado.

- A10-2017- Registro y Monitoreo Insuficientes: El sistema tiene un registro de logs de aplicación donde se registran los errores junto con un detalle del contexto en el que se produce el error.

La composición en Docker en la que funciona el sistema gestiona todas las salidas de log en una forma unificada y permite una salida hacia una aplicación de monitoreo como Kibana donde se realiza el monitoreo desde un punto centralizado. (Fernandez Gines X. 2018)

Con respecto a la seguridad de los usuarios de la aplicación se realizan una serie de consideraciones que se desarrollan a continuación.

La aplicación se encuentra configurada con un esquema de roles que delimita el alcance de las funcionalidades que pueden realizar los usuarios finales dejando solo disponibles las que corresponden al objetivo funcional. La gestión de usuarios y configuración se realiza con un rol de administrador que habilita acceso a un nivel más avanzado y es una tarea que se ejecuta solo por personal calificado bajo el rol de administrador.

Los archivos y objetos resultantes del procesamiento se encuentran asociados a los usuarios en forma independiente. Cada usuario ve solo los archivos que le corresponden asegurando de esa manera la privacidad de su trabajo.

Los usuarios utilizan una combinación de usuario y contraseña para ingresar y son dados de alta por el administrador del sistema. Los usuarios deben apegarse a una política de contraseñas que incluye la rotación mensual, no repetición de 5 claves anteriores y complejidad que incluye 8 caracteres mínimo incluyendo obligatoriamente mayúsculas, minúsculas, números y al menos 1 carácter especial compuesto por un símbolo visible. Las conexiones con los usuarios finales se

establecen mediante canales seguros protegidos con SSL y el almacenamiento se realiza mediante el uso de una función que genera un resumen (hash) (Lucena López, 2010, p. 195) de la contraseña mediante librerías de criptografía, en particular se utiliza *bcrypt* que ofrece una salida muy resistente a ataques de fuerza bruta (Yii2, Working with passwords, 2020) donde los atacantes prueban en forma sistemática una contraseña detrás de otra. De esta manera si se comprometiera el hash no se podría derivar la contraseña original dado que es un mecanismo de única vía.

Los usuarios poseen mecanismos para solicitar el restablecimiento de la contraseña en el caso de olvidar la clave y se utiliza un correo electrónico como medio de recuperación. El correo electrónico utilizado es el establecido en el perfil del usuario. Como mecanismo adicional de cambio de contraseña el administrador del sistema puede asignar una nueva contraseña al usuario a través de la interfaz de administración de usuarios.

En el aspecto de la seguridad del servidor y de la infraestructura se utiliza una composición con Docker que genera un entorno separado dentro del sistema operativo anfitrión que funciona en el servidor. En el sistema operativo anfitrión se debe mantener al día en cuanto a las actualizaciones de seguridad descargadas en forma periódica. Sobre el sistema operativo se realiza un proceso de endurecimiento (*hardening*) que “es la acción que se toma para aumentar la seguridad al reducir la vulnerabilidad de una estación de trabajo, servidor y red frente a la explotación”. (Price, R, 2019)

El proceso que se realiza sobre el servidor difiere sobre el que se realizaría sobre una estación de trabajo. En palabras de Price:

Asegurar un sistema operativo es muy importante en el proceso de proteger su computadora. El proceso de endurecimiento de un SO es ligeramente diferente para un SO cliente que para un SO de servidor. La mayoría de los pasos son los mismos en cualquier caso, pero centrémonos en el refuerzo del sistema operativo de un servidor de red. Esencialmente, fortalecer un sistema operativo implica cuatro pasos principales: deshabilitar los servicios innecesarios, deshabilitar los puertos TCP / UDP no utilizados, instalar solo el software necesario para las tareas que son compatibles y mantener el sistema operativo actualizado mediante la aplicación de parches y actualizaciones. (Price, 2019)

La composición Docker genera un entorno aislado dentro del servidor donde los procesos que se ejecutan como parte del ciclo de procesos quedan inaccesibles desde la red externa y la comunicación entre los servicios se mantiene en un ambiente libre de interferencias. El acceso desde

la red externa al servicio de composición se realiza solo mediante los puertos configurados que son los que corresponden a la aplicación web visible al usuario (frontend), en este caso el 443 que corresponde a SSL. Hacia el interior de la composición debido al aislamiento los servicios se pueden comunicar sin SSL mejorando la performance dado que no se debe realizar procesamiento de encriptado. (Fernández Gines, X. 2018)

En cuanto a la seguridad de los datos se gestiona desde el motor de la base de datos que tiene configurado un usuario de acceso exclusivo a la base de datos con los permisos mínimos para poder realizar las operaciones que utiliza el sistema. La gestión y configuración se realiza desde la misma interfaz del motor de base de datos mediante un rol de Administrador de Bases de Datos (DBA) que permite realizar las modificaciones y configuraciones correspondientes para mantener el buen funcionamiento de este. El acceso como DBA solo se puede realizar desde un equipo específico ubicado en la red interna con el propósito de impedir accesos indebidos.

La seguridad de los datos se encuentra establecida mediante una política de copia de seguridad que incluye una copia diaria en caliente de los datos generados durante el día y una copia en frío de la base completa en forma semanal. Los resguardos son almacenados en forma local con una política de rotación de 8 semanas. Como medida de resguardo adicional, el sistema permite que cada usuario realice una descarga de sus archivos y objetos lógicos en formato de archivo comprimido, además protegido con una contraseña, en el caso de que quiera tener una copia de seguridad personal que pueden almacenar en medios propios.

Desde el punto de vista de las comunicaciones, el servidor se encuentra conectado a un Switch (dispositivo concentrador de comunicaciones) administrable en el que se le habilita acceso al puerto por dirección física y se lo asocia a una Red Privada Virtual (VLAN) a la que tienen acceso solo los usuarios que harán uso del servicio. (Leischner & Tews, 2007)

Se utiliza SSL para las comunicaciones con los clientes web y Shell Seguro (SSH) para la gestión remota por lo que se mitiga el riesgo de interceptación de datos. (Stamp, 2011)

### **Análisis de Costos**

Se presenta el costo del proyecto para los aspectos de Recursos Humanos y Equipamiento e Infraestructura necesarios para cubrir todas las necesidades de desarrollo del sistema, puesta en marcha y asegurar la conexión y procesamiento desde los clientes hasta el servidor.

Desde el punto de vista del desarrollo se dimensiona una participación de tres personas durante un período de 6 meses (26 semanas) realizando actividades durante unas 8 horas diarias en días laborables (lunes a viernes). Las personas actuarán bajo 3 roles: Analista Funcional, Desarrollador y Tester e Implementador.

Según el cuadro de honorarios del (*Consejo Profesional de Ciencias Informáticas de la Provincia de Córdoba, 2020*) se consideran los siguientes importes máximos mensuales en pesos argentinos según el rol de este proyecto:

*Tabla 2. Análisis de Costos de Recursos Humanos*

Rol	Costo (en pesos argentinos)	
	Mensual (Máximo si hay 2 roles)	Total (6 meses)
Analista Funcional	96.235.-	577.410.-
Desarrollador (Programador Páginas web / Programador ambientes Unix/Linux)	145.903.-	875.418.-
Tester / Implementador	129.399.-	776.394.-
<b>Total</b>		<b>2.229.222.-</b>

Fuente: Elaboración Propia

Con respecto al hardware necesario para el proyecto, se presupuestó una estructura consistente en un servidor, cinco puestos de trabajo, un puesto de impresión y los dispositivos de conectividad para ofrecer alternativamente acceso a la red por cable o por conexión inalámbrica. En dicha estructura se alojó y puso en marcha el conjunto de servicios web implementado y el servicio de impresión.

La estructura de hardware consignada quedó compuesta por los siguientes dispositivos:

*Tabla 3. Costos de Hardware*

Dispositivo	Características	Cant.	Costo (en pesos argentinos)	
			Unitario	Total
Servidor	Dell Poweredge R240 Procesador Intel® Xeon® E-2124, 3.3 GHz, 8M Cache, 4C/4T turbo (71W) Memoria Ram 8GB Memory, 2666MT/s, DDR4 ECC UDIMM	1	179.099.-	179.099.-

	<p>2TB 7.2K RPM SATA 6Gps 3.5" Hot-Plug HD Broadcom® BCM5720   Gigabit LAN 10/100/1000 Mbps. ( RJ-45) Disposición física: Rackeable</p> <p><i>(oportutek.com Tecnología para Crecer. Servidor Rack Dell Poweredge R240 Xeon E-2124 8 GB 2TB., s. f.)</i></p>			
Puesto de trabajo	<p>Notebook Dell Inspiron 3493 Intel® Core™ i5-1035G1 10a. Generación (6MB Caché, hasta 3.6 GHz) RAM 16GB (2 x 8GB) DDR4 2666MHz HD SSD 256GB M2 Placa de Red 10/100 mbps Conector RJ45 Tarjeta inalámbrica Dell (DW1810/QCA9377 (1x1 AC,BT 4.1)</p> <p><i>(oportutek.com Tecnología para Crecer. Notebook Dell Inspiron 3493 14", s.f.)</i></p>	5	141.299.-	706.495.-
Impresora	<p>Impresora Láser Lexmark CS417dn Impresión color full duplex Conectividad USB + Red (RJ45) Resolución Negro: 1200 x 1200 dpi Color: 4800 (2400 x 600 ppp) Papel soportado A4, Legal, A5, Carta, Ejecutivo, Universal, Folio, Oficio, JIS-B5 Ciclo de trabajo de hasta 75.000 páginas por mes</p> <p><i>(oportutek.com Tecnología para Crecer. Impresora Láser Lexmark CS417dn, s.f.)</i></p>	1	46.899.-	46.899.-
Concentrador de comunicaciones local (Switch)	<p>Switch Cisco SG350-10-K9 Full Administrable 10 Puertos 10/100/1000 Transferencia 20 GBPS Disposición física: rackeable</p>	1	24.699.-	24.699.-

	(oportutek.com Tecnología para Crecer. Switch Full Administrable Cisco 10 Puertos 10/100/1000 - SG350-10-K9, s.f.)			
Punto de Acceso inalámbrico (Acces Point)	Access Point Cisco Aironet 1852I 802.11i, Acceso protegido Wi-Fi 2 (WPA2), WPA 802.1X Estándar de cifrado avanzado (AES) Velocidad de datos hasta 1,7 Gbps en la radio de 5 GHz Puertos 1 x 10/100 / 1000BASE-T autosensing (RJ-45), Power over Ethernet (PoE) 1 x 10/100 / 1000BASE-T con detección automática (RJ-45), AUX (utilizado para la agregación de enlaces) Puerto de consola de gestión (RJ-45) USB 2.0 (habilitado a través de software futuro) Dispisición física: Interior  (oportutek.com Tecnología para Crecer. Access Point Cisco Aironet 1852I, s.f.)	1	25.299.-	25.299.-
Puesto de red (cableado)	Consiste en el cableado estructurado cable UTP Cat-5, soporte externo para el cable y Jack RJ 45 para los 5 puestos de trabajo, el puesto de red y el Access Point (mano de obra y materiales)  (Electroinstalador.com, Costos mano de obra, 2020) (It Buy, Accesorios para cables de red, 2020)	7	4.000.-	28.000.-
<b>Total</b>				<b>1.010.491.-</b>

Fuente: Elaboración Propia

Sobre los costos presentados valen las siguientes aclaraciones:

- El servidor se instaló en un centro de datos organizacional donde se benefició de los servicios de refrigeración y tensión eléctrica estabilizada y monitoreada por lo que no se

incluyen costos de sistemas ininterrumpidos de energía eléctrica ni sistema de refrigeración.

- Se ha priorizado la compra en un único proveedor por resultar conveniente en el aspecto comercial y en la unificación del servicio postventa en un solo agente.

Con respecto al software no se han considerado costos debido a los siguientes motivos:

- El desarrollo se realizó utilizando software libre: Como sistema operativo de escritorio se utilizó Ubuntu Desktop sobre el que se realizó la codificación y puesta en marcha del entorno de desarrollo. Se utilizó Netbeans como entorno de desarrollo integrado para la escritura de código y Docker con imágenes conteniendo como sistema operativo de base Ubuntu Server con el servicio web Apache y los intérpretes de lenguajes de programación PHP y Python sin recurrir a librerías que requirieran un licenciamiento comercial. Como motor de base de datos se utilizó Mysql y la gestión de la base se realizó con PHPMyadmin. Como navegador web se utilizó Mozilla Firefox y para la ejecución dinámica de las páginas se utilizó javascript con el apoyo de JQuery y D3.js.
- Las estaciones de trabajo sobre las que realizarán su trabajo los investigadores son notebooks con el sistema operativo Windows Home Edition que viene preinstalado y licenciado. El uso del sistema se realiza mediante un navegador web por lo que los usuarios pueden optar por utilizar el navegador web Edge que viene preinstalado o descargar uno libre como Mozilla Firefox.
- El servidor utiliza software libre: Como sistema operativo utiliza Ubuntu Server y los servicios del sistema son ejecutados utilizando la tecnología Docker ejecutando la misma composición utilizada en el entorno de desarrollo.

Finalmente, el costo resultante de todos los ítems asciende a la suma de pesos argentinos tres millones doscientos treinta y nueve mil setecientos trece (\$ 3.239.713.-).

### **Análisis de Riesgos**

Como se trata de un sistema web que está compuesto por un conjunto de servicios de software y dispositivos físicos que incluyen el servidor, las terminales y la conectividad se analizó la probabilidad de ocurrencia fallos sobre cada uno y el impacto que tendría en caso de manifestarse el evento. También se consideró el hecho de que el sistema es dependiente de factores externos que

podrían cambiar en el futuro a causa de que el proveedor del servicio (WhatsApp) introduzca variaciones en el modo de exportar los archivos. Asimismo, por ser un sistema de información en el que intervienen factores humanos se podrían haber omitido requerimientos, como también existe la posibilidad de cometer errores por parte de los operadores al utilizarlo obteniendo resultados no deseados o mal interpretados.

Analizados los riesgos fueron representados y evaluados mediante un diagrama de Pareto:

El Diagrama de Pareto es una gráfica en donde se organizan diversas clasificaciones de datos por orden descendente, de izquierda a derecha por medio de barras sencillas después de haber reunido los datos para calificar las causas. De modo que se pueda asignar un orden de prioridades. (Sales, 2009).

Esta visualización permite concentrarse el esfuerzo en reducir los problemas más significativos.

Análisis de Pareto es una técnica que separa los “pocos vitales” de los “muchos triviales”. Una gráfica de Pareto es utilizada para separar gráficamente los aspectos significativos de un problema desde los triviales de manera que un equipo sepa dónde dirigir sus esfuerzos para mejorar. Reducir los problemas más significativos (las barras más largas en una Gráfica Pareto) servirá más para una mejora general que reducir los más pequeños. Con frecuencia, un aspecto tendrá el 80% de los problemas. En el resto de los casos, entre 2 y 3 aspectos serán responsables por el 80% de los problemas. (Sales, 2009).

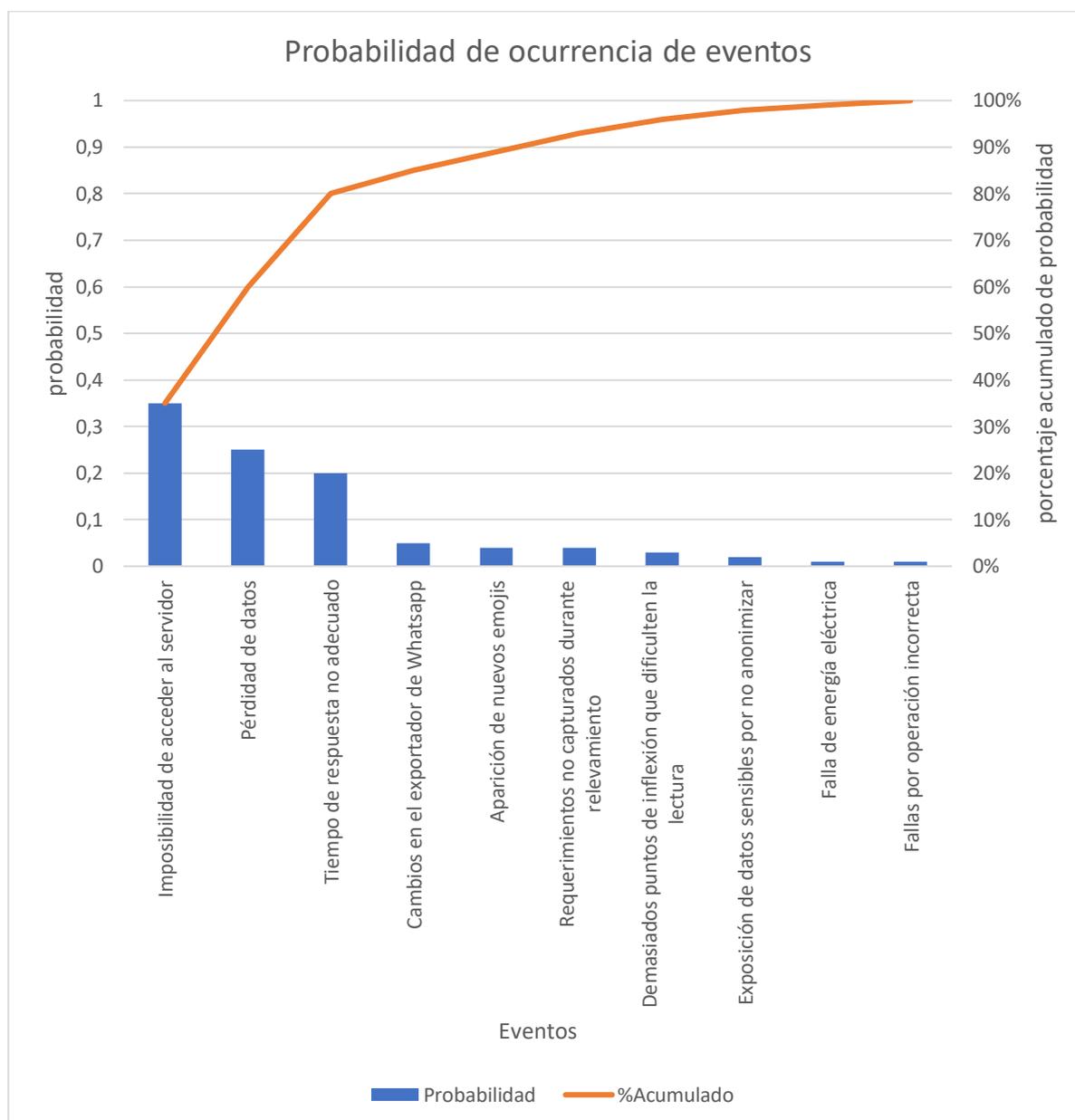
Se presenta a continuación una serie de riesgos valorados con la probabilidad de ocurrencia para el proyecto desarrollado. Se consideraron riesgos relacionados con el funcionamiento, contexto

*Tabla 4. Análisis de Riesgos*

<b>Riesgo</b>	<b>Probabilidad %</b>		<b>Acumulado %Acumulado</b>	
Imposibilidad de acceder al servidor	0,35	35%	0,35	35%
Pérdida de datos	0,25	25%	0,6	60%
Tiempo de respuesta no adecuado	0,2	20%	0,8	80%
Cambios en el exportador de WhatsApp	0,05	5%	0,85	85%
Aparición de nuevos emojis	0,04	4%	0,89	89%
Requerimientos no capturados durante relevamiento	0,04	4%	0,93	93%
Demasiados puntos de inflexión que dificulten la lectura	0,03	3%	0,96	96%
Exposición de datos sensibles por no anonimizar	0,02	2%	0,98	98%
Falla de energía eléctrica	0,01	1%	0,99	99%
Fallas por operación incorrecta	0,01	1%	1	100%
Total	1	100%		

Fuente: Elaboración Propia

Ilustración 12. Probabilidad de ocurrencia de eventos



Fuente: Elaboración Propia

Se desarrollan estrategias de mitigación para los tres eventos que acumulan el 80% de probabilidades de ocurrencia.

- Imposibilidad de acceder al servidor: Las estrategias de mitigación implican utilizar otro servidor, instalar la aplicación y restaurar el resguardo. Si es inaceptable la pérdida de tiempo de esperar una instalación se puede mudar el servicio que está en una

composición a una nube temporal. También se puede mantener un servicio en paralelo que recibe las modificaciones que ocurren en el servicio principal y está siempre listo para ser utilizado. La estrategia más razonable es la de tener más de un servidor y desplegar una nube en la que el acceso a los recursos se distribuye en forma dinámica y ante la falla de un servidor físico reacciona redireccionando las peticiones a los servidores que siguen funcionando.

- Pérdida de datos o archivos: La estrategia de resolución es restaurar la copia de seguridad y se pierde la actividad del día. En el caso de que corresponda a un usuario en particular se puede recurrir a una restauración de una copia de seguridad personal del usuario si es que éste la había realizado.
- Tiempo de respuesta no adecuado para la finalización de los procesos: La estrategia de mitigación implica incrementar los recursos físicos del servidor, agregar memoria o procesadores, en caso de ser posible aumentar el paralelismo.

## **Conclusiones**

El presente trabajo tuvo como objetivo el catalogar los mensajes de texto intercambiados por los participantes de una conversación de mensajería instantánea para luego ser analizados con técnicas de procesamiento de lenguaje natural y a partir de lo cual, poder detectar las participaciones que provocan cambios en el estado emocional de las intervenciones siguientes.

Sobre la base de las ideas expuestas, se considera que se logró satisfactoriamente el cumplimiento del objetivo propuesto, gracias a la implementación de un sistema que posibilitó cargar las conversaciones exportadas desde el sistema de mensajería WhatsApp, aplicando las funcionalidades desarrolladas. A partir de las cuales, se pudo obtener una importante estadística para poder determinar: el contexto de esas comunicaciones; conocer el número de participantes; saber qué cantidad de mensajes envió cada interviniente; detectar qué valor emocional aportaban las palabras y los emojis en cada mensaje; identificar cuál fue la variación de emociones a lo largo de una línea de tiempo; y finalmente, poder visualizar cuáles fueron los mensajes más importantes que produjeron un cambio en la situación emocional de los de los participantes.

En virtud de los resultados, las funcionalidades del sistema muestran un gran valor para la comunidad científica específicamente para los investigadores de las comunicaciones digitales. Ya que permite una considerable reducción del tiempo, además, aportan nuevas herramientas con las

que se pueden realizar análisis, que antes eran inviables por la carga de trabajo manual que implicaban.

En lo personal el trabajo realizado me aportó una gran satisfacción por haber podido convertir en una realidad un proyecto que surgió de conversaciones con investigadores que expresaban su frustración por no poder realizar ciertas tareas y que gracias al diálogo y buena predisposición permitió que se fueran construyendo claramente los requisitos que debía cumplir una aplicación que resolviera sus problemas.

En lo profesional me genera un nuevo valor, ya que es una herramienta que me permite aportar una novedad en el campo de las humanidades digitales que está ávida de herramientas útiles. También me ha permitido profundizar mis conocimientos en el campo de las comunicaciones digitales, el procesamiento del lenguaje natural y deja abierta una oportunidad de mejora continua mediante el agregado de nuevas funcionalidades aplicadas a la educación, la psicología, la medicina y toda ciencia en la que sea valioso analizar el intercambio de conceptos a través de comunicaciones digitales.

#### *Demo*

El código fuente del proyecto junto con sus instrucciones de instalación se encuentran disponibles para su descarga en la siguiente dirección pública

<https://bitbucket.org/lmbsoft/tfg-leandro-boisselier-ue21>

## Referencias

- A. Mohta, A. Jain, A. Saluja and S. Dahiya, "Pre-Processing and Emoji Classification of WhatsApp Chats for Sentiment Analysis," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020, pp. 514-519, doi: 10.1109/I-SMAC49090.2020.9243443.
- Alcántara-Pla, M. (2014). Las unidades discursivas en los mensajes instantáneos de wasap. *Estudios de Lingüística Del Español*, 35, 223–242.
- Anthony, L. (2004). AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit. En L.A. (Ed.), *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning* (pp. 7-13). Recuperado de [https://www.researchgate.net/profile/Laurence\\_Anthony/publication/267631346\\_Proceeding\\_of\\_IWLeL\\_2004\\_An\\_Interactive\\_Workshop\\_on\\_Language\\_Elearning\\_2004/links/5458cd870cf26d5090acf212/Proceedings-ofIWLeL2004AnInteractiveWorkshoponLanguageElearning-2004.pdf#page=7](https://www.researchgate.net/profile/Laurence_Anthony/publication/267631346_Proceeding_of_IWLeL_2004_An_Interactive_Workshop_on_Language_Elearning_2004/links/5458cd870cf26d5090acf212/Proceedings-ofIWLeL2004AnInteractiveWorkshoponLanguageElearning-2004.pdf#page=7).
- Apache HTTP (s.f.) recuperado 15 de octubre de 2020 de <https://httpd.apache.org/>
- Balbachan, F. (2014). *Técnicas de Clustering para Inducción de Categorías Sintácticas en Español. Tesis de doctorado*. Buenos Aires: Universidad de Buenos Aires.
- Baviera, T. (2017). Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado. *Revista Dígitos*, 1(3), 33-50.
- Bolaños Cuéllar, S. (2015). Corpus Linguistics: Aproxes for contemporary linguistics research. *Forma y Función*, 28(1), 31-54.
- Bootstrap (s.f.) recuperado 15 de octubre de 2020 de <https://getbootstrap.com/docs/4.3/getting-started/introduction/>
- Briz, A. (2000a). *¿Cómo se comenta un texto coloquial?* Barcelona: Ariel.
- Cantamutto, L. (2018). *Estrategias pragmáticas de la comunicación por SMS en español bonaerense (Tesis de doctorado)*, Universidad Nacional del Sur. Bahía Blanca.
- Cantamutto, L. M. y Vela Delfa, C. (2015). Problemas de recogida y fijación de muestras del discurso digital, *CHIMERA: Romance Corpora and Linguistic Studies* (2), 131-155
- Cantamutto, L., y Vela Delfa, C. (2016a). "El discurso digital como objeto de estudio: de la descripción de interfaces a la definición de propiedades". *Aposta. Revista de Ciencias Sociales*, (69), 296–323.

Cantamutto, L. y Vela Delfa, C. CoDiCE. Comunicación Digital. Corpus del español. Universidad Nacional del Comahue/ Universidad Nacional del Sur/ Universidad de Valladolid. Disponible en: [codice.aplicacionesonline.com.ar](http://codice.aplicacionesonline.com.ar)

*Consejo Profesional de Ciencias Informáticas de la Provincia de Córdoba.* (s. f.). CPCIPC. Recuperado 15 de octubre de 2020, de <https://www.cpcipc.org.ar/content/honorarios>.

Bostock, M. (s. f.). D3.js - Data-Driven Documents. <https://d3js.org/>. Recuperado 22 de noviembre de 2020, de <https://d3js.org/>

De-Matteis, L (2014). Ejes para un debate sobre el uso ético de datos interaccionales escritos y orales obtenidos en línea. *I Jornadas Nacionales de Humanidades Digitales*. Conferencia llevada a cabo en las Jornadas de Asociación Argentina de Humanidades Digitales, Buenos Aires.

Danesi, M. (2016). *The semiotics of emoji: The rise of visual language in the age of the internet*. London: Bloomsbury Publishing.

De-Matteis, L (2016). Ejes para un debate sobre el uso ético de datos interaccionales escritos y orales obtenidos en línea. En: Riande, G., L. Cantamutto y G. Striker, *I Jornadas Argentinas de Humanidades Digitales*, Bs. As.: FFyL-UBA, 235-247.

Diagrams (s.f.) recuperado 15 de octubre de 2020 de <https://www.diagrams.net/>

Docker (s.f), recuperado 15 de octubre de 2020 de <https://www.docker.com/resources/what-container>

Electroinstalador.com, Costos de mano de obra, 2020, recuperado 15 de octubre de 2020 de <https://www.electroinstalador.com/contenidos/costos-mano-obra.html>

Gimp (s.f.) recuperado 15 de octubre de 2020 de <https://www.gimp.org/>

Ginés, F (2018). Seguridad en Docker. (Tesis de Máster). Universidad Oberta de Catalunya. Barcelona.  
<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81276/6/xfernangTFM0618memoria.pdf>

Git (S.f.), recuperado 15 de octubre de 2020 de <https://git-scm.com/>

GNU (s.f.) recuperado 15 de octubre de 2020 de <https://www.gnu.org/philosophy/free-sw.es.html>

Gupta, P., & Nene, M. J. (2017, March). Analysis of Text Messages in Social Media to Investigate CyberPsycho Attack. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 581-587). Springer, Cham.

It Buy, Accesorios para cables de red (2020) recuperado 15 de octubre de 2020 de <https://tienda.itbuy.com.ar/categorias/66-adaptadores-para-cables-de-red>

Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). Natural language processing: python and NLTK. Packt Publishing Ltd.

Kaplan, N. (2004). Nuevos desarrollos en el estudio de la evaluación en el lenguaje: la teoría de la valoración. Boletín de Lingüística (22), 52-78.

Leischner, G., & Tews, C. (2007, April). Security through VLAN segmentation: Isolating and securing critical assets without loss of usability. In proceedings of the 9th Annual Western Power Delivery and Automation Conference, Spokane, WA.

LibreOffice (s.f), recuperado 15 de octubre de 2020 de <https://es.libreoffice.org/>

Lucena López, M. (2010). Criptografía y seguridad en computadores. 4º Edición. Versión 0.7.0. Jaén, España: Universidad de Jaen.

Mayer-Schönberger, V., y Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Boston. Houghton Mifflin Harcourt.

Martin, J. R. y White, P. R. (2003). The language of evaluation. Basingstoke: Palgrave Macmillan.

Moreno Ortiz, A. (2019). Mi opinión cuenta: La expresión del sentimiento en la red. En Moreno Ortiz, A. y Robles Ávila, S. (Ed.) Comunicación mediada por ordenador: la lengua, el discurso y la imagen (pp. 38-74). Madrid: Cátedra.

Netbeans (s.f) recuperado 15 de octubre de 2020 de [https://netbeans.org/index\\_es.html](https://netbeans.org/index_es.html)

Nixon, R (2018) Learning PHP, MySQL & JavaScript, 5th Edition

Openmoji (s.f.) recuperado 15 de octubre de 2020 de <https://www.openmoji.org/about/>

*oportutek.com Tecnología para Crecer.* (s. f.). Servidor Rack Dell Poweredge R240 Xeon E-2124 8GB 2TB. Recuperado 15 de octubre de 2020, de <https://oportutek.com/products/servidor-rack-dell-poweredge-r240-xeon-e-2124-8gb-2tb>.

*oportutek.com Tecnología para Crecer.* (s. f.). Notebook Dell Inspiron 3493 14". Recuperado 15 de octubre de 2020, de <https://oportutek.com/products/notebook-dell-inspiron-3493-14-intel%C2%AEcore-i5-16gb-ssd-256gb-windows-10-home>.

*oportutek.com Tecnología para Crecer.* (s. f.). Impresora Láser Lexmark CS417dn. Recuperado 15 de octubre de 2020, de <https://oportutek.com/products/impresora-laser-lexmark-cs417dn-color-duplex>.

*oportutek.com Tecnología para Crecer.* (s. f.). Switch Full Administrable Cisco 10 Puertos 10/100/1000 - SG350-10-K9. Recuperado 15 de octubre de 2020, de <https://oportutek.com/products/switch-full-administrable-cisco-10-puertos-10-100-1000-sg350-10-k9>.

*oportutek.com Tecnología para Crecer.* (s. f.). Access Point Cisco Aironet 1852I. Recuperado 15 de octubre de 2020, de <https://oportutek.com/products/access-point-cisco-aironet-1850-series>.

OWASP, 2020. Recuperado 19 de noviembre de 2020, de <https://owasp.org/www-pdf-archive/OWASP-Top-10-2017-es.pdf>.

Pencil (s.f.) recuperado 15 de octubre de 2020 de <https://pencil.evolus.vn/>

PHP (s.f), recuperado 15 de octubre de 2020 de <https://www.php.net/manual/es/intro-what-is.php>

Phpmyadmin (s.f.) recuperado 15 de octubre de 2020 de <https://www.phpmyadmin.net/>

Price, R. (2019). *CompTIA Server+ Certification Guide: A comprehensive, end-to-end study guide for the SK0-004 certification, along with mock exams.* Packt Publishing Ltd.

Python (S.F), recuperado 15 de octubre de 2020 de <https://wiki.python.org/moin/SpanishLanguage>

RedHat (S.f.), ¿qué es el open source?, recuperado 15 de octubre de 2020 de <https://www.redhat.com/es/topics/open-source/what-is-open-source>

Reddy, C. B. S., Kowshik, S., Kumar, R., & Kumar, O. N. (2020). *Analysing and Predicting the Emotion of WhatsApp Chats Using Sentiment Analysis.*

Sales, M. (2009). *Diagrama de pareto.* Recuperado el, 15 de octubre de 2020, de <https://www.gestiopolis.com/diagrama-de-pareto>.

Sommerville, I. (2011). *Ingeniería del Software.* Madrid: Pearson Educación.

Smith, R. (2017). *Docker Orchestration.* Packt Publishing Ltd.

Stamp, M. (2011). *Information security: principles and practice.* John Wiley & Sons.

Tanenbaum A, Wetherall D. (2012), *Redes de Computadoras, 5ta edición*

Thurlow, C. y Mroczek, K. (2011). *Digital Discourse. Language in the New Media.* New York: Oxford University Press.

Umbrello (s.f.) recuperado 15 de octubre de 2020 de <https://umbrello.kde.org/>

Unicode.org (s.f) recuperado 15 de octubre de 2020 de <https://home.unicode.org/emoji/about-emoji/>

Vázquez-Cano, E., Mengual-Andrés, S., & Roig-Vila, R. (2015). ANÁLISIS LEXICOMÉTRICO DE LA ESPECIFICIDAD DE LA ESCRITURA DIGITAL DEL ADOLESCENTE EN WHATSAPP. *RLA. Revista de lingüística teórica y aplicada*, 53(1), 83-105. <https://doi.org/10.4067/s0718-48832015000100005>.

Vázquez-Cano, E., Mengual-Andrés, S., y Roig-Vila, R. (2015). Análisis lexicométrico de la especificidad de la escritura digital del adolescente en WhatsApp, *RLA. Revista de lingüística teórica y aplicada*, 53(1), 83-105.

Vela, Cantamutto y Nuñez (2021), Análisis del hashtag #MeQuedoEnCasa: la conversación digital ante la emergencia de la COVID-19, *Revista Íkala*.

Verschueren, J. (2002). Para entender la pragmática. Gredos.

W3 (s.f) recuperado 15 de octubre de 2020 de <https://www.w3.org/International/articles/definitions-characters/index.es>

Websecurity, 2020, Recuperado 19 de noviembre de 2020, de <https://www.websecurity.digicert.com/es/es/security-topics/what-is-ssl-tls-https>.

Winston, P. (1992) *Artificial Intelligence* (3a ed.). Reading. Addison-Wesley.

Yii2, Working with passwords, 2020, recuperado de <https://www.yiiframework.com/doc/guide/2.0/en/security-passwords>

Yii2 Framework (2020). Introduction: About Yii. Recuperado de: <https://www.yiiframework.com/doc/guide/2.0/en/intro-yii>

## Anexos

### Anexo 1 Formulario recepción de archivos

El formulario se encuentra en la dirección: <https://forms.gle/GQHLLF7wmAwqGd7eT9>

## Recepción de archivos TFG Boisselier Leandro

El presente formulario permite la colaboración desinteresada en el proyecto de TFG "Sistema clasificador de emociones para las comunicaciones digitales privadas por medio de técnicas de inteligencia artificial"

La recolección de muestras reales de comunicación digital es una de las tareas más arduas de los estudios sobre la comunicación digital. Es por ello que con una pequeña colaboración de cada uno de ustedes (exportando conversaciones que encuentres en tu Whastapp) será posible acrecentar significativamente nuestro corpus actual. Los datos serán utilizados exclusivamente para los fines de la investigación y los nombres de las personas participantes serán anonimizados.

Para generar el archivo debe ingresar en la conversación a exportar, elegir (...) en la esquina superior derecha y luego ingresar en Más. Elegir Exportar Chat y seleccionar la opción SIN ARCHIVOS  
Por último podrá elegir la opción de Guardar en Drive o enviar por email. Luego descargarlo y estará listo para subir en éste formulario.

Gracias por tu participación!

El nombre y la foto asociados a tu cuenta de Google se registrarán cuando subas archivos y envíes este formulario. ¿No es tuya la dirección [imbsoft@gmail.com](mailto:imbsoft@gmail.com)? [Cambiar de cuenta](#)

**\*Obligatorio**

---

**Fecha de exportación \***

DD MM AAAA

\_\_ / \_\_ / \_\_\_\_

---

**Cantidad de personas que interactúan \***

Tu respuesta

---

**Archivo \***

[📁 Añadir archivo](#)

---

**Observaciones**

Tu respuesta

---

**Enviar**

Nunca envíes contraseñas a través de Formularios de Google.

Este contenido no ha sido creado ni aprobado por Google. [Notificar uso inadecuado](#) - [Términos del Servicio](#) - [Política de Privacidad](#)

## Anexo 2 - Fragmento de un archivo anonimizado

Tomado a partir de un archivo exportado desde el sistema de mensajería WhatsApp

30/04/2019 09:48:28 a.m.: Gisela: <imagen omitida> Mi primera granola. Quedó bueniisima. Gracias! 🍌

30/04/2019 09:48:36 a.m.: Romina: uff

30/04/2019 09:48:37 a.m.: Romina: alta pinta

30/04/2019 09:48:41 a.m.: Romina: que bueno Gisegi

30/04/2019 09:48:47 a.m.: Romina: es lo más rico

30/04/2019 09:48:54 a.m.: Romina: voy a probar en sartén

30/04/2019 09:49:00 a.m.: Romina: máss fácil para que no se quemee

30/04/2019 09:49:31 a.m.: Gisela: Si... la verdad que no que acordaba como lo habías hecho. Jaja solo la lista de ingredientes.

30/04/2019 09:49:53 a.m.: Gisela: Lo doré con aceite de coco y le agregué miel y jengibre en polvo

30/04/2019 09:50:07 a.m.: Romina: rico rico rico

30/04/2019 09:50:10 a.m.: Romina: aguante la granola casera

30/04/2019 09:50:15 a.m.: Romina: tenemos uqe incursionar en barritas

30/04/2019 09:50:31 a.m.: Gisela: Siii próximamente

30/04/2019 09:50:36 a.m.: Gisela: Jajaja

05/05/2019 04:37:01 p.m.: Gisela: <documento omitido>

05/05/2019 04:37:07 p.m.: Gisela: Hola rooo

05/05/2019 04:37:13 p.m.: Gisela: Después te cuento más sobre eso

05/05/2019 04:37:43 p.m.: Romina: 😊 😊 😊

10/05/2019 09:45:21 a.m.: Romina: gisegi

10/05/2019 09:45:23 a.m.: Romina: te invité

10/05/2019 09:45:25 a.m.: Romina: pero no te invité

10/05/2019 09:45:26 a.m.: Romina: je

10/05/2019 09:45:29 a.m.: Romina: mañana al muerzo en casa

10/05/2019 09:45:30 a.m.: Romina: con mer

10/05/2019 09:45:31 a.m.: Romina: lore

10/05/2019 09:45:32 a.m.: Romina: jul

10/05/2019 09:45:34 a.m.: Romina: pipo

10/05/2019 09:45:35 a.m.: Romina: catalu{a

10/05/2019 09:45:39 a.m.: Romina: cataluña

10/05/2019 09:45:41 a.m.: Romina: mati

10/05/2019 09:45:43 a.m.: Romina: ivancito

10/05/2019 09:45:45 a.m.: Romina: et moi

10/05/2019 09:46:25 a.m.: Gisela: Qué lindo!!

10/05/2019 09:46:52 a.m.: Gisela: Buenísimo

10/05/2019 09:47:07 a.m.: Gisela: Qué aporito?

10/05/2019 09:47:31 a.m.: Romina: ehh tengo que distribuir entre los presentes

10/05/2019 09:47:35 a.m.: Romina: yo voy a hacer tacos

10/05/2019 09:47:39 a.m.: Romina: alguienes bebidas

10/05/2019 09:47:43 a.m.: Romina: alguienes postre

10/05/2019 09:48:02 a.m.: Gisela: Bien. Qué rico.

10/05/2019 09:48:34 a.m.: Gisela: Yo para los postres no soy la más indicada, probablemente...

10/05/2019 09:48:40 a.m.: Gisela: Si puedo, me apunto en bebida

10/05/2019 09:48:44 a.m.: Romina: dale

10/05/2019 09:48:46 a.m.: Romina: listo

10/05/2019 09:48:48 a.m.: Romina: bebida entonces

10/05/2019 09:48:58 a.m.: Romina: y tirame esto en inglés

10/05/2019 09:49:04 a.m.: Romina: jjaj

10/05/2019 09:49:05 a.m.: Romina: no

10/05/2019 09:49:07 a.m.: Romina: eso no es para vos

10/05/2019 09:49:14 a.m.: Gisela: Estoy en cama, también... jaa! Pero para mañana repunto eh

10/05/2019 09:49:17 a.m.: Romina: uyyy

10/05/2019 09:49:20 a.m.: Romina: fuerza  
10/05/2019 09:49:21 a.m.: Romina: te  
10/05/2019 09:49:24 a.m.: Romina: y jenjibre  
10/05/2019 09:49:31 a.m.: Gisela: Sii ❤️  
10/05/2019 09:49:54 a.m.: Gisela: Y tú granola power que es la base de me mi subsistencia  
10/05/2019 09:51:33 a.m.: Romina: ayy siiii  
10/05/2019 09:51:38 a.m.: Romina: yo me la hago y me la como a cada rato  
11/05/2019 12:17:52 p.m.: Gisela: Alohaaa  
11/05/2019 12:18:07 p.m.: Gisela: A qué hora les caigo?  
11/05/2019 12:51:11 p.m.: Romina: A las 13  
11/05/2019 01:04:00 p.m.: Romina: Entra por el quincho  
12/05/2019 12:06:23 a.m.: Gisela: Hola Romchi  
12/05/2019 12:06:33 a.m.: Gisela: Por casualidad aún despiertos?  
12/05/2019 06:51:16 a.m.: Romina: Giiiii estaba durmiendo!!!! Todo bien??  
12/05/2019 09:07:54 a.m.: Gisela: Jaja me matan las diferencias de nuestros horarios  
12/05/2019 09:08:10 a.m.: Gisela: Sii tenía a los vecinos de alta fiesta al lado y estaba buscando asilo  
12/05/2019 09:08:30 a.m.: Gisela: Pero ya está. He sobrevivido  
12/05/2019 09:15:15 a.m.: Romina: Ughh  
12/05/2019 09:15:46 a.m.: Romina: <mensaje de voz omitido>  
12/05/2019 09:28:47 a.m.: Gisela: No pro! Logre descansar igual. Gracias. Tengo en cuenta lo de Mati. Buen tip 🙏  
12/05/2019 09:28:53 a.m.: Gisela: Jaja besote a los tres  
12/05/2019 10:28:27 a.m.: Romina: Sii

(Cantamutto, L. y Vela Delfa, C. CoDiCE)