



Universidad Autónoma de Querétaro  
Facultad de Informática  
Especialidad en Maestría en Sistemas Computacionales

**Analizador semántico de emociones y sentimientos en textos contenidos en  
letras de canciones en español en tiempo real.**

**TESIS**

Que como parte de los requisitos para obtener el Grado de  
Maestría en Sistemas Computacionales

**Presenta:**

L.I. Juan Carlos Mancera Barrera

**Dirigido por:**

Mtro. Alberto Lamadrid Álvarez

**SINODALES**

Mtro. Alberto Lamadrid Álvarez  
Presidente

\_\_\_\_\_

Firma

Dr. Alberto Lara Guevara  
Secretario

\_\_\_\_\_

Firma

Mtro. Ricardo Chaparro Sánchez  
Vocal

\_\_\_\_\_

Firma

Dra. María Teresa García Ramírez  
Suplente

\_\_\_\_\_

Firma

Mtro. Carlos Alberto Olmos Trejo  
Suplente

\_\_\_\_\_

Firma

\_\_\_\_\_

Mtro. Juan Salvador Hernández Valerio  
Director de la Facultad

\_\_\_\_\_

Dra. Ma. Guadalupe Flavia Loarca Piña  
Directora de Investigación y Posgrado

Centro Universitario, Querétaro, Qro. México.  
24 de septiembre del 2020

“La creatividad se trata de conectar cosas. Cuando le preguntas a personas creativas cómo hicieron algo, se sienten un poco culpables porque realmente no lo hicieron, sólo lo vieron. Pareció obvio para ellos después de un tiempo.”

-Steve Jobs

“No puedes conectar los puntos viendo hacia adelante, sólo puedes conectarlos viendo hacia atrás. Así que tienes que confiar que de alguna manera se conectarán en un futuro. Debes confiar en algo, tu instinto, destino, vida, karma, lo que sea.”

-Steve Jobs

Dirección General de Bibliotecas UAQ

## **DEDICATORIA**

A mi hermosa familia, amigos, maestros y todas las personas que me han brindado parte de su conocimiento a lo largo de mi vida.

## AGRADECIMIENTOS

Al amor de vida mi esposa Vicky Ramírez gracias por apoyarme siempre en todos mis proyectos y locuras, sin ti esto no sería posible. A mis hijos Demián Rusell e Iker Donovan son mi motor e inspiración para lograr lo que sea, los amo.

A mis padres, hermanas y sobrinos gracias por estar conmigo en todas mis decisiones y por comprender que a veces me alejan temporalmente de ustedes, pero siempre están conmigo a donde vaya.

A mi amigo, socio y asesor el Ing. David Ramírez Reyes por apoyarme en todos los aspectos durante esta etapa de posgrado.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT), a la Universidad Autónoma de Querétaro y a la empresa EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV por todo el apoyo que me ofrecieron para el desarrollo de este proyecto de investigación.

A mis profesores que forman parte de la maestría en sistemas computacionales de la facultad de informática gracias por su apoyo que me brindaron durante esta etapa de posgrado, principalmente a mi asesor de tesis el Mtro. Alberto Lamadrid Álvarez por su guía, consejo y enseñanza durante toda la etapa de investigación.

Al Dr. Alberto Lara Guevara, al Mtro. Ricardo Chaparro Sánchez y al Mtro. Carlos Alberto Olmos Trejo por sus aportaciones, orientación y observaciones las cuales siempre fueron oportunas y atinadas, ayudándome a dar una apropiada estructura a este trabajo.

Agradezco a la Dra. María Teresa García Ramírez por todas las enseñanzas en mi formación profesional y por invitarme a cursar el posgrado en sistemas computacionales.

A mis compañeros de carrera, fue un placer y privilegio el haber cursado el posgrado y coincidir en esta vida con ustedes.

## RESUMEN

En este trabajo de investigación se presenta un analizador semántico de emociones y sentimientos en texto contenido en letras de canciones en español en tiempo real. Mediante un algoritmo de procesamiento de lenguaje natural se etiqueta gramaticalmente el texto de las letras de las canciones para darle sentido y estructura antes de analizarlo. El analizador propuesto utiliza una base de datos léxica que esta compuesta por unigramas clasificados en ocho emociones básicas (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y dos sentimientos (positivo y negativo). El texto de la letra de la canción es comparado y etiquetado gramaticalmente con los unigramas de la base de datos para llevar a cabo el análisis de emociones y sentimientos. Se determina el porcentaje de cada emoción y la polaridad del sentimiento para clasificar las letras de las canciones de acuerdo con su carga emocional y sentimental. Se detecta la situación emocional del usuario en un momento determinado. El analizador también se aplicó en una publicación para identificar situaciones emocionales y de sentimientos en mensajes de textos de correos electrónicos que infieren en la satisfacción de los clientes de una empresa de servicios de facturación electrónica. De esta forma, los usuarios del servicio son atendidos con acciones apropiadas de acuerdo con su estado emocional.

**(Palabras clave:** Análisis de emociones y sentimientos, procesamiento de lenguaje natural, base de datos léxica.)

## ABSTRACT

This research presents a real time Spanish song word lyrics semantic analyzer of emotions and feelings. Through a natural language processing algorithm, the unstructured words of the lyrics is grammatically labeled to make sense and structure before being analyzed. The proposed analyzer uses a lexical database which is composed by unigrams, classified in eight basic emotions (anger, fear, anticipation, confidence, surprise, sadness, joy and disgust) and two feelings (positive and negative). The text in the lyrics is compared and grammatically labeled through the unigrams of the database to perform the analysis of emotions and feelings, to determine emotion percentages and feeling polarity. Lyrics are classified according to the emotional and sentimental load. The emotional situation of the user is identified. The analyzer was also applied in a publication to identify emotional and sentimental situations in e-mail texts that infer in the satisfaction of the clients of an electronic billing services company. In this way, the users of the service are taken care of with appropriate actions according to their emotional state.

**(Key words:** Analysis of emotions and feelings, natural language processing, lexical database.)

## TABLA DE CONTENIDOS

1.	INTRODUCCIÓN.....	12
1.1	Definición del proyecto de investigación.....	12
1.2	EmoLex.....	12
1.3	Natural Language Cloud de Google .....	13
1.4	Justificación.....	14
1.5	Objetivos de la investigación .....	19
2.	ANTECEDENTES.....	20
2.1	Marco Histórico .....	20
2.1.1	Procesamiento del lenguaje natural.....	20
2.1.2	Niveles del lenguaje natural.....	22
2.1.3	Minería de opiniones y análisis de sentimientos .....	25
2.1.4	Dos enfoques .....	26
2.1.5	Pros y contras.....	27
2.2	Trabajos relacionados.....	28
3.	METODOLOGÍA.....	29
3.1	Etiquetado gramatical.....	30
3.2	Filtrado de datos .....	30
3.3	Comparación de datos .....	31
3.4	Análisis de emociones y sentimientos .....	31
3.5	Clasificación de las letras de las canciones por su polaridad sentimental ...	33
3.6	Precisión .....	34
4.	RESULTADOS .....	34
4.1	Etiquetado gramatical.....	36
4.2	Filtrado de datos .....	43

4.3	Comparación de datos .....	47
4.4	Análisis de emociones y sentimientos .....	49
4.5	Clasificación de las letras de las canciones por su polaridad sentimental ...	57
4.6	Precisión .....	60
5.	CONCLUSIONES Y TRABAJO FUTURO.....	60
6.	REFERENCIAS .....	62
7.	APÉNDICE .....	65

Dirección General de Bibliotecas UAG

## ÍNDICE DE TABLAS

Tabla 1. Niveles y herramientas del procesamiento del lenguaje natural. ....	22
Tabla 2. Tipos de etiquetas gramaticales. ....	30
Tabla 3. Palabras filtradas que coinciden con la base de datos léxica. ....	31
Tabla 4. Valencias de emociones y sentimientos.....	32
Tabla 5. Porcentaje de emociones y sentimientos. ....	33
Tabla 6. Muestra de cuatro canciones.....	34
Tabla 7. Etiquetado gramatical del texto de la letra de la canción 1. ....	36
Tabla 8. Etiquetado gramatical del texto de la letra de la canción 2. ....	38
Tabla 9. Etiquetado gramatical del texto de la letra de la canción 3. ....	39
Tabla 10. Etiquetado gramatical del texto de la letra de la canción 4. ....	41
Tabla 11. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 1.....	43
Tabla 12. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 2.....	44
Tabla 13. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 3.....	45
Tabla 14. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 4.....	46
Tabla 15. Coincidencias de las palabras filtradas del texto de la letra de la canción 1 y los unigramas de la base de datos léxica.....	47
Tabla 16. Coincidencias de las palabras filtradas del texto de la letra de la canción 2 y los unigramas de la base de datos léxica.....	47

Tabla 17. Coincidencias de las palabras filtradas del texto de la letra de la canción 3 y los unigramas de la base de datos léxica.....	48
Tabla 18. Coincidencias de las palabras filtradas del texto de la letra de la canción 4 y los unigramas de la base de datos léxica.....	48
Tabla 19. Valencias de emociones y sentimientos del texto de la letra de la canción 1.....	49
Tabla 20. Porcentaje de emociones y sentimientos del texto de la letra de la canción 1.....	50
Tabla 21. Valencias de emociones y sentimientos del texto de la letra de la canción 2.....	51
Tabla 22. Porcentaje de emociones y sentimientos del texto de la letra de la canción 2.....	52
Tabla 23. Valencias de emociones y sentimientos del texto de la letra de la canción 3.....	53
Tabla 24. Porcentaje de emociones y sentimientos del texto de la letra de la canción 3.....	54
Tabla 25. Valencias de emociones y sentimientos del texto de la letra de la canción 4.....	55
Tabla 26. Porcentaje de emociones y sentimientos del texto de la letra de la canción 4.....	56

## ÍNDICE DE FIGURAS

Figura 1. Metodología propuesta para el análisis de emociones y sentimientos contenidos en letras de canciones en español en tiempo real. ....	29
Figura 2. Clasificación del análisis de emociones y sentimientos de la letra de la canción 1.....	57
Figura 3. Clasificación del análisis de emociones y sentimientos de la letra de la canción 4.....	58
Figura 4. Clasificación del análisis de emociones y sentimientos de la letra de la canción 3.....	59
Figura 5. Clasificación del análisis de emociones y sentimientos de la letra de la canción 2.....	60

## 1. INTRODUCCIÓN

### 1.1 Definición del proyecto de investigación

En esta investigación se presenta una base de datos léxica llamada Emolex (Mohammad & Turney, 2013; Mohammad *et al.*, 2017, 2018), la cual contiene 14,182 unigramas (palabras) que permite identificar ocho emociones básicas (ira, disgusto, tristeza, sorpresa, miedo, confianza, alegría y anticipación) y la polaridad del sentimiento (positivo y negativo) y un algoritmo cognitivo de procesamiento de lenguaje natural llamado Natural Language Cloud (Google, 2019b) que incluye un analizador semántico que permite utilizar un etiquetador gramatical llamado Part of Speech (POS) que forma parte del discurso para el análisis de texto no estructurado en letras de canciones en español, para clasificarlas de acuerdo con su carga emocional y sentimental. Se detecta la situación emocional del usuario en un momento determinado. Posteriormente se utiliza el analizador en la identificación de situaciones emocionales y de sentimientos en mensajes de textos de correos electrónicos que infieren en la satisfacción del cliente en la empresa de servicios de facturación electrónica EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV. Se gestionan efectivamente las relaciones con los clientes y se toman acciones apropiadas en función de su estado emocional.

### 1.2 EmoLex

Es un recurso léxico desarrollado por el National Research Council (NRC), y fue creado a través de un servicio en línea de Amazon llamado Amazon Mechanical Turk, que requiere de inteligencia humana para desarrollar diferentes tareas. Se basa en el conjunto de ocho emociones básicas y dos sentimientos que consideraron que la cualidad bipolar de estas emociones resultaría de gran utilidad para el proyecto de creación de EmoLex. Respecto al número de entradas, este lexicón consta de 14,182 unigramas (palabras).

El número de términos por emoción que presenta es:

- Ira: 1247 términos
- Anticipación: 839 términos
- Disgusto: 1058 términos
- Miedo: 1476 términos
- Alegría: 688 términos
- Tristeza: 1191 términos
- Sorpresa: 534 términos
- Confianza: 1230 términos

El NRC eligió la plataforma de Amazon por varias razones, siendo las más destacable la posibilidad de obtener una gran cantidad de anotaciones humanas de forma eficiente. A través de esta fórmula se genera la base de datos léxica EmoLex (Mohammad & Turney, 2013; Mohammad *et al.*, 2017, 2018), cuyo tamaño es mayor que muchas otras herramientas utilizadas en este campo como lo son el lexicón SenticNet, SO-CAL, General Inquirer entre otros (Crossley *et al.*, 2017).

### 1.3 Natural Language Cloud de Google

Natural Language Cloud es una interfaz de programación de aplicaciones (API) de inteligencia artificial de Google que implementa procesamiento de lenguaje general para descubrir la estructura y el significado del texto a través de modelos de aprendizaje automático. Con esta API, se puede extraer información sobre personas, lugares, eventos y mucho más, independientemente de si los datos se encuentran en documentos de texto, artículos de noticias, entradas de blog, etc. (Google,2019b).

Las características de la API Natural Language Cloud de Google son:

- Análisis sintáctico: extrae componentes léxicos y frases, identifica categorías gramaticales y crea árboles de análisis de dependencias para cada frase.
- Reconocimiento de entidades: identifica entidades y las etiqueta por tipo: persona, organización, ubicación, evento, producto o medio.
- Análisis de opinión: muestra la opinión general que se desprende de un bloque de texto.
- Clasificación de contenido: clasifica documentos en las más de 700 categorías predefinidas.
- Multilingüe: analiza texto fácilmente en varios idiomas, como alemán, chino (tanto simplificado como tradicional), coreano, español, francés, inglés, italiano, japonés y portugués.

#### 1.4 Justificación

El lenguaje es una herramienta para expresar los sentimientos a los demás. Los seres humanos han mejorado esta habilidad en forma hablada y escrita durante milenios hasta el punto en que el oyente o el lector puede experimentar los mismos sentimientos que el hablante o el escritor. La psicología ha identificado que diferentes palabras expresan diversas emociones en diferentes grados. Por ejemplo, la palabra dulce expresa una emoción positiva y crea un sentimiento de felicidad, mientras que la palabra matar crea una emoción negativa (Miller & Charles, 2016).

La extracción automática de información semántica relacionada con sentimientos y opiniones humanas y el análisis de textos basados en esta información se clasifican en varios términos generales, incluida la subjetividad, minería de opinión, emoción, afecto y análisis de sentimientos (Crossley *et al.*, 2017). El sentimiento está ampliamente asociado con las emociones y la opinión, y el término análisis de sentimiento se usa comúnmente como un término general relacionado con la extracción de información subjetiva relacionada con sentimientos humanos y opiniones de textos del lenguaje natural (Hutto & Gilbert, 2014; Liu, 2012; Pang & Lee, 2008). El análisis de sentimientos es un enfoque útil

para una serie de problemas diferentes que se plantean en distintas disciplinas, como psicología, educación, sociología, negocios, ciencias políticas y economía (Hutto & Gilbert, 2014), así en campos de investigación como el procesamiento de lenguaje natural, minería de datos y recuperación de información (Zhang *et al.*, 2014).

El análisis de sentimientos es un área de investigación popular que utiliza el procesamiento del lenguaje natural, el análisis de textos y la lingüística computacional para identificar y extraer información subjetiva en los materiales de origen (Miller & Charles, 2016). El manejo de grandes volúmenes de datos de texto no estructurados es una tarea desafiante que se está volviendo cada vez más necesaria en una variedad de situaciones. (Jayathilaka *et al.*, 2016).

La cantidad de texto no estructurado ha aumentado con la rápida adopción de información digitalizada. Leer y extraer información significativa de grandes colecciones de datos no estructurados es cada vez más una necesidad. Hoy en día, dar sentido a los grandes volúmenes de datos no estructurados se está volviendo esencial (Jayathilaka *et al.*, 2016).

Los datos se están convirtiendo en el foco de atención en los escenarios de negocios actuales debido al espectacular crecimiento de Internet y los almacenamientos en línea. Estos datos pueden ser estructurados, semiestructurados y no estructurados. Los datos estructurados están organizados y se pueden analizar fácilmente. Los datos semiestructurados y no estructurados no están predefinidos y son complejos, lo que hace que obtener conocimiento de ellos sea un gran desafío (Mahlawi & Sasi, 2017).

Tipos de datos estructurados, semiestructurados y no estructurados.

- Datos estructurados: tienen perfectamente definido la longitud, el formato y el tamaño de sus datos. Se almacenan en formato tabla, hojas de cálculo o en bases de datos relacionales.

- Datos no estructurados: se caracterizan por no tener un formato específico. Se almacenan en múltiples formatos como documentos pdf o word, correos electrónicos, archivos multimedia de imagen, audio o video.
- Datos semiestructurados: son una mezcla de los dos anteriores no presenta una estructura perfectamente definida como los datos estructurados, pero si presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención, como por ejemplo los formatos html, xml o json.

La mayoría de las organizaciones prefieren trabajar con datos estructurados porque tiene la ventaja de que se pueden almacenar, administrar y analizar fácilmente. Sin embargo, la mayoría de los datos de toma de decisiones de hoy provienen de fuentes externas desestructuradas que representan más del 90% de los datos actuales. La mayoría de estos datos vienen en forma de texto tales como documentos, publicaciones en redes sociales (Lee *et al.*, 2018), correos electrónicos, etc. (Mahlawi & Sasi, 2017).

Extraer el conocimiento de textos no estructurados es difícil, debido a la naturaleza desorganizada de los datos y la ambigüedad del lenguaje natural. La minería de textos y el procesamiento de lenguaje natural se pueden aplicar a los textos para estructurar y extraer información significativa de ellos. La minería de textos es el proceso que estructura y deriva las percepciones y el significado del texto. Procesamiento de lenguaje natural es el proceso que puede aplicarse al lenguaje humano para ayudar a la computadora a obtener resultados valiosos del texto (Singh *et al.*, 2017). Una de las muchas fuentes de datos de texto no estructurados externos para las organizaciones es el correo electrónico. (Mahlawi & Sasi, 2017).

El análisis de sentimientos es un enfoque útil para una serie de problemas diferentes que se plantean en diversas disciplinas, como psicología (Hoogendoorn *et al.*, 2016), educación, sociología, negocios, ciencias políticas y economía, así

como en campos de investigación como el procesamiento de lenguaje natural, minería de datos y recuperación de información (Jiang *et al.*, 2018).

El análisis de sentimientos también es útil para automatizar la toma de decisiones al ayudar a las organizaciones a comprender mejor los efectos de los problemas específicos en las percepciones de las personas y responder a estos efectos de manera adecuada a través del marketing y la comunicación (Jiang *et al.*, 2018). Por lo tanto, medir las características del sentimiento automáticamente en un texto es valioso, para comprender mejor cómo las emociones, los sentimientos, el afecto y las opiniones que influyen en el usuario (Crossley *et al.*, 2017).

Al utilizar en este trabajo de investigación como banco de datos el texto de letras de canciones en español permite analizar textos no estructurados más largos y obtener una mejor validación emocional por la carga considerable de emociones y sentimientos que contienen las letras de las canciones en español, posteriormente el analizador podrá ser implementado a cualquier tipo de texto no estructurado como: correos electrónicos, chats, redes sociales, documentos, etc.

El problema se asocia con empresas de servicios por Internet que manejan atención a clientes, y que dicha atención mayormente se realiza a través de medios electrónicos, pues el total de información se maneja en modo texto. En este caso, se presenta el problema de vinculación con la empresa que apoya en el desarrollo de este trabajo de tesis, por lo que es importante describir las condiciones actuales que se presentan.

La empresa de servicios en tecnologías de la información EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV con RFC: EST101217592 inició actividades en la ciudad de Querétaro en el año 2010, cuenta con 9 años de presencia y trayectoria en el mercado. Desde el inicio de sus actividades se ha enfocado en el desarrollo de sistemas, consultoría informática y servicios por Internet. La línea principal de negocios de la empresa es la venta, soporte y

distribución en toda la República Mexicana del sistema de facturación electrónica DigiFact – Sicofi (Edigital, 2019).

Actualmente EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV cuenta con una cartera de clientes activos de 2,582, de los cuales se tiene una actividad constante mensual con al menos 1,600 clientes. El soporte que se ofrece a esta población de clientes es por tres medios de comunicación: teléfono, correo electrónico y chat. Para la empresa el servicio es uno de los procesos de comunicación más importante de los que se disponen y el poder identificar las emociones y sentimientos de los clientes es un factor que determina qué tipo de soporte y la prioridad de servicio que se le va a ofrecer. El detectar la carga emocional y de sentimientos por parte del personal en llamadas telefónicas es más simple de realizar y de identificar, pero al trasladar este tipo de análisis emocional a textos no estructurados como lo son correo electrónico y chat se vuelve una tarea difícil de llevar a cabo, ya que se requiere que se analice uno por uno los correos y los chats para determinar la prioridad de atención y soporte que requiere el cliente de acuerdo a su carga emocional que expresa al estar solicitando ayuda por estos canales.

Se realizó un análisis manual de los correos electrónicos para clasificarlos de acuerdo con su tipo, prioridad de servicio y el nivel de sentimiento que emiten estos mensajes de acuerdo con su polaridad (positiva, neutra o negativa). En promedio se reciben 450 correos electrónicos semanales, lo cual al trasladarlo por mes son 1,950 y al año 23,400 correos electrónicos que el personal encargado de servicio y soporte tiene que analizar uno por uno para su clasificación.

Gran parte de las investigaciones de análisis de sentimientos y emociones sobre datos no estructurados se centran solamente en la polaridad de las opiniones del texto para clasificarlo de acuerdo con su carga sentimental en tres tipos de valores: positivo, neutro y negativo, incluso aplicaciones de correo electrónico de compañías como Outlook (Microsoft, 2019), Gmail (Google, 2019a), Inbenta (Inbenta, 2019), así como Emailparser ,(FrozenFrog Software, 2019) solo

calculan la polaridad sentimental de los textos de los correos electrónicos. Esto representa a las organizaciones un problema al intentar identificar emociones como la ira, tristeza, miedo, anticipación, sorpresa, alegría, etc., en sus textos no estructurados que se emiten mediante correos electrónicos. Al identificar sentimientos y emociones, como en este trabajo se propone, se puede garantizar a las organizaciones una mejor forma de elevar los índices de calidad en el servicio y mantener una relación cercana con los clientes para la toma de acciones apropiadas en función de sus necesidades de acuerdo con su estado emocional y sentimental. Así se podrá incrementar la competitividad y el crecimiento de la entidad corporativa en un mercado determinado.

### 1.5 Objetivos de la investigación

Objetivo general:

Desarrollar un analizador semántico de emociones y sentimientos en tiempo real con el propósito de determinar el léxico emocional de las palabras contenidas en las letras de canciones en español con la finalidad de detectar la situación emocional del usuario en un momento determinado. El uso posterior del analizador es en la identificación de situaciones emocionales y de sentimientos en mensajes de textos de correos electrónicos que infieren en la satisfacción del cliente en la empresa EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV.

Objetivos específicos:

- Implementar el algoritmo de inteligencia artificial de procesamiento de lenguaje natural llamado Natural Language Cloud (Google, 2019b) para obtener las palabras que son parte del discurso Part of Speech (POS) con la finalidad de separar las palabras que si emiten emociones de las que no.
- Analizar e implementar la base de datos léxico emocional Emolex (Mohammad & Turney, 2013; Mohammad *et al.*, 2017, 2018) la cual contiene 14,182 unigramas (una sola palabra) para determinar emociones y sentimientos de las letras de las canciones en español.

- Normalizar los resultados que se extraen de la base de datos Emolex a datos discretos para su cálculo, análisis y clasificación de las emociones y sentimientos de las letras de las canciones en español.
- Desarrollar una plataforma web que permita agregar un listado de canciones en español, las cuales se analizarán para determinar y mostrar el porcentaje y palabras de cada emoción básica (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y sentimientos (negativo y positivo) que emiten los textos de las letras de las canciones en tiempo real para clasificarlas de acuerdo con su polaridad sentimental.
- Utilizar la metodología propuesta del analizador a textos de correos electrónicos de la empresa EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV, para identificar situaciones emocionales y de sentimientos que infieren en la satisfacción del cliente.

## **2. ANTECEDENTES**

### 2.1 Marco Histórico

#### 2.1.1 Procesamiento del lenguaje natural

La disciplina que se ocupa del tratamiento automático del lenguaje natural en textos se llama procesamiento del lenguaje natural. El procesamiento del lenguaje natural es parte del área de investigación de la inteligencia artificial, que se define como “la ciencia y la ingeniería de hacer máquinas inteligentes” (McCarthy, 1959, p.77) mediante la simulación de los mecanismos de la inteligencia humana.

“El procesamiento del lenguaje natural es un conjunto de técnicas computacionales teóricamente motivadas para analizar y representar naturalmente textos de origen natural en uno o más niveles de análisis lingüísticos para lograr el propósito de procesar el lenguaje humano por una serie de tareas o aplicaciones”. Esta definición fue aportada en Liddy (2001, p.2126), y hace hincapié en la

capacidad de estos sistemas para llevar a cabo diferentes niveles de análisis lingüísticos.

Hasta la década de 1980, la mayoría de los sistemas de procesamiento del lenguaje natural se basaban en un complejo conjunto de reglas diseñadas a mano. A partir de finales de 1980 hubo una revolución en el procesamiento del lenguaje natural con la introducción de algoritmos de aprendizaje automático. Durante los últimos años, esta área de investigación ha seguido en constante movimiento. En primer lugar, los modelos probabilísticos y los modelos conducidos por datos se encuentran bastante estandarizados en todo el procesamiento del lenguaje natural. Los algoritmos de análisis, etiquetado gramatical, resolución de referencia y discurso han comenzado a incorporar modelos probabilísticos y a emplear metodologías de evaluación en el reconocimiento de voz y recuperación de la información. Estos años se han caracterizado por el incremento del nivel de procesamiento y la rapidez de las computadoras, lo que ha permitido la explotación de diferentes áreas del procesamiento del lenguaje natural, como el reconocimiento de la voz y la detección de errores ortográficos y gramaticales (Liddy, 2001). Por último, el auge de la web ha resaltado la necesidad de innovación en los procesos de minería de opinión y extracción de información de los sistemas de almacenamiento.

El procesamiento del lenguaje natural es una técnica esencial de la inteligencia artificial que tiene como propósito el modelado y procesamiento computacional del lenguaje humano. Todo sistema de procesamiento del lenguaje natural intenta simular el comportamiento lingüístico humano. Para ello debe tomar conciencia tanto de las estructuras propias del lenguaje, como de su dominio de aplicación. Es decir, un sistema de procesamiento del lenguaje natural deberá conocer los significados de las palabras y cómo estas afectan al significado global del discurso.

La información no estructurada la encontramos en fuentes tales como documentos, la web y cualquier otro medio que contenga información expresada

en lenguaje natural y en formatos muy diversos como texto, videos, audio o imágenes. Este tipo de información ocupa más del 94 % del universo digital (Gantz *et al.*, 2008). Con la llegada de Internet, la información no estructurada (como el correo electrónico, los informes, contratos, facturas, formularios, hojas de cálculo, imágenes, presentaciones, etc.) es relevante para su análisis.

### 2.1.2 Niveles del lenguaje natural

El lenguaje humano puede ser dividido en distintos niveles de cara a su análisis y entendimiento. En Feldman (1999) y Liddy (2001) se incluyen los siete niveles del lenguaje que utilizan los humanos para extraer el significado del lenguaje natural. La Tabla 1 contiene un cuadro resumen con los distintos niveles del lenguaje natural, su ámbito de actuación, las herramientas de procesamiento que se pueden aplicar para analizarlos de forma automática y las estructuras lingüísticas obtenidas en cada nivel.

**Tabla 1. Niveles y herramientas del procesamiento del lenguaje natural.**

Niveles del procesamiento del lenguaje natural	Ámbito de actuación	Herramientas Utilizadas	Resultados del procesamiento
Fonético	Sonidos	Corpus de aprendizaje. Modelos de acústicos. Diccionario de unidades de síntesis.	Fonemas
Léxico	Formas	Lematizador. Etiquetado POS. Lexicón computacional.	Palabras
Morfológico	Formas	Lematizador. Etiquetado POS. Lexicón computacional.	Palabras
Sintáctico	Estructuras	Bases de datos sintácticas. Treebank.	Frases
Semántico	Significados	Bases de datos semánticas. Lexicón computacional. Ontologías.	Relaciones
Discurso	Comunicación	Bases de datos semánticas. Ontologías.	Textos

Pragmático	Comunicación	Bases de datos semánticas. Ontologías.	Textos
------------	--------------	---	--------

Fuente: Liddy (2001, p.2127).

Los niveles del lenguaje humano no son niveles aislados, sino que están interrelacionados. Además, existen niveles que, para desarrollar su proceso de análisis, requieren conocimiento de niveles anteriores, o incluso posteriores. Por ejemplo, para los procesos de desambiguación de las categorías morfológicas es necesario recurrir a la sintaxis para determinar qué función desempeña un término en una frase.

A continuación, se explican en detalle las características más importantes de cada uno de los niveles del lenguaje humano. Según Feldman (1999) y Liddy (2001), los siete niveles son los siguientes:

1. Fonológico. La fonética estudia los sonidos físicos del discurso humano. Los sistemas que procesan a un nivel fonológico analizan la voz humana codificando las ondas producidas en señales digitales para interpretarlas y poder establecer comparaciones con diferentes modelos de lenguaje.
2. Morfológico. La morfología estudia la estructura interna y la composición de las palabras. Los sistemas que procesan a un nivel morfológico separan y clasifican los caracteres en secuencias de morfemas, utilizando técnicas como lematizadores, o etiquetadores POS. Con estas técnicas es posible clasificar las palabras en categorías gramaticales como tiempo, sufijo, prefijo, género, sinónimo, sustantivo, verbo, adjetivo, adverbio, etc.
3. Léxico. La lexicología estudia el significado de las unidades léxicas de una lengua junto con las relaciones sistemáticas y las relaciones entre ellas. Los sistemas que procesan a un nivel léxico interpretan el significado de las palabras de forma individual. Los sistemas de procesamiento de lenguaje natural requieren de la lexicología para interpretar y conocer las categorías léxicas en un lenguaje. En este nivel existen técnicas que analizan el léxico de las palabras, pero de las técnicas más extendidas que se conocen etiquetan de forma individual cada palabra del texto. Estas etiquetas

definen la categoría léxica de las palabras. Si una palabra desempeñara diferentes funciones dentro de la oración se organiza y etiqueta de acuerdo con el contexto de la sentencia.

4. Sintáctico. La sintaxis estudia las reglas y el orden que existe en las palabras dentro de una oración para determinar la función que desempeña cada palabra. Los sistemas que procesan a un nivel sintáctico analizan las secuencias y las estructuras gramaticales de las palabras. Este análisis cuenta los patrones sintácticos que son más frecuentes para generar una estructura gramatical de acuerdo con las unidades léxicas y poder representarlas una forma de red o de árbol.
5. Semántico. La semántica estudia el significado de las palabras, expresiones y oraciones. Los sistemas que procesan a un nivel semántico determinan los significados posibles de las palabras. Este tipo de procesamiento requiere de técnicas de separación de ambigüedad semántica para palabras que tengan varios significados.
6. Discurso o contextual. Los sistemas que procesan a este nivel interpretan los textos como grupos independientes de oraciones concatenadas y analizan las propiedades del texto en un conjunto que transmite significado al conectar las oraciones. El propósito de este nivel es usar una estructura semántica para formular una interpretación final de la oración de acuerdo con el contexto. El contexto es considerado un conjunto de palabras, relaciones sintácticas y categorías semánticas (Vázquez *et al.*, 2003). Existen varios tipos de procesamiento, los más conocidos son la resolución anafórica y el reconocimiento de la estructura del texto o discurso (Carrión, 2014).
7. Pragmático. La pragmática estudia la relación con los usuarios y las circunstancias de la comunicación en un contexto sociocultural. El análisis en este nivel se relaciona con factores extralingüísticos que condicionan el lenguaje en ciertas situaciones comunicativas concretas. La pragmática

utiliza el contexto o discurso, por encima de los contenidos para entender el significado del texto.

### 2.1.3 Minería de opiniones y análisis de sentimientos

La minería de opiniones permite valorar cuantitativamente expresiones subjetivas como sentimientos o sensaciones (Chen & Zimbra, 2010; Pang & Lee, 2008). Permitiendo conocer si se está hablando positivamente, negativamente o de forma neutra sobre productos o servicios. Así como medir la intensidad de dicha opinión. Debido a la importancia actual de foros y redes sociales para algunos sectores de actividad, la posibilidad de disponer de herramientas que permitan monitorizar y valorar las opiniones vertidas en estos foros y redes sociales es fundamental (Rushdi-Saleh *et al.*, 2011).

Cuando se hace referencia a la minería de opiniones, también llamada análisis del sentimiento, se refiere a la aplicación de una serie de técnicas del procesamiento del lenguaje natural, lingüística computacional y minería de textos, que tienen como objetivo la extracción de información subjetiva a partir de contenidos generados por los usuarios, como comentarios en blogs u opiniones en revistas de productos (Rushdi-Saleh *et al.*, 2011; Cruz *et al.*, 2013). Con este tipo de tecnologías es posible extraer un valor tangible y directo a partir de un comentario o texto en lenguaje natural.

La minería de opinión tiene distintos campos de aplicación (Qiong & Songbo, 2011). Por ejemplo, con el uso de esta técnica se podrían monitorizar y gestionar las opiniones de los usuarios acerca de productos y servicios en redes sociales y foros. Se podría intentar identificar qué se opina sobre un producto o servicio, o cuál es la opinión sobre determinadas características del producto o servicio concreto (Min & Park, 2012). También se podrían monitorizar y gestionar las opiniones sobre personas o entidades en foros y redes sociales concretas. Así como abordar temas como las tendencias y hot topics. Esto se traduce en datos subjetivos con gran potencial. De hecho, pueden ser explotados por

administraciones públicas, empresas y particulares para, por ejemplo, conocer la opinión sobre personalidades públicas, elegir la propaganda idónea según las preferencias u opiniones de la gente o para encontrar el producto mejor valorado por los usuarios (Liu, 2007).

Además, la minería de opiniones puede cubrir diversas necesidades, por lo que, dependiendo de cuál sea la necesidad concreta, es posible adaptar la tecnología para cubrir dicha necesidad.

En la última década, el análisis de sentimientos, también conocido como minería de opiniones, ha despertado un creciente interés. Resulta un gran reto para las tecnologías del lenguaje, ya que obtener buenos resultados es mucho más difícil de lo que muchos creen (Villena, 2015). La tarea de clasificar automáticamente un texto escrito en un lenguaje natural en un sentimiento positivo o negativo, opinión o subjetividad (Pang & Lee, 2008), es a veces tan complicada que incluso es difícil poner de acuerdo con diferentes anotadores humanos sobre la clasificación a asignar a un texto dado. La interpretación personal de un individuo es diferente de la de los demás, y además se ve afectada por factores culturales y experiencias propias de cada persona. Y la tarea es aún más difícil cuanto más corto sea el texto, y peor escrito esté, como es el caso de los mensajes en redes sociales como Twitter o Facebook.

#### 2.1.4 Dos enfoques

El problema se ha abordado principalmente desde dos enfoques diferentes (Liu, 2012): técnicas de aprendizaje computacional (Pang *et al.*, 2002) y aproximaciones semánticas (Turney, 2002).

Los enfoques semánticos se caracterizan por el uso de diccionarios de términos (lexicones) con orientación semántica de polaridad u opinión. Típicamente los sistemas pre procesan el texto y lo dividen en palabras, con la apropiada eliminación de las palabras de parada (stopwords) y una normalización lingüística o lematización, y luego comprueban la aparición de los términos del

lexicón para asignar el valor de polaridad del texto mediante la suma de valores de polaridad de los términos.

Por otra parte, los enfoques basados en aprendizaje computacional consisten en entrenar un clasificador usando un algoritmo de aprendizaje supervisado a partir de una colección de textos anotados, donde cada texto habitualmente se representa con un vector de palabras (bag of words), n-gramas o skip-grams, en combinación con otro tipo de características semánticas que intentan modelar la estructura sintáctica de las frases, la intensificación, la negación, la subjetividad o la ironía. Los sistemas utilizan diversas técnicas, aunque las más populares son los clasificadores basados en Support Vector Machines (SVM), Naive Bayes y K-Nearest Neighbor (KNN). En las investigaciones más recientes se han empezado a utilizar otras técnicas más avanzadas, como Latent Semantic Analysis (LSA) e incluso Deep Learning.

#### 2.1.5 Pros y contras

La ventaja principal de los enfoques semánticos es que los errores son relativamente sencillos de corregir, añadiendo cuantos términos fuera necesario, y se podría obtener una precisión tan alta como se quisiera, simplemente invirtiendo más tiempo en la construcción del lexicón. En este sentido, los enfoques basados en aprendizaje automático suelen ser una caja negra en la que corregir errores o añadir nuevo conocimiento es más complicado, y muchas veces sólo es posible ampliando la colección de textos y volviendo a entrenar el modelo.

Por otra parte, la ventaja de los enfoques basados en aprendizaje automático es que cuesta muy poco construir un analizador de sentimientos a partir de la colección de textos etiquetados, ya que la tarea de modelado reside en el algoritmo. Por ello, es relativamente fácil construir clasificadores adaptados a un dominio determinado. En contraposición, el esfuerzo para construir un lexicón para un cierto dominio, empezando de cero, es muy elevado, porque se basa en mucho trabajo manual, así que en general son menos adaptables.

## 2.2 Trabajos relacionados

Miller y Charles (2016) proponen un análisis semántico, descriptivo y observacional basado en la psicología de marketing en las líneas de asunto del correo electrónico, lo cual es un factor decisivo para abrir y leer el mensaje.

Jayathilaka *et al.* (2016) abordan como dar sentido a 500,000 envíos de correos electrónicos en formato HTML recibidos en respuesta a un debate sobre la neutralidad de la red en la India utilizando un enfoque basado en el léxico para clasificar la polaridad del sentimiento de las respuestas como positiva, negativa y neutral.

Mahlawi y Sasi (2017) presentan un proceso para extraer datos estructurados: palabras clave, entidades, expresiones regulares y análisis de sentimientos en correos electrónicos sobre un dominio, como en un proyecto, servicio o producto.

Jiang *et al.* (2018) demuestran cómo se puede crear un perfil psicológico del usuario basado en el análisis de sentimiento de su navegación por la web y el contenido de su correo electrónico.

Hoogendoorn *et al.* (2016) proponen modelos de predicción del resultado del tratamiento de ansiedad social basado en conversaciones terapéuticas por correo electrónico centrados en los siguientes aspectos: comportamiento de envió del correo, uso de las palabras que utiliza el paciente, estilo de escritura, el sentimiento del texto y el tema a que se refiere el mensaje.

Crossley *et al.* (2017) presentan una herramienta automática para el análisis de sentimiento, la cognición y el orden social y cómo se pueden usar los elementos léxicos relacionados con la cognición, el afecto y las preocupaciones personales para comprender mejor el comportamiento humano en diferentes disciplinas como: la psicología, la educación, la sociología, los negocios, las ciencias políticas y la economía.

### 3. METODOLOGÍA

La metodología consta de cinco fases: 1) etiquetado gramatical, 2) filtrado de datos, 3) comparación de datos, 4) análisis de emociones y sentimientos y 5) clasificación de las letras de las canciones por su polaridad sentimental. La Figura 1 muestra la metodología propuesta.



**Figura 1. Metodología propuesta para el análisis de emociones y sentimientos contenidos en letras de canciones en español en tiempo real.**

Fuente: Elaboración propia.

### 3.1 Etiquetado gramatical

Se implementa el algoritmo cognitivo Natural Language Cloud de Google (Google, 2019b), al cual se envía el texto de la letra de la canción mediante un servicio web para ser tratada por técnicas de procesamiento de lenguaje natural con el objetivo de etiquetar gramaticalmente cada palabra del texto contenido en la canción. En la Tabla 2 se muestra el tipo de etiquetas gramaticales con las que se clasifica el texto de la letra.

**Tabla 2. Tipos de etiquetas gramaticales.**

Etiqueta gramatical	Descripción
Unknown	Desconocido
Adj	Adjetivo
Adp	Adposición
Adv	Adverbio
Conj	Conjunción
Det	Determinante
Noun	Sustantivo
Num	Número cardinal
Pron	Pronombre
Prt	Partícula u otra palabra funcional
Punct	Puntuación
Verb	Verbo
X	Otros
Affix	Afijo

Fuente: Elaboración propia.

### 3.2 Filtrado de datos

Al obtener el texto de la letra de la canción procesado por el algoritmo cognitivo Natural Language Cloud y etiquetado de acuerdo con su categoría gramatical, se requiere filtrar y extraer las palabras del mensaje que emiten sentimientos y emociones, las cuales para esta investigación son: sustantivos, verbos, adjetivos y adverbios y separar las que no emiten emociones ni sentimientos como: desconocido, adposición, conjunción, determinante, número cardinal, pronombre, partícula u otra palabra funcional, puntuación, otros (palabras

extranjerías, errores tipográficos, abreviaturas y afijos), de acuerdo con los tipos de etiquetas gramaticales de la Tabla 2.

### 3.3 Comparación de datos

Una vez que se obtuvieron las palabras filtradas del texto de la letra de la canción, se etiquetaron de acuerdo con su categoría gramatical, se realiza una consulta a la base de datos léxica Emolex (Mohammad & Turney, 2013; Mohammad *et al.*, 2017, 2018) la cual contiene 14,182 unigramas (palabras), que están categorizados en ocho emociones (ira, disgusto, tristeza, sorpresa, miedo, confianza, alegría y anticipación) y dos sentimientos (positivo y negativo) para comparar los datos y obtener las palabras que coincidan con el texto que se filtró del mensaje y los unigramas de la base de datos como se muestra en la Tabla 3.

**Tabla 3. Palabras filtradas que coinciden con la base de datos léxica.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
P1	1	0	0	0	1	0	0	0	0	1
P2	0	0	0	0	0	1	1	1	1	0
P3	0	0	0	0	0	0	1	0	1	0
P4	0	0	0	0	0	1	0	0	1	0
PN	0	0	0	1	0	0	1	1	1	0

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza, Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

### 3.4 Análisis de emociones y sentimientos

Las palabras de la letra de la canción pueden tener más de una emoción y sentimiento al mismo tiempo. Por ejemplo, la palabra “atraso” genera emociones de sorpresa, anticipación y un sentimiento negativo. La palabra “tratamiento” genera emociones de ira, anticipación, disgusto, miedo, alegría y tristeza y un sentimiento positivo y negativo.

Las emociones y sentimientos tienen un valor binario dentro de la base de datos léxica, donde 1 determina que la palabra emite esa emoción o sentimiento, si es un 0 es que no lo emite.

Para calcular la valencia de cada emoción y sentimiento se suman los valores binarios de la columna correspondiente como se muestra en la Tabla 4.

**Tabla 4. Valencias de emociones y sentimientos.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
P1	1	0	0	0	1	0	0	0	0	1
P2	0	0	0	0	0	1	1	1	1	0
P3	0	0	0	0	0	0	1	0	1	0
P4	0	0	0	0	0	1	0	0	1	0
PN	0	0	0	1	0	0	1	1	1	0
Valencia	1	0	0	1	1	2	3	2	4	1

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza, Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Al obtener las valencias de cada emoción y sentimiento, estas se suman para obtener la valencia total de emociones y la valencia total de sentimientos.

- Valencia total de emociones =  $v_{ir} + v_{di} + v_{tr} + v_{so} + v_{mi} + v_{co} + v_{al} + v_{an}$
- Valencia total de sentimientos =  $v_{po} + v_{ne}$

Para determinar el porcentaje de cada emoción se divide la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira =  $v_{ira} / v_{te}$
- Porcentaje disgusto =  $v_{disgusto} / v_{te}$
- Porcentaje tristeza =  $v_{tristeza} / v_{te}$
- Porcentaje sorpresa =  $v_{sorpresa} / v_{te}$
- Porcentaje miedo =  $v_{miedo} / v_{te}$

- Porcentaje confianza =  $v_{\text{confianza}} / v_{\text{te}}$
- Porcentaje alegría =  $v_{\text{alegría}} / v_{\text{te}}$
- Porcentaje anticipación =  $v_{\text{anticipacion}} / v_{\text{te}}$

Para determinar el porcentaje de cada sentimiento se divide la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo =  $v_{\text{positivo}} / v_{\text{ts}}$
- Porcentaje negativo =  $v_{\text{negativo}} / v_{\text{ts}}$

En la Tabla 5 se muestra que, una vez aplicadas las fórmulas, se obtienen los niveles de porcentaje de emociones y sentimientos de cada letra de las canciones.

**Tabla 5. Porcentaje de emociones y sentimientos.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
P1	1	0	0	0	1	0	0	0	0	1
P2	0	0	0	0	0	1	1	1	1	0
P3	0	0	0	0	0	0	1	0	1	0
P4	0	0	0	0	0	1	0	0	1	0
PN	0	0	0	1	0	0	1	1	0	1
Valencia	1	0	0	1	1	2	3	2	3	2
Porcentaje	0.10	0.0	0.0	0.10	0.10	0.20	0.30	0.20	0.60	0.40

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza, Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

### 3.5 Clasificación de las letras de las canciones por su polaridad sentimental

La clasificación de canciones se realiza por la polaridad de sentimientos que emite cada letra analizada.

Las letras de las canciones se ordenan de acuerdo con el porcentaje de polaridad negativa, las más negativas primero y así sucesivamente, en orden

descendente. Este ordenamiento deja las letras de las canciones mas positivas abajo de la clasificación.

El conjunto de datos analizados y clasificados fue de 1,014 canciones. El periodo de recolección comprende del 09 de mayo del 2019 al 16 de diciembre del 2019.

### 3.6 Precisión

La evaluación de la precisión del análisis de emociones y sentimientos de las letras de las canciones se determina dividiendo las canciones clasificadas correctamente entre el total de canciones analizadas.

Precisión = canciones clasificadas correctamente / total canciones analizadas

## 4. RESULTADOS

Una muestra de cuatro canciones de las 1,014 letras analizadas aparece en la Tabla 6.

**Tabla 6. Muestra de cuatro canciones.**

	Artista	Caifanes
	Canción	La célula que explota
1	Letra	<p>Hay veces que no tengo ganas de verte          Hay veces que no quiero ni tocarte          Hay veces que quisiera ahogarte en un grito          Y olvidarme de esa imagen tuya          Pero no me atrevo          Hay veces que no dejo de soñarte          De acariciarte hasta que ya no pueda          Hay veces que quisiera morir contigo          Y olvidarme de toda materia          Pero no me atrevo          Hay veces que no sé lo que me pasa          Ya no puedo saber que es lo que pasa adentro          Somos como gatos en celo          Somos una célula que explota          Y esa no la paras, no, no la paras          No, no</p>

2	Artista	Tatiana
	Canción	Los reyes magos
	Letra	<p>Un buen día todo comenzó la historia lo describe así llegaron los Reyes Magos a saludarte a ti Muchos juguetes han traído con alegría y felicidad tu premio por ser tan bueno bueno con tus papás Él es Melchor, él es Gaspar y el otro Baltazar Él es Melchor, él es Gaspar y el otro va saltar Él es Melchor, él es Gaspar y el otro Baltazar Él es Melchor, él es Gaspar y el otro va saltar Quiero un trenecito que me diga así Pu pu y también una muñeca ideal que siempre me diga así Mama</p>
3	Artista	José Alfredo Jiménez
	Canción	El rey
	Letra	<p>Yo sé bien que estoy afuera Pero el día que yo me muera Sé que tendrás que llorar Dirás que no me quisiste Pero vas a estar muy triste Y así te vas a quedar Con dinero y sin dinero Hago siempre lo que quiero Y mi palabra es la ley No tengo trono ni reina Ni nadie que me comprenda Pero sigo siendo el rey Una piedra del camino Me enseñó que mi destino Era rodar y rodar Después me dijo un arriero Que no hay que llegar primero Pero hay que saber llegar</p>
4	Artista	Banda MS
	Canción	c
	Letra	<p>El color de tus ojos Despertó mi interés Y sólo tengo ganas de Verte otra vez Dime que no está prohibido</p>

	<p>Quizás me animo y te pido verte el sábado a las diez  El color de tus ojos  Se robó mi atención  Te vas metiendo dentro de mi corazón  Perfecta en cualquier sentido  Con pantalón o vestido robas mi respiración  Que más quisiera que fueras  El sueño que se vuelve realidad  Me gustas tanto y eso es toda la verdad  Me siento emocionado  No sé si te ha pasado  Que si pudiera te viera  De lunes a domingo sin parar  Esto que siento no se puede comparar  Y si ves que me sonrojo  Si te burlas, no me enojo  Yo sólo sé  Que de ti me enamoré</p>
--	---

Fuente: Elaboración propia.

#### 4.1 Etiquetado gramatical

Al aplicar el procesamiento de lenguaje natural con el algoritmo cognitivo Natural Language Cloud de Google al texto de las letras de las canciones de la muestra, se obtiene el etiquetado gramatical de cada palabra que forma parte del discurso como se observa en las Tablas Tabla 7, Tabla 8, Tabla 9 y Tabla 10.

**Tabla 7. Etiquetado gramatical del texto de la letra de la canción 1.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	Hay	Verbo	59	contigo	Pronombre
2	veces	Sustantivo	60	Y	Conjunción
3	que	Pronombre	61	olvidar	Verbo
4	no	Adverbio	62	me	Pronombre
5	tengo	Verbo	63	de	Adposición
6	ganas	Sustantivo	64	toda	Determinante
7	de	Adposición	65	materia	Sustantivo
8	ver	Verbo	66	Pero	Conjunción
9	te	Pronombre	67	no	Adverbio

10	Hay	Verbo	68	me	Pronombre
11	veces	Sustantivo	69	atrevo	Verbo
12	que	Pronombre	70	Hay	Verbo
13	no	Adverbio	71	veces	Sustantivo
14	quiero	Verbo	72	que	Pronombre
15	ni	Conjunción	73	no	Adverbio
16	tocar	Verbo	74	sé	Verbo
17	te	Pronombre	75	lo	Determinante
18	Hay	Verbo	76	que	Pronombre
19	veces	Sustantivo	77	me	Pronombre
20	que	Pronombre	78	pasa	Verbo
21	quisiera	Verbo	79	Ya	Adverbio
22	ahogar	Verbo	80	no	Adverbio
23	te	Pronombre	81	puedo	Verbo
24	en	Adposición	82	saber	Verbo
25	un	Determinante	83	que	Adposición
26	grito	Sustantivo	84	es	Verbo
27	Y	Conjunción	85	lo	Determinante
28	olvidar	Verbo	86	que	Pronombre
29	me	Pronombre	87	pasa	Verbo
30	de	Adposición	88	adentro	Adverbio
31	esa	Determinante	89	Somos	Verbo
32	imagen	Sustantivo	90	como	Adposición
33	tuya	Adjetivo	91	gatos	Sustantivo
34	Pero	Conjunción	92	en	Adposición
35	no	Adverbio	93	celo	Sustantivo
36	me	Pronombre	94	Somos	Verbo
37	atrevo	Verbo	95	una	Determinante
38	Hay	Verbo	96	célula	Sustantivo
39	veces	Sustantivo	97	que	Pronombre
40	que	Pronombre	98	explota	Verbo
41	no	Adverbio	99	Y	Conjunción
42	dejo	Verbo	100	esa	Determinante
43	de	Adposición	101	no	Adverbio
44	soñar	Verbo	102	la	Determinante
45	te	Pronombre	103	paras	Sustantivo
46	De	Adposición	104	,	Puntuación
47	acariciar	Verbo	105	no	Adverbio
48	te	Pronombre	106	,	Puntuación
49	hasta	Adposición	107	no	Adverbio

50	que	Adposición	108	la	Determinante
51	ya	Adverbio	109	paras	Sustantivo
52	no	Adverbio	110	No	Adverbio
53	pueda	Verbo	111	,	Puntuación
54	Hay	Verbo	112	no	Adverbio
55	veces	Sustantivo			
56	que	Pronombre			
57	quisiera	Verbo			
58	morir	Verbo			

Fuente: Elaboración propia.

**Tabla 8. Etiquetado gramatical del texto de la letra de la canción 2.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	Un	Determinante	54	es	Verbo
2	buen	Adjetivo	55	Gaspar	Sustantivo
3	día	Sustantivo	56	y	Conjunción
4	todo	Pronombre	57	el	Determinante
5	comenzó	Verbo	58	otro	Pronombre
6	la	Determinante	59	va	Verbo
7	historia	Sustantivo	60	saltar	Verbo
8	lo	Pronombre	61	El	Determinante
9	describe	Verbo	62	es	Verbo
10	así	Adverbio	63	Melchor	Sustantivo
11	llegaron	Verbo	64	,	Puntuación
12	los	Determinante	65	el	Determinante
13	Reyes	Sustantivo	66	es	Verbo
14	Magos	Sustantivo	67	Gaspar	Sustantivo
15	a	Adposición	68	y	Conjunción
16	saludar	Verbo	69	el	Determinante
17	te	Pronombre	70	otro	Determinante
18	a	Adposición	71	Baltazar	Sustantivo
19	ti	Pronombre	72	El	Determinante
20	Muchos	Determinante	73	es	Verbo
21	juguetes	Sustantivo	74	Melchor	Sustantivo
22	han	Verbo	75	,	Puntuación
23	traído	Verbo	76	el	Determinante
24	con	Adposición	77	es	Verbo
25	alegría	Sustantivo	78	Gaspar	Sustantivo

26	y	Conjunción	79	y	Conjunción
27	felicidad	Sustantivo	80	el	Determinante
28	tu	Pronombre	81	otro	Pronombre
29	premio	Sustantivo	82	va	Verbo
30	por	Adposición	83	saltar	Verbo
31	ser	Verbo	84	Quiero	Verbo
32	tan	Adverbio	85	un	Determinante
33	bueno	Adjetivo		trenecit	
34	bueno	Adjetivo	86	o	Sustantivo
35	con	Adposición	87	que	Pronombre
36	tus	Pronombre	88	me	Pronombre
37	Papás	Sustantivo	89	diga	Verbo
38	El	Determinante	90	así	Adverbio
39	es	Verbo	91	Pu	Sustantivo
40	Melchor	Sustantivo	92	pu	Otros
41	,	Puntuación	93	y	Conjunción
42	el	Determinante	94	también	Adverbio
43	es	Verbo	95	una	Determinante
44	Gaspar	Sustantivo	96	muñeca	Sustantivo
45	y	Conjunción	97	ideal	Adjetivo
46	el	Determinante	98	que	Pronombre
47	otro	Determinante	99	siempre	Adverbio
48	Baltazar	Sustantivo	100	me	Pronombre
49	El	Determinante	101	diga	Verbo
50	es	Verbo	102	así	Adverbio
51	Melchor	Sustantivo	103	Mama	Sustantivo
52	,	Puntuación			
53	el	Determinante			

Fuente: Elaboración propia.

**Tabla 9. Etiquetado gramatical del texto de la letra de la canción 3.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	Yo	Pronombre	49	es	Verbo
2	sé	Verbo	50	la	Determinante
3	bien	Adverbio	51	ley	Sustantivo
4	que	Adposición	52	No	Adverbio
5	estoy	Verbo	53	tengo	Verbo

6	afuera	Adverbio	54	trono	Sustantivo
7	Pero	Conjunción	55	ni	Conjunción
8	el	Determinante	56	reina	Sustantivo
9	día	Sustantivo	57	Ni	Adposición
10	que	Pronombre	58	nadie	Pronombre
11	yo	Pronombre	59	que	Pronombre
12	me	Pronombre	60	me	Pronombre
13	muera	Verbo		comprend	
14	Sé	Verbo	61	a	Verbo
15	que	Pronombre	62	Pero	Conjunción
16	tendrás	Verbo	63	siglo	Verbo
17	que	Adposición	64	siendo	Verbo
18	llorar	Verbo	65	el	Determinante
19	Dirás	Verbo	66	rey	Sustantivo
20	que	Pronombre	67	Una	Determinante
21	no	Adverbio	68	piedra	Sustantivo
22	me	Pronombre	69	del	Adposición
23	quisiste	Verbo	70	camino	Sustantivo
24	Pero	Conjunción	71	Me	Pronombre
25	vas	Verbo	72	enseño	Verbo
26	a	Adposición	73	que	Adposición
27	estar	Verbo	74	mi	Pronombre
28	muy	Adverbio	75	destino	Sustantivo
29	triste	Adjetivo	76	Era	Verbo
30	Y	Conjunción	77	rodar	Verbo
31	así	Adverbio	78	y	Conjunción
32	te	Pronombre	79	rodar	Verbo
33	vas	Verbo	80	Después	Adverbio
34	a	Adposición	81	me	Pronombre
35	quedar	Verbo	82	dijo	Verbo
36	Con	Adposición	83	un	Determinante
37	dinero	Sustantivo	84	arriero	Sustantivo
38	y	Conjunción	85	Que	Pronombre
39	sin	Adposición	86	no	Adverbio
40	dinero	Sustantivo	87	hay	Verbo
41	Hago	Verbo	88	que	Adposición
42	siempre	Adverbio	89	llegar	Verbo
43	lo	Determinante	90	primero	Adverbio
44	que	Pronombre	91	Pero	Conjunción
			92	hay	Verbo

45	quiero	Verbo	93	que	Adposición
46	Y	Conjunción	94	saber	Verbo
47	mi	Pronombre	95	llegar	Verbo
48	palabra	Sustantivo			

Fuente: Elaboración propia.

**Tabla 10. Etiquetado gramatical del texto de la letra de la canción 4.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	El	Determinante	67	fueras	Verbo
2	color	Sustantivo	68	El	Determinante
3	de	Adposición	69	sueño	Sustantivo
4	tus	Pronombre	70	que	Pronombre
5	ojos	Sustantivo	71	se	Pronombre
6	Despertó	Verbo	72	vuelve	Verbo
7	mi	Pronombre	73	realidad	Sustantivo
8	interés	Sustantivo	74	Me	Pronombre
9	Y	Conjunción	75	gustas	Verbo
10	sólo	Adverbio	76	tanto	Adverbio
11	tengo	Verbo	77	y	Conjunción
12	ganas	Sustantivo	78	eso	Determinante
13	de	Adposición	79	es	Verbo
14	Verte	Verbo	80	toda	Determinante
15	otra	Determinante	81	la	Determinante
16	vez	Sustantivo	82	verdad	Sustantivo
17	Di	Verbo	83	Me	Pronombre
18	me	Pronombre	84	siento	Verbo
19	que	Adposición	85	emocionad	Verbo
20	no	Adverbio	86	o	Verbo
21	está	Verbo	87	No	Adverbio
22	prohibido	Verbo	88	sé	Verbo
23	Quizás	Adverbio	89	si	Adposición
24	me	Pronombre	90	te	Pronombre
25	animo	Verbo	91	ha	Verbo
26	y	Conjunción	92	pasado	Verbo
27	te	Pronombre	93	Que	Adposición
28	pido	Verbo	94	si	Adposición
29	ver	Verbo	95	podiera	Verbo
				te	Pronombre

30	te	Pronombre	96	viera	Verbo
31	el	Determinante	97	De	Adposición
32	sábado	Sustantivo	98	lunes	Sustantivo
33	a	Adposición	99	a	Adposición
34	las	Determinante	100	domingo	Sustantivo
35	diez	Número cardinal	101	sin	Adposición
36	El	Determinante	102	parar	Verbo
37	color	Sustantivo	103	Esto	Determinante
38	de	Adposición	104	que	Pronombre
39	tus	Pronombre	105	siento	Verbo
40	ojos	Sustantivo	106	no	Adverbio
41	Se	Pronombre	107	se	Pronombre
42	robó	Verbo	108	puede	Verbo
43	mi	Pronombre	109	comparar	Verbo
44	atención	Sustantivo	110	Y	Conjunción
45	Te	Pronombre	111	si	Adposición
46	vas	Verbo	112	ves	Verbo
47	metiendo	Verbo	113	que	Adposición
48	dentro	Adverbio	114	me	Pronombre
49	de	Adposición	115	sonrojo	Verbo
50	mi	Pronombre	116	Si	Adposición
51	corazón	Sustantivo	117	te	Pronombre
52	Perfecta	Adjetivo	118	burlas	Sustantivo
53	en	Adposición	119	,	Puntuación
54	cualquier	Determinante	120	no	Adverbio
55	sentido	Sustantivo	121	me	Pronombre
56	Con	Adposición	122	enojo	Verbo
57	pantalón	Sustantivo	123	Yo	Pronombre
58	o	Conjunción	124	sólo	Adverbio
59	vestido	Verbo	125	sé	Verbo
60	robas	Adjetivo	126	Que	Adposición
61	mi respiración	Pronombre	127	de	Adposición
62	n	Sustantivo	128	ti	Pronombre
63	Que	Pronombre	129	me	Pronombre
64	más	Adverbio	130	enamorado	Verbo
65	quisiera	Verbo			
66	que	Adposición			

Fuente: Elaboración propia.

## 4.2 Filtrado de datos

En las Tablas Tabla 11, Tabla 12, Tabla 13 y Tabla 14 se muestra que al haber etiquetado gramaticalmente las palabras de las letras de las canciones se filtraron y extrajeron las palabras que emitieron emociones y sentimientos que son las correspondientes a sustantivos, verbos, adjetivos y adverbios (Mohammad & Turney, 2013).

**Tabla 11. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 1.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	Hay	Verbo	34	morir	Verbo
2	veces	Sustantivo	35	olvidar	Verbo
3	no	Adverbio	36	materia	Sustantivo
4	tengo	Verbo	37	no	Adverbio
5	ganas	Sustantivo	38	atrevo	Verbo
6	ver	Verbo	39	Hay	Verbo
7	Hay	Verbo	40	veces	Sustantivo
8	veces	Sustantivo	41	no	Adverbio
9	no	Adverbio	42	sé	Verbo
10	quiero	Verbo	43	pasa	Verbo
11	tocar	Verbo	44	Ya	Adverbio
12	Hay	Verbo	45	no	Adverbio
13	veces	Sustantivo	46	puedo	Verbo
14	quisiera	Verbo	47	saber	Verbo
15	ahogar	Verbo	48	es	Verbo
16	grito	Sustantivo	49	pasa	Verbo
17	olvidar	Verbo	50	adentro	Adverbio
18	imagen	Sustantivo	51	Somos	Verbo
19	tuya	Adjetivo	52	gatos	Sustantivo
20	no	Adverbio	53	celo	Sustantivo
21	atrevo	Verbo	54	Somos	Verbo
22	Hay	Verbo	55	célula	Sustantivo
23	veces	Sustantivo	56	explota	Verbo
24	no	Adverbio	57	no	Adverbio
25	dejo	Verbo	58	paras	Sustantivo

26	soñar	Verbo	59	no	Adverbio
27	acariciar	Verbo	60	no	Adverbio
28	ya	Adverbio	61	paras	Sustantivo
29	no	Adverbio	62	No	Adverbio
30	pueda	Verbo	63	no	Adverbio
31	Hay	Verbo			
32	veces	Sustantivo			
33	quisiera	Verbo			

Fuente: Elaboración propia.

**Tabla 12. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 2.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	buen	Adjetivo	31	va	Verbo
2	día	Sustantivo	32	saltar	Verbo
3	comenzó	Verbo	33	es	Verbo
4	historia	Sustantivo	34	Melchor	Sustantivo
5	describe	Verbo	35	es	Verbo
6	así	Adverbio	36	Gaspar	Sustantivo
7	llegaron	Verbo	37	Baltazar	Sustantivo
8	Reyes	Sustantivo	38	es	Verbo
9	Magos	Sustantivo	39	Melchor	Sustantivo
10	saludar	Verbo	40	es	Verbo
11	juguetes	Sustantivo	41	Gaspar	Sustantivo
12	han	Verbo	42	va	Verbo
13	traído	Verbo	43	saltar	Verbo
14	alegría	Sustantivo	44	Quiero	Verbo
15	felicidad	Sustantivo	45	trenecito	Sustantivo
16	premio	Sustantivo	46	diga	Verbo
17	ser	Verbo	47	así	Adverbio
18	tan	Adverbio	48	Pu	Sustantivo
19	bueno	Adjetivo	49	también	Adverbio
20	bueno	Adjetivo	50	muñeca	Sustantivo
21	Papás	Sustantivo	51	ideal	Adjetivo
22	es	Verbo	52	siempre	Adverbio
23	Melchor	Sustantivo	53	diga	Verbo
24	es	Verbo	54	así	Adverbio
25	Gaspar	Sustantivo	55	Mama	Sustantivo
26	Baltazar	Sustantivo			
27	es	Verbo			

28	Melchor	Sustantivo
29	es	Verbo
30	Gaspar	Sustantivo

Fuente: Elaboración propia.

**Tabla 13. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 3.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	bien	Adverbio	26	No	Adverbio
2	estoy	Verbo	27	tengo	Verbo
3	afuera	Adverbio	28	trono	Sustantivo
4	día	Sustantivo	29	reina	Sustantivo
5	muera	Verbo	30	comprenda	Verbo
6	tendrás	Verbo	31	sigo	Verbo
7	llorar	Verbo	32	siendo	Verbo
8	Dirás	Verbo	33	rey	Sustantivo
9	no	Adverbio	34	piedra	Sustantivo
10	quisiste	Verbo	35	camino	Sustantivo
11	vas	Verbo	36	enseñó	Verbo
12	estar	Verbo	37	destino	Sustantivo
13	muy	Adverbio	38	Era	Verbo
14	triste	Adjetivo	39	rodar	Verbo
15	así	Adverbio	40	rodar	Verbo
16	vas	Verbo	41	Después	Adverbio
17	quedar	Verbo	42	dijo	Verbo
18	dinero	Sustantivo	43	arriero	Sustantivo
19	dinero	Sustantivo	44	no	Adverbio
20	Hago	Verbo	45	hay	Verbo
21	siempre	Adverbio	46	llegar	Verbo
22	quiero	Verbo	47	primero	Adverbio
23	palabra	Sustantivo	48	hay	Verbo
24	es	Verbo	49	saber	Verbo
25	ley	Sustantivo	50	llegar	Verbo

Fuente: Elaboración propia.

**Tabla 14. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del texto de la letra de la canción 4.**

Número	Palabra	Etiquetado gramatical	Número	Palabra	Etiquetado gramatical
1	color	Sustantivo	34	quisiera	Verbo
2	ojos	Sustantivo	35	fueras	Verbo
3	Despertó	Verbo	36	sueño	Sustantivo
4	interés	Sustantivo	37	vuelve	Verbo
5	sólo	Adverbio	38	realidad	Sustantivo
6	tengo	Verbo	39	gustas	Verbo
7	ganas	Sustantivo	40	tanto	Adverbio
8	Verte	Verbo	41	es	Verbo
9	vez	Sustantivo	42	verdad	Sustantivo
10	Di	Verbo	43	siento emocionad	Verbo
11	no	Adverbio	44	o	Verbo
12	está	Verbo	45	No	Adverbio
13	prohibido	Verbo	46	sé	Verbo
14	Quizás	Adverbio	47	ha	Verbo
15	animo	Verbo	48	pasado	Verbo
16	pido	Verbo	49	pudiera	Verbo
17	ver	Verbo	50	viera	Verbo
18	sábado	Sustantivo	51	lunes	Sustantivo
19	color	Sustantivo	52	domingo	Sustantivo
20	ojos	Sustantivo	53	parar	Verbo
21	robó	Verbo	54	siento	Verbo
22	atención	Sustantivo	55	no	Adverbio
23	vas	Verbo	56	puede	Verbo
24	metiendo	Verbo	57	comparar	Verbo
25	dentro	Adverbio	58	ves	Verbo
26	corazón	Sustantivo	59	sonrojo	Verbo
27	Perfecta	Adjetivo	60	burlas	Sustantivo
28	sentido	Sustantivo	61	no	Adverbio
29	pantalón	Sustantivo	62	enojo	Verbo
30	vestido	Verbo	63	sólo	Adverbio
31	robas	Adjetivo	64	sé	Verbo
	respiración				
32	n	Sustantivo	65	enamorado	Verbo
33	más	Adverbio			

Fuente: Elaboración propia.

### 4.3 Comparación de datos

En la Tablas Tabla 15, Tabla 16, Tabla 17 y Tabla 18 se muestra el resultado de la consulta de la comparación de las coincidencias de las palabras filtradas y los unigramas de la base de datos léxica Emolex por emociones y sentimientos del texto de las letras de las canciones de la muestra.

**Tabla 15. Coincidencias de las palabras filtradas del texto de la letra de la canción 1 y los unigramas de la base de datos léxica.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
ahogar	1	0	1	0	1	0	0	0	0	1
celo	0	0	0	1	0	1	1	1	1	0
grito	1	0	1	1	1	0	0	0	0	1
morir	0	0	1	0	1	0	0	0	0	1
no	0	0	0	0	0	0	0	1	0	1
olvidar	0	0	0	0	0	0	0	0	0	1

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,

Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

**Tabla 16. Coincidencias de las palabras filtradas del texto de la letra de la canción 2 y los unigramas de la base de datos léxica.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
alegría	0	0	0	0	0	0	1	0	1	0
bueno	0	0	0	1	0	1	1	1	1	0
felicidad	0	0	0	0	0	0	1	1	1	0
mamá	0	0	0	0	0	1	0	0	0	0
muñeca	0	0	0	0	0	0	1	0	0	0
premio	0	0	0	1	0	1	1	1	1	0
saltar	0	0	0	0	0	0	1	0	1	0

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,

Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

**Tabla 17. Coincidencias de las palabras filtradas del texto de la letra de la canción 3 y los unigramas de la base de datos léxica.**

Palabra	Emociones							Sentimientos		
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
destino	0	0	1	1	1	0	1	1	1	0
dinero	1	0	0	1	0	1	1	1	1	0
enseñó	0	0	0	0	0	1	0	0	0	0
ley	0	0	0	0	0	1	0	0	0	0
llegar	0	0	0	0	0	0	0	1	0	0
llorar	0	0	1	0	0	0	0	0	0	1
no	0	0	0	0	0	0	0	1	0	1
palabra	0	0	0	0	0	1	0	0	1	0
piedra	1	0	0	0	0	0	0	0	0	1
rey	0	0	0	0	0	0	0	0	1	0
triste	0	1	1	0	1	0	0	0	0	1
trono	0	0	0	0	0	1	0	0	1	0

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,

Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

**Tabla 18. Coincidencias de las palabras filtradas del texto de la letra de la canción 4 y los unigramas de la base de datos léxica.**

Palabra	Emociones							Sentimientos		
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
ánimo	0	0	0	0	0	1	1	1	0	1
atención	0	0	0	0	0	0	0	0	0	1
burlas	1	0	0	0	1	0	0	0	1	0
emocionado	0	0	0	1	0	1	1	1	0	1
no	0	0	0	0	0	0	0	1	1	0
prohibido	1	1	0	0	1	0	0	0	1	0
robo	1	1	1	0	1	0	0	0	1	0
sentido	0	0	0	0	0	0	0	0	0	1
sonrojo	0	0	0	0	0	0	0	0	1	0
verdad	0	0	0	0	0	1	0	0	0	1

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,  
 Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

#### 4.4 Análisis de emociones y sentimientos

Se determinó la valencia de cada emoción y sentimiento del texto de la letra de la canción 1 sumando las columnas de acuerdo con su valor binario, como se muestra en la Tabla 19.

**Tabla 19. Valencias de emociones y sentimientos del texto de la letra de la canción 1.**

Palabra	Emociones						Sentimientos			
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
ahogar	1	0	1	0	1	0	0	0	0	1
celo	0	0	0	1	0	1	1	1	1	0
grito	1	0	1	1	1	0	0	0	0	1
morir	0	0	1	0	1	0	0	0	0	1
no	0	0	0	0	0	0	0	1	0	1
olvidar	0	0	0	0	0	0	0	0	0	1
Valencia	2	0	3	2	3	1	1	2	1	5

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,  
 Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Al obtener la valencia de cada emoción y sentimiento se calculó la valencia total de emociones sumando la valencia de cada emoción y la valencia total de sentimientos sumando la valencia de cada sentimiento.

- Valencia total de emociones =  $2 + 0 + 3 + 2 + 3 + 1 + 1 + 2 = 14$
- Valencia total de sentimientos =  $1 + 5 = 6$

Para calcular el porcentaje de cada emoción se dividió la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira =  $2 / 14 = 0.14$

- Porcentaje disgusto =  $0 / 14 = 0.00$
- Porcentaje tristeza =  $3 / 14 = 0.21$
- Porcentaje sorpresa =  $2 / 14 = 0.14$
- Porcentaje miedo =  $3 / 14 = 0.21$
- Porcentaje confianza =  $1 / 14 = 0.07$
- Porcentaje alegría =  $1 / 14 = 0.07$
- Porcentaje anticipación =  $2 / 14 = 0.14$

Para calcular el porcentaje de cada sentimiento se dividió la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo =  $1 / 6 = 0.17$
- Porcentaje negativo =  $5 / 6 = 0.83$

En la Tabla 20 se muestra el resultado del análisis de emociones y sentimientos que forma parte del texto de la letra de la canción 1 y sus porcentajes de cada emoción y sentimiento para su posterior clasificación.

**Tabla 20. Porcentaje de emociones y sentimientos del texto de la letra de la canción 1.**

Palabra	Emociones							Sentimientos		
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
ahogar	1	0	1	0	1	0	0	0	0	1
celo	0	0	0	1	0	1	1	1	1	0
grito	1	0	1	1	1	0	0	0	0	1
morir	0	0	1	0	1	0	0	0	0	1
no	0	0	0	0	0	0	0	1	0	1
olvidar	0	0	0	0	0	0	0	0	0	1
Valencia	2	0	3	2	3	1	1	2	1	5
Porcentaje	0.14	0.00	0.21	0.14	0.21	0.07	0.07	0.14	0.17	0.83

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza, Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Se determinó la valencia de cada emoción y sentimiento del texto de la letra de la canción 2 sumando las columnas de acuerdo con su valor binario, como se muestra en la Tabla 21.

**Tabla 21. Valencias de emociones y sentimientos del texto de la letra de la canción 2.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
alegría	0	0	0	0	0	0	1	0	1	0
bueno	0	0	0	1	0	1	1	1	1	0
felicidad	0	0	0	0	0	0	1	1	1	0
mamá	0	0	0	0	0	1	0	0	0	0
muñeca	0	0	0	0	0	0	1	0	0	0
premio	0	0	0	1	0	1	1	1	1	0
saltar	0	0	0	0	0	0	1	0	1	0
Valencia	0	0	0	2	0	3	6	3	5	0

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,

Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Al obtener la valencia de cada emoción y sentimiento se calculó la valencia total de emociones sumando la valencia de cada emoción y la valencia total de sentimientos sumando la valencia de cada sentimiento.

- Valencia total de emociones =  $0 + 0 + 0 + 2 + 0 + 3 + 6 + 3 = 14$
- Valencia total de sentimientos =  $5 + 0 = 5$

Para calcular el porcentaje de cada emoción se dividió la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira =  $0 / 14 = 0.00$
- Porcentaje disgusto =  $0 / 14 = 0.00$
- Porcentaje tristeza =  $0 / 14 = 0.00$
- Porcentaje sorpresa =  $2 / 14 = 0.14$
- Porcentaje miedo =  $0 / 14 = 0.00$

- Porcentaje confianza =  $3 / 14 = 0.21$
- Porcentaje alegría =  $6 / 14 = 0.43$
- Porcentaje anticipación =  $3 / 14 = 0.21$

Para calcular el porcentaje de cada sentimiento se dividió la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo =  $5 / 5 = 1.00$
- Porcentaje negativo =  $0 / 5 = 0.00$

En la Tabla 22 se muestra el resultado del análisis de emociones y sentimientos que forma parte del texto de la letra de la canción 2 y sus porcentajes de cada emoción y sentimiento para su posterior clasificación.

**Tabla 22. Porcentaje de emociones y sentimientos del texto de la letra de la canción 2.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
alegría	0	0	0	0	0	0	1	0	1	0
bueno	0	0	0	1	0	1	1	1	1	0
felicidad	0	0	0	0	0	0	1	1	1	0
mamá	0	0	0	0	0	1	0	0	0	0
muñeca	0	0	0	0	0	0	1	0	0	0
premio	0	0	0	1	0	1	1	1	1	0
saltar	0	0	0	0	0	0	1	0	1	0
Valencia	0	0	0	2	0	3	6	3	5	0
Porcentaje	0.00	0.00	0.00	0.14	0.00	0.21	0.43	0.21	1.00	0.00

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza, Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Se determinó la valencia de cada emoción y sentimiento del texto de la letra de la canción 3 sumando las columnas de acuerdo con su valor binario, como se muestra en la Tabla 23.

**Tabla 23. Valencias de emociones y sentimientos del texto de la letra de la canción 3.**

Palabra	Emociones							Sentimientos		
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
destino	0	0	1	1	1	0	1	1	1	0
dinero	1	0	0	1	0	1	1	1	1	0
enseñó	0	0	0	0	0	1	0	0	0	0
ley	0	0	0	0	0	1	0	0	0	0
llegar	0	0	0	0	0	0	0	1	0	0
llorar	0	0	1	0	0	0	0	0	0	1
no	0	0	0	0	0	0	0	1	0	1
palabra	0	0	0	0	0	1	0	0	1	0
piedra	1	0	0	0	0	0	0	0	0	1
rey	0	0	0	0	0	0	0	0	1	0
triste	0	1	1	0	1	0	0	0	0	1
trono	0	0	0	0	0	1	0	0	1	0
Valencia	2	1	3	2	2	5	2	4	5	4

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza, Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Al obtener la valencia de cada emoción y sentimiento se calculó la valencia total de emociones sumando la valencia de cada emoción y la valencia total de sentimientos sumando la valencia de cada sentimiento.

- Valencia total de emociones =  $2 + 1 + 3 + 2 + 2 + 5 + 2 + 4 = 21$
- Valencia total de sentimientos =  $5 + 4 = 9$

Para calcular el porcentaje de cada emoción se dividió la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira =  $2 / 21 = 0.10$
- Porcentaje disgusto =  $1 / 21 = 0.05$
- Porcentaje tristeza =  $3 / 21 = 0.14$
- Porcentaje sorpresa =  $2 / 21 = 0.10$
- Porcentaje miedo =  $2 / 21 = 0.10$

- Porcentaje confianza =  $5 / 21 = 0.24$
- Porcentaje alegría =  $2 / 21 = 0.10$
- Porcentaje anticipación =  $4 / 21 = 0.19$

Para calcular el porcentaje de cada sentimiento se dividió la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo =  $5 / 9 = 0.56$
- Porcentaje negativo =  $4 / 9 = 0.44$

En la Tabla 24 se muestra el resultado del análisis de emociones y sentimientos que forma parte del texto de la letra de la canción 3 y sus porcentajes de cada emoción y sentimiento para su posterior clasificación.

**Tabla 24. Porcentaje de emociones y sentimientos del texto de la letra de la canción 3.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
destino	0	0	1	1	1	0	1	1	1	0
dinero	1	0	0	1	0	1	1	1	1	0
enseñó	0	0	0	0	0	1	0	0	0	0
ley	0	0	0	0	0	1	0	0	0	0
llegar	0	0	0	0	0	0	0	1	0	0
llorar	0	0	1	0	0	0	0	0	0	1
no	0	0	0	0	0	0	0	1	0	1
palabra	0	0	0	0	0	1	0	0	1	0
piedra	1	0	0	0	0	0	0	0	0	1
rey	0	0	0	0	0	0	0	0	1	0
triste	0	1	1	0	1	0	0	0	0	1
trono	0	0	0	0	0	1	0	0	1	0
Valencia	2	1	3	2	2	5	2	4	5	4
Porcentaje	0.10	0.05	0.14	0.10	0.10	0.24	0.10	0.19	0.56	0.44

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,

Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Se determinó la valencia de cada emoción y sentimiento del texto de la letra de la canción 4 sumando las columnas de acuerdo con su valor binario, como se muestra en la Tabla 25.

**Tabla 25. Valencias de emociones y sentimientos del texto de la letra de la canción 4.**

Palabra	Emociones								Sentimientos	
	Ir	Di	Tr	So	Mi	Co	Al	An	Po	Ne
ánimo	0	0	0	0	0	1	1	1	0	1
atención	0	0	0	0	0	0	0	0	0	1
burlas	1	0	0	0	1	0	0	0	1	0
emocionado	0	0	0	1	0	1	1	1	0	1
no	0	0	0	0	0	0	0	1	1	0
prohibido	1	1	0	0	1	0	0	0	1	0
robo	1	1	1	0	1	0	0	0	1	0
sentido	0	0	0	0	0	0	0	0	0	1
sonrojo	0	0	0	0	0	0	0	0	1	0
verdad	0	0	0	0	0	1	0	0	0	1
Valencia	3	2	1	1	3	3	2	3	5	5

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,

Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

Al obtener la valencia de cada emoción y sentimiento se calculó la valencia total de emociones sumando la valencia de cada emoción y la valencia total de sentimientos sumando la valencia de cada sentimiento.

- Valencia total de emociones =  $3 + 2 + 1 + 1 + 3 + 3 + 2 + 3 = 18$
- Valencia total de sentimientos =  $5 + 5 = 10$

Para calcular el porcentaje de cada emoción se dividió la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira =  $3 / 18 = 0.17$
- Porcentaje disgusto =  $2 / 18 = 0.11$

- Porcentaje tristeza =  $1 / 18 = 0.06$
- Porcentaje sorpresa =  $1 / 18 = 0.06$
- Porcentaje miedo =  $3 / 18 = 0.17$
- Porcentaje confianza =  $3 / 18 = 0.17$
- Porcentaje alegría =  $2 / 18 = 0.11$
- Porcentaje anticipación =  $3 / 18 = 0.17$

Para calcular el porcentaje de cada sentimiento se dividió la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo =  $5 / 10 = 0.50$
- Porcentaje negativo =  $5 / 10 = 0.50$

En la Tabla 26 se muestra el resultado del análisis de emociones y sentimientos que forma parte del texto de la letra de la canción 4 y sus porcentajes de cada emoción y sentimiento para su posterior clasificación.

**Tabla 26. Porcentaje de emociones y sentimientos del texto de la letra de la canción 4.**

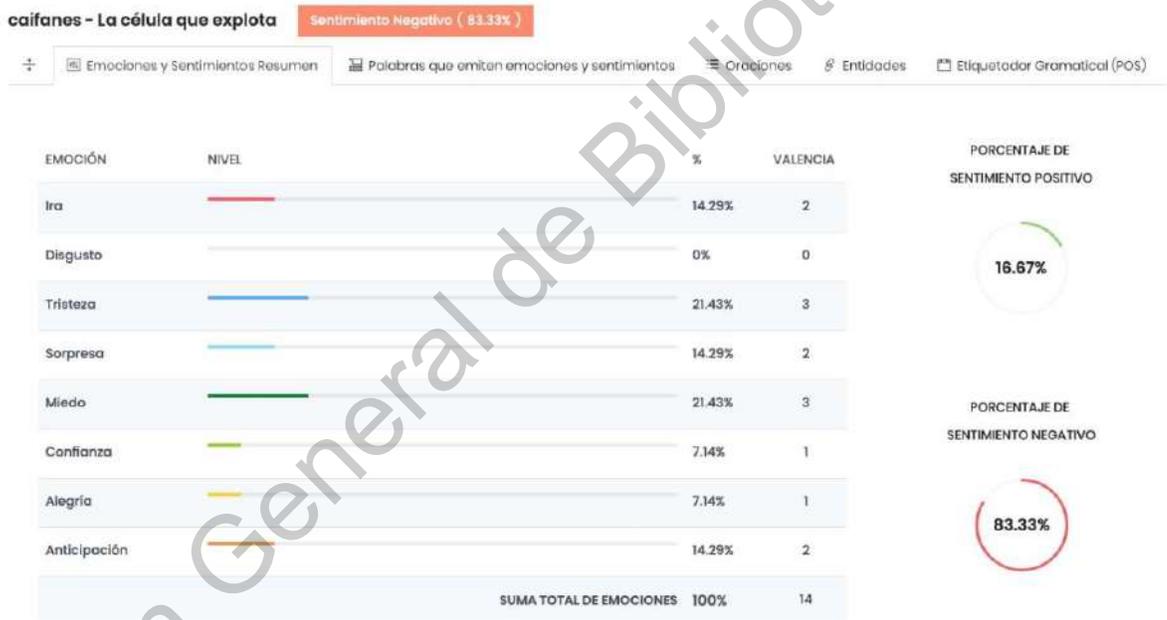
Palabra	Emociones								Sentimientos	
	Ir	Dí	Tr	So	Mi	Co	Al	An	Po	Ne
ánimo	0	0	0	0	0	1	1	1	0	1
atención	0	0	0	0	0	0	0	0	0	1
burlas	1	0	0	0	1	0	0	0	1	0
emocionado	0	0	0	1	0	1	1	1	0	1
no	0	0	0	0	0	0	0	1	1	0
prohibido	1	1	0	0	1	0	0	0	1	0
robo	1	1	1	0	1	0	0	0	1	0
sentido	0	0	0	0	0	0	0	0	0	1
sonrojo	0	0	0	0	0	0	0	0	1	0
verdad	0	0	0	0	0	1	0	0	0	1
Valencia	3	2	1	1	3	3	2	3	5	5
Porcentaje	0.17	0.11	0.06	0.06	0.17	0.17	0.11	0.17	0.50	0.50

Ir = Ira, Di = Disgusto, Tr = Tristeza, So = Sorpresa, Mi = Miedo, Co = Confianza,  
 Al = Alegría, An = Anticipación, Po = Positivo, Ne = Negativo.

Fuente: Elaboración propia.

#### 4.5 Clasificación de las letras de las canciones por su polaridad sentimental

La clasificación y el orden se establece de acuerdo con el análisis de emociones y sentimientos que se realizó al texto de las letras de las canciones con el resultado del porcentaje de polaridad positiva y negativa, los más negativos primero y así sucesivamente, en un orden descendente como se muestra en las Figuras Figura 2, Figura 3, Figura 4 y Figura 5.



**Figura 2. Clasificación del análisis de emociones y sentimientos de la letra de la canción 1.**

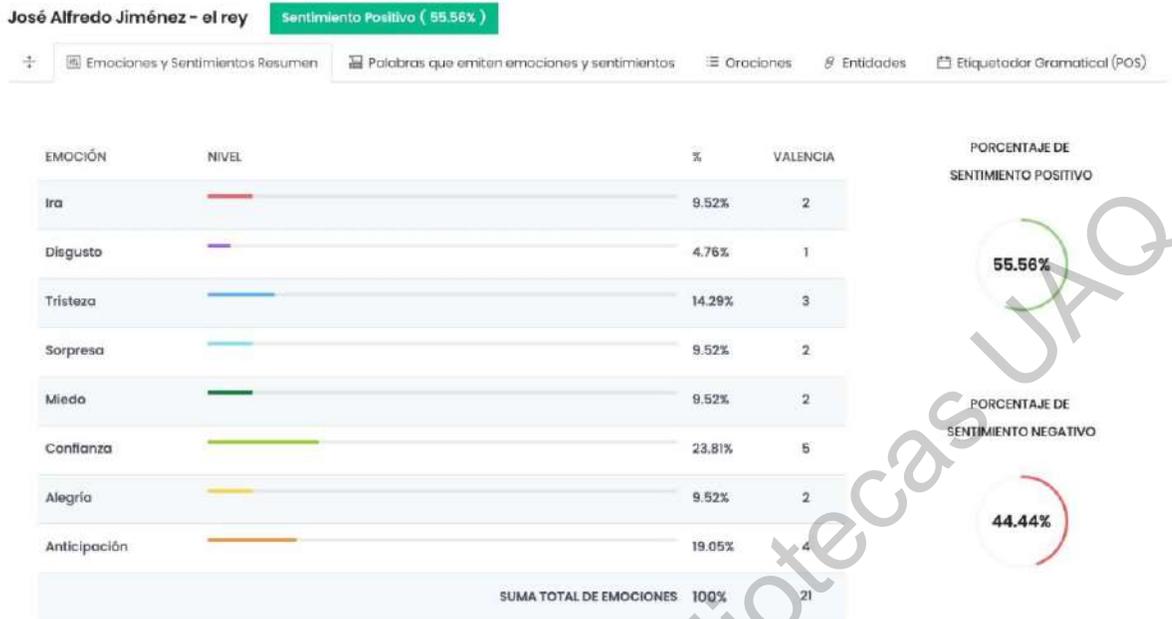
Fuente: Elaboración propia.



**Figura 3. Clasificación del análisis de emociones y sentimientos de la letra de la canción 4.**

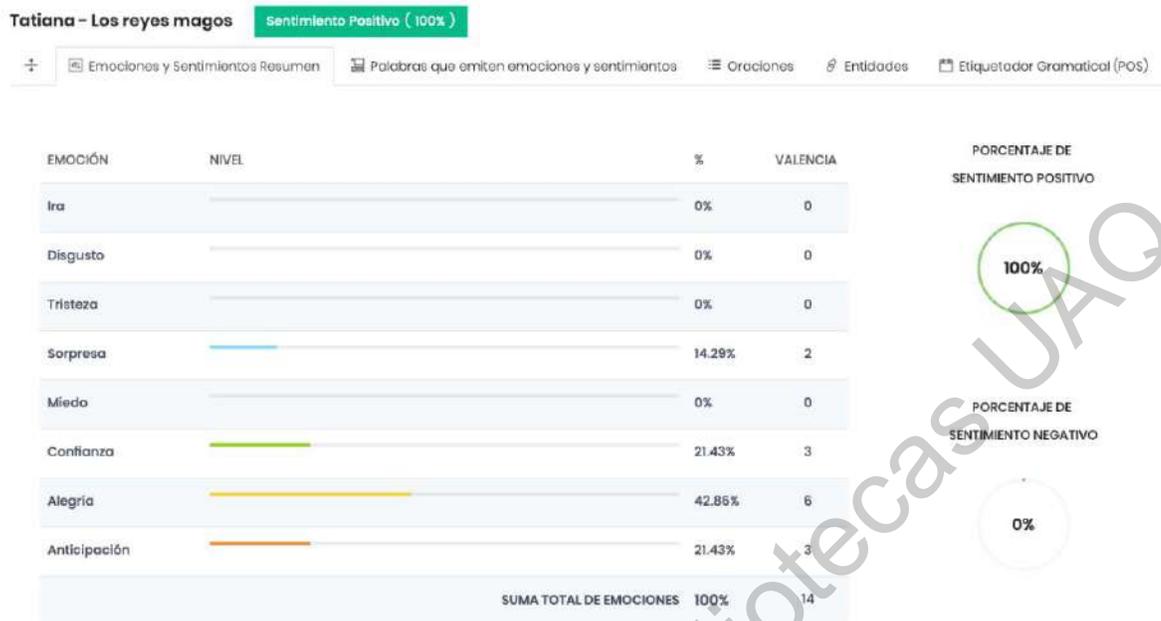
Fuente: Elaboración propia.

Dirección General de Bibliotecas UAQ



**Figura 4. Clasificación del análisis de emociones y sentimientos de la letra de la canción 3.**

Fuente: Elaboración propia.



**Figura 5. Clasificación del análisis de emociones y sentimientos de la letra de la canción 2.**

Fuente: Elaboración propia.

#### 4.6 Precisión

La evaluación de la precisión del análisis de emociones y sentimientos de las letras de las canciones se determinó dividiendo las canciones clasificadas correctamente entre el total de canciones analizadas.

$$\text{Precisión} = 912 / 1,014 = 0.8994\%$$

- Canciones analizadas correctamente: 912
- Canciones analizadas incorrectamente: 102

### 5. CONCLUSIONES Y TRABAJO FUTURO

El desempeño del algoritmo de procesamiento de lenguaje natural de Google Natural Language Cloud es eficiente al etiquetar gramaticalmente cada palabra de la letra de la canción que forma parte del discurso, dando formato y

estructura al texto de la letra. La base de datos léxica Emolex contiene un vasto diccionario de palabras en español que emiten emociones y sentimientos, de gran importancia para esta investigación.

Cabe mencionar que los recursos mencionados son ajenos una vez que se integran con el algoritmo del analizador es posible: filtrar palabras, compararlas con la base de datos léxica, calcular las valencias y porcentajes de las emociones y sentimientos para la clasificación las letras de las canciones.

En este trabajo de investigación, el etiquetado gramatical se realizó para obtener datos estructurados del texto de las letras de las canciones, el filtrado de datos se realizó para garantizar que se detecten o ignoren los datos irrelevantes o incorrectos. La comparación de datos se realizó mediante una combinación de procesamiento de lenguaje natural y la base de datos léxica. La comparación de datos se realizó para combinar los datos extraídos y obtener datos estructurados. Estos datos estructurados ayudaron a que el proceso de análisis de emociones y sentimientos sea fácil de administrar y analizar, para finalmente obtener la clasificación de las letras de las canciones de acuerdo con su polaridad negativa y positiva.

El porcentaje de efectividad de el análisis de emociones y sentimientos de las letras de las canciones mediante procesamiento de lenguaje natural, la base de datos léxica y el analizador se llevo acabo examinando las 1,014 letras de las canciones cualitativa y cuantitativamente arrojando un 89.94% de precisión.

El procedimiento aplicado para analizar y clasificar este volumen de conjunto de datos se puede utilizar como marco para otras tareas que involucren grandes colecciones de textos (correos electrónicos, chats, mensajes, blogs, redes sociales, etc.), cuyo procesamiento sea impráctico por medios manuales.

Como trabajo futuro la base de datos léxica Emolex se puede extender agregando nuevos unigramas para el análisis de texto en general. Esta adición

permitirá una extracción de conocimiento más amplia para distintos ámbitos y un análisis de emociones y sentimientos más preciso.

## 6. REFERENCIAS

- Carrión, M. (2014). Resolución de anáforas que requieren conocimiento cultural con la herramienta FunGramKB. *Revista de Lingüística y Lenguas Aplicadas*, (9) 1-13.
- Chen, H., & Zimbra, D. (2010). Ai and opinion mining. *IEEE Intelligent Systems*, 74-80.
- Crossley, S. A., Kyle, K., & McNamara, D.S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behav Res*, 49: 803. doi: 10.3758/s13428-016-0743-z
- Cruz, F. L., Troyano, J. A., Enríquez, F., Ortega, F. J. & Vallejo, C. G. (2013). Long autonomy or long delay? The importance of domain in opinion mining. *Expert Systems with Applications*, 3174-3184.
- Edigital. (2019). ¿Quiénes somos? Consultado el 15 de enero del 2019. Edigital Soluciones Tecnológicas SA de CV. Recuperado de <https://www.edigital.com.mx/portal/>
- Feldman, S. (1999). NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval. *ONLINE-WESTON THEN WILTON*, 62-73.
- FrozenFrog Software. (2019). *Emailparser*. Consultado el 9 de febrero del 2019. FrozenFrog Software. Recuperado de <https://www.emailparser.com/>
- Gantz, J.F., Reinsel, D., Chute, C., Schlichting, W., McArthur, J., Minton, S., Xheneti, I., Toncheva, A. & Manfrediz, A. (2008), The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2008, *IDC & EMC*.
- Google. (2019a). *Gmail*. Consultado el 9 de febrero del 2019. Google. Recuperado de <https://www.gmail.com/mail/help/intl/es/about.html?iframe>
- Google. (2019b). *Natural Language Cloud*. Consultado el 14 de febrero del 2019. Google. Recuperado de <https://cloud.google.com/natural-language/?hl=es>
- Hoogendoorn, M., Berger, T., Schulz, A., Stolz, T., & Szolovits P. (2016). Predicting Social Anxiety Treatment Outcome Based on Therapeutic Email Conversations. *IEEE*, 21(5), 1449-1459. doi: 10.1109 / JBHI.2016.2601123

- Hutto, C. J. & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In E. Adar & P. Resnick (Eds.), *Proceedings of the eighth international AAAI conference on weblogs and social media* (pp. 216-225). Palo Alto: AAAI Press.
- Inbenta. (2019). *Descubre si las opiniones de tu cliente son positivas o negativas*. Consultado el 9 de febrero del 2019. Inbenta Technologies Inc. Recuperado de <https://www.inbenta.com/mx/caracteristicas/autoservicio/analisis-de-sentimientos/>
- Jayathilaka, K.M.P.N., Weerasinghe A.R., & Wijsekara W.M.L.K.N. (2016). Making sense of large volumes of unstructured email responses. *IEEE*, 35-40. doi: 10.1109 / ICTER.2016.7829896
- Jiang, J., Chen, J., Choo, K.K.R., Liu, K., Liu C., Yu, M., & Mohapatra P. (2018). Prediction and Detection of Malicious Insiders' Motivation based on Sentiment Profile on Webpages and Emails. *IEEE*, 225-229. doi: 10.1109 / MILCOM.2018.8599790
- Lee J., Thabsuwan C., Pongpaichet S., & Kim K. (2018). Towards Building a Human Perception Knowledge for Social Sensation Analysis. *IEEE*, 668-671. doi: 10.1109/WI.2018.00-15
- Liddy, E. D. (2001). *Natural language processing en Encyclopedia of Library and Information Science*, segunda edición, Nueva York: Marcel Decker, (3), 2126-2136.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5, 1-167.
- Liu, B. (2007). *Web Data Mining. Exploring Hyperlinks, Contents and Usage Data*. Springer, first edition.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- Mahlawi, A., Q. & Sasi, S. (2017). Structured Data Extraction from Emails. *IEEE*, 20-22. doi: 10.1109 / NETACT.2017.8076789
- McCarthy, J. (1959). Programs with Common Sense. In *Mechanisation of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory*, 77-84.
- Microsoft. (2019). *Microsoft Outlook*. Consultado el 9 de febrero del 2019. Microsoft. Recuperado de <https://products.office.com/es-mx/outlook/email-and-calendar-software-microsoft-outlook>
- Miller, R. & Charles, E. (2016). A psychological based analysis of marketing email subject lines. *IEEE*, 58-65. doi: 10.1109/ICTER.2016.7829899

- Min, H. J., & Park, J. C. (2012). Identifying helpful reviews based on customers mentions about experiences. *Expert Systems with Applications*, 39(15), 11830-11838.
- Mohammad, S. & Turney, P. (2013). *Computational Intelligence*, 29 (3), 436-465.
- Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in tweets. Special Section of the *ACM Transactions on Internet Technology on Argumentation in Social Media*, 17(3), 1-19. doi: 10.1145/3003433
- Mohammad, S. M., Marquez, F., Mohammad, S., & Kiritchenko S. (2018). *Semeval 2018 Task 1: Affect in tweets*. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA. 1-17. doi: 10.18653/v1/S18-1001
- Pang, B. & Lee, L. (2008). *Opinion mining and sentiment analysis*. *Foundations and Trends in Information Retrieval*, 2, 1-135.
- Pang, B., Lillian L., & Shivakumar V. (2002). Thumbs up Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 79-86.
- Qiong W., & Songbo T. (2011). A two stage framework for cross-domain sentiment classification. *Expert Systems with Applications*, 14269-14275.
- Rushdi-Saleh, M., Martín-Validvia, M. T., Montejo-Ráez, A. & Ureña-López, L.A. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*, 14799-14804.
- Singh, J., Singh, G.S., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human Centric Computing and Information Sciences*, 7: 32. doi: 10.1186/s13673-017-0116-3
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics -ACL 02*, 417, Philadelphia, Pennsylvania.
- Vázquez, S., A. Montoyo, & G. Rigau. (2003). Método de desambiguación léxica basada en el recurso léxico: dominios relevantes. *Procesamiento del Lenguaje Natural*, 141-148.
- Villena, J. (2015). *Introducción al análisis de sentimientos (minería de opiniones)*. Consultado el 17 de mayo del 2018. Meaningcloud. Recuperado de <https://www.meaningcloud.com/es/blog/introduccion-al-analisis-de-sentimientos-mineria-de-opinion>
- Zhang, H., Gan, W., & Jiang, B. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. In X. Yuan & X. Meng (Eds.),

*Proceedings of the 11th Web information system and application conference* (pp. 262 - 265). Piscataway: IEEE Press.

## **7. APÉNDICE**

En el desarrollo del trabajo de investigación se publicó un artículo en la revista salud y administración de la Universidad de la Sierra Sur de Oaxaca en México con ISSN en papel 2007-7971, en electrónico 2448-6159 y Latindex 21940, con el título de: “Clasificador de urgencia de correos electrónicos en empresas de servicios mediante análisis de emociones y sentimientos”.

Dirección General de Bibliotecas UNAO



**UNIVERSIDAD DE LA SIERRA SUR**  
Revista Salud y Administración  
ISSN papel 2007-7971 / ISSN electrónico 2448-6159

**Oficio UNSIS/SA/005/2019**

Juan Carlos Mancera-Barrera  
Alberto Lamadrid-Álvarez  
Alberto Lara-Guevara

La presente es para comunicarles que su escrito **“Clasificador de Urgencia de Correos Electrónicos en Empresas de Servicios mediante Análisis de Emociones y Sentimientos”** ha sido aceptado para su publicación en la Revista *Salud y Administración* de la Universidad de la Sierra Sur. El artículo aparecerá próximamente publicado en el **número 16** de la revista mencionada, **correspondiente a enero-abril 2019**, en la sección de **Artículos Originales**.

A nombre del Comité Editorial, les agradecemos haber enviado su trabajo a nuestra revista y les invitamos a hacernos llegar nuevas aportaciones.

**Atentamente**  
*“DocendoDiscimus”*  
Iur rluuaria rsēedaa  
Miahuatlán de Porfirio Díaz, Oaxaca  
3 de julio de 2019

M.C. Teresita de Jesús Mijangos Martínez  
Directora



c.c.p. Expediente.

---

Guillermo Rojas Mijangos s/n. Esq. Avenida Universidad. Col. Ciudad Universitaria,  
Miahuatlán de Porfirio Díaz, Oaxaca. C.P. 70800 Tel. 951 57 24 100 Ext. 205  
<https://revista.unsis.edu.mx>

## Clasificador de Urgencia de Correos Electrónicos en Empresas de Servicios mediante Análisis de Emociones y Sentimientos

### Emergency Email Sorter in Service Companies by Analysis of Emotions and Feelings

Juan Carlos Mancera-Barrera\*, Alberto Lamadrid-Álvarez\*, Alberto Lara-Guevara\*

#### Resumen

En este trabajo se presenta un clasificador de urgencias de correos electrónicos en empresas de servicios, que mediante un algoritmo de procesamiento de lenguaje natural etiqueta gramaticalmente el texto no estructurado de los mensajes para darle sentido y estructura antes de analizarlo. El clasificador propuesto utiliza una base de datos léxica que está compuesta por unigramas clasificados en ocho emociones básicas (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y dos sentimientos (positivo y negativo), la cual permite comparar el texto del mensaje etiquetado gramaticalmente con los unigramas de la base de datos para llevar a cabo el análisis de emociones y sentimientos y determinar el porcentaje de cada emoción y la polaridad del sentimiento para clasificar los correos electrónicos más negativos como urgentes y canalizarlos a los departamentos correspondientes para ser atendidos. El trabajo experimental se ha realizado utilizando un conjunto de datos tomados de una empresa dedicada al servicio web de facturación electrónica en México.

**Palabras clave:** Análisis de emociones y sentimientos, procesamiento de lenguaje natural, base de datos léxica, empresas de servicios.

#### Abstract

In this paper, a classifier of emails by level of urgency in service companies is presented, using a natural language processing algorithm to grammatically label the unstructured text of messages to give it meaning and structure before analyzing it. The proposed classifier uses a lexical database that is composed of unigrams classified into eight basic emotions (anger, fear, anticipation, confidence, surprise, sadness, joy and disgust) and two feelings (positive and negative). This allows to compare the text of the grammatically labeled messages with the unigrams of the database, in order to conduct the analysis of emotions and feelings. The analysis determines the percentage of each emotion and the polarity of feeling in order to classify the most negative emails as urgent and channel them to the corresponding departments to be attended to. The research has been implemented using a set of data taken from a company dedicated to electronic invoicing in Mexico.

**Keywords:** Analysis of emotions and feelings, natural language processing, lexical database, service companies.

\* Universidad Autónoma de Querétaro, Facultad de Informática, Av. de las Ciencias s/n, Juriquilla, C.P. 76230, Querétaro, Querétaro, México.

Correspondencia: Juan Carlos Mancera Barrera  
Universidad Autónoma de Querétaro  
Correo electrónico: jcmancera@e-digital.com.mx

## Introducción

El lenguaje es una herramienta para expresar los sentimientos a los demás. Los seres humanos han mejorado esta habilidad en forma hablada y escrita durante milenios hasta el punto en que el oyente o el lector puede experimentar los mismos sentimientos que el hablante o el escritor. La psicología ha identificado que diferentes palabras expresan diversas emociones en diferentes grados. Por ejemplo, la palabra "dulce" expresa una emoción positiva y crea un sentimiento de felicidad, mientras que la palabra "matar" crea una emoción negativa (Miller & Charles, 2016).

El correo electrónico (e-correo) ha sido un medio de comunicación efectivo y popular desde su nacimiento en 1971, hasta ahora. La cantidad de usuarios de e-correo ha aumentado de unos pocos a 2672 millones a partir de 2016 y se prevé que aumentará aún más (Miller & Charles, 2016). El e-correo juega un papel importante en las comunicaciones efectivas dentro y fuera de las organizaciones (Mahlawi & Sasi, 2017).

El análisis de sentimientos es un área de investigación popular que utiliza el procesamiento del lenguaje natural (PLN), el análisis de textos y la lingüística computacional para identificar y extraer información subjetiva en los materiales de origen (Zhang et al., 2014; Miller & Charles, 2016). El manejo de grandes volúmenes de datos de texto no estructurados es una tarea desafiante que se está volviendo cada vez más necesaria en una variedad de situaciones (Jayathilaka et al., 2016).

La cantidad de texto no estructurado ha aumentado con la rápida adopción de información digitalizada. Leer y extraer información significativa de grandes colecciones de datos no estructurados es cada vez más una necesidad. Hoy en día, dar sentido a los grandes volúmenes de datos no estructurados se está volviendo esencial (Jayathilaka et al., 2016).

Los datos se están convirtiendo en el foco de atención en los escenarios de negocios actuales debido al espectacular crecimiento de Internet y los almacenamientos en línea. Estos datos pueden ser estructurados, semiestructurados y no estructurados. Los datos estructurados están organizados y se pueden analizar fácilmente. Los datos semiestructurados y no estructurados no están predefinidos y son complejos, lo que hace que obtener conocimiento de ellos sea un gran desafío (Mahlawi & Sasi, 2017).

Tipos de datos estructurados, semiestructurados y no estructurados.

- Datos estructurados: tienen perfectamente definidos la longitud, el formato y el tamaño de sus datos. Se almacenan en formato tabla, hojas de cálculo o en bases de datos relacionales.
- Datos no estructurados: se caracterizan por no tener un formato específico. Se almacenan en múltiples formatos como documentos pdf o word, e-correo, archivos multimedia de imagen, audio o video.
- Datos semiestructurados: son una mezcla de los dos anteriores no presentan una estructura perfectamente definida como los datos estructurados pero sí presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención, como por ejemplo, los formatos html, xml o json.

La mayoría de las organizaciones prefiere trabajar con datos estructurados porque tienen la ventaja de que se pueden almacenar, administrar y analizar fácilmente. Sin embargo, la mayoría de los datos de toma de decisiones de hoy provienen de fuentes externas desestructuradas que representan más del 90% de los datos actuales. La mayoría de estos datos vienen en forma de texto tales como documentos, publicaciones en redes sociales (Lee et al., 2018), e-correos, etc. (Mahlawi & Sasi, 2017).

Extraer el conocimiento de textos no estructurados es difícil, debido a la naturaleza desorganizada de los datos y a la ambigüedad del lenguaje natural. La minería de textos y el PLN se pueden aplicar a los textos para estructurar y extraer información significativa de ellos. La minería de textos es el proceso que estructura y deriva las percepciones y el significado del texto. PLN es el proceso que puede aplicarse al lenguaje humano para ayudar a la computadora a obtener resultados valiosos del texto (Singh et al., 2017). Una de las muchas fuentes de datos de texto no estructurados externos para las organizaciones es el e-correo (Mahlawi & Sasi, 2017).

El análisis de sentimientos es un enfoque útil para una serie de problemas diferentes que se plantean en diversas disciplinas, como psicología (Hoogendoorn et al., 2016), educación, sociología, negocios, ciencias políticas y economía, así como en campos de investigación como el PLN, minería de datos y recuperación de información (Jianget al., 2018).

El análisis de sentimientos también es útil para automatizar la toma de decisiones al ayudar a las organizaciones a comprender mejor los efectos de los problemas específicos en las percepciones de las personas y responder a estos efectos de manera adecuada a través del marketing y la comunicación (Jiang et al., 2018). Por lo tanto, medir las características del sentimiento automáticamente en un texto es valioso, para comprender mejor cómo las emociones, los sentimientos, el afecto y las opiniones influyen en el usuario (Crossley et al., 2017).

Gran parte de las investigaciones de análisis de emociones y sentimientos sobre datos no estructurados se centran solamente en la polaridad de las opiniones del texto para clasificarlo de acuerdo con su carga sentimental en tres tipos de valores: positivo, neutro y negativo, incluso aplicaciones de e-correo de compañías como Outlook (Microsoft, 2019), Gmail (Google, 2019a), Inbenta (Inbenta, 2019), así como Emailparser, (FrozenFrog

Software, 2019), solo calculan la polaridad sentimental de los textos de los e-correos. Esto significa para las organizaciones un problema al intentar identificar emociones como la ira, tristeza, miedo, anticipación, sorpresa, alegría, etc., en sus textos no estructurados que se emiten mediante e-correo. Al identificar sentimientos y emociones, como en este trabajo se propone, se puede garantizar a las organizaciones una mejor forma de elevar los índices de calidad en el servicio y mantener una relación cercana con los clientes para la toma de acciones apropiadas en función de sus necesidades de acuerdo con su estado emocional y sentimental. Así se podrá incrementar la competitividad y el crecimiento de la entidad corporativa en un mercado determinado.

En este artículo se presenta una base de datos léxica llamada EmoLex (Mohammad & Turney, 2013; Mohammad et al., 2017, 2018), la cual contiene 14,182 unigramas (palabras) que permiten identificar ocho emociones básicas (ira, disgusto, tristeza, sorpresa, miedo, confianza, alegría y anticipación), la polaridad del sentimiento (positivo y negativo) y un algoritmo cognitivo de PLN llamado "Natural Language Cloud" (NLC): nube de lenguaje natural (Google, 2019b) que incluye un analizador semántico que permite utilizar un etiquetador gramatical llamado "Part of Speech" (parte del discurso) para el análisis de texto no estructurado en mensajes de e-correos, para así clasificarlos de acuerdo con su carga emocional y sentimental.

## EmoLex

Es un recurso léxico desarrollado por el "National Research Council" (NRC): Consejo Nacional de Investigación de Canadá, y fue creado a través de un servicio en línea de Amazon llamado "Amazon Mechanical Turk" (turco mecánico de Amazon), que requiere de inteligencia humana para desarrollar diferentes tareas. Se basa en el conjunto de ocho emociones básicas y dos sentimientos que

consideraron que la cualidad bipolar de estas emociones resultaría de gran utilidad para el proyecto de creación de EmoLex. Respecto al número de entradas, este lexicón consta de 14,182 unigramas (palabras).

El número de términos por emoción que presenta es:

- Ira: 1,247 términos
- Anticipación: 839 términos
- Disgusto: 1,058 términos
- Miedo: 1,476 términos
- Alegría: 688 términos
- Tristeza: 1,191 términos
- Sorpresa: 534 términos
- Confianza: 1,230 términos

El NRC eligió la plataforma de Amazon por varias razones, siendo las más destacable la posibilidad de obtener una gran cantidad de anotaciones humanas de forma eficiente. A través de esta fórmula se genera la base de datos léxica EmoLex (Mohammad & Turney, 2013; Mohammad et al., 2017, 2018), cuyo tamaño es mayor que muchas otras herramientas utilizadas en este campo como lo son el lexicón SenticNet, SO-CAL, General Inquirer entre otros (Crossley et al., 2017).

## NLC

NLC es una interfaz de programación de aplicaciones (IPA) de inteligencia artificial de Google que implementa PLN para descubrir la estructura y el significado del texto a través de modelos de aprendizaje automático. Con IPA, se puede extraer información sobre personas, lugares, eventos y mucho más, independientemente de si los datos se encuentran en documentos de texto, artículos de noticias, entradas de blog, etc. (Google, 2019b).

Las características de la IPA de la NLC de Google son:

- Análisis sintáctico: extrae componentes léxicos y frases, identifica categorías gramaticales y crea árboles de análisis de dependencias para cada frase.
- Reconocimiento de entidades: identifica entidades y las etiqueta por tipo: persona, organización, ubicación, evento, producto o medio.
- Análisis de opinión: muestra la opinión general que se desprende de un bloque de texto.
- Clasificación de contenido: clasifica documentos en las más de 700 categorías predefinidas.
- Multilingüe: analiza texto fácilmente en varios idiomas, como alemán, chino (tanto simplificado como tradicional), coreano, español, francés, inglés, italiano, japonés y portugués.

## Metodología

La metodología consta de cinco fases: 1) etiquetado gramatical, 2) filtrado de datos, 3) comparación de datos, 4) análisis de emociones y sentimientos y 5) clasificación de e-correos. La Figura 1 muestra la metodología propuesta.

## Etiquetado gramatical

Se implementa el algoritmo cognitivo NLC de Google (Google, 2019b), al cual se envía el texto del mensaje del e-correo mediante un servicio web para ser tratado por técnicas de PLN con el objetivo de etiquetar gramaticalmente cada palabra del texto contenido en el e-correo. En la Tabla 1 se muestra el tipo de etiquetas gramaticales con las que se clasifica el texto del mensaje.

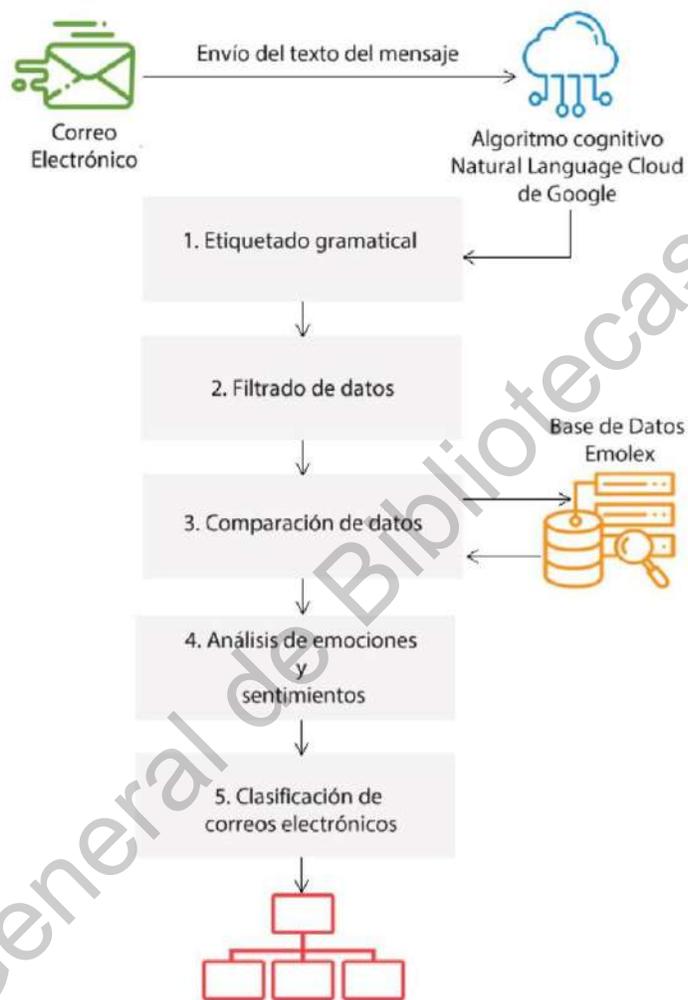


Figura 1. Metodología propuesta para la elaboración de e-correos.  
Fuente: Elaboración propia.

Tabla 1. Tipos de etiquetas gramaticales.

Etiquetas gramatical	Descripción
Unknown	Desconocido
Adj	Adjetivo
Adp	Adposición
Adv	Adverbio
Conj	Conjunción
Det	Determinante
Noun	Sustantivo
Num	Número cardinal
Pron	Pronombre
Prt	Partícula u otra palabra funcional
Punct	Puntuación
Verb	Verbo
X	Otros
Affix	Afijo

Fuente: Elaboración propia.

### Filtrado de datos

Al obtener el texto del mensaje del e-correo procesado por el algoritmo cognitivo NLC y etiquetado de acuerdo con su categoría gramatical, se requiere filtrar y extraer las palabras del mensaje que emiten sentimientos y emociones, las cuales para esta investigación son: sustantivos, verbos, adjetivos y adverbios y separar las que no emiten emociones ni sentimientos como: desconocido, adposición, conjunción, determinante, número cardinal, pronombre, partícula u otra palabra funcional, puntuación y otros (palabras extranjeras, errores tipográficos, abreviaturas y afijos), de acuerdo con los tipos de etiquetas gramaticales de la Tabla 1.

### Comparación de datos

Una vez que se obtuvieron las palabras filtradas del texto del mensaje del e-correo y se etiquetaron de acuerdo con su categoría gramatical, se realiza una consulta a la base de datos léxica Emolex la cual contiene 14,182 unigramas (palabras), que están categorizados en ocho emociones (ira, disgusto, tristeza, sorpresa, miedo, confianza, alegría y anticipación) y dos sentimientos (positivo y negativo) para

comparar los datos y obtener las palabras que coincidan con el texto que se filtró del mensaje y los unigramas de la base de datos.

### Análisis de emociones y sentimientos

Las palabras del mensaje del e-correo pueden tener más de una emoción y sentimiento al mismo tiempo. Por ejemplo, la palabra "atraso" genera emociones de sorpresa, anticipación y un sentimiento negativo. La palabra "tratamiento" genera emociones de ira, anticipación, disgusto, miedo, alegría y tristeza y un sentimiento positivo y negativo.

Las emociones y sentimientos tienen un valor binario dentro de la base de datos léxica, donde 1 determina que la palabra emite esa emoción o sentimiento y 0 que no la emite.

Para calcular la valencia de cada emoción y sentimiento se suman los valores binarios de la columna correspondiente como se muestra en la Tabla 2.

Al obtener la valencia de cada emoción y sentimiento, estas se suman para obtener la valencia total de emociones y la valencia total de sentimientos.

- Valencia total de emociones = vir + vdi + vtr + vso + vmi + vco + val + van
- Valencia total de sentimientos = vpo + vne

Para determinar el porcentaje de cada emoción se divide la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira = vira / vte
- Porcentaje disgusto = vdisgusto / vte
- Porcentaje tristeza = vtristeza / vte
- Porcentaje sorpresa = vsorpresa / vte
- Porcentaje miedo = vmiedo / vte
- Porcentaje confianza = vconfianza / vte

- Porcentaje alegría = valegría / vte
- Porcentaje anticipación = vanticipacion / vte

Para determinar el porcentaje de cada sentimiento se divide la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo = vpositivo / vts
- Porcentaje negativo = vnegativo / vts

En la Tabla 3 se muestra que, una vez aplicadas las fórmulas, se obtienen los niveles de porcentaje de emociones y sentimientos de cada mensaje de e-correo.

Tabla 2. Valencias de emoción y sentimiento.

Palabra	Emociones								Sentimientos	
	Ira	Disgusto	Tristeza	Sorpresa	Miedo	Confianza	Alegría	Anticipación	Positivo	Negativo
P1	1	0	0	0	1	0	0	0	0	1
P2	0	0	0	0	0	1	1	1	1	0
P3	0	0	0	0	0	0	1	0	1	0
P4	0	0	0	0	0	1	0	0	1	0
PN	0	0	0	1	0	0	1	1	1	0
Valencia	1	0	0	1	1	2	3	2	4	1

Fuente: Elaboración propia.

Tabla 3. Porcentaje de emociones y sentimientos.

Palabra	Emociones								Sentimientos	
	Ira	Disgusto	Tristeza	Sorpresa	Miedo	Confianza	Alegría	Anticipación	Positivo	Negativo
P1	1	0	0	0	1	0	0	0	0	1
P2	0	0	0	0	0	1	1	1	1	0
P3	0	0	0	0	0	0	1	0	1	0
P4	0	0	0	0	0	1	0	0	1	0
PN	0	0	0	1	0	0	1	1	0	1
Valencia	1	0	0	1	1	2	3	2	3	2
%	0.10	0.00	0.00	0.10	0.10	0.20	0.30	0.20	0.60	0.40

Fuente: Elaboración propia.

## Clasificación de e-correos

La clasificación de los e-correos se realiza por la polaridad de sentimientos que emite cada mensaje analizado.

Los mensajes de los e-correos se ordenan de acuerdo con el porcentaje de polaridad negativa, los más negativos primero y así sucesivamente, en orden descendente. Este ordenamiento deja los mensajes más positivos abajo de la clasificación.

El conjunto de datos se tomó de una empresa dedicada al servicio web de facturación electrónica en México con una cartera de 2,544 clientes. El periodo de recolección de e-correos comprende del 12 de noviembre del 2018 al 14 de marzo del 2019. El conjunto de datos analizados y clasificados fue de 2,000 mensajes.

## Precisión

La evaluación de la precisión del análisis de emociones y sentimientos para la clasificación de los e-correos se determina dividiendo los e-correos clasificados correctamente entre el total de e-correos.

Precisión = e-correos clasificados correctamente / total e-correos electrónicos

## Resultados

Para el desarrollo de los resultados se tomó una muestra de cuatro e-correos de los 2,000 mensajes analizados, los cuales se ilustran en la Tabla 4.

Tabla 4. Muestra de los cuatro mensajes de los e-correos.

	<b>Asunto</b>	<b>Comprobante de pago agrícola zafu</b>
1	Mensaje	Buenas tardes envío adjunto a este correo el comprobante de pago de la compra del plan de folios pyme de 150 folios. Mi RFC es: AZA1702026J5 Saludos
	<b>Asunto</b>	<b>RFC OCV1404116UA</b>
2	Mensaje	Buenos días Hace unos momentos realice el pago de 50 folios, pero no puedo sacar el recibo. Pueden checar su cuenta? Les mando evidencia del cargo. Saludos. Muchas gracias
	<b>Asunto</b>	<b>Error en el sistema</b>
3	Mensaje	Me marca el siguiente error al intentar timbrar facturas "Usted no puede generar facturas por que no tiene con folios disponibles" no sé si es un problema del programa y me urge enviar una factura como se puede solucionar? Por favor requiero una respuesta rápida ha esta situación
	<b>Asunto</b>	<b>Certificado de sellos</b>
4	Mensaje	Que tal buenos días Le comento que al momento de tratar de cargar el Sello Digital el sistema marca que la contraseña es incorrecta, la contraseña que se requiere es la que se le asignó a los archivos en este caso al Certificado de Sello Digital (CSD), no la contraseña para acceder al portal del SAT. Seguimos pendientes de cualquier detalle, Saludos!

Fuente: Elaboración propia.

### Etiquetado gramatical

Al aplicar el PLN con el algoritmo cognitivo NLC de Google al texto del mensaje 3, se obtiene el etiquetado gramatical de cada palabra que forma parte del discurso.

### Filtrado de datos

En la Tabla 5 se muestra que al haber etiquetado gramaticalmente las palabras del texto del mensaje 3 del e-correo se filtraron y extrajeron las palabras que emitieron emociones y sentimientos que son las correspondientes a sustantivos, verbos, adjetivos y adverbios (Mohammad & Turney, 2013).

Tabla 5. Filtrado de datos: sustantivos, verbos, adjetivos y adverbios del mensaje 3.

Mensaje 3	
Palabra	Etiqueta gramatical
marca	Verbo
siguiente	Adjetivo
error	Sustantivo
intentar	Verbo
timbrar	Verbo
facturas	Sustantivo
no	Adverbio
puede	Verbo
generar	Verbo
facturas	Sustantivo
no	Adverbio
tiene	Verbo
folios	Sustantivo
disponibles	Adjetivo
no	Adverbio
es	Verbo
problemas	Sustantivo
programa	Sustantivo
urge	Verbo
enviar	Verbo
factura	Sustantivo
puede	Verbo
solucionar	Verbo
favor	Sustantivo
requiero	Verbo
respuesta	Sustantivo
rápida	Adjetivo
ha	Verbo
situación	Sustantivo

Fuente: Elaboración propia.

## Comparación de datos

En la Tabla 6 se muestra el resultado de la consulta de la comparación de las coincidencias de las palabras filtradas y los unigramas de la base de datos léxica Emolex por emociones y sentimientos del mensaje 3.

Tabla 6. Mensaje 3, coincidencias de las palabras filtradas y los unigramas de la base de datos léxica.

Palabra	Emociones								Sentimientos	
	Ira	Disgusto	Tristeza	Sorpresa	Miedo	Confianza	Alegría	Anticipación	Positivo	Negativo
enviar	0	0	0	0	0	0	0	1	1	0
error	0	0	1	0	0	0	0	0	0	1
intentar	0	0	0	0	0	1	0	0	0	0
no	0	0	0	0	0	0	0	1	0	1
problema	0	0	1	0	1	0	0	0	0	1
situación	0	1	1	0	1	0	0	1	0	1

Fuente: Elaboración propia.

### Análisis de emociones y sentimientos

Se determinó la valencia de cada emoción y sentimiento del mensaje 3 sumando las columnas de acuerdo con su valor binario.

Al obtener la valencia de cada emoción y sentimiento se calculó la valencia total de emociones sumando la valencia de cada emoción y la valencia total de sentimientos sumando la valencia de cada sentimiento.

- Valencia total de emociones =  $0 + 1 + 3 + 0 + 2 + 1 + 0 + 3 = 10$
- Valencia total de sentimientos =  $1 + 4 = 5$

Para calcular el porcentaje de cada emoción se dividió la valencia de la emoción entre la valencia total de las emociones.

- Porcentaje ira =  $0 / 10 = 0.00$
- Porcentaje disgusto =  $1 / 10 = 0.10$
- Porcentaje tristeza =  $3 / 10 = 0.30$
- Porcentaje sorpresa =  $0 / 10 = 0.00$

- Porcentaje miedo =  $2 / 10 = 0.20$
- Porcentaje confianza =  $1 / 10 = 0.10$
- Porcentaje alegría =  $0 / 10 = 0.00$
- Porcentaje anticipación =  $3 / 10 = 0.30$

Para calcular el porcentaje de cada sentimiento se dividió la valencia del sentimiento entre la valencia total de sentimientos.

- Porcentaje positivo =  $1 / 5 = 0.20$
- Porcentaje negativo =  $4 / 5 = 0.80$

En la Tabla 7 se muestran las valencias, los porcentajes de cada emoción y sentimiento y el resultado del análisis de emociones y sentimientos que forma parte del texto del mensaje 3 del e-correo para su posterior clasificación.

### Clasificación de e-correo

En las Figuras 2, 3, 4 y 5 se presenta el resultado del análisis de emociones y sentimientos de los cuatro mensajes de la muestra de los e-correos.

Tabla 7. Valencia y porcentaje de emociones y sentimientos del mensaje 3.

Palabra	Emociones								Sentimientos	
	Ira	Disgusto	Tristeza	Sorpres	Miedo	Confianza	Alegría	Anticipación	Positivo	Negativo
enviar	0	0	0	0	0	0	0	1	1	0
error	0	0	1	0	0	0	0	0	0	1
intentar	0	0	0	0	0	1	0	0	0	0
no	0	0	0	0	0	0	0	1	0	1
problema	0	0	1	0	1	0	0	0	0	1
situación	0	1	1	0	1	0	0	1	0	1
Valencia	0	1	3	0	2	1	0	3	1	4
%	0.00	0.10	0.30	0.00	0.20	0.10	0.00	0.30	0.20	0.80

Fuente: Elaboración propia.

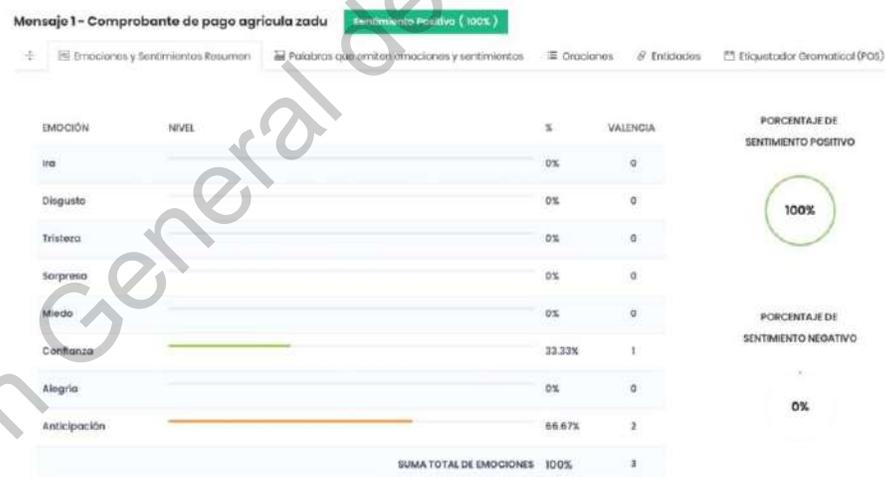


Figura 2. Análisis de emociones y sentimientos del mensaje 1.

Fuente: Elaboración propia.

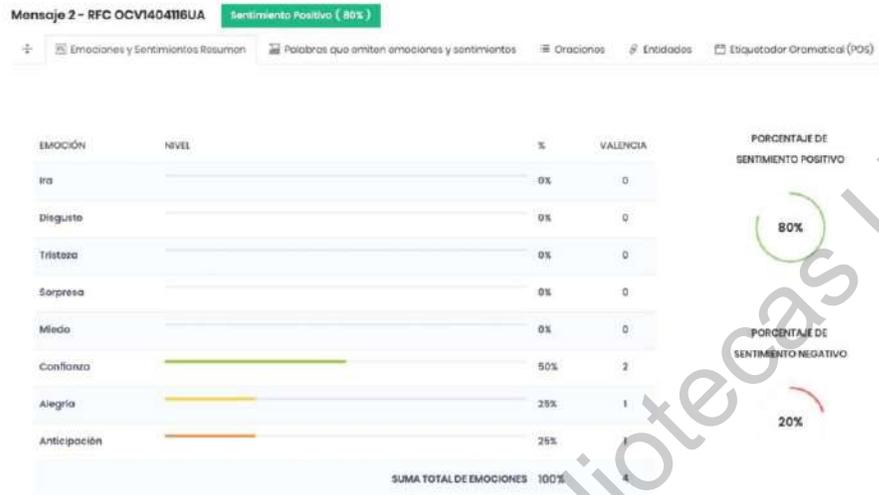


Figura 3. Análisis de emociones y sentimientos del mensaje 2.  
Fuente: Elaboración propia.

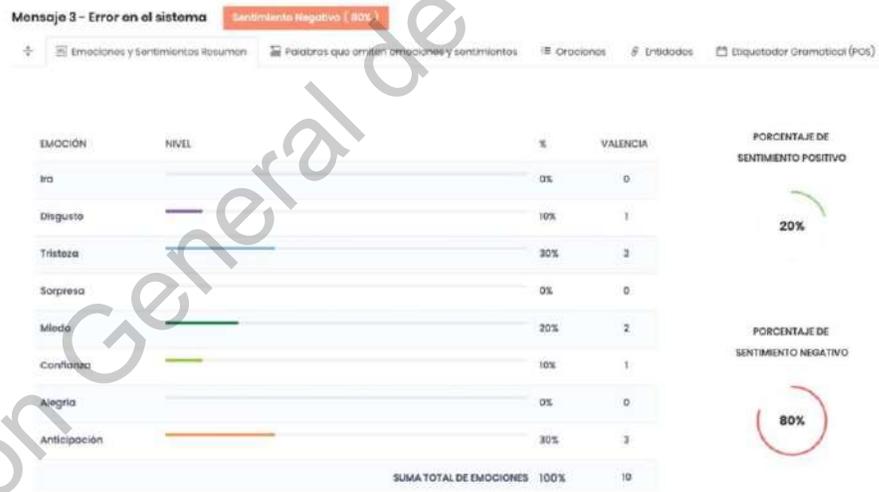


Figura 4. Análisis de emociones y sentimientos del mensaje 3.  
Fuente: Elaboración propia.

La clasificación y el orden de los e-correos se establece de acuerdo con el análisis de emociones y sentimientos que se realizó al texto de los mensajes con el resultado del porcentaje de polaridad negativa. Los más negativos primero y así sucesivamente, en un orden descendente como se muestra en la Tabla 8.

### Precisión

Se determinó que de los 2,000 e-correos, 1,849 se analizaron y se clasificaron de manera correcta y solo 151 mensajes no entraron en la clasificación esperada, dando como resultado una precisión del 92%.

$$\text{Precisión} = 1,849 / 2,000 = 0.92$$

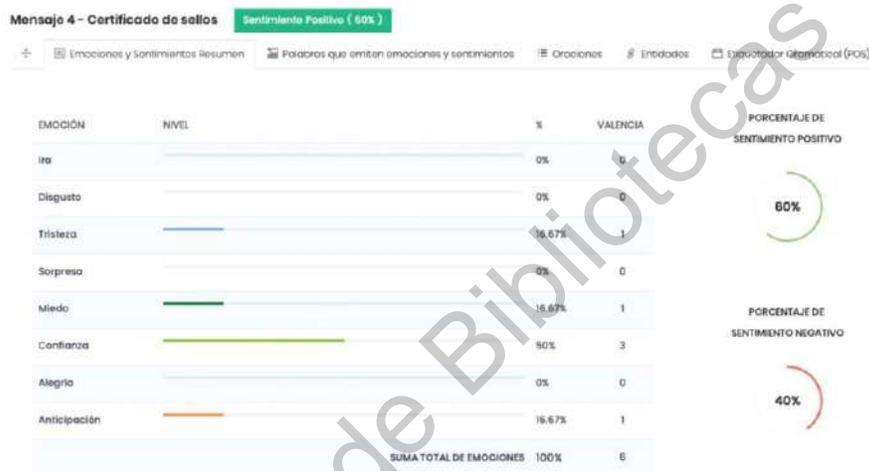


Figura 5. Análisis de emociones y sentimientos del mensaje 4.  
Fuente: Elaboración propia.

Tabla 8. Clasificación y orden de los mensajes de los e-correos.

Mensaje	Emociones								Sentimientos	
	Ira	Disgusto	Tristeza	Sorpresa	Miedo	Confianza	Alegría	Anticipación	Positivo	Negativo
3	0.00	0.10	0.30	0.00	0.20	0.10	0.00	0.30	0.20	0.80
4	0.00	0.00	0.17	0.00	0.17	0.50	0.00	0.17	0.60	0.40
2	0.00	0.00	0.00	0.00	0.00	0.50	0.25	0.25	0.80	0.20
1	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.67	1.00	0.00

Fuente: Elaboración propia.

### Impacto del clasificador de e-correo en la organización

La atención y el soporte al cliente normalmente se realizan por medios electrónicos, pues el total de la información se maneja en modo de texto. Actualmente la empresa dedicada al servicio web de facturación electrónica en México cuenta con una cartera de 2,544 clientes activos, de los cuales se tiene una actividad constante mensual con al menos 1,600 clientes. El 80% del soporte que se ofrece a esta población de clientes es por e-correo. Para la empresa el servicio es uno de los

procesos de comunicación más importantes. El poder identificar las emociones y sentimientos de los clientes es un factor que determina qué tipo de soporte y la prioridad de servicio que se le va a ofrecer.

La lectura y el análisis de los e-correos se realizaban manualmente para clasificarlos de acuerdo con su tipo y prioridad de servicio. En promedio se han recibido 450 e-correos semanalmente, 1,950 mensualmente y 23,400 anualmente que el personal encargado de servicio y soporte tenía que analizar uno por uno para su clasificación.

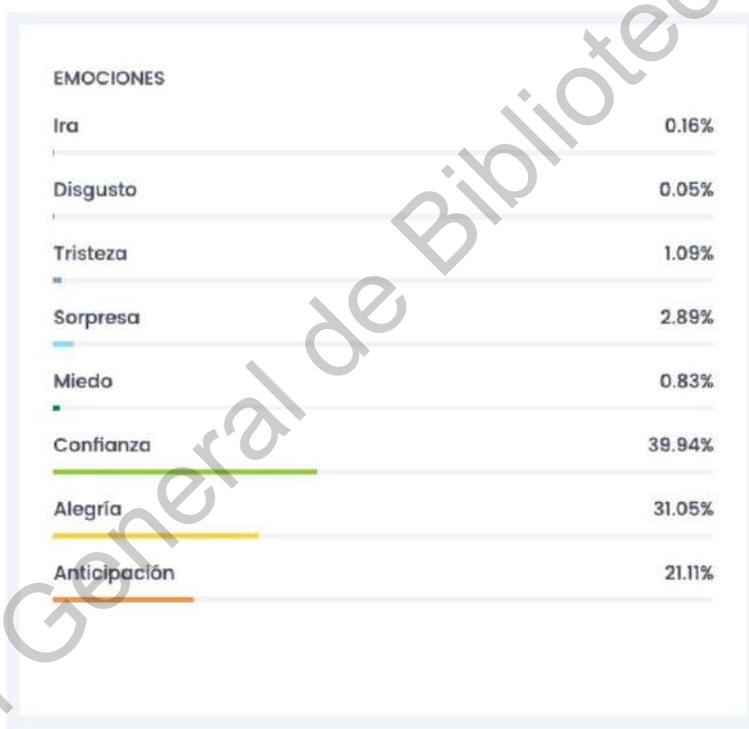


Figura 6. Nivel de emociones acumuladas del total de e-correos.  
Fuente: Elaboración propia.

Con el clasificador, el análisis de emociones y sentimientos de los mensajes de los e-correos es instantáneo. Los e-correos se pueden canalizar a los departamentos correspondientes de acuerdo con su polaridad positiva o negativa y con el contenido en el texto, dando prioridad a los mensajes más negativos para ser atendidos cuanto antes.

Dentro de la organización, con el clasificador se ha podido medir el nivel de satisfacción de los clientes por medio del canal de comunicación que representa el e-correo. Una vez que se obtuvo el acumulado total del análisis de los 2,000 e-correos, se determinó el nivel de satisfacción general de los clientes por la polaridad del sentimiento que fue del 97.85% positivo y 2.15% negativo. El nivel de emociones acumuladas se muestra en la Figura 6. Como se observa en la Figura 6, las tres emociones que impactan sobre la satisfacción del cliente por medio del canal de comunicación de e-correos son positivas, con un 39.94% de confianza, 31.05% de alegría y un 21.11% de anticipación.

#### Conclusiones

El agregar estructura al texto extraído del mensaje del e-correos proporciona contenido significativo para la toma de decisiones. Ignorar estos datos en el mundo de las empresas podría llevar a la insatisfacción del cliente o a la pérdida del mismo. La información resultante permite a las empresas identificar las preferencias de los consumidores, sus comentarios y el comportamiento relacionado con algún producto o servicio.

El desempeño del algoritmo de PLN NLC de Google es eficiente al etiquetar gramaticalmente cada palabra del mensaje del e-correo que forma parte del discurso, dando formato y estructura al texto del mensaje. La base de datos léxica Emolex1 contiene un vasto diccionario de palabras en español que emiten emociones y sentimientos, de gran importancia para esta investigación.

Cabe mencionar que los recursos mencionados son ajenos una vez que se integran con el algoritmo del clasificador de urgencias de e-correos es posible: filtrar palabras, compararlas con la base de datos léxica, calcular las valencias y porcentajes de las emociones y sentimientos para la clasificación de los e-correos.

Los resultados obtenidos demuestran que es significativa la utilidad de manejar características emocionales y sentimentales sobre e-correos, ya que se pueden clasificar y canalizar a los departamentos correspondientes de acuerdo con su polaridad de sentimiento para ser atendidos a la brevedad. También se mide la satisfacción del cliente sobre el canal de comunicación de e-correos sobre la atención y soporte del servicio de facturación electrónica. La metodología de esta investigación arroja una precisión del 92% sobre el análisis de los e-correos, la cual es bastante aceptable.

El procedimiento aplicado para clasificar y analizar este volumen de conjunto de datos se puede utilizar como marco para otras tareas que involucren grandes colecciones de textos (chats, mensajes, blogs, redes sociales, etc.), cuyo procesamiento sea impráctico por medios manuales.

<sup>1</sup> El uso de lexicones o base de datos léxicas en español no es tan común ya que la mayoría de estos recursos vienen en el idioma inglés. Uno de los lexicones más completo y que ha sido traducido a diferentes idiomas es el Emolex, el cual fue utilizado para el desarrollo de este artículo. Existen otros lexicones que se han vuelto populares, los cuales son evaluados y comparados en diferentes artículos. Con la finalidad de orientar al lector, en las Referencias se comparte la ficha de la lectura: Inserción de Conocimiento Emocional Externo a un Clasificador Lineal de Emociones (Arco et al., 2018).

### Agradecimientos

A la empresa EDIGITAL SOLUCIONES TECNOLÓGICAS SA DE CV por la aportación del conjunto de datos.

### Referencias

- Arco, F.M., Martínez-Cámara, E., Valdivia, M.T., & López, L.A. (2018). SINA en TASS 2018: Inserción de Conocimiento Emocional Externo a un Clasificador Lineal de Emociones (SINA) at TASS 2018: Lineal Classification System with Emotional External Knowledge). TASS@SEPLN. 125-130.
- Crossley, S. A., Kyle, K., & McNamara, D.S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behav Res*, 49: 803. doi: 10.3758/s13428-016-0743-z
- FrozenFrog Software. (2019). Emailparser. Consultado el 9 de febrero del 2019. FrozenFrog Software. Recuperado de <https://www.emailparser.com/>
- Google. (2019a). Gmail. Consultado el 9 de febrero del 2019. Google. Recuperado de <https://www.gmail.com/mail/help/intl/es/about.html?frame>
- Google. (2019b). Natural Language Cloud. Consultado el 14 de febrero del 2019. Google. Recuperado de <https://cloud.google.com/natural-language/?hl=es>
- Hoogendoorn, M., Berger, T., Schulz, A., Stolz, T., & Szolovits P. (2016). Predicting Social Anxiety Treatment Outcome Based on Therapeutic Email Conversations. *IEEE*, 21(5), 1449-1459. doi: 10.1109 / JBHI.2016.2601123
- Inbenta. (2019). Descubre si las opiniones de tu cliente son positivas o negativas. Consultado el 9 de febrero del 2019. Inbenta Technologies Inc. Recuperado de <https://www.inbenta.com/mx/características/autoservicio/analisis-de-sentimientos/>
- Jayathilaka, K.M.P.N., Weerasinghe A.R., & Wijesekara W.M.L.K.N. (2016). Making sense of large volumes of unstructured email responses. *IEEE*, 35-40. doi: 10.1109 / ICTER.2016.7829896
- Jiang, J., Chen, J., Choo, K.K.R., Liu, K., Liu C., Yu, M., & Mohapatra P. (2018). Prediction and Detection of Malicious Insiders' Motivation based on Sentiment Profile on Webpages and Emails. *IEEE*, 225-229. doi: 10.1109 / MILCOM.2018.8599790
- Lee J., Thabsuwan C., Pongpaichet S., & Kim K. (2018). Towards Building a Human Perception

- Knowledge for Social Sensation Analysis. IEEE, 668-671. doi: 10.1109/WI.2018.00-15
- Mahlawi, A., Q. & Sasi, S. (2017). Structured Data Extraction from Emails. IEEE, 20-22. doi: 10.1109 / NETACT.2017.8076789
- Microsoft. (2019). Microsoft Outlook. Consultado el 9 de febrero del 2019. Microsoft. Recuperado de <https://products.office.com/es-mx/outlook/email-and-calendar-software-microsoft-outlook>
- Miller, R. & Charles, E. (2016). A psychological based analysis of marketing email subject lines. IEEE, 58-65. doi: 10.1109/ICTER.2016.7829899
- Mohammad, S. & Turney, P. (2013). Computational Intelligence, 29 (3), 436-465.
- Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in tweets. Special Section of the ACM Transactions on Internet Technology on Argumentation in Social Media, 17(3), 1-19. doi: 10.1145/3003433
- Mohammad, S. M., Marquez, F., Mohammad, S., & Kiritchenko S. (2018). Semeval 2018 Task 1: Affect in tweets. In Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA. 1-17. doi: 10.18653/v1/S18-1001
- Singh, J., Singh, G.S., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. Human Centric Computing and Information Sciences, 7: 32. doi: 10.1186/s13673-017-0116-3
- Zhang, H., Gan, W., & Jiang, B. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. In X. Yuan & X. Meng (Eds.), Proceedings of the 11th Web information system and application conference. 262-265. Piscataway, IEEE Press.
- Recibido:** 23 de mayo de 2019  
**Corregido:** 2 de julio de 2019  
**Aceptado:** 3 de julio de 2019
- Conflicto de interés:** No existe conflicto de interés