

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

OPCIÓN I.- TESIS

TRABAJO PROFESIONAL

“DESARROLLO DE UN MÓDULO DE OBTENCIÓN DE EMOCIONES
BASADO EN TÉCNICAS HÍBRIDAS DE ANÁLISIS SENTIMENTAL”.

QUE PARA OBTENER EL GRADO DE:
MAESTRO EN SISTEMAS
COMPUTACIONALES

PRESENTA:

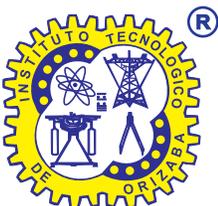
I.S.C. Francisco Javier Ramírez Tinoco

DIRECTOR DE TESIS:

Dr. Giner Alor Hernández

CODIRECTOR DE TESIS:

Dr. José Luis Sánchez Cervantes



FECHA: 05/12/2018
DEPENDENCIA: POSGRADO
ASUNTO: Autorización de Impresión
OPCIÓN: I

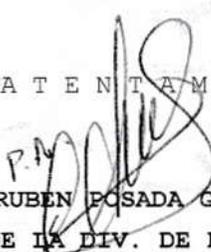
C. FRANCISCO JAVIER RAMIREZ TINOCO
CANDIDATO A GRADO DE MAESTRO EN:
SISTEMAS COMPUTACIONALES

De acuerdo con el Reglamento de Titulación vigente de los Centros de Enseñanza Técnica Superior, dependiente de la Dirección General de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

**"DESARROLLO DE UN MODULO DE OBTENCION DE EMOCIONES BASADO EN
TECNICAS HIBRIDAS DE ANALISIS SENTIMENTAL".**

Comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

A T E N T A M E N T E


DR. RUBEN POSADA GOMEZ
JEFE DE LA DIV. DE ESTUDIOS DE POSGRADO

C.A. TITULACIÓN



SECRETARIA DE
EDUCACIÓN PÚBLICA
INSTITUTO
TECNOLÓGICO
DE ORIZABA

ggc



FECHA : 27/11/2018

ASUNTO: Revisión de Trabajo Escrito

C. DR. RUBEN POSADA GOMEZ
JEFE DE LA DIVISION DE ESTUDIOS
DE POSGRADO E INVESTIGACION.
P R E S E N T E

Los que suscriben, miembros del jurado, han realizado la revisión de la Tesis del (la) C. :

FRANCISCO JAVIER RAMIREZ TINOCO

la cual lleva el titulo de:

"DESARROLLO DE UN MODULO DE OBTENCION DE EMOCIONES BASADO EN TECNICAS HIBRIDAS DE ANALISIS SENTIMENTAL".

Y concluyen que se acepta.

A T E N T A M E N T E

PRESIDENTE : DR. GINER ALOR HERNANDEZ



FIRMA

SECRETARIO : DR. JOSE LUIS SANCHEZ CERVANTES



FIRMA

VOCAL : MCE BEATRIZ ALEJANDRA OLIVARES



FIRMA

VOCAL SUP. : DRA. LISBETH RODRIGUEZ MAZAHUA



FIRMA

EGRESADO(A) DE LA MAESTRIA EN **SISTEMAS COMPUTACIONALES**

OPCION: I **Tesis**



Índice General

Índice de figuras.....	iii
Índice de listados.....	iv
Índice de tablas.....	v
Índice de fórmulas.....	vi
Agradecimientos	vii
Resumen.....	viii
Abstract	ix
Introducción	x
Capítulo I. Antecedentes	1
1.1 Marco teórico	1
1.1.1 Análisis sentimental	1
1.1.2 APIs para análisis sentimental sobre texto	3
1.1.3 Tareas de análisis sentimental	6
1.1.4 Enfoques de análisis sentimental.....	7
1.1.5 Lenguajes de programación para la Web	13
1.1.6 Sistemas Gestores de Bases de Datos.....	14
1.1.7 Redes Sociales.....	15
1.1.8 APIs para redes sociales	16
1.1.9 API para la construcción de gráficas.....	17
1.2 Situación tecnológica, económica y operativa de la empresa	17
1.3 Planteamiento del problema	18
1.4 Objetivo general y específico.....	19
1.4.1 Objetivo general	19
1.4.2 Objetivos específicos.....	19
1.5 Justificación.....	19
Capítulo II. Estado de la práctica	21
2.1 Trabajos relacionados con análisis sentimental.....	21
2.2 Análisis comparativo de trabajos relacionados con análisis sentimental	29
2.3 Propuesta de solución.....	35

2.3.1 Justificación de la solución	36
Capítulo III. Aplicación de la metodología	38
3.1 Análisis de requerimientos del módulo	38
3.2 Fuentes de información seleccionadas para analizar.....	40
3.3 Esquema de almacenamiento para el resultado del análisis sentimental.....	43
3.4 Flujo de trabajo	45
3.5 Diseño de <i>mockups</i> para el módulo.....	47
3.6 Iconografía para emociones y entidades	51
3.7 Principales funcionalidades del módulo.....	52
3.7.1 Inicio de sesión.....	52
3.7.2 Extracción de información de redes sociales.....	54
3.7.3 Extracción de sentimientos y emociones.....	58
3.7.4 Representación gráfica del análisis de sentimientos y emociones	62
3.7.5 Interpretación de la información	64
Capítulo IV. Resultados	67
4.1 Casos de estudio.....	67
4.1.1 Caso de estudio: Análisis sentimental sobre comentarios de médicos especialistas en diabetes para seleccionar el más adecuado para atender dicha enfermedad.....	68
4.1.2 Caso de estudio: Análisis sentimental sobre comentarios de medicamentos para tratar la hipertensión	76
Capítulo V. Conclusiones.....	83
5.1 Recomendaciones.....	84
Productos académicos	85
Referencias.....	86

Índice de figuras

Figura 3.1 Organización del Análisis Sentimental de acuerdo con Ravi y colaboradores.	2
Figura 3.2 Ejemplo de separadores lineales empleando SVM.	10
Figura 3.3 Ejemplo de red bayesiana.	12
Figura 3.1 Diagrama de Casos de Uso para el módulo de obtención de emociones.	39
Figura 3.2 Diagrama de Base de Datos para el módulo de extracción de emociones.	44
Figura 3.3 Flujo de trabajo para la extracción de sentimientos y emociones mediante el módulo. ..	47
Figura 3.4 Gráfica de emociones por género.	65
Figura 3.5 Gráfica de emociones por edad.	66
Figura 3.6 Gráfica que muestra la polaridad de entidades.	66
Figura 4.1 Perfil de un médico en donde es posible emplear las funcionalidades del módulo.	69
Figura 4.2 Ventana emergente que muestra los comentarios analizados.	70
Figura 4.3 Pestaña que muestra el porcentaje de polaridad y emociones para cada comentario.	70
Figura 4.4 Pestaña que muestra las entidades encontradas junto con su polaridad y emociones.	71
Figura 4.5 Gráfica que muestra la polaridad expresada hacia el médico.	72
Figura 4.6 Gráfica que muestra los porcentajes de cada emoción expresados hacia el médico.	72
Figura 4.7 Gráfica que muestra la distribución de emociones por género.	73
Figura 4.8 Gráfica que muestra la distribución de emociones por edad.	74
Figura 4.9 Gráfica que muestra la distribución de emociones por ubicación.	74
Figura 4.10 Mapa de calor que muestra la polaridad de entidades.	75
Figura 4.11 Detalle de un medicamento en donde integraron las funcionalidades del módulo.	77
Figura 4.12 Pestaña de una ventana modal que muestra los comentarios analizados.	78
Figura 4.13 Pestaña que muestra el porcentaje de polaridad y emociones para cada comentario. ...	78
Figura 4.14 Pestaña que muestra las entidades encontradas junto con su polaridad y emociones.	79
Figura 4.15 Gráfica que muestra la polaridad expresada hacia el medicamento.	80
Figura 4.16 Gráfica que muestra los porcentajes de cada emoción hacia el medicamento.	80
Figura 4.17 Gráfica que muestra la distribución de emociones por ubicación.	81
Figura 4.18 Mapa de calor que muestra la polaridad de entidades.	82

Índice de listados

Listado 3.1 Fragmento de código de inicio de sesión en Facebook®	53
Listado 3.2 Fragmento de código de inicio de sesión en Twitter®	54
Listado 3.3 Función para establecer la conexión con Facebook®	55
Listado 3.4 Petición para extraer comentarios a partir de un identificador en Facebook®	55
Listado 3.5 Petición para extraer datos de usuarios en Facebook®	56
Listado 3.6 Función para establecer la conexión con Twitter®	56
Listado 3.7 Extracción de tuits a partir de un identificador en Twitter®	57
Listado 3.8 Extracción de datos de usuarios en Twitter®	58
Listado 3.9 Fragmento de código para extraer sentimientos, emociones y entidades	60
Listado 3.10 Método que consume el servicio Web para extraer sentimientos y emociones.	61
Listado 3.11 Función que trata el resultado del análisis cuando se trata de un arreglo de objetos. ..	61
Listado 3.12 Agrupamiento de polaridad y emociones por género	63
Listado 3.13 Función que permite desarrollar la gráfica de emociones por género	64

Índice de tablas

Tabla 1.1 APIs para análisis sentimental sobre texto.....	5
Tabla 1.2 Principales tareas del análisis sentimental.	6
Tabla 1.3 Técnicas empleadas para el enfoque basado en corpus.....	8
Tabla 1.4 Redes sociales con mayor impacto en la actualidad.....	15
Tabla 2.1 Análisis comparativo de los trabajos relacionados con análisis sentimental	34
Tabla 2.2 Tecnologías involucradas en la solución propuesta	37
Tabla 3.1 Funcionalidades requeridas para el módulo de obtención de emociones.....	39
Tabla 3.2 Fuentes de información adecuadas para trabajar en Facebook®.	42
Tabla 3.3 Fuentes de información adecuadas para trabajar en Twitter®.	43
Tabla 3.4 Diccionario de datos de la Base de Datos del módulo de extracción de emociones.	45
Tabla 3.5 Diseño de mockups para la representación gráfica de los resultados.....	50
Tabla 3.6 Iconografía para las diferentes emociones y entidades	51
Tabla 3.7 Valores obtenidos tras analizar un comentario en diferentes niveles.....	59

Índice de fórmulas

Fórmula 1.1 Teorema de Bayes.....	11
Fórmula 1.2 Teorema de Bayes reescrito.....	12

Agradecimientos

A Dios, por guiarme siempre y por permitirme cumplir una meta más, la cual implicó grandes retos, gratas experiencias y conocer a personas muy valiosas.

A mis padres, Patricia Tinoco Tapia y Javier Ramírez Sánchez, por ser los pilares fundamentales en mi vida, por apoyarme en todas y cada una de las decisiones que he tomado, por enseñarme a afrontar cada reto, por su guía, por sus consejos, por su cariño, por sus cuidados y porque sin ellos, ninguno de mis logros sería posible.

A mi director de tesis, el Dr. Giner Alor Hernández, por su apoyo, su guía, sus consejos, sus recomendaciones, su confianza, y por transmitirme sus conocimientos y las experiencias de su trayectoria, los cuales permitieron desarrollar este trabajo adecuadamente y me permitieron a mí, crecer personal y profesionalmente.

A la MCE. Beatriz Alejandra Olivares Zepahua, por estar siempre al pendiente de mí, por la motivación para estudiar una maestría, por su apoyo incondicional, sus consejos y sugerencias, su confianza, sus amenas y divertidas pláticas siempre llenas de conocimiento y experiencias increíbles, y por todo el conocimiento adquirido de su persona.

Al Dr. José Luis Sánchez Cervantes, por su orientación, su apoyo, sus consejos, su confianza, sus divertidas pláticas, por estar siempre al pendiente y por los conocimientos y experiencias transmitidos que aportaron en gran medida a este trabajo.

A la Dra. Lisbeth Rodríguez Mazahua, por todos sus consejos, su apoyo, su guía y por los conocimientos transmitidos que me permitieron enriquecer este proyecto. Asimismo, agradezco al resto de docentes de la Maestría en Sistemas Computacionales, quienes mediante sus sugerencias, comentarios y conocimientos aportaron a este proyecto.

A todos los amigos y familiares que sinceramente me brindaron su apoyo. No es posible mencionarlos a todos; sin embargo, no pueden faltar mis compañeros de maestría: Carlos, Betia, Daniel y Mariela, gracias porque juntos superamos distintos retos, crecimos profesionalmente y también personalmente.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado para la realización de mis estudios de maestría, así como al Instituto Tecnológico de Orizaba y a la División de Estudios de Posgrado e Investigación (DEPI) por brindarme las instalaciones necesarias para la realización de esta tesis.

Resumen

Actualmente existe una gran cantidad de personas que emplea las redes sociales de manera cotidiana para expresar sus ideas, opiniones o sentimientos. La interacción con las redes sociales genera una cantidad exponencial de información subjetiva que si se trata y analiza adecuadamente tiene la capacidad de generar grandes beneficios para organizaciones, empresas o particulares. Por ejemplo, para una empresa es posible conocer la aceptación de un nuevo producto y crear estrategias que mejoren la reputación e incrementen las ventas de éste con base en lo que las personas opinan sobre él.

En este sentido, existen algunos trabajos sobre análisis sentimental, un campo de estudio en auge que permite analizar información subjetiva proveniente de diversas fuentes. Sin embargo, éstos se enfocan en su mayoría en determinar si un texto tiene una connotación positiva o negativa, con lo que se deja de lado la extracción de emociones, la cual revela un mayor detalle acerca del sentir de las personas. Además, aunque hay varias técnicas para análisis sentimental, no todas se adaptan al tipo de información que ofrecen las redes sociales. Por estos motivos, el proyecto que se propone tiene por objetivo desarrollar un módulo que permita detectar/obtener sentimientos y emociones mediante técnicas híbridas de análisis sentimental.

El desarrollo de un módulo para la obtención de sentimientos y emociones resulta benéfico porque permite obtener información más detallada y precisa acerca del sentir de las personas respecto a algún producto, servicio, persona o situación, con lo cual se posibilita que diferentes interesados tales como empresas, organizaciones o personas en particular, tomen mejores decisiones, planeen estrategias más exitosas o lleven a cabo acciones que les representen diversos beneficios, tanto económicos como de otra índole.

Abstract

Currently there are a large number of people using Social Networks daily to express ideas, opinions or sentiments. The interaction process with Social Networks generates an exponential amount of subjective information having the capacity of generating great benefits for organizations, companies or individuals if this information is properly treated and analyzed. For example, for a company, it is possible to know the acceptance of a new product and to create strategies that improve the reputation and increase the sales of the product based on what people think about it.

In this context, there are research works on sentiment analysis, a field of study in rise that allows analyzing subjective information from different data sources. However, these works are mostly focused on whether a text has a positive or negative connotation, and they forget the extraction of emotions, which shows a deeper detail about people's sentiments. In addition, although there are several techniques for sentiment analysis, not all of them are adapted to the type of information offered in Social Networks. For these reasons, this project aims to develop a software component that allows obtaining/detecting sentiments and emotions through hybrid techniques of sentiment analysis.

The development of a software component to obtain sentiments and emotions is beneficial because it allows obtaining more detailed and precise information about the people's sentiments regarding some product, service, person or situation. This information allows that different stakeholders as companies, organizations or individuals make better decisions, plan more successful strategies or carry out actions that represent them various benefits, economic as well as other.

Introducción

El SA (*Sentiment Analysis*, Análisis Sentimental) es un campo de estudio que se encarga de extraer y analizar información subjetiva proveniente de diversas fuentes mediante la utilización de procesamiento de lenguaje natural, lingüística computacional, minería de datos y recuperación de información. Por otro lado, con el crecimiento exponencial de los usuarios en la Web y de la información que se genera diariamente, el SA es cada vez más importante para analizar los sentimientos y las emociones que las personas expresan, por ejemplo, en redes sociales, blogs, foros, entre otros.

Además, el análisis sentimental genera beneficios en varios dominios, por ejemplo, conocer el nivel de satisfacción de un cliente con algún producto y tomar medidas en caso de que éste no sea bueno, averiguar cómo se encuentra la opinión pública respecto a algún partido político y planificar mejores campañas o estrategias con base en ello, ajustar el contenido de un sitio Web respecto a los intereses que demuestran sus usuarios, entre otros.

La problemática que se aborda en este trabajo se relaciona con la gran necesidad que se tiene, sobre todo por empresas y organizaciones, de analizar la información subjetiva de plataformas en Internet de manera eficiente para alcanzar resultados más exitosos, así como con el requerimiento constante de nuevas aplicaciones y técnicas para extraer emociones. El problema se atiende mediante el desarrollo de un módulo que utiliza técnicas de análisis sentimental para extraer emociones que le permitan a las empresas, organizaciones o particulares, llevar a cabo acciones o estrategias que generen una mejora en su desempeño diario.

La organización de este documento consta de cinco capítulos. El primer capítulo aborda los antecedentes y los conceptos generales relacionados con este proyecto, así como el planteamiento del problema, los objetivos establecidos y la justificación. Por su parte, en el segundo capítulo se presenta un análisis comparativo de trabajos relacionados y se explica la solución propuesta para resolver el problema planteado. En el tercer capítulo se aborda todo lo relacionado con el desarrollo del trabajo y en el cuarto capítulo se ejemplifica la validez del módulo mediante dos casos de estudio. Finalmente, en el quinto capítulo se presentan las conclusiones y recomendaciones.

Capítulo I. Antecedentes

En este capítulo se presenta la información más relevante que se relaciona con este proyecto. Asimismo, se describe el problema a resolver, el objetivo general y los objetivos específicos del proyecto, así como la justificación de éste.

1.1 Marco teórico

En esta sección se describen los conceptos fundamentales relacionados con el presente proyecto con la finalidad de facilitar la comprensión de éste.

1.1.1 Análisis sentimental

De acuerdo con Ravi y Ravi [1] el análisis sentimental o minería de opiniones es el estudio de las opiniones, sentimientos y actitudes que expresan las personas hacia diferentes tópicos de interés. Dicho análisis utiliza procesamiento de lenguaje natural, análisis de textos, lingüística computacional, recuperación de información y minería de datos. Además, involucra tareas de detección, extracción y clasificación de sentimientos sobre diferentes recursos en diversos formatos como foros de discusión en Internet, blogs, redes sociales, sitios Web de comercio electrónico, noticias, entre otros. Por otra parte, el análisis sentimental permite que empresas, organizaciones y particulares lleven a cabo acciones como: 1) Observar el estado de ánimo de las personas respecto a movimientos sociales o políticos; 2) Medir la satisfacción de clientes sobre productos o servicios; 3) Predecir ventas, entre muchas otras.

Asimismo, es importante mencionar que el análisis sentimental conlleva diferentes dimensiones, tareas, subtareas y enfoques. Básicamente, las dimensiones que se identifican son: 1) Clasificación de subjetividad; 2) Clasificación del sentimiento; 3) Revisión de medida de utilidad; 4) Detección de *spam*; 5) Creación de léxico; 6) Extracción de aspectos, y 7) Aplicaciones de análisis sentimental. Las primeras cinco dimensiones se consideran también como tareas, dentro de las cuales, la clasificación del sentimiento involucra subtareas como la determinación de la polaridad, la resolución de la vaguedad en textos, la clasificación de diferentes lenguajes y la clasificación entre dominios. Para llevar a cabo las primeras cuatro tareas existen enfoques clasificados como: 1) Basados en léxico; 2)

Aprendizaje automático, e 3) Híbridos. El resto de tareas emplean enfoques basados en ontologías. La Figura 1.1 muestra la organización de tareas, subtareas y enfoques de análisis sentimental.

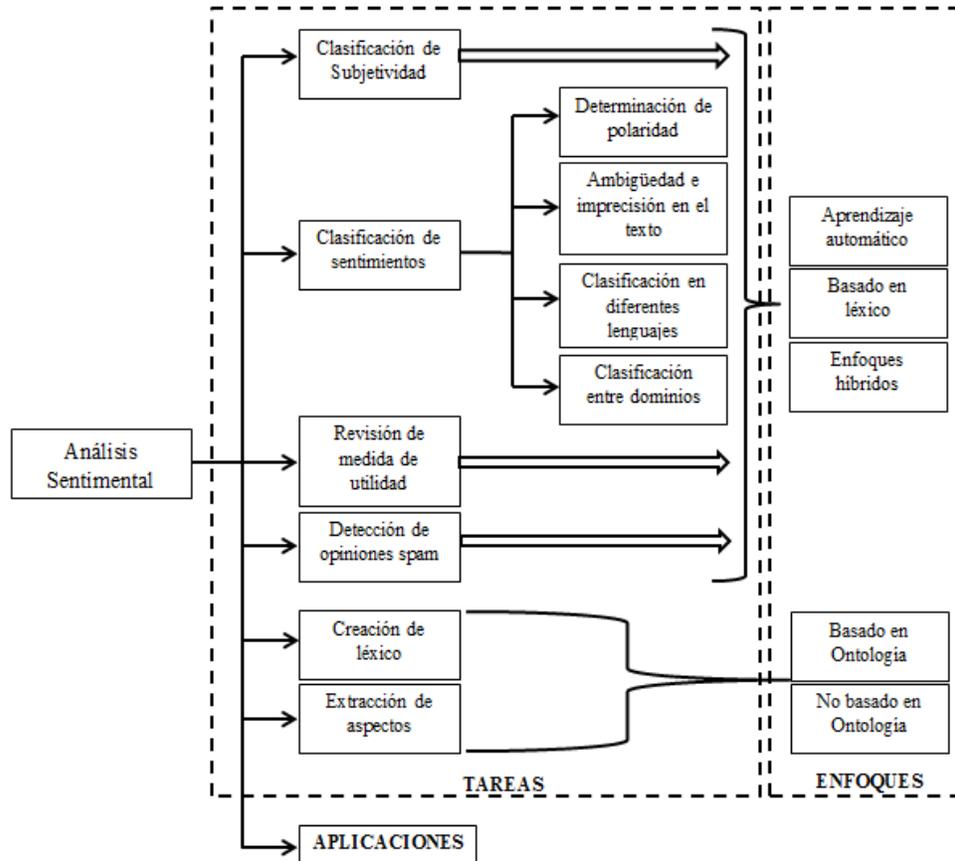


Figura 3.1 Organización del Análisis Sentimental de acuerdo con Ravi y colaboradores.

Por otro lado, el análisis sentimental tiene la capacidad de ser aplicado a diferentes niveles de análisis dependiendo del objetivo de estudio. Aunque existen varias clasificaciones de los niveles de análisis propuestas por varios autores, Serrano et al. [2] mencionó que en su mayoría, dichos autores concuerdan en los siguientes niveles:

- **Nivel de documento:** Se trabaja sobre un documento completo indicando si éste muestra sentimiento positivo, negativo o neutro, para lo cual se considera que el documento trata sobre un mismo tópico.
- **Nivel de oración:** Identifica para cada una de las oraciones que conforman un texto si éstas son positivas, negativas o neutras.

- **Nivel de entidad/aspecto:** Es el tipo de análisis más fino, opera identificando la positividad o negatividad de entidades o aspectos dentro de oraciones.

1.1.1.1 Análisis sentimental aplicado a texto

El análisis sentimental aplicado a texto consiste en la clasificación de documentos de texto en función de la positividad o negatividad del lenguaje utilizado en ellos, así como en la extracción e identificación de los sentimientos, emociones y actitudes reflejadas por el autor de dichos documentos. El análisis sentimental sobre texto es capaz de aplicarse para detectar la actitud que refleja un documento completo, o bien la que refleja cada una de las oraciones o entidades presentes en éste.

En la actualidad, el análisis sentimental aplicado a textos es muy utilizado porque por un lado, se cuenta con una cantidad ingente de información subjetiva proveniente de diferentes plataformas en Internet, como por ejemplo, redes sociales, foros, blogs, reseñas, entre otras; mientras que por otro lado, existe un gran interés por parte de empresas y organizaciones por utilizar dicha información, ya que ésta tiene la capacidad de ser muy útil para que tomen medidas o realicen acciones que les favorezcan. Por otra parte, se cuenta con varios enfoques para realizar análisis sentimental sobre textos. Existen enfoques basados en la presencia de palabras clave no ambiguas como “feliz”, “triste”, “enojado”, entre otras, que representan un sentimiento para clasificar textos, mientras que algunos se basan en la detección de palabras con una afinidad probable con ciertos sentimientos, y otros emplean métodos probabilísticos para encontrar la mejor clasificación [3].

1.1.2 APIs para análisis sentimental sobre texto

Existen muchas APIs (*Application Programming Interface*, Interfaz de Programación de Aplicaciones) que facilitan la realización de las tareas de SA sobre texto, imágenes o voz. Sin embargo, las APIs para SA sobre texto son de las más utilizadas gracias a la gran cantidad de información textual existente en la actualidad. A continuación, en la Tabla 1.1 se presenta el análisis de algunas de ellas.

API	Descripción
Natural Language Understanding	Esta API identifica la polaridad del sentimiento (positivo o negativo) de cualquier texto a nivel de documento, de objetivos específicos, de entidad o de palabra clave. Asimismo, indica a qué persona, lugar o cosa se refiere el texto, encuentra emociones, palabras clave, relaciones, conceptos y autores. Por otra parte, se integra con proyectos en Node.js, Python, Java y Unity; sin embargo, requiere al menos cinco palabras para funcionar y soporta solo mil transacciones por día en su versión sin costo [4].
Bitext	Bitext detecta la polaridad del sentimiento de un texto a nivel de documento o de tuit, así como la intensidad de éste e identifica el tema principal del texto. Se basa en el enfoque de análisis lingüístico profundo, con lo cual el análisis se realiza a nivel de oración y a nivel de frases dentro de la oración. Además, ofrece una precisión de hasta 70% al realizar análisis sentimental y soporta idiomas como: inglés, español, francés, portugués, italiano, alemán, holandés y catalán [5].
Synesketch	Synesketch analiza el contenido emocional de las frases de algún texto en términos de: 1) Emociones, tales como felicidad, tristeza, ira, miedo, asco y sorpresa; 2) Pesos, es decir, cuán intensa es la emoción, y 3) Valencia, lo cual indica si la emoción es positiva o negativa. También, genera gráficos y animaciones que representan las emociones identificadas y facilitan su visualización. Por otra parte, se integra bien con proyectos en Java, sin embargo, el trabajo sobre esta API se detuvo [6].
ToneAPI	Esta API es capaz de detectar más de veinte emociones en un texto, lo cual hace posible la calificación del contenido con una intención emocional en cuestión de segundos. De la misma forma, ToneAPI da automáticamente sugerencias sobre cómo mejorar el contenido para alinearlo con la respuesta emocional deseada [7].
Repustate API	Repustate [®] extrae información semántica y realiza análisis sentimental a nivel de tópico en medios sociales, noticias, blogs, foros, entre otros. Además, reconoce hasta trece idiomas, identifica las entidades a las cuales hace referencia un sentimiento (personas, lugares o negocios) y se integra con proyectos en C#, Java, PHP, Ruby y Python [8].
Aylien	Aylien emplea el procesamiento del lenguaje natural y la recuperación de información para extraer el significado y la percepción de un texto. Permite realizar análisis sentimental en términos de polaridad y subjetividad a nivel de documento y de tuit, clasifica textos, extrae entidades y conceptos, detecta lenguajes y genera resúmenes. Por otra parte, realiza hasta diez mil transacciones por mes en su versión sin costo, pero solo soporta el idioma inglés [9].

API	Descripción
Text-Processing	Esta API lleva a cabo análisis sentimental, extracción de frases, reconocimiento de entidades, etiquetado y lematización. Opera sobre textos escritos en inglés, francés u holandés, y se integra bien con proyectos en Python, Java, Ruby, PHP y Objective-C. Por otra parte, permite hasta cien transacciones diarias y soporta un máximo de ochenta mil caracteres en cada texto. Además, es de código abierto [10].
TheySay PreCeive	La API aborda el análisis de textos, es escalable e independiente de la plataforma. Realiza análisis sentimental capturando una variedad de sentimientos a nivel de documento, de oraciones y de entidades. Asimismo, detecta emociones multifacéticas tales como el enojo, la calma, el miedo, la felicidad, el gusto, la vergüenza, la sorpresa y la certeza. Por otra parte, soporta los idiomas inglés, español y alemán, y se integra bien con proyectos en C#, Java, Node.js, PHP, Python, Ruby, R y Scala [11].
Semantria	Semantria realiza análisis sentimental sobre páginas Web o texto. Se basa en la evaluación semántica, sintáctica y de contexto mediante la combinación de técnicas de aprendizaje automático no supervisado, procesamiento del lenguaje natural y diccionarios. Permite extraer entidades, temas, categorías, intenciones y sentimientos. Por otro lado, soporta veinte mil transacciones en su versión de prueba y se integra con proyectos en Ruby, JavaScript, Node.js, Python, Java, PHP y C++ [12].
Skyttle API	Es un sistema SaaS (<i>Software as a Service</i> , Software como un Servicio) que proporciona servicios de análisis de texto para extraer patrones y almacenarlos en un formato estructurado para un análisis en profundidad. Además, integra complejas funcionalidades de procesamiento de lenguaje natural y extrae palabras clave, entidades con nombre y sentimientos a nivel de frases. Cubre los idiomas inglés, francés, alemán y ruso [13].
DatumBox Machine Learning API	Esta API ofrece una función de análisis sentimental que clasifica documentos o tuits como positivos, negativos o neutros. También, analiza textos e indica si éstos son subjetivos u objetivos, lleva a cabo una clasificación de tópicos, detecta información irrelevante, contenido para adultos, comercial o educativo, identifica el lenguaje, el género, extrae palabras clave y compara la similitud entre documentos [14].

Tabla 1.1 APIs para análisis sentimental sobre texto.

1.1.3 Tareas de análisis sentimental

Para la aplicación del análisis sentimental es necesario llevar a cabo varias tareas esenciales. Aunque las tareas varían dependiendo del autor que las propone, en su mayoría éstas implican el tratamiento de la subjetividad, clasificación sentimental, detección de *spam*, creación de léxico, entre otras. A continuación, en la Tabla 1.2 se presenta la clasificación de tareas propuesta por Ravi y Ravi [1].

Tarea	Descripción
Clasificación de subjetividad	Se define como la clasificación de documentos dependiendo de si expresan subjetividad u objetividad. Por subjetividad se entiende al tipo de información que depende de la percepción de cada persona, por ejemplo, sentimientos, opiniones, emociones, creencias y especulaciones. Por su parte, la objetividad expresa información independiente del sentir o pensar de cada persona.
Clasificación sentimental	Clasifica documentos en dos o más clases con base en la orientación del sentimiento de éste. Para identificar aspectos que permiten llevar a cabo la clasificación se llevan a cabo sub tareas tales como: <ul style="list-style-type: none"> ➤ Determinación de polaridad: Decide si un texto expresa un sentimiento positivo, negativo o neutro. ➤ Resolución de vaguedad: Trata textos que tienen falta de claridad o precisión, y detecta si contienen sarcasmo, ironía, retórica o metáfora. ➤ Clasificación multilingüe: Identifica el idioma en el que se encuentra escrito un texto. ➤ Clasificación de muchos dominios: Detecta el o los dominios sobre los que trata un texto.
Medida de utilidad de las opiniones	Se encarga de verificar y medir la utilidad de las opiniones expresadas, ya que en muchos casos se trata de recursos de mala calidad que no sirven para obtener información subjetiva valiosa.
Detección de <i>Spam</i>	Se refiere a la detección de comentarios falsos, fraudulentos, no genuinos o irrelevantes que en muchas ocasiones se emplean para promover productos de baja calidad o para desacreditar productos de buena calidad.
Creación de léxico	Construcción de un conjunto de términos mediante un proceso iterativo que comienza con una lista de palabras semilla que se va ampliando mediante la incorporación de los sinónimos y antónimos de éstas. Cada vez que se agregan términos a la lista el proceso se repite hasta que no se encuentren términos nuevos.

Tabla 1.2 Principales tareas del análisis sentimental.

1.1.4 Enfoques de análisis sentimental

Los principales enfoques de análisis sentimental se clasifican en enfoques basados en léxico, enfoques de aprendizaje automático e híbridos.

1.1.4.1 Enfoques basados en léxico

Los enfoques basados en léxico se centran en la utilización de una colección de términos sentimentales conocidos y precompilados, a la que se le conoce como léxico. Dicha colección, no solamente incluye términos, sino también frases, expresiones y modismos que se desarrollaron para los géneros tradicionales de comunicación. Además, existen estructuras complejas que se aplican con este enfoque, tales como ontologías o diccionarios que miden la orientación semántica de palabras y frases. Por otra parte, algunas de las desventajas que posee este enfoque son: 1) Muchas veces el léxico empleado es difícil de construir; 2) No consideran el contexto de las palabras, y 3) Generalmente tienen un rendimiento inferior al del aprendizaje automático, aunque superan varias limitantes de éste. Los enfoques basados en léxico se dividen en los que usan diccionarios y los que usan un corpus [2].

1.1.4.1.1 Basado en diccionarios

Se basan en el uso de un conjunto de términos que se recogen y anotan de manera manual, a los cuales se les conoce como semillas. Posteriormente, dicho conjunto crece mediante la búsqueda de sinónimos y antónimos de sus términos en un diccionario [2]. Las palabras que se encuentran se añaden a la lista de semillas y de manera iterativa se vuelve a realizar una búsqueda hasta que ya no se encuentran palabras nuevas. Además, permite inspeccionar manualmente el diccionario formado para eliminar o corregir errores; sin embargo, este tipo de enfoque no es bueno manejando las orientaciones específicas del dominio y del contexto [15].

1.1.4.1.2 Basado en corpus

A diferencia de los enfoques basados en diccionarios, proporciona un conjunto de términos relacionados con un dominio determinado, lo cual resuelve el problema de encontrar palabras con orientaciones específicas del contexto. Su conjunto de términos se

genera mediante patrones sintácticos y semillas que aumentan el número de términos conforme se realiza una búsqueda de palabras relacionadas en un corpus grande. El enfoque basado en corpus se lleva a cabo empleando técnicas estadísticas o semánticas, las cuales se describen con mayor detalle en la Tabla 1.3.

Técnica	Descripción
Estadística	Encuentra patrones de co-ocurrencia y palabras semilla derivando polaridades de adjetivos en un corpus. Dicha polaridad se identifica estudiando la frecuencia de aparición de palabras en un gran corpus anotado, por ejemplo, si la palabra aparece con mayor frecuencia en textos positivos, su polaridad es positiva; si su frecuencia es mayor en textos negativos, su frecuencia es negativa, y si su frecuencia es igual en ambos casos, entonces es neutra [15].
Semántica	Calcula la similitud entre palabras mediante un principio que da valores de sentimientos parecidos a palabras cercanas, por ejemplo, WordNet, una base de datos léxica que agrupa palabras en inglés, calcula las polaridades del sentimiento empleando diferentes tipos de relaciones semánticas entre las palabras [15].

Tabla 1.3 Técnicas empleadas para el enfoque basado en corpus.

1.1.4.2 Enfoques basados en aprendizaje automático

El aprendizaje automático es una rama de la inteligencia artificial que tiene por objetivo conseguir que las computadoras desarrollen cierto comportamiento mediante el suministro de información en forma de ejemplos. El enfoque de aprendizaje automático trata el análisis sentimental como un problema de clasificación de texto regular, empleando características sintácticas y/o lingüísticas de éste. Considerando que se cuenta con un conjunto de datos $D = \{X_1, X_2, \dots, X_n\}$ donde cada tupla pertenece a una clase en particular, el aprendizaje automático emplea algoritmos que generan un modelo que permite predecir la clase a la que pertenecen las tuplas nuevas o desconocidas [15].

Además, el aprendizaje automático tiene una fuerte relación con el procesamiento del lenguaje natural porque considera aspectos como los términos y la frecuencia con la que aparecen éstos en el texto, las palabras con propiedades gramaticales similares, las negaciones, las dependencias sintácticas, entre otros. Por otra parte, se divide en aprendizaje automático supervisado y no supervisado. Es posible obtener éxito con ambas

categorías en análisis sentimental, sin embargo, es necesario tener consideraciones como el tipo de información, el dominio, el idioma, entre otros [2].

1.1.4.2.1 Aprendizaje supervisado

El aprendizaje automático supervisado se distingue de su contraparte no supervisada, porque requiere de la utilización de conjuntos de datos de entrenamiento previamente etiquetados, es decir, requiere que existan clases bien definidas e identificables en el conjunto de datos. Generalmente, este tipo de aprendizaje da mejores resultados para análisis sentimental en cuanto a precisión; sin embargo, la necesidad que tiene de contar con datos etiquetados y la dependencia hacia el dominio que presenta, son algunas de sus desventajas. Los principales clasificadores que utiliza el aprendizaje automático supervisado son los árboles de decisión, los clasificadores lineales, basados en reglas, y los probabilísticos [15].

1.1.4.2.1.1 Árboles de decisión

Son clasificadores que descomponen un conjunto de datos en una estructura jerárquica que usualmente se representa en forma de grafo, para lo cual, emplea como criterio de división una condición sobre el valor de los datos. Dicha condición, generalmente se refiere a la presencia o ausencia de una o más palabras determinadas. De esta forma, el conjunto de datos se divide una y otra vez de manera recursiva, hasta que los nodos inferiores del grafo cuentan con un número mínimo de tuplas con las que se lleva a cabo la clasificación. Además, se cuenta con diferentes tipos de divisiones, las cuales son:

- **División de un solo atributo:** Se usa la presencia o ausencia de ciertas palabras o frases en un nodo del árbol como criterio de división.
- **División de atributos múltiples basada en la similitud:** Centra su criterio de división en la similitud entre documentos y grupos de palabras frecuentes.
- **División de atributos múltiples basada en la discriminación:** Realiza la división empleando discriminantes.

Por otra parte, es importante mencionar que los árboles de decisión que se usan con más frecuencia son ID3 (*Iterative Dichotomiser 3*, Dicotomizador Iterativo 3), C4.5 y CART (*Classification And Regression Tree*, Árbol de Clasificación y Regresión) [15].

1.1.4.2.1.2 Clasificadores lineales

Identifican la clase a la que pertenece un determinado objeto usando las características de éste, para lo cual, llevan a cabo la clasificación basándose en el valor de una combinación lineal de sus características. Una combinación lineal se refiere a una expresión matemática que toma pares de elementos de algunos conjuntos en particular, los multiplica entre sí y los suma con el resto de los pares de elementos. Se cuenta con muchos tipos de clasificadores lineales, pero los más conocidos son SVM (*Support Vector Machine*, Máquina de Soporte Vectorial) y Redes Neuronales [15].

SVM es un clasificador lineal que dado un conjunto de datos con dos clases identificables, se encarga de encontrar un separador lineal o hiperplano que divida dichas clases de la mejor forma posible [16]. SVM selecciona el hiperplano que consigue el mayor margen de separación entre ambas clases, pues se considera el más preciso. En la Figura 1.2 es posible apreciar varios hiperplanos, donde el que consiguió el mayor margen de separación fue el hiperplano A [15].

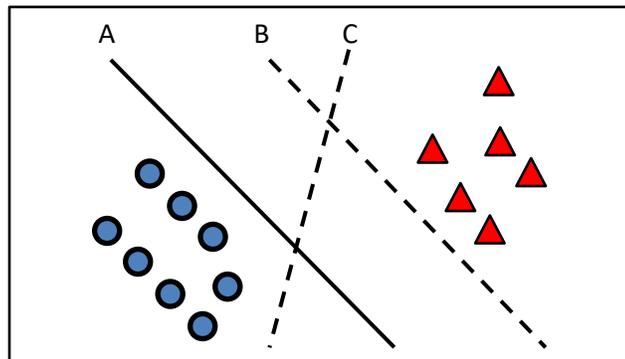


Figura 3.2 Ejemplo de separadores lineales empleando SVM.

Por otra parte, las redes neuronales se forman por componentes conectados ponderados a los que se les conoce como neuronas. En su fase inicial, los pesos asociados a cada neurona se ajustan para hacer una predicción correcta de las clases. Este proceso se repite hasta obtener el menor margen de error posible. Las redes neuronales son muy difíciles de interpretar por el ser humano y requieren mucho tiempo de entrenamiento, sin embargo, son muy utilizadas gracias a su capacidad de clasificar datos no entrenados y a su alta tolerancia al ruido [16].

1.1.4.2.1.3 Clasificadores basados en reglas

Utilizan un conjunto de entrenamiento con datos previamente etiquetados para calcular un grupo de reglas mediante las cuales es posible clasificar documentos. Si las reglas generadas se entienden como una igualdad, la parte izquierda de ésta consiste en condiciones expresadas disyuntivamente, es decir, de manera que se tiene que elegir entre una u otra, mientras que la parte derecha representa una clase existente en el conjunto de entrenamiento. Las condiciones empleadas en estas reglas se relacionan con la presencia de un término en particular.

Existen varios criterios para generar reglas durante la fase de entrenamiento; sin embargo, los dos criterios más comunes son el soporte y la confianza. El soporte es el número absoluto de instancias relevantes para una regla que existen en el conjunto de entrenamiento, mientras que la confianza es la probabilidad de que se cumpla la parte derecha de la regla si se cumple la parte izquierda [15].

1.1.4.2.1.4 Clasificadores probabilísticos

Los clasificadores probabilísticos estiman un conjunto de parámetros que muestran la probabilidad de que un documento pertenezca a una determinada clase dado un conjunto de datos [17]. Dichos parámetros permiten asignar nuevos datos a las clases que correspondan. Este tipo de clasificadores también se conoce como clasificadores generativos y algunos de los más populares son NB (*Naive Bayes*, Bayesiano Ingenuo), BN (*Bayesian Network*, Red Bayesiana) y ME (*Maximum Entropy*, Entropía Máxima) [18]. El clasificador NB calcula la probabilidad de que un documento pertenezca a una clase basándose en la distribución de las palabras de éste. Trabaja con el método BOW (*Bag Of Words*, Bolsa de Palabras), el cual ignora la posición de las palabras en el documento. NB emplea el teorema de *Bayes*, el cual se aprecia mediante la Fórmula 1.1.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Fórmula 1.1 Teorema de Bayes.

P (A) es la probabilidad de una etiqueta, P (B|A) es la probabilidad de que un registro se clasifique como una etiqueta en particular y P (B) es la probabilidad de que se produzca un

conjunto de características determinado. Por otra parte, a pesar de que NB se considera el clasificador más simple y uno de los más utilizados, tiene el inconveniente de asumir siempre que las variables que se usan para clasificar son independientes entre sí, por lo cual el teorema de *Bayes* se reescribe como muestra la Fórmula 1.2 [15].

$$P(A|B) = \frac{P(A) * P(B1|A) * ... * P(Bn|A)}{P(B)}$$

Fórmula 1.2 Teorema de Bayes reescrito.

Por su lado, las redes bayesianas son un grafo acíclico dirigido, que posee nodos que representan variables aleatorias, y aristas que representan dependencias condicionales (Figura 1.3). Las redes bayesianas enfrentan la suposición que hace NB de que las variables empleadas para clasificar son independientes entre sí, por lo que se considera un modelo completo. No obstante, pese a que obtienen buenos resultados, la complejidad computacional que tienen es muy costosa, lo cual genera un bajo índice de utilización [15].

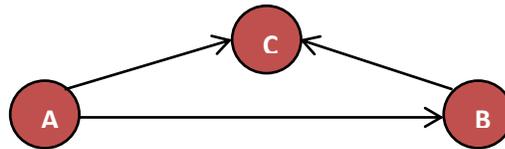


Figura 3.3 Ejemplo de red bayesiana.

De manera similar, otro clasificador importante es el clasificador ME o también conocido como clasificador exponencial condicional, el cual utiliza codificación para transformar en vectores un conjunto de características etiquetadas. Posteriormente, calcula los pesos de cada una de las características localizadas en dicho vector que son capaces de combinarse para encontrar la clase o etiqueta más probable para un conjunto de características [15]. Por otra parte, es importante mencionar que ME permite combinar evidencias estadísticas provenientes de diversas fuentes de manera flexible [17].

1.1.4.2.2 Aprendizaje no supervisado

El aprendizaje automático no supervisado supera la dependencia de los conjuntos de datos etiquetados de su contraparte supervisada, ya que es capaz de categorizar nuevos registros o documentos mediante técnicas de agrupamiento cuando se cuenta con un conjunto de datos inicial sin etiquetas de clase. No obstante, el aprendizaje automático no supervisado generalmente da resultados inferiores a los de los métodos supervisados [15].

1.1.4.3 Enfoques híbridos

Como se mostró en secciones anteriores, tanto el enfoque de aprendizaje automático como el enfoque basado en léxico generan buenos resultados para análisis sentimental dependiendo de diferentes circunstancias, por ejemplo, la naturaleza de los datos, el dominio, los recursos existentes, entre otros; sin embargo, tienen limitantes o dificultades. Por ello, surgieron los enfoques híbridos, los cuales combinan el enfoque de aprendizaje automático con el basado en léxico. Su objetivo es emplear las ventajas de ambos enfoques, mitigando sus inconvenientes [15].

1.1.5 Lenguajes de programación para la Web

El desarrollo de páginas y aplicaciones Web requiere de la utilización de algún lenguaje de programación para la Web. Los lenguajes de programación para la Web básicamente se describen como lenguajes formales a través de los cuales es posible definir un conjunto de instrucciones que la computadora es capaz de procesar. Existen dos tipos de lenguajes de programación para la Web: 1) Lenguajes de programación de lado del cliente, que son lenguajes que se interpretan directamente por un *browser*, y 2) Lenguajes de programación de lado del servidor, que llevan a cabo sus diferentes procesos en un servidor para posteriormente enviar el resultado del proceso al *browser* en un formato entendible por éste.

1.1.5.1 PHP

PHP es un lenguaje de programación de alto nivel, de propósito general, de código abierto y que posee tipificación dinámica y débil. PHP es especialmente adecuado para el desarrollo Web y es posible incrustar su código en HTML (*HyperText Markup Language*, Lenguaje de Marcado de Hipertexto). Además, PHP se ejecuta del lado del servidor, por lo que la respuesta que se genera y se presenta al cliente no conlleva el código subyacente de ésta [19]. Dado que PHP está principalmente orientado a la programación del lado del servidor, permite llevar a cabo acciones como recopilar datos de formularios, generar páginas con contenidos dinámicos, enviar o recibir *cookies*, entre otras. Algunas de sus características restantes se listan a continuación:

- Capacidad de conexión con la mayoría de Sistemas Gestores de Bases de Datos.
- Permite añadir módulos para extender su potencial.
- Posee una amplia documentación en su sitio Web oficial.
- Permite aplicar técnicas de programación orientada a objetos.
- Posee manejo de excepciones.

1.1.5.2 Java

Java es un lenguaje de programación de alto nivel, orientado a objetos, multiplataforma con tipificación estática fuerte. Además, se caracteriza por ser un lenguaje robusto, rápido, seguro, fiable, distribuido y de alto rendimiento. También, Java es uno de los lenguajes de programación más populares actualmente, especialmente para el desarrollo de aplicaciones Web basadas en la arquitectura Cliente-Servidor. Por otro lado, existe una amplia disponibilidad de bibliotecas y otros recursos para este lenguaje dado que constantemente se prueba, ajusta y amplía por una comunidad de desarrolladores y arquitectos de aplicaciones Java [20].

1.1.6 Sistemas Gestores de Bases de Datos

Los DBMS (*DataBase Management System*, Sistema Manejador de Bases de Datos) son un tipo de software a gran escala que permite controlar y administrar bases de datos. Los DBMS se utilizan para establecer, usar y mantener bases de datos. Además, permiten administrar y controlar las bases de datos de manera uniforme con el fin de garantizar la seguridad y la integridad. Este tipo de software proporciona una variedad de funciones para garantizar que las aplicaciones múltiples y los usuarios tengan la capacidad de construir, modificar e interrogar la base de datos con diferentes métodos [21].

1.1.6.1 MySQL[®]

Es el Sistema Gestor de Bases de Datos relacional de código abierto más popular del mundo. De hecho, muchas de las más grandes organizaciones a nivel mundial y organizaciones que presentan un rápido crecimiento, tales como Facebook[®], Google[®], Adobe[®] y Alcatel[®] emplean MySQL[®], ya que permite ahorrar tiempo y dinero en sitios

Web con gran volumen, sistemas críticos y software empaquetado. MySQL[®] proporciona un servidor de Bases de Datos SQL (*Structured Query Language*, Lenguaje Estructurado de Consultas) que se diseñó para entornos de producción críticos y con alta carga de trabajo. Posee una versión comunitaria que se desarrolló bajo la licencia GPL (*General Public License*, Licencia Pública General), es multiplataforma, tiene un bajo consumo de recursos, veloz, seguro y cuenta con el respaldo de una gran comunidad de desarrolladores [22].

1.1.7 Redes Sociales

Las redes sociales son sitios de Internet que se conforman por comunidades o grupos de personas con intereses similares, en donde es posible que los miembros intercambien información entre sí. El objetivo principal de las redes sociales es interrelacionar a las personas a través de las diferentes herramientas de las que disponen. Muchas veces las redes sociales están formadas con personas que tienen parentesco entre sí, que pertenecen al mismo trabajo o tienen los mismos pasatiempos. Además, mientras algunas redes sociales son generales, es decir, que no están dirigidas a un tipo específico de usuarios o un tema en concreto, algunas otras son específicas y se dirigen a un público determinado. En la Tabla 1.4 se describen brevemente las dos redes sociales con mayor impacto en la actualidad [23].

Red social	Descripción
Facebook[®]	Facebook [®] es la red social más popular en la actualidad [23] y consiste en un sitio en línea en el que las personas disponen de un perfil donde comparten sus actividades diarias con amigos o familiares, así como sus fotografías, su información personal, se unen a grupos, indican si les gusta algo, comentan diferentes contenidos y envían mensajes privados [24].
Twitter[®]	Twitter [®] es una red social que permite que sus usuarios compartan mensajes de texto no mayores a 140 caracteres a los que se les conoce como tuits. También, es posible que algún usuario comparta tuits de otros usuarios, marque los tuits que le gustan e indique de qué cuentas le interesa observar contenido [25]. Además, identifica temas que demuestran un comportamiento de tendencia entre los usuarios [26].

Tabla 1.4 Redes sociales con mayor impacto en la actualidad.

1.1.8 APIs para redes sociales

Muchas de las redes sociales que existen en la actualidad proporcionan APIs que facilitan el desarrollo relacionado con éstas. Aunque las APIs existentes son distintas entre sí, ofrecen diferentes funcionalidades y tienen distintas limitantes, de manera general permiten enviar y recibir información proveniente de las redes sociales. A continuación, se describen APIs proporcionadas por las dos redes sociales con mayor impacto en la actualidad, Facebook® y Twitter®.

1.1.8.1 Graph API

Es una API basada en HTTP (*Hypertext Transfer Protocol*, Protocolo de Transferencia de Hipertextos) de nivel inferior que permite publicar y extraer información de Facebook®. Entre las distintas tareas que permite realizar se encuentran consultar datos, publicar información, administrar anuncios, subir fotos y llevar a cabo varias tareas más en Facebook®. La Graph API consta de: 1) Nodos, elementos como usuarios, fotos, páginas, comentarios, entre otros; 2) Perímetros, que son las conexiones entre los nodos, y 3) Campos, que es la información sobre los nodos, por ejemplo, el cumpleaños de una persona o el nombre de una página.

Además, es importante mencionar que dado que la Graph API se basa en HTTP, cualquier lenguaje que cuente con una biblioteca HTTP es compatible con ella. Existen SDKs (*Software Developer Kits*, Kit de Desarrollo de Software) oficiales desarrollados por Facebook® para trabajar con la Graph API mediante iOS®, Android®, JavaScript y PHP, así como algunos otros desarrollados por terceros [27].

1.1.8.2 TwitterOAuth

TwitterOAuth es la biblioteca de PHP más popular para usar con la API de Twitter® que proporciona un par de funciones genéricas mediante las cuales es posible recibir y enviar información a dicha red social. A través de estas funciones garantiza que no sea necesario actualizar la biblioteca cada vez que se le añadan nuevas características a la API de Twitter®. Entre las acciones que son posibles llevar a cabo se encuentran obtener la lista de seguidores de una persona, publicar tuits, enviar mensajes directos, extraer los

comentarios de un tuit en específico, obtener información del perfil de un determinado usuario, buscar tuits, entre muchas otras. También, es importante aclarar que dada la popularidad de esta API para trabajar con PHP existe numerosa documentación y ejemplos sobre ella [28].

1.1.9 API para la construcción de gráficas

Las gráficas tienen una gran importancia porque facilitan la comprensión, interpretación y utilización de la información, por lo cual, son ampliamente utilizadas en diversos ámbitos como lo son la ingeniería, finanzas, estadística, entre otras. Dada la importancia de las gráficas, existen distintos tipos de APIs que ofrecen a los programadores facilidades para la creación de distintos tipos de gráficas, por ejemplo, gráficas de barras, de líneas, de puntos, de pastel, de columnas, entre otras.

1.1.9.1 Plotly.js

Plotly.js es una biblioteca de JavaScript gratuita y de código abierto que permite crear gráficos sofisticados e interactivos para finanzas, ingeniería y ciencias. Plotly.js permite construir hasta veinte diferentes tipos de gráficas de alto nivel, entre los que se incluyen gráficos tridimensionales, gráficos estadísticos y mapas SVG (*Scalable Vector Graphics*, Gráficos Vectoriales Escalables). Algunos ejemplos de gráficas que es posible construir mediante esta biblioteca son las gráficas de líneas, barras, pastel, histogramas, mapas de calor, entre otras. Además, facilita la interacción con las gráficas al permitir llevar a cabo acciones como aumentar y disminuir el tamaño de la gráfica, mostrar u ocultar ejes, descargar la gráfica como imagen, entre otras. Asimismo, cuenta con amplia documentación que orienta no solo sobre la creación de las gráficas, sino también sobre la posibilidad de cambiar aspectos como el color, la posición de la leyenda, el título de los ejes, entre otros [29].

1.2 Situación tecnológica, económica y operativa de la empresa

El Instituto Tecnológico de Orizaba ofrece carreras a nivel licenciatura tales como Ingeniería Eléctrica, Ingeniería Electrónica, Ingeniería en Sistemas Computacionales, Ingeniería Química, Ingeniería Mecánica, Ingeniería en Gestión Empresarial, Ingeniería

Industrial e Ingeniería Informática. Esto limita el acceso a las tecnologías de información, sin embargo, existe la posibilidad de solicitar el uso de algún laboratorio de forma temporal sin garantizar la petición.

A su vez, el instituto ofrece carreras a nivel maestría y doctorado, como son Maestría en Ingeniería Electrónica, Maestría en Ingeniería Industrial, Maestría en Sistemas Computacionales, Maestría en Ingeniería Administrativa, Maestría en Ciencias en Ingeniería Química y Doctorado en Ciencias de la Ingeniería. Particularmente, la Maestría en Sistemas Computacionales cuenta con un laboratorio para cada una de sus líneas de trabajo, las cuales son: 1) Ingeniería de Software, y 2) Tecnologías Web.

1.3 Planteamiento del problema

En la actualidad, la realización de actividades en línea tales como videoconferencias, comercio electrónico, interacción en redes sociales, participación en foros de discusión, entre otros, aumenta de manera exponencial. Particularmente, las actividades en redes sociales como Twitter® y Facebook® generan una cantidad ingente de información subjetiva, la cual expresa opiniones, actitudes, emociones y sentimientos de las personas hacia algún producto, servicio o tema en particular.

Por otro lado, existe una gran necesidad de contar con medios para analizar y tratar la información subjetiva (emociones y sentimientos) por parte de empresas, organizaciones y particulares, ya que comprender este tipo de información les brinda beneficios tales como encontrar nuevas estrategias empresariales, planificar campañas publicitarias más exitosas, mejorar la reputación de productos, personas o servicios, y ajustarlos de acuerdo con las actitudes que muestra la gente respecto a ellos. Sin embargo, el tratamiento de este tipo de información mediante herramientas y técnicas tradicionales es ineficiente. Por ello, se propone la utilización de técnicas de análisis sentimental para el desarrollo de un módulo capaz de tratar textos, extraer las emociones que contienen, determinar si son positivos, negativos o neutros, así como medir la intensidad de su sentimiento.

1.4 Objetivo general y específico

En esta sección se presentan el objetivo general y los principales objetivos específicos de este trabajo.

1.4.1 Objetivo general

Desarrollar un módulo que permita detectar/obtener sentimientos y emociones mediante técnicas híbridas de análisis sentimental.

1.4.2 Objetivos específicos

- Analizar las diferentes APIs para análisis sentimental para el desarrollo del módulo.
- Diseñar el conjunto de funcionalidades o prestación de servicios que ofrecerá el módulo para la obtención/detección de emociones.
- Desarrollar las funcionalidades o prestación de servicios del módulo para la obtención/detección de emociones basados en técnicas híbridas de análisis sentimental.
- Realizar la prueba de concepto mediante la validación del módulo a partir de dos casos de estudio en el ámbito de la salud.

1.5 Justificación

Diariamente una gran cantidad de personas expresan las opiniones o sentimientos que tienen respecto a diversos temas mediante la utilización de plataformas en línea como Facebook® y Twitter®, por ejemplo, realizan comentarios respecto a un nuevo celular que salió al mercado, al servicio de televisión de paga, a un partido político, a una reforma energética o a un cantante, por citar algunos. Este tipo de información, conocida como información subjetiva, es muy valiosa porque permite conocer cómo se encuentra el mercado, cuál es la posición política de una región del país, qué contenido tiene mayor demanda o el grado de satisfacción sobre algún producto, entre muchos otros.

El análisis sentimental tiene la capacidad de tratar la información subjetiva proveniente de diversas fuentes y en diferentes formatos para extraer la polaridad sentimental que contiene, así como la intensidad del sentimiento y las emociones que

expresa a través del uso de diversas técnicas. De esta manera, es posible utilizar dicha información para llevar a cabo acciones que representen un beneficio.

Con el desarrollo de un módulo capaz de detectar, extraer y clasificar las emociones y sentimientos que las personas expresan en plataformas como Twitter® y Facebook® respecto a algún tópico en particular, se brinda la posibilidad de obtener múltiples beneficios a corto, mediano y largo plazo. Dichos beneficios, de manera general implican la posibilidad de que empresas, organizaciones o particulares, tomen medidas o realicen acciones que les favorezcan con base en el análisis de la información subjetiva, lo cual conlleva, por ejemplo, a que propongan nuevas estrategias de negocios o campañas publicitarias que resulten más exitosas, a optimizar los servicios que ofrecen, a incrementar su clientela, a mejorar la satisfacción de clientes o usuarios, a enriquecer la reputación positiva de algún producto o inclusive de alguna persona, entre otros. Asimismo, es importante mencionar que los beneficios que proporciona el módulo van en relación del ámbito donde se utilice.

Capítulo II. Estado de la práctica

En este capítulo se presenta un resumen de los trabajos más significativos relacionados con análisis sentimental aplicado a textos, los cuales permitieron obtener una gran cantidad de información valiosa, relacionada con este proyecto.

2.1 Trabajos relacionados con análisis sentimental

Diariamente las personas expresan sus sentimientos a través de redes sociales, blogs y reseñas en línea, por lo cual los propietarios de negocios y empresas de publicidad emplean análisis sentimental sobre dichos medios para descubrir nuevas estrategias de negocio y campañas publicitarias. Sin embargo, con el crecimiento exponencial del uso de los medios mencionados anteriormente, surgió la necesidad de encontrar enfoques más precisos para análisis sentimental. Por esta razón, Tripathy et al. [30] propusieron un enfoque capaz de extraer, clasificar y predecir sentimientos, el cual se basó en técnicas de aprendizaje automático supervisado sobre un conjunto de datos etiquetados y utilizó algoritmos como NB y SVM. El enfoque propuesto alcanzó una precisión superior en 3% y 8% a la de otros enfoques populares empleando NB y SVM, respectivamente.

De manera similar, Ceron et al. [31] consideraron que con la gran cantidad de datos en la Web con que se cuenta en la actualidad, se tiene la capacidad de realizar tareas como explicar y anticipar diferentes eventos. Sin embargo, aunque el análisis sentimental permite llevar a cabo dichas tareas, ningún método de análisis sentimental es confiable al 100%, por lo cual, se requiere encontrar nuevos métodos que obtengan mejores resultados. Por ello, se propuso iSA (*Integrated Sentiment Analysis*, Análisis de Sentimiento Integrado), un algoritmo diseñado para análisis sentimental en redes sociales y Web 2.0, que se basa en la codificación manual supervisada y explota un enfoque de reducción dimensional que lo hace escalable, rápido, eficiente, estable y preciso. Los resultados de la evaluación de iSA mostraron que pese a que su precisión fue aceptable para muestras de datos grandes en comparación con otros algoritmos conocidos, disminuyó con muestras de datos pequeñas.

El análisis sentimental emplea métodos basados en aprendizaje automático o en léxico; no obstante, así como ambos métodos poseen fortalezas, también tienen muchas debilidades. En consecuencia, surgió la necesidad de crear enfoques para maximizar las fortalezas de estos métodos y suprimir sus debilidades. Como solución Khan et al. [32] propusieron eSAP (*Enhanced Sentiment Analysis and Polarity*, Análisis de Sentimientos y Polaridad Mejorados), un marco de trabajo que permite clasificar la polaridad de las opiniones empleando un vocabulario con puntajes de sentimientos y un modelo de aprendizaje de máquinas de soporte vectorial. Dicho marco de trabajo consta de dos pasos: 1) Adquisición y pre-procesamiento de datos, y 2) Clasificación basada en el vocabulario. Tras evaluar eSAP sobre siete conjuntos de datos de diversos dominios, se alcanzó una precisión promedio de 80.83%, un resultado superior al del clasificador para minería de opinión llamado SentiWordNet [33].

Asimismo, Khan et al. [34] consideraron que el aprendizaje automático da mejores resultados para el análisis sentimental, pero tiene obstáculos importantes para su aplicación, mientras que los algoritmos basados en léxico tienen menos limitantes, pero su rendimiento es bajo. Como consecuencia se propuso SWIMS (*Semi-supervised feature Weighting and Intelligent Model Selection*, Función Semi-supervisada de Ponderación y Selección Inteligente de Modelos), un marco de trabajo que enfrenta los obstáculos de los algoritmos de aprendizaje supervisado y mejora el desempeño de los algoritmos basados en léxico. Además, emplea SVM para la detección de la subjetividad y un enfoque léxico para la selección simultánea de la polaridad. SWIMS se evaluó sobre tres conjuntos de datos y obtuvo una precisión de hasta 85.50%, un porcentaje superior al obtenido por otros algoritmos sobre los mismos conjuntos de datos.

De la misma forma, cada vez más empresas centran sus campañas de mercadotecnia en la utilización de aprendizaje automático supervisado y/o no supervisado para realizar análisis sentimental sobre textos en plataformas en línea. De acuerdo con Fernández-Gavilanes et al. [35], el aprendizaje automático supervisado es el más utilizado. Sin embargo, éste tiene dependencia del dominio y necesita conjuntos de datos etiquetados que son difíciles de obtener. Por lo cual, se propuso un método no supervisado para análisis sentimental en tuits y reseñas que no necesita entrenamiento previo o específico. Además, dicho método se basó en

el análisis de dependencias para determinar la polaridad de los textos y en un léxico de sentimiento que considera la estructura especial y el contenido lingüístico de los mensajes. El método propuesto se evaluó con dos conjuntos de datos y obtuvo resultados superiores a otros métodos no supervisados sobre los mismos conjuntos de datos.

No obstante, Appel et al. [36] discutieron que aunque en su mayoría se abordan técnicas de aprendizaje automático para el análisis sentimental, existen otros enfoques capaces de tener éxito también, por ejemplo, el enfoque basado en conjuntos difusos. Dicho enfoque es capaz de realizar análisis sentimental por su capacidad para lidiar con la vaguedad y la incertidumbre; sin embargo, la utilización de este enfoque muchas veces se ignora. Por ello, se propuso un enfoque híbrido para análisis sentimental en el nivel de oración que utiliza conjuntos difusos, técnicas de procesamiento del lenguaje natural y un léxico sentimental para estimar la polaridad de la orientación semántica y la intensidad sentimental de las oraciones. El enfoque propuesto se evaluó mediante conjuntos de datos integrados por tuits y reseñas de películas. Los resultados fueron superiores a los alcanzados por los algoritmos NB y ME usados aisladamente.

Un entorno muy importante para el análisis sentimental que requiere atención especial en la actualidad es YouTube[®]. Severyn et al. [37] discutieron que YouTube[®] es un entorno multimodal, multilingüe, multicultural y de múltiples dominios, en donde existe una gran cantidad de contenido generado diariamente. Dicho contenido a menudo expresa opiniones que afectan la reputación de una persona, organización o producto específico. Por lo tanto, se necesitan modelos robustos de minería de opiniones que manejen el ruido de los datos y se adapten fácilmente a un nuevo dominio o lenguaje. Por ello, se presentó STRUCT, un modelo basado en una estructura sintáctica que predice tipos de comentarios, polaridad (positivo, negativo o neutral), distingue si la polaridad se dirige hacia el video o hacia su contenido, se adapta a diferentes dominios y funciona para los idiomas inglés e italiano. En comparación con los modelos supervisados más usados, STRUCT obtuvo una precisión superior cuando se aplicó en un mismo dominio hasta de 2.6% y 3% para italiano e inglés, respectivamente, así como hasta en un 4% cuando se aplicó en diferentes dominios.

En el mismo sentido, un entorno bastante importante para análisis sentimental en la actualidad es Twitter®; sin embargo, los principales enfoques existentes para el análisis sentimental tienen las siguientes limitaciones: 1) Los enfoques de aprendizaje supervisado requieren datos de entrenamiento, dependen del dominio y necesitan re-entrenamiento para nuevos datos, por lo cual son poco adecuados para Twitter® dada la gran variedad de temas surgiendo constantemente, y 2) Los enfoques basados en léxico no consideran el contexto de las palabras y se basan en diccionarios costosos de construir, por lo cual son poco adecuados, ya que Twitter® contiene muchas expresiones mal formadas y/o coloquiales. Por ello, Saif et al. [38] propusieron SentiCircles, un enfoque basado en léxico para análisis sentimental a nivel de entidad (tuits sobre un tema) y de tuit individual. SentiCircles identifica el sentimiento de las palabras considerando el contexto de éstas mediante patrones de co-ocurrencia de palabras en un texto, sugiriendo que las palabras que cumplen dicho patrón tienen cierta relación semántica. SentiCircles se evaluó en tres conjuntos de datos y a nivel de entidad superó significativamente a otros enfoques, mientras que a nivel de tuit fue superior en un 4-5% en dos conjuntos de datos e inferior en un 1% en el tercero.

Sin duda, un reto importante para el análisis sentimental es explicar la brecha semántica entre las polaridades de los términos capturados por un léxico y las polaridades de los términos en un contexto específico (polaridad contextual). Existen varios recursos basados en léxico para análisis sentimental, sin embargo, su rendimiento es inferior al de los niveles aceptables, ya que la polaridad de un término difiere de la polaridad según su contexto. Como solución, Muhammad et al. [39] presentaron SmartSA, un sistema de clasificación de sentimientos basado en léxico, para diferentes tipos de medios sociales, el cual integra estrategias para capturar la polaridad contextual considerando la interacción de los términos con su contexto local y global. Los resultados de SmartSA sobre tres conjuntos de datos obtenidos de redes sociales fueron superiores en promedio a otros algoritmos de aprendizaje automático supervisado. Asimismo, se obtuvo un rendimiento superior sobre dos de los tres conjuntos de datos en comparación con SentiStrength [40], un sistema de vanguardia para clasificación de sentimientos en redes sociales.

Por otro lado, en la actualidad el texto transcrito, es decir, el lenguaje convertido a texto mediante traducción automática, se investiga ampliamente para análisis sentimental. No obstante, este tipo de texto presenta dificultades como ruido, gramáticas erróneas o falta de signos de puntuación, que afectan la precisión de los enfoques tradicionales de análisis sentimental. Para atender las problemáticas anteriores, Katz et al. [41] presentaron ConSent (*Context-based Sentiment Analysis*, Análisis Sentimental basado en Contexto), un enfoque para análisis sentimental aplicable a textos regulares (adheridos a reglas gramaticales) y a textos con un alto nivel de ruido. ConSent considera el contexto de los términos, no se apoya en estructuras gramaticales y permite una fácil integración de información de fuentes adicionales. El enfoque propuesto se evaluó en tres conjuntos de datos y, en comparación con enfoques tradicionales, obtuvo una precisión superior en dos conjuntos de datos y ligeramente inferior en el tercero. Asimismo, se evaluó en dos conjuntos de datos de llamadas transcritas, en donde obtuvo una precisión superior a los enfoques tradicionales.

Las opiniones expresadas en la Web tienen mucha importancia en el dominio turístico, por ejemplo, las opiniones sobre servicios hoteleros ayudan tanto a los clientes que buscan alojamiento, como a los hoteles que desean mejorar sus servicios. La utilidad de las opiniones generó la necesidad de sistemas que faciliten a los clientes la toma de decisiones, y que ayuden a los hoteles a que mejoren sus servicios mediante la extracción de opiniones sobre hoteles desde la Web. Por ello, Bucur [42] propuso una plataforma para extraer y clasificar opiniones de reseñas de hoteles del popular sitio Web de viajes TripAdvisor®. Dicha plataforma utiliza un método no supervisado y un recurso léxico para extraer opiniones. Además, consta de dos módulos: 1) Adquisición de contenidos, el cual recolecta las reseñas de TripAdvisor®, y 2) Análisis, que pre-procesa los datos extraídos e implementa la minería de opinión. La plataforma se evaluó utilizando tres mil reseñas de hoteles y alcanzó una precisión de hasta 76.5%, lo cual refleja un rendimiento aceptable para un método no supervisado, pero bajo en comparación con algoritmos supervisados.

Por otro lado, Ortigosa et al. [43] comentaron que los factores emocionales afectan la motivación de los estudiantes y el resultado del proceso de aprendizaje. Por lo tanto, detectar y gestionar información sobre las emociones de los estudiantes es bastante útil para los entornos

de aprendizaje electrónico, ya que les permite a dichos entornos proporcionar recomendaciones sobre actividades o contenidos adaptados al estado emocional de los estudiantes. Comúnmente se utilizan cuestionarios para obtener información; sin embargo, los estudiantes consideran que este tipo de procedimiento es aburrido, tardado y manipulable. Por ello, dado que en la actualidad las redes sociales son el medio donde los estudiantes expresan mayormente sus sentimientos, se planteó un método para análisis sentimental en Facebook[®] que extrae información sobre el estado emocional de los estudiantes. Dicho método se basó en un enfoque híbrido que combina técnicas de aprendizaje automático y técnicas basadas en léxico. Los resultados obtenidos a través de este enfoque mostraron que es factible realizar análisis sentimental en Facebook[®] con alta precisión (83.27%).

Otro contexto de uso del análisis sentimental en los años recientes es la política, ya que de acuerdo con Gull et al. [44] en Pakistán algo muy común es que las personas muestren su opinión sobre partidos políticos a través de las redes sociales. Conocer la opinión pública es trascendental para los partidos políticos, ya que ésta es su mayor indicador de éxito o fracaso, por lo cual, se requieren enfoques que faciliten la obtención y clasificación de opiniones en redes sociales y que muestren el escenario político actual respecto a dichas opiniones. Como respuesta a lo anterior, se propuso un método que facilita la minería de opinión en Twitter[®] con ayuda de clasificadores de opiniones y análisis lingüístico. Dicho método pre-procesa los datos brutos de Twitter[®], clasifica los sentimientos hacia los partidos políticos de Pakistán de acuerdo con su polaridad y captura el escenario político actual para promover una mejora de los políticos paquistaníes. El método obtuvo resultados positivos; no obstante, la precisión que alcanzó usando NB fue inferior a la obtenida empleando el clasificador SVM.

Por otro lado, Eirinaki et al. [45] abordaron que con la expansión del comercio electrónico y las redes sociales, las personas desean conocer las opiniones de otras personas sobre un producto antes de comprarlo. Desafortunadamente, para un motor de búsqueda típico resulta difícil clasificar los comentarios que representan opiniones de los que no lo hacen. Como solución, se presentó un algoritmo de minería de opinión que además de analizar y clasificar el sentimiento general de un documento o comentario (positivo, negativo o neutral), identifica la orientación semántica de componentes específicos que conducen al sentimiento

en particular. El algoritmo se integró en un motor de búsqueda de opinión que presenta los resultados de una consulta, su clase general y un resumen de los sentimientos de las características más importantes de la consulta. Se realizaron experimentos sobre diferentes conjuntos de datos de diversos tamaños para evaluar el algoritmo propuesto. Los resultados obtenidos fueron superiores a otros algoritmos simples de conteo de palabras, pues se alcanzó hasta 97% de precisión.

Sin embargo, pese a que el análisis sentimental es cada vez más popular, de acuerdo con Sánchez-Rada e Iglesias [46], existen pocos recursos léxicos y semánticos disponibles que fomenten el desarrollo de nuevos servicios y aplicaciones conscientes de las emociones. Además, existen obstáculos como la diversidad de las teorías de la emoción y la ausencia de un vocabulario común. Por otra parte, aunque *Linked Data* (Datos enlazados) tiene la capacidad de reducir dichas problemáticas, para hacerlo requiere una representación sentimental ampliamente aceptada que aún no existe. Como solución se propuso Onyx, un vocabulario semántico de emociones centrado en recursos léxicos y servicios de análisis emocional, el cual sigue un enfoque lingüístico de *Linked Data*. Onyx se alineó con la Ontología Provenance, se integró con el modelo para representar entradas léxicas conocido como *lemon* y se construyó con una taxonomía SKOS (*Simple Knowledge Organization System*, Sistema de Organización Simple de Conocimiento). Asimismo, se evaluó y se obtuvieron resultados aceptables de manera general, sin embargo, se consideró que necesita seguir evolucionando y probándose en entornos más apegados a la realidad.

También, es importante mencionar que Twitter[®] es una de las redes sociales más utilizadas en la actualidad, por lo cual tiene la capacidad de utilizarse para encontrar los puntos de vista y sentimientos de los usuarios mediante el uso de métodos basados en *clústeres*. Sin embargo, a causa de la naturaleza subjetiva de los conjuntos de datos de Twitter[®], los métodos de agrupamiento basados en meta-heurísticas superan a los métodos tradicionales para el análisis de sentimientos. Por lo tanto, Chandra et al. [47] propusieron un nuevo método meta-heurístico que se basa en K-means y cuckoo search. El método propuesto se utilizó para encontrar los grupos de cabeceras óptimos a partir de los contenidos sentimentales de un conjunto de datos de Twitter[®]. El método propuesto se probó en diferentes conjuntos de datos

y se comparó con otros métodos existentes. Los resultados experimentales y el análisis estadístico validaron que el método propuesto superó a los métodos existentes.

Además, Twitter[®] es una de las redes sociales más útiles para el análisis de tendencias dada la cantidad de datos y su popularidad en diferentes países. Sin embargo, la percepción pública de éstas a menudo cambia de una cultura a otra y el análisis es más difícil cuando los tuits se escriben en diferentes idiomas. Por ello, Vilares et al. [48] abordaron la problemática de clasificar polaridad en Twitter[®] de forma multilingüe, para lo cual evaluaron y compararon: 1) Un modelo multilingüe; 2) Un doble modelo monolingüe, y 3) Un modelo monolingüe que actúa en función de la decisión de una herramienta de identificación de idiomas. Asimismo, desarrollaron un sistema capaz de determinar el sentimiento presente en textos escritos en diferentes idiomas. Los resultados obtenidos tras la evaluación mostraron que el enfoque multilingüe es sólido y supera a los modelos monolingües en algunos conjuntos de datos, sin embargo, ni los enfoques monolingües ni los multilingües basados en el cambio de idioma obtuvieron buenos resultados, lo que plantea un nuevo desafío para el análisis sentimental.

De acuerdo con Neppalli et al. [49] el análisis sentimental se estudia ampliamente en diferentes sitios de Internet con el objetivo de generar opiniones resumidas de los usuarios sobre diferentes aspectos. Sin embargo, la identificación de la polaridad sentimental expresada por los usuarios durante los desastres es escasa. Dicha información tiene una gran relevancia porque tiene la capacidad de ayudar a los que responden emergencias a comprender, por ejemplo, las preocupaciones de los usuarios, el pánico y los impactos emocionales entre los miembros. Para atender la problemática descrita, se realizó un análisis sentimental de los tuits publicados durante el huracán Sandy y se visualizaron los sentimientos de los usuarios en un mapa geográfico centrado en torno al huracán. Tras el análisis se concluyó que los sentimientos de los usuarios cambian de acuerdo con sus ubicaciones y con la distancia a la que se encuentran del desastre. Además, se corroboró que la extracción de sentimientos permite obtener información valiosa para ayudar a quienes responden emergencias a desarrollar un mayor conocimiento sobre la situación en la zona de desastre.

2.2 Análisis comparativo de trabajos relacionados con análisis sentimental

La Tabla 2.1 comprende un análisis comparativo de los trabajos relacionados con análisis sentimental que se presentaron en este capítulo.

Artículo	Objetivo	Tópico	Tipo de SA	Contexto de uso	Tecnologías	Resultados
Tripathy et al. [30]	Proponer un enfoque de análisis sentimental para extraer, clasificar y predecir sentimientos con gran precisión.	➤ Análisis Sentimental	➤ Texto	➤ Comercio ➤ Publicidad	➤ Algoritmo NB ➤ Algoritmo SVM	Se propuso un enfoque basado en técnicas de aprendizaje supervisado sobre datos etiquetados que permite extraer, clasificar y predecir sentimientos con una precisión superior a otros enfoques.
Ceron et al. [31]	Desarrollar un método de análisis sentimental que extraiga datos desde la Web y obtenga precisión superior a otras alternativas para análisis sentimental agregado.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	➤ Biblioteca Libstemmer ➤ Método de Montecarlo ➤ Porter Stemmer	Se desarrolló un algoritmo para análisis sentimental basado en codificación manual supervisada, que empleando muestras grandes de datos obtuvo una precisión superior a otros algoritmos conocidos, pero inferior con muestras pequeñas de datos.
Khan et al. [32]	Proponer un enfoque para análisis sentimental que utilice las fortalezas de los métodos basados en aprendizaje supervisado y/o en léxico y suprima sus debilidades.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	➤ Diccionario de WordNet ➤ Porter Stemmer ➤ WordNet Lemmatizer ➤ Stanford POS tagger ➤ SVMlight	Se propuso un marco de trabajo capaz de pre-procesar datos y clasificar la polaridad de opiniones con una precisión promedio de 80.83%.

Artículo	Objetivo	Tópico	Tipo de SA	Contexto de uso	Tecnologías	Resultados
Khan et al. [34]	Desarrollar un marco de trabajo para análisis sentimental que reduzca los obstáculos de los algoritmos de aprendizaje supervisado y mejore el desempeño de los algoritmos basados en léxico.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	<ul style="list-style-type: none"> ➤ Diccionario de WordNet ➤ JSpell ➤ Porter Stemmer ➤ Stanford POS tagger ➤ SVMlight ➤ SentiWordNet 	Se desarrolló un marco de trabajo que emplea algoritmos de aprendizaje supervisado y basados en léxico, el cual es capaz de detectar la subjetividad y clasificar la polaridad sentimental con una precisión de hasta 85.50%.
Fernández-Gavilanes et al. [35]	Promover un nuevo método no supervisado para análisis sentimental aplicable a textos cortos en línea.	➤ Análisis Sentimental	➤ Texto	➤ Mercado-tecnia	<ul style="list-style-type: none"> ➤ FreeLing ➤ Diccionario WordNet ➤ SentiWordNet ➤ Amazon Mechanical Turk 	Se construyó un método no supervisado para determinar la polaridad sentimental de tuits y comentarios que no necesita entrenamiento previo o específico. Obtuvo resultados superiores a otros métodos no supervisados.
Appel et al. [36]	Demostrar que un enfoque basado en conjuntos difusos es capaz de tener éxito para análisis sentimental en el nivel de oración.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	<ul style="list-style-type: none"> ➤ SentiWordNet ➤ LISP 	Se propuso un enfoque híbrido que utiliza conjuntos difusos para análisis sentimental a nivel de oración. Estima la polaridad y la intensidad sentimental de las oraciones. Además, obtuvo resultados superiores a otros algoritmos.
Severyn et al. [37]	Proporcionar un modelo robusto que permita manejar el ruido de los	➤ Minería de Opiniones	➤ Texto	➤ Indefinido	<ul style="list-style-type: none"> ➤ SVMlight ➤ CMU Twitter POS tagger 	Se desarrolló un modelo que predice e identifica tipos de comentarios, polaridad y

Artículo	Objetivo	Tópico	Tipo de SA	Contexto de uso	Tecnologías	Resultados
	datos de YouTube® y se adapte fácilmente a nuevos dominios e idiomas.				<ul style="list-style-type: none"> ➤ TextPro ➤ SentiWordNet 	entidades en YouTube®. También, se adapta a diferentes dominios y funciona para inglés e italiano con una precisión superior a otros modelos hasta en 4%.
Saif et al. [38]	Proponer un enfoque para realizar análisis sentimental sobre Twitter® que minimice las limitantes de los principales enfoques existentes.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	<ul style="list-style-type: none"> ➤ AlchemyAPI ➤ Tweepster ➤ SentiWordNet 	Se propuso un enfoque basado en léxico para análisis sentimental a nivel de entidad y de tuit que identifica la polaridad sentimental de las palabras. A nivel de entidad superó a otros enfoques y a nivel de tuit fue superior en el 66% de los casos.
Muhammad et al. [39]	Realizar análisis sentimental sobre diferentes medios sociales empleando un algoritmo basado en léxico que permita capturar la polaridad contextual.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	<ul style="list-style-type: none"> ➤ TweetNLP ➤ SentiWordNet ➤ SentiStrength 	Se construyó un sistema de clasificación de sentimientos para medios sociales, el cual es capaz de capturar la polaridad contextual usando un enfoque basado en léxico. Obtuvo una precisión promedio superior a otros algoritmos de aprendizaje automático supervisado.
Katz et al. [41]	Realizar análisis sentimental sobre textos regulares y sobre textos con alto nivel de ruido,	➤ Análisis Sentimental	<ul style="list-style-type: none"> ➤ Texto ➤ Audio 	➤ Indefinido	<ul style="list-style-type: none"> ➤ Weka ➤ LibSVM 	Se presentó un enfoque para análisis sentimental aplicable a textos regulares y a textos con un alto nivel de ruido que

Artículo	Objetivo	Tópico	Tipo de SA	Contexto de uso	Tecnologías	Resultados
	considerando el contexto de los términos.					considera el contexto de los términos. El enfoque se probó en tres conjuntos de datos y en dos de ellos obtuvo una precisión superior a otros enfoques.
Bucur [42]	Desarrollar un sistema que permita extraer opiniones sobre hoteles desde la Web, facilitar a los clientes la toma de decisiones y ayudar a los hoteles a que mejoren sus servicios.	➤ Minería de Opiniones	➤ Texto	➤ Turismo	<ul style="list-style-type: none"> ➤ MySQL® ➤ SentiWordNet ➤ PHP 	Se creó una plataforma para extraer y clasificar opiniones de reseñas de hoteles de TripAdvisor® que recolecta las reseñas, pre-procesa los datos extraídos e identifica su polaridad sentimental con una precisión de hasta 76.5%.
Ortigosa et al. [43]	Desarrollar un método para análisis sentimental capaz de extraer emociones mediante el perfil de Facebook® de estudiantes.	➤ Análisis Sentimental	➤ Texto	➤ Aprendizaje electrónico	<ul style="list-style-type: none"> ➤ GNU Aspell ➤ Facebook® API ➤ Weka ➤ NoSQL 	Se construyó un método para análisis sentimental en Facebook® que emplea un enfoque híbrido. Dicho método extrae información sobre el estado emocional de estudiantes y la clasifica con una precisión de 83.27%.
Gull et al. [44]	Obtener y clasificar opiniones sobre partidos políticos paquistanés de las redes sociales y mostrar el escenario político actual respecto a dichas opiniones.	➤ Minería de Opiniones	➤ Texto	➤ Política	<ul style="list-style-type: none"> ➤ Twitter® API ➤ Python ➤ PyCharm ➤ jQuery ➤ SQLite 	Se desarrolló un método de minería de opinión para Twitter® que extrae, pre-procesa y clasifica los sentimientos hacia partidos políticos paquistanés y captura el escenario político

Artículo	Objetivo	Tópico	Tipo de SA	Contexto de uso	Tecnologías	Resultados
						actual. Los resultados obtenidos fueron aceptables pero inferiores a los del clasificador SVM.
Eirinaki et al. [45]	Analizar y clasificar el sentimiento de un documento o comentario en la Web e identificar la orientación semántica de componentes específicos de éste que conducen a un sentimiento particular.	➤ Minería de Opiniones	➤ Texto	➤ Comercio electrónico	➤ Porter Stemmer	Se construyó un algoritmo de minería de opinión que analiza y clasifica el sentimiento de documentos o comentarios en la Web, identificando la orientación semántica que conduce a dicho sentimiento. También se desarrolló un motor de búsqueda de opinión en el que se integró el algoritmo y se obtuvo una precisión de 97%.
Sánchez-Rada e Iglesias [46]	Proveer recursos léxicos y semánticos que sigan un enfoque de <i>Linked Data</i> y permitan fomentar el desarrollo de nuevos servicios y aplicaciones conscientes de las emociones.	➤ Análisis Sentimental ➤ Minería de Opiniones	➤ Texto	➤ Indefinido	➤ Provenance ➤ Lemon ➤ EmotionML ➤ WordNet-Affect ➤ SKOS	Se creó un vocabulario semántico de emociones centrado en recursos léxicos y servicios de análisis sentimental que sigue un enfoque lingüístico de <i>Linked Data</i> . Los resultados fueron aceptables pero el vocabulario continúa en mejoras.
Chandra et al. [47]	Proponer un nuevo método meta-heurístico basado en K-means y cuckoo search para	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	➤ K-means ➤ Cuckoo Search	Los resultados experimentales y el análisis estadístico demostraron que el método propuesto superó a otros

Artículo	Objetivo	Tópico	Tipo de SA	Contexto de uso	Tecnologías	Resultados
	encontrar grupos de cabeceras óptimos en Twitter® a partir del sentimiento.					métodos existentes.
Vilares et al. [48]	Abordar la problemática de clasificar tuits de forma multilingüe con base en su polaridad sentimental, evaluando diferentes enfoques.	➤ Análisis Sentimental	➤ Texto	➤ Indefinido	➤ SentiStrength	El enfoque multilingüe superó al enfoque monolingüe en algunos conjuntos de datos; sin embargo, ninguno obtuvo buenos resultados al evaluar textos que cambian de idioma.
Neppalli et al. [49]	Identificar la polaridad sentimental expresada en tuits durante un desastre para ayudar a comprender mejor la situación al responder emergencias.	➤ Análisis Sentimental	➤ Texto	➤ Meteorología	➤ SentiStrength ➤ ArcGIS	Se concluyó que con el análisis sentimental realizado se tiene la capacidad de ayudar a quienes responden emergencias a obtener un mayor conocimiento sobre la situación en una zona de desastre.

Tabla 2.1 Análisis comparativo de los trabajos relacionados con análisis sentimental.

En los diferentes proyectos de investigación que se abordaron, se realizaron pruebas que demostraron la factibilidad de aplicar análisis sentimental a plataformas como las redes sociales, así como las aportaciones que dicho análisis proporciona en diversos dominios. Sin embargo, la mayoría de estos proyectos de investigación coincidieron en que se requiere seguir trabajando en nuevas aplicaciones, enfoques o métodos que obtengan mejores resultados. Por otro lado, la mayor parte de dichos proyectos de investigación realizaron análisis sentimental considerando únicamente la identificación de la polaridad y la intensidad del sentimiento, dejando de lado la detección y extracción de emociones tales como el enojo, la alegría o la tristeza, por mencionar algunas. Dichas emociones son significativas porque permiten obtener resultados más específicos del sentimiento que refleja algún recurso, por ejemplo, si alguna empresa u organización conoce que un texto relacionado con algún producto o servicio que ofrece tiene polaridad negativa, adquiere la capacidad de llevar a cabo alguna acción para mejorar la polaridad, pero si además es capaz de saber que esa negatividad expresa enojo o miedo, las acciones que se lleven a cabo serán más específicas y efectivas.

2.3 Propuesta de solución

Se propone que para solucionar el problema planteado se utilice Natural Language Understanding[®] como API de análisis sentimental, PHP (*PHP Hypertext Preprocessor*, Preprocesador de Hipertexto PHP) como lenguaje de programación para el desarrollo del módulo, y Java como lenguaje de programación para el desarrollo de un servicio Web que funja de intermediario entre la API de análisis sentimental y el módulo. También, se consideró utilizar Netbeans como IDE (*Integrated Development Environment*, Entorno de Desarrollo Integrado), las redes sociales Twitter[®] y Facebook[®] como fuentes de información, así como las APIs TwitterOAuth de Twitter[®] y Graph API de Facebook[®] para manipular los datos de dichas redes sociales.

Asimismo, se propone utilizar un enfoque de clasificación híbrido para la detección de emociones, MySQL[®] como Sistema Gestor de Base de Datos para almacenar los resultados obtenidos del análisis sentimental y Plotly.js para llevar a cabo la representación gráfica de los resultados de la clasificación.

2.3.1 Justificación de la solución

Las tecnologías mencionadas anteriormente a través de sus diferentes características y ventajas proporcionan las herramientas necesarias para el desarrollo del módulo. La Tabla 2.2 describe brevemente la justificación de la elección de las tecnologías que forman parte de la solución propuesta.

Tecnología	Justificación
Natural Language Understanding [®]	Se seleccionó Natural Language Understanding [®] porque permite extraer la polaridad y la intensidad sentimental de un texto, así como sus emociones, lo cual es una diferencia notable respecto a otras APIs que solo se enfocan en los dos primeros aspectos. Además, ofrece varios niveles de SA, soporta ocho idiomas, entiende modificadores y negaciones, permite un número elevado de transacciones por día y tiene una versión sin costo.
PHP	Se eligió PHP para el desarrollo del módulo porque más allá de las características que posee y que se mencionaron en el Capítulo 1, es uno de los lenguajes de programación para la Web de lado del servidor más utilizados actualmente [50], lo que garantiza que al desarrollar el módulo con este lenguaje, existe un gran número de aplicaciones para integrarlo. Además, la integración es menos compleja dado que ambos están escritos en el mismo lenguaje.
Java	Para el desarrollo de un servicio Web que permita proveer al módulo con los resultados de la extracción de emociones se propone usar Java, ya que es compatible con Natural Language Understanding [®] y es un lenguaje concurrente, robusto, seguro y con una amplia cantidad de documentación.
NetBeans	La elección del IDE se fundamentó en aquellos compatibles con los lenguajes de programación seleccionados. Se escogió NetBeans porque es el IDE oficial para Java 8 y proporciona un gran conjunto de herramientas y soporte para PHP. Además, es un IDE libre que facilita y guía el desarrollo, destaca el código fuente de forma sintáctica y semántica, proporciona autocompletado de código, herramientas para generar documentación y para desarrollar y probar servicios Web.
Twitter [®]	La elección de Twitter [®] se fundamentó en que es una de las redes sociales más populares en la actualidad, trabaja con diferentes dominios de información, ofrece una API que permite manipular sus datos y tiene menos limitantes respecto a la privacidad al extraer datos en comparación con otras redes sociales.

Tecnología	Justificación
TwitterOAuth	Como API para trabajar con Twitter® se eligió TwitterOAuth porque existe abundante documentación sobre ella y su utilización se basa en solo dos funciones genéricas que garantizan que no se obtenga un gran impacto cuando la API de Twitter® añada nuevas funciones o características. Además, es una de las APIs que recomienda Twitter® para trabajar con PHP.
Facebook®	Facebook® se eligió como una de las fuentes de información porque es la red social más popular en la actualidad. Además, ofrece una gran diversidad de dominios de información en comparación con otras redes sociales, permite que los usuarios expresen sus sentimientos y emociones de diferentes formas, cuenta con una API para manipular sus datos y proporciona un mayor número de datos que el resto de las redes sociales.
Graph API	Como API para trabajar con Facebook® se eligió Graph API porque es la API oficial de dicha red social, lo cual implica que existe abundante documentación y soporte.
Enfoque híbrido	Se seleccionó el enfoque híbrido porque comparado con otros enfoques es mejor trabajando con datos que tienen alto índice de ruido. Además, logra mitigar problemas, tales como la dependencia del dominio y de los datos etiquetados. Por otra parte, el enfoque híbrido se adecua al trabajo con Natural Language Understanding®.
MySQL®	Se eligió MySQL® porque es de los Sistemas Gestores de Base de Datos más populares en el mundo. Además, se caracteriza por ser veloz, poseer una versión sin costo y por requerir un bajo consumo de recursos en comparación con algunos otros Sistemas Gestores similares.
Plotly.js	Se seleccionó Plotly.js porque permite desarrollar diferentes tipos de gráficas como lo son las gráficas de pastel, barras, columnas, mapas de calor, entre otras. Además, permite añadir algunas funcionalidades como lo son la descarga de las gráficas como imágenes, el despliegue de información emergente, el aumento y disminución del tamaño de las gráficas, entre otros.

Tabla 2.2 Tecnologías involucradas en la solución propuesta.

Capítulo III. Aplicación de la metodología

En este capítulo se presenta todo lo relacionado al análisis, diseño y desarrollo del proyecto. En primer lugar se muestra un análisis sobre las distintas funcionalidades que conforman el módulo. Posteriormente, se presenta un análisis de diferentes recursos a través de los cuales el módulo obtiene información proveniente de redes sociales para analizar. También, se describe el repositorio de información necesario para almacenar los resultados del análisis llevado a cabo y se detalla el flujo de trabajo con la finalidad de mostrar las etapas involucradas en el funcionamiento del módulo. Se presenta un conjunto de *mockups* diseñados para representar gráficamente los resultados obtenidos, se describen con detalle los procesos de extracción de información de redes sociales, obtención de sentimientos y emociones del texto analizado y la representación gráfica de éstos.

3.1 Análisis de requerimientos del módulo

El módulo comprende un conjunto de funcionalidades a través de las cuales es posible cumplir su propósito general, el cual consiste en la detección/extracción de sentimientos y emociones expresadas mediante texto. Las funcionalidades mencionadas se describen en la Tabla 3.1.

Funcionalidad	Descripción
Iniciar Sesión con Facebook®	Un usuario tiene la posibilidad de autenticarse mediante su cuenta de Facebook® para hacer uso del módulo.
Iniciar Sesión con Twitter®	Un usuario tiene la posibilidad de autenticarse mediante su cuenta de Twitter® para hacer uso del módulo.
Analizar Texto	El análisis de texto involucra un conjunto de tareas tales como: extraer información de redes sociales, obtener sentimientos, obtener emociones, extraer entidades y mostrar gráficas.
Extraer Información	Es necesario extraer textos provenientes de redes sociales, por lo cual incluye las funcionalidades de: extraer comentarios de Facebook® y extraer tuits de Twitter®.
Extraer comentarios de Facebook®	Se requiere obtener el conjunto de comentarios asociados a una publicación de Facebook® a partir de su enlace, así como datos básicos (nombre, género, fecha de nacimiento y ubicación) de los usuarios que realizaron los comentarios.
Extraer tuits de Twitter®	Se requiere obtener el conjunto de comentarios asociados a un tuit de Twitter® a partir de su enlace, así como datos básicos (nombre,

Funcionalidad	Descripción
	ubicación) de los usuarios que realizaron los comentarios.
Obtener Sentimientos	Es necesario identificar y obtener la polaridad sentimental que expresa individualmente y grupalmente un conjunto de textos.
Obtener Emociones	Se necesita identificar y calcular el porcentaje de emociones (enojo, desagrado, miedo, alegría y tristeza) que expresa individualmente y grupalmente un conjunto de textos.
Extraer Entidades	Se requiere extraer las entidades (objetos, lugares, personas) que forman parte de un conjunto de comentarios e identificar y obtener la polaridad sentimental y el porcentaje de emociones (enojo, desagrado, miedo, alegría y tristeza) asociadas a cada una de dichas entidades.
Mostrar Gráficas	Es necesario presentar gráficas que permitan resumir y facilitar la comprensión y utilización de los resultados obtenidos tras el análisis y obtención de sentimientos y emociones.

Tabla 3.1 Funcionalidades requeridas para el módulo de obtención de emociones.

Asimismo, es posible visualizar las diferentes funcionalidades y la relación que existe entre ellas mediante el diagrama de casos de uso que se muestra en la Figura 3.1.

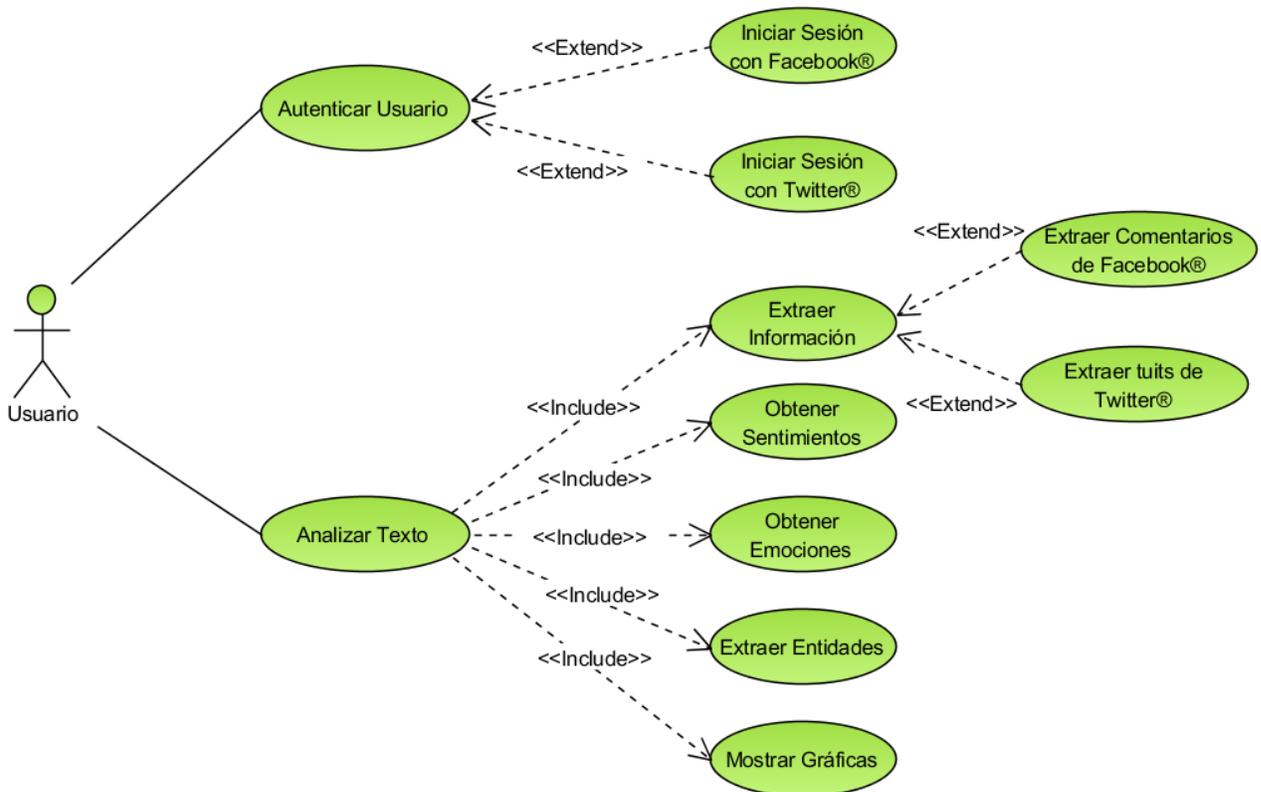


Figura 3.1 Diagrama de Casos de Uso para el módulo de obtención de emociones.

3.2 Fuentes de información seleccionadas para analizar

El módulo que se describe en este trabajo requiere fuentes de información que le proporcionen los datos necesarios para iniciar el proceso de análisis y extracción de sentimientos y emociones. Particularmente, se eligió utilizar como fuentes de información algunas páginas y grupos públicos de Twitter® y Facebook®, ya que son el medio por excelencia en el que las personas expresan sus opiniones y emociones actualmente; sin embargo, muchos de estos recursos son falsos o no son oficiales, lo que a su vez ocasiona que en muchos de los casos la información que comparten tenga carencia de veracidad e interacción por parte de otros usuarios.

Dada la situación mencionada anteriormente, una parte importante que se consideró previo al desarrollo del módulo fue la realización de un análisis de las fuentes de información a utilizar. El análisis consistió en la búsqueda de diferentes páginas y grupos públicos de Facebook®, así como perfiles de Twitter® relacionados con enfermedades crónico-degenerativas con la finalidad de identificar aspectos en éstos, tales como el idioma de las publicaciones, si son oficiales o no, y la cantidad de personas interesadas y que interactúan habitualmente con éstos, entre otros.

La realización del análisis permitió encontrar las páginas y grupos más adecuados para fungir como fuentes de información para el módulo. También, es importante mencionar que la selección de recursos relacionados con el dominio médico se determinó en función de los objetivos específicos del proyecto. La Tabla 3.2 muestra el resultado del análisis de las fuentes de información en Facebook®, en donde es posible apreciar las páginas y grupos seleccionados que tienen las características para utilizarse por el módulo.

Análisis de fuentes de información en Facebook® al día 19/11/2017					
	Recurso	Idioma	Tipo	Miembros	Enlace
Cáncer	The Breast Cancer Site	Inglés	Página	6, 163, 526	https://www.facebook.com/TheBreastCancerSite
	Breast Cancer Awareness	Inglés	Página	4, 163, 929	https://www.facebook.com/BreastCancerAwareness
	Cancer Research UK	Inglés	Página	1, 476, 057	https://www.facebook.com/cancerresearchuk

Análisis de fuentes de información en Facebook® al día 19/11/2017					
	Recurso	Idioma	Tipo	Miembros	Enlace
Cáncer	American Cancer Society	Inglés	Página	1, 157, 368	https://www.facebook.com/AmericanCancerSociety
	Stop Cancer. Start Praying	Inglés	Página	728, 894	https://www.facebook.com/stopcancerstartpraying
	Cure Childhood Cancer	Inglés	Página	344, 273	https://www.facebook.com/curechildhoodcancer
Diabetes	The Diabetes Awareness Ribbon	Inglés	Página	1, 604, 945	https://www.facebook.com/TheDiabetesSite
	Diabetes Support	Inglés	Página	1, 336, 934	https://www.facebook.com/Diabetes-Support-119703998103102
	American Diabetes Association	Inglés	Página	717, 008	https://www.facebook.com/AmericanDiabetesAssociation
	Managing Diabetes	Inglés	Página	525, 135	https://www.facebook.com/ManagingDiabetesQH
	Diabetes Community	Inglés	Página	107, 778	https://www.facebook.com/TheDiabetesCommunity
Obesidad	Obesity NOT 4ME	Inglés	Página	748, 893	https://www.facebook.com/obesitynot4me
	ObesityHelp.com	Inglés	Página	22, 272	https://www.facebook.com/obesityhelp
	Obesity, Weight loss & Fitness	Inglés	Grupo	16, 603	https://www.facebook.com/groups/obesity2weightloss
Hepatitis	HepatitisC.net	Inglés	Página	29, 319	https://www.facebook.com/HepatitisCDotNet
	Hepatitis Awareness	Inglés	Página	18, 723	https://www.facebook.com/behepatitisaware.com.ng
	Hepatitis B Foundation	Inglés	Grupo	4, 678	https://www.facebook.com/groups/hepatitisbfoundation
Hipertensión	Life With Idiopathic Intracranial Hypertension	Inglés	Página	14, 505	https://www.facebook.com/LIFEWITHIIH
	National Trust for Hypertension and Stroke Prevention Research	Inglés	Página	11, 941	https://www.facebook.com/Ammar.trust/
	Intercranial Hypertension Awareness	Inglés	Página	2, 697	https://www.facebook.com/IntercranialHypertensionAwareness

Análisis de fuentes de información en Facebook® al día 19/11/2017					
	Recurso	Idioma	Tipo	Miembros	Enlace
Alzheimer	Alzheimer's Team	Inglés	Página	283, 310	https://www.facebook.com/AlzheimersTeam
	Remember For Alzheimer's	Inglés	Página	277, 167	https://www.facebook.com/RememberForAlzheimers
	Alzheimer's Awareness	Inglés	Página	41, 458	https://www.facebook.com/alzheimersstore

Tabla 3.2 Fuentes de información adecuadas para trabajar en Facebook®.

De manera similar, la Tabla 3.3 presenta el resultado del análisis de las fuentes de información en Twitter®, en donde es posible apreciar los perfiles de dicha red social que tienen las condiciones adecuadas para trabajar con el módulo.

Análisis de fuentes de información en Twitter® al día 19/11/2017					
	Recurso	Idioma	Miembros	Enlace	
Cáncer	American Cancer Soc	Inglés	1, 057, 066	https://twitter.com/AmericanCancer	
	Stand Up To Cancer	Inglés	309, 904	https://twitter.com/SU2C	
	Cancer Research UK	Inglés	308, 200	https://twitter.com/CR_UK	
	National Cancer Inst	Inglés	137, 106	https://twitter.com/theNCI	
Diabetes	Diabetes UK	Inglés	163, 835	https://twitter.com/DiabetesUK	
	Amer. Diabetes Assn.	Inglés	129, 517	https://twitter.com/AmDiabetesAssn	
	World Diabetes Day	Inglés	65, 603	https://twitter.com/WDD	
	Int. Diabetes Fed.	Inglés	44, 002	https://twitter.com/IntDiabetesFed	
Obesidad	The Obesity Society	Inglés	25, 745	https://twitter.com/ObesitySociety	
	CDC Obesity	Inglés	24, 114	https://twitter.com/CDCObesity	
	Campaign To End Obesity	Inglés	12, 974	https://twitter.com/ToEndObesity	
Hepatitis	CDC Hepatitis	Inglés	34, 531	https://twitter.com/cdchep	
	Hep B Foundation	Inglés	6, 367	https://twitter.com/HepBFoundation	
	Hepatitis C Trust	Inglés	4, 573	https://twitter.com/HepatitisCTrust	
Hipertensión	Hypertension	Inglés	2, 023	https://twitter.com/HyperAHA	
	Cochrane Hypertension	Inglés	1, 299	https://twitter.com/CochraneHTN	
	Hypertension Talk	Inglés	1, 063	https://twitter.com/HypertensionTalk	

Análisis de fuentes de información en Twitter® al día 19/11/2017				
	Recurso	Idioma	Miembros	Enlace
Alzheimer	Alzheimer's Society	Inglés	157, 518	https://twitter.com/alzheimerssoc
	Alzheimer's Assoc.	Inglés	119, 075	https://twitter.com/alzassociation
	Alzheimer's Research UK	Inglés	54, 506	https://twitter.com/ARUKnews

Tabla 3.3 Fuentes de información adecuadas para trabajar en Twitter®.

3.3 Esquema de almacenamiento para el resultado del análisis sentimental

Un aspecto importante relacionado con el módulo es la necesidad de un medio para almacenar el resultado del análisis de textos llevado a cabo, para con ello tener la capacidad de utilizar los resultados posteriormente, lo cual requiere un respaldo de información en caso de que la API de análisis sentimental no esté disponible, esté en mantenimiento, o para controlar la realización de peticiones por parte de los usuarios. Dada la necesidad anterior, se realizó el diseño de una Base de Datos para almacenar el resultado del análisis y algunos otros datos asociados a éste.

La Base de Datos diseñada cuenta con las siguientes tablas: 1) *Comment*; 2) *User*; 3) *SentimentAnalysis*; 4) *Entity*, y 5) *User_analysis*. La primera tabla almacena todos los comentarios o textos relacionados con el análisis que se esté realizando. Por su parte, la segunda tabla almacena información básica de cada usuario que efectuó cada uno de los diferentes comentarios, por ejemplo, el nombre del usuario, el género, la ubicación y la fecha de nacimiento. Cabe mencionar que estos datos no se almacenan obligatoriamente dado que no siempre es posible contar con ellos.

Por otro lado, la tercera tabla almacena el identificador de la publicación analizada, así como la fecha y hora en que se realizó el análisis. La cuarta tabla almacena las entidades (persona, lugar u objeto hacia el cual se expresa una opinión, crítica o sentimiento) junto con la polaridad sentimental y el porcentaje de cada emoción asociado a éstas. La última tabla almacena el porcentaje de polaridad, enojo, desagrado, miedo, alegría y tristeza calculado para cada comentario efectuado por cada usuario. La Figura 3.2 muestra el diagrama de Base de Datos descrito anteriormente, en donde es posible apreciar las diferentes tablas, campos y relaciones que lo conforman.

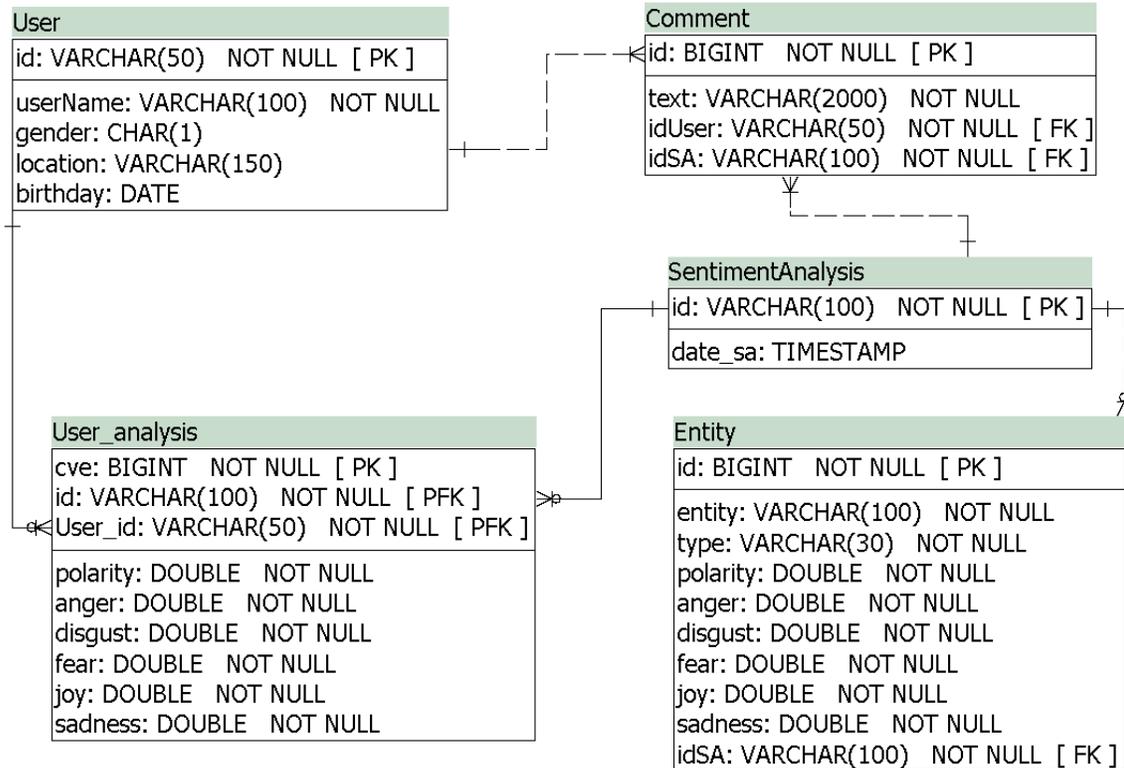


Figura 3.2 Diagrama de Base de Datos para el módulo de extracción de emociones.

De igual manera, mediante la Tabla 3.4 es posible apreciar el diccionario de datos correspondiente a la base de datos para el módulo de extracción de emociones, en donde se describen los campos de cada tabla que conforma la base de datos.

Diccionario de datos de la base de datos del módulo de obtención de emociones				
Tabla Comment				
Campo	Tipo	Tamaño	Nulos	Descripción
Id	Numérico	-	No	Identificador del comentario.
Text	Texto	2000	No	Texto que compone el comentario realizado.
IdUser	Texto	50	No	Identificador del usuario que realizó el comentario.
IdSA	Texto	100	Sí	Identificador de la publicación a la que pertenece el comentario.
Tabla User				
Campo	Tipo	Tamaño	Nulos	Descripción
Id	Texto	50	No	Identificador del usuario.
UserName	Texto	100	No	Nombre del usuario que realizó el comentario.
Gender	Caracter	1	Sí	Género del usuario (M, F) que comentó.
Location	Texto	150	Sí	Ubicación del usuario que comentó.
Birthday	Fecha	-	Sí	Fecha de nacimiento del usuario que comentó.

Diccionario de datos de la base de datos del módulo de obtención de emociones				
Tabla SentimentAnalysis				
Campo	Tipo	Tamaño	Nulos	Descripción
Id	Texto	100	No	Identificador de la publicación analizada.
Date_sa	Fecha	-	No	Fecha de realización del análisis.
Tabla Entity				
Campo	Tipo	Tamaño	Nulos	Descripción
Id	Numérico	-	No	Identificador de la entidad.
Entity	Texto	100	No	Texto que compone la entidad.
Type	Texto	30	No	Tipo de entidad de la que se trata.
Polarity	Numérico	-	No	Polaridad obtenida tras analizar una entidad.
Anger	Numérico	-	No	Enojo obtenido tras analizar una entidad.
Disgust	Numérico	-	No	Desagrado obtenido tras analizar una entidad.
Fear	Numérico	-	No	Miedo obtenido tras analizar una entidad.
Joy	Numérico	-	No	Alegría obtenida tras analizar una entidad.
Sadness	Numérico	-	No	Tristeza obtenida tras analizar una entidad.
IdSA	Texto	100	No	Identificador de la publicación a la que pertenece la entidad.
Tabla User_analysis				
Campo	Tipo	Tamaño	Nulos	Descripción
Cve	Numérico	-	No	Identificador de la tabla.
User_id	Texto	50	No	Identificador del usuario.
Id	Texto	100	No	Identificador de la publicación analizada.
Polarity	Numérico	-	No	Polaridad obtenida tras analizar un comentario.
Anger	Numérico	-	No	Enojo obtenido tras analizar un comentario.
Disgust	Numérico	-	No	Desagrado obtenido tras analizar un comentario.
Fear	Numérico	-	No	Miedo obtenido tras analizar un comentario.
Joy	Numérico	-	No	Alegría obtenida tras analizar un comentario.
Sadness	Numérico	-	No	Tristeza obtenida tras analizar un comentario.

Tabla 3.4 Diccionario de datos de la Base de Datos del módulo de extracción de emociones.

3.4 Flujo de trabajo

El módulo de extracción de emociones se compone por diferentes capas interrelacionadas, por lo cual, a continuación, se describen dichas capas junto con el flujo de trabajo que permite llevar a cabo la extracción de sentimientos y emociones a través de cada una de ellas:

1. **Capa de inicio de sesión:** El usuario que requiere extraer sentimientos y emociones a partir de un conjunto de comentarios de una publicación, inicia sesión mediante su cuenta de Twitter® o Facebook®.
2. **Capa de extracción de datos:** Una vez que se inició sesión, el usuario tiene dos posibilidades: 1) Proporcionar información de una publicación de Facebook® que desee analizar, con lo cual, mediante la Graph API de Facebook® se extraen los comentarios asociados a la publicación junto con la información básica (nombre, género, fecha de nacimiento y ubicación) de los usuarios que escribieron dichos comentarios, y 2) Proporcionar información de un tuit en específico que se desee analizar, con lo cual, mediante TwitterOAuth se extraen los tuits escritos como respuesta a dicho tuit junto con la información básica (nombre, ubicación) de los usuarios que escribieron dichas respuestas. Independientemente de lo que el usuario seleccione, una vez que se tenga la información necesaria se pasa a la siguiente capa del flujo de trabajo.
3. **Capa de obtención de emociones:** En primer lugar, se realiza un procesamiento de texto que consiste en eliminar palabras irrelevantes para el análisis de cada comentario, por ejemplo, los enlaces de Internet, ya que éstos no expresan alguna emoción. En segundo lugar, se realiza el análisis sentimental y la extracción de emociones de los comentarios, en donde se tienen dos opciones para obtener el resultado del análisis: 1) Llevar a cabo la obtención de sentimientos y emociones mediante una petición a un servicio Web que emplea la API Natural Language Understanding®, y 2) Obtener el resultado almacenado de la Base de Datos en caso de que éste exista. Cuando una publicación se analiza por primera vez se emplea el servicio Web para obtener el resultado y posteriormente se almacena en la Base de Datos. Sin embargo, si la misma publicación se analiza nuevamente, el módulo decide entre usar el servicio Web o usar el resultado almacenado en la Base de Datos. Si pasó suficiente tiempo como para que cambiaran los comentarios se usa el servicio Web porque esto implica que cambiaron los resultados, de lo contrario se usa la Base de Datos. Como tiempo suficiente se consideraron ocho horas.
4. **Capa de representación gráfica:** Tras obtener el resultado del análisis, éste se representa a través del uso de distintos tipos de gráficas mediante la utilización de

Plotly.js. La representación gráfica tiene la finalidad de facilitar la comprensión y utilización del resultado por parte de los interesados. Dichas gráficas constituyen la parte principal de la respuesta que obtiene el usuario que inició el proceso.

La Figura 3.3 permite visualizar cada una de las diferentes capas que se describieron anteriormente, así como el flujo de trabajo y las relaciones entre éstas.

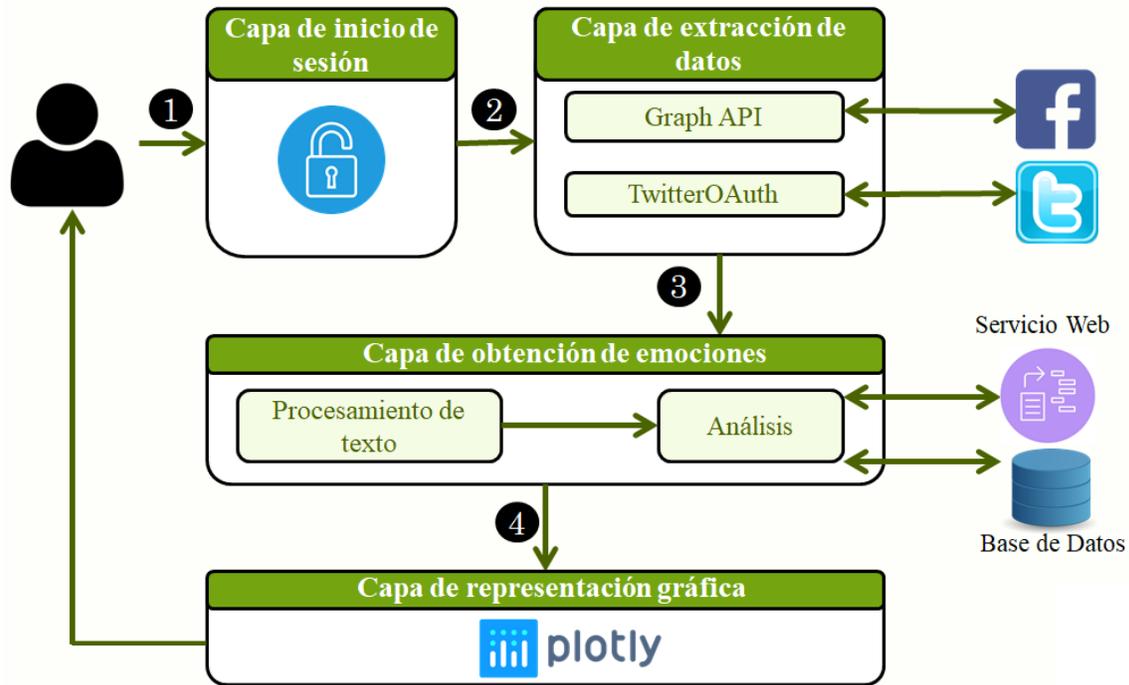
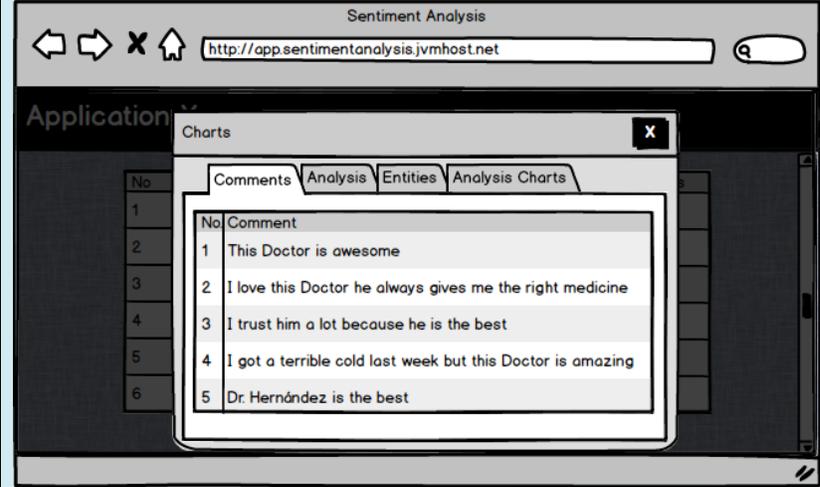
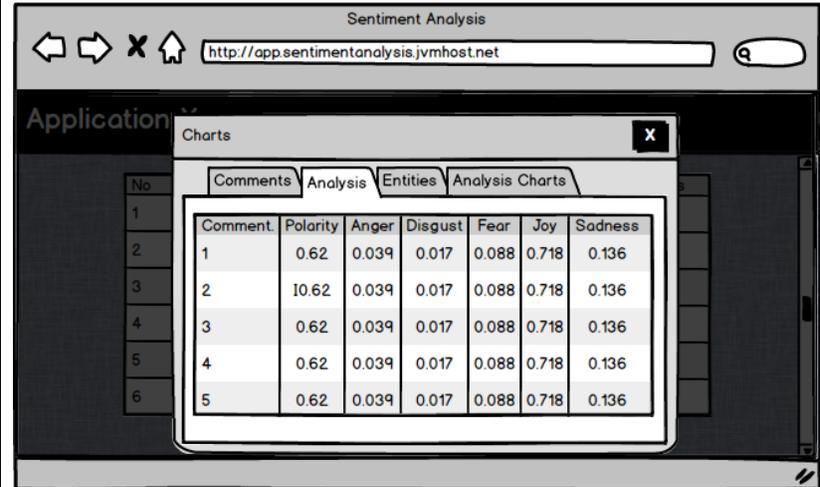
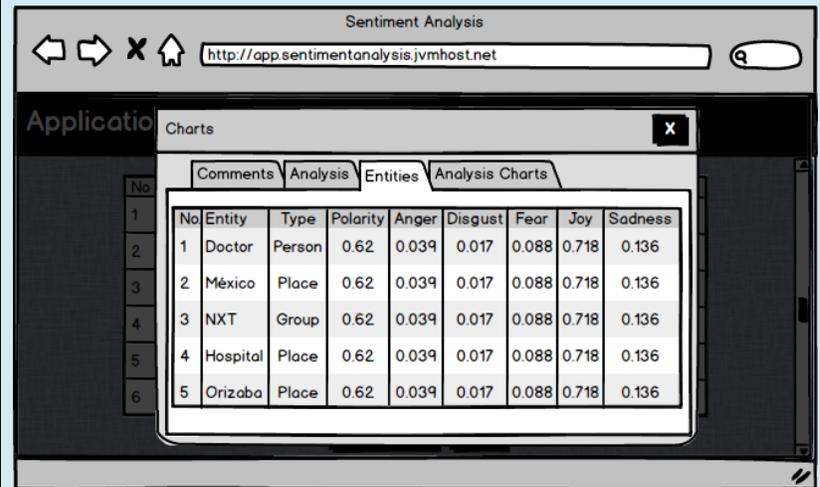
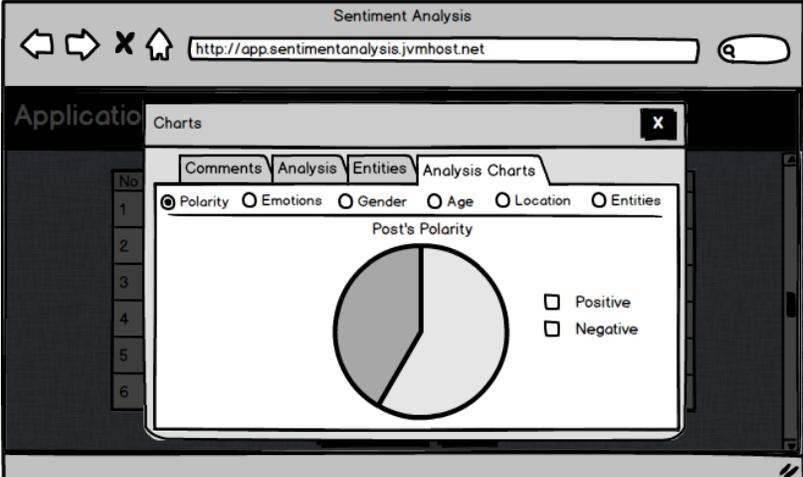
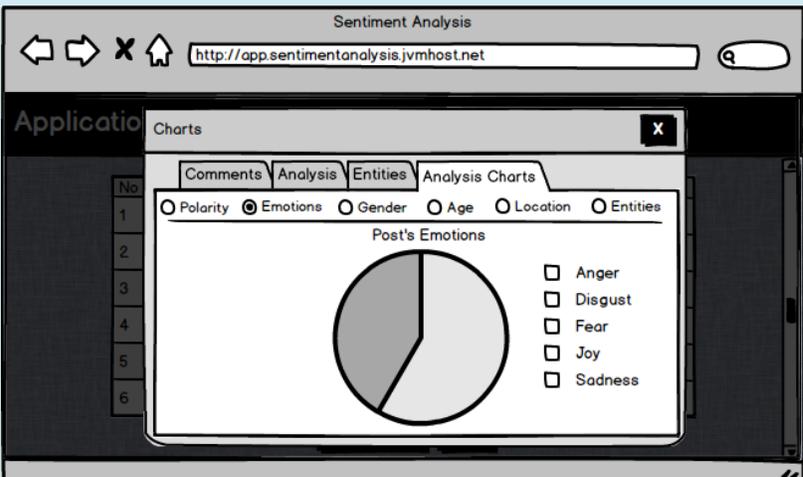
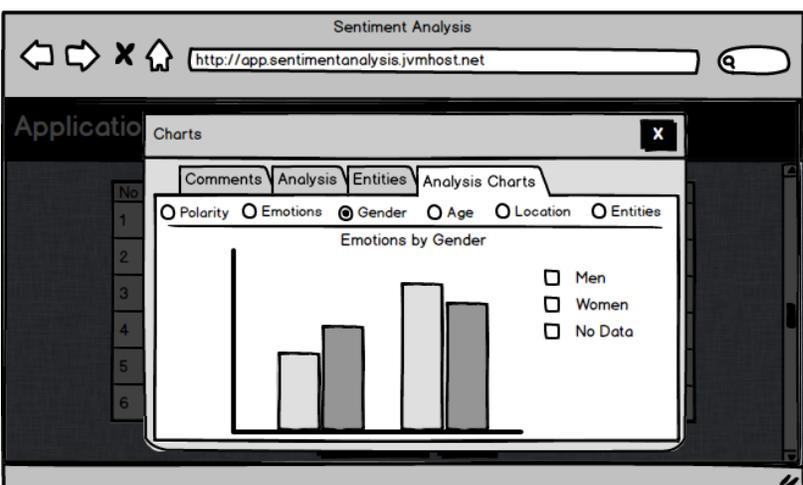


Figura 3.3 Flujo de trabajo para la extracción de sentimientos y emociones mediante el módulo.

3.5 Diseño de *mockups* para el módulo

Para clarificar la forma de visualizar la respuesta que se obtiene como resultado del análisis de comentarios de publicaciones de Facebook® o Twitter® se diseñaron algunos *mockups*. En dichos *mockups* se incluyen los comentarios obtenidos de las redes sociales, la obtención de sentimientos y emociones, las entidades encontradas y algunas gráficas que representan los resultados del análisis. En lo concerniente a las gráficas, para Facebook® se graficó la polaridad sentimental, las emociones, la distribución de emociones por género, edad, ubicación y la polaridad correspondiente a las entidades encontradas. Para Twitter® las gráficas son similares, excepto por las que muestran la distribución de emociones por género y edad, las cuales se descartaron para dicha red social porque no permite obtener tales datos. La Tabla 3.5 muestra los *mockups* diseñados.

Mockup	Descripción																																																						
 <p>The screenshot shows a web browser window titled "Sentiment Analysis" with the URL "http://app.sentimentanalysis.jvmhost.net". A modal window titled "Charts" is open, displaying a table of comments. The table has two columns: "No" and "Comment".</p> <table border="1" data-bbox="389 472 860 703"> <thead> <tr> <th>No</th> <th>Comment</th> </tr> </thead> <tbody> <tr><td>1</td><td>This Doctor is awesome</td></tr> <tr><td>2</td><td>I love this Doctor he always gives me the right medicine</td></tr> <tr><td>3</td><td>I trust him a lot because he is the best</td></tr> <tr><td>4</td><td>I got a terrible cold last week but this Doctor is amazing</td></tr> <tr><td>5</td><td>Dr. Hernández is the best</td></tr> </tbody> </table>	No	Comment	1	This Doctor is awesome	2	I love this Doctor he always gives me the right medicine	3	I trust him a lot because he is the best	4	I got a terrible cold last week but this Doctor is amazing	5	Dr. Hernández is the best	<p style="text-align: center;">Comentarios</p> <p>Se muestran los diferentes comentarios obtenidos de la publicación o tuit analizados mediante una tabla.</p>																																										
No	Comment																																																						
1	This Doctor is awesome																																																						
2	I love this Doctor he always gives me the right medicine																																																						
3	I trust him a lot because he is the best																																																						
4	I got a terrible cold last week but this Doctor is amazing																																																						
5	Dr. Hernández is the best																																																						
 <p>The screenshot shows the same application with the "Analysis" tab selected in the "Charts" modal. It displays a table with sentiment analysis results for each comment.</p> <table border="1" data-bbox="389 997 860 1207"> <thead> <tr> <th>Comment</th> <th>Polarity</th> <th>Anger</th> <th>Disgust</th> <th>Fear</th> <th>Joy</th> <th>Sadness</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>2</td><td>10.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>3</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>4</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>5</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> </tbody> </table>	Comment	Polarity	Anger	Disgust	Fear	Joy	Sadness	1	0.62	0.039	0.017	0.088	0.718	0.136	2	10.62	0.039	0.017	0.088	0.718	0.136	3	0.62	0.039	0.017	0.088	0.718	0.136	4	0.62	0.039	0.017	0.088	0.718	0.136	5	0.62	0.039	0.017	0.088	0.718	0.136	<p style="text-align: center;">Análisis de textos</p> <p>Se presenta mediante una tabla el resultado del análisis de los diferentes comentarios, el cual consiste en la polaridad sentimental y el porcentaje de enojo, desagrado, miedo, alegría y tristeza de cada uno.</p>												
Comment	Polarity	Anger	Disgust	Fear	Joy	Sadness																																																	
1	0.62	0.039	0.017	0.088	0.718	0.136																																																	
2	10.62	0.039	0.017	0.088	0.718	0.136																																																	
3	0.62	0.039	0.017	0.088	0.718	0.136																																																	
4	0.62	0.039	0.017	0.088	0.718	0.136																																																	
5	0.62	0.039	0.017	0.088	0.718	0.136																																																	
 <p>The screenshot shows the application with the "Entities" tab selected in the "Charts" modal. It displays a table with extracted entities for each comment.</p> <table border="1" data-bbox="389 1522 860 1732"> <thead> <tr> <th>No</th> <th>Entity</th> <th>Type</th> <th>Polarity</th> <th>Anger</th> <th>Disgust</th> <th>Fear</th> <th>Joy</th> <th>Sadness</th> </tr> </thead> <tbody> <tr><td>1</td><td>Doctor</td><td>Person</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>2</td><td>México</td><td>Place</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>3</td><td>NXT</td><td>Group</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>4</td><td>Hospital</td><td>Place</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> <tr><td>5</td><td>Orizaba</td><td>Place</td><td>0.62</td><td>0.039</td><td>0.017</td><td>0.088</td><td>0.718</td><td>0.136</td></tr> </tbody> </table>	No	Entity	Type	Polarity	Anger	Disgust	Fear	Joy	Sadness	1	Doctor	Person	0.62	0.039	0.017	0.088	0.718	0.136	2	México	Place	0.62	0.039	0.017	0.088	0.718	0.136	3	NXT	Group	0.62	0.039	0.017	0.088	0.718	0.136	4	Hospital	Place	0.62	0.039	0.017	0.088	0.718	0.136	5	Orizaba	Place	0.62	0.039	0.017	0.088	0.718	0.136	<p style="text-align: center;">Extracción de entidades</p> <p>Se muestra una tabla con el conjunto de entidades encontradas en los textos analizados, incluyendo aspectos como el tipo, la polaridad y el porcentaje de enojo, desagrado, miedo, alegría y tristeza de cada entidad.</p>
No	Entity	Type	Polarity	Anger	Disgust	Fear	Joy	Sadness																																															
1	Doctor	Person	0.62	0.039	0.017	0.088	0.718	0.136																																															
2	México	Place	0.62	0.039	0.017	0.088	0.718	0.136																																															
3	NXT	Group	0.62	0.039	0.017	0.088	0.718	0.136																																															
4	Hospital	Place	0.62	0.039	0.017	0.088	0.718	0.136																																															
5	Orizaba	Place	0.62	0.039	0.017	0.088	0.718	0.136																																															

Mockup	Descripción
	<p>Polaridad</p> <p>Se presenta el porcentaje de positividad y negatividad que se demuestra hacia la publicación analizada mediante una gráfica de pastel.</p>
	<p>Emociones</p> <p>Se muestra el porcentaje de enojo, desagrado, miedo, alegría y tristeza que se demuestra hacia la publicación analizada mediante una gráfica de pastel.</p>
	<p>Emociones por género</p> <p>Se presenta la distribución de las emociones obtenidas por género mediante una gráfica de barras.</p>

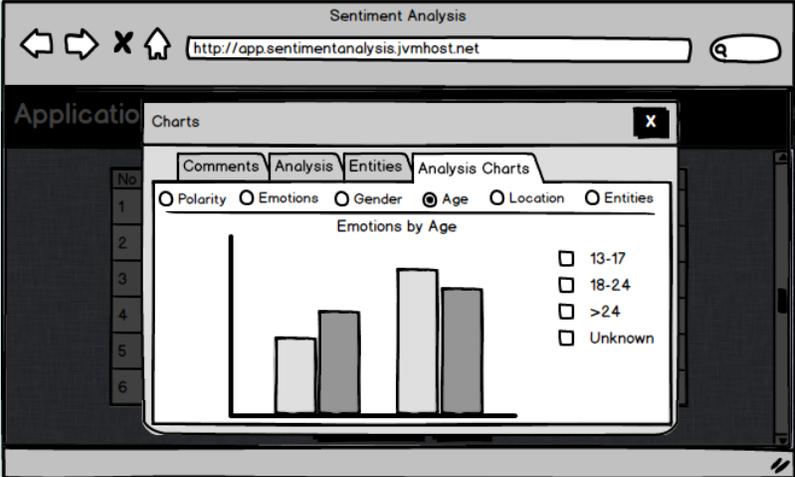
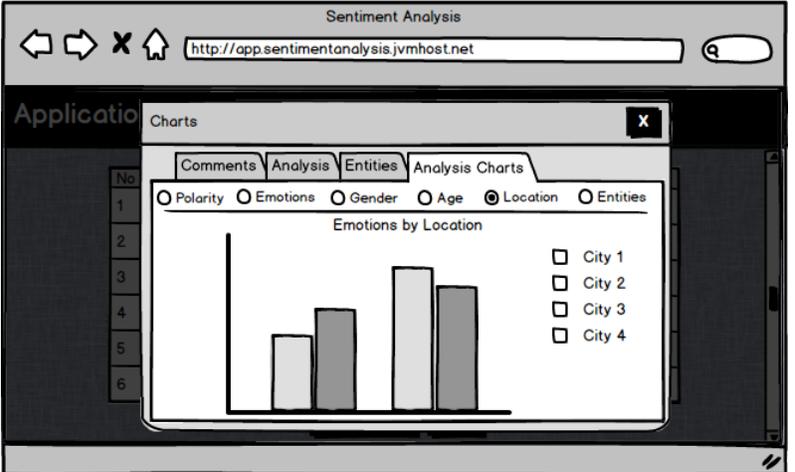
Mockup	Descripción
	<p>Emociones por edad</p> <p>Se muestra la distribución de las emociones obtenidas por rangos de edad mediante una gráfica de barras.</p>
	<p>Emociones por ubicación</p> <p>Se presenta la distribución de las emociones obtenidas por ubicación mediante una gráfica de barras.</p>
	<p>Entidades</p> <p>Se muestra la positividad o negatividad que se expresa hacia cada una de las entidades identificadas entre los comentarios de la publicación mediante un mapa de calor.</p>

Tabla 3.5 Diseño de *mockups* para la representación gráfica de los resultados.

3.6 Iconografía para emociones y entidades

Con la finalidad de que los usuarios identifiquen y diferencien rápidamente los distintos tipos de emociones en los resultados del análisis se diseñaron íconos que representen cada una de éstas. Asimismo, se identificaron distintos tipos de entidades y se diseñaron íconos que los representen para facilitar la comprensión del tipo de entidad del que se trata. Los íconos se diseñaron tomando como base imágenes que poseen una licencia que permite su utilización con modificaciones. La Tabla 3.6 muestra los íconos diseñados.

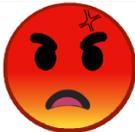
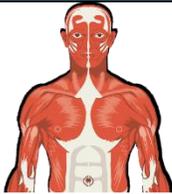
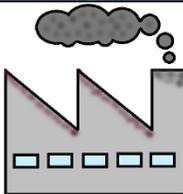
Iconografía para las diferentes emociones				
Enojo	Desagrado	Miedo	Alegría	Tristeza
				
Iconografía para los diferentes tipos de entidades				
Anatomía	Objeto	Empresa	Crimen	
				
Droga	Característica geográfica	Estado de salud	Hashtag	
				
Evento natural	Persona	Cantidad	Televisión	
				
Twitter	Grupo	Lugar	Deportes	
				

Tabla 3.6 Iconografía para las diferentes emociones y entidades.

3.7 Principales funcionalidades del módulo

En esta sección se describen las principales funcionalidades que ofrece el módulo de extracción de emociones, tales como el inicio de sesión, la extracción de información de redes sociales, la obtención de sentimientos y emociones, así como la representación gráfica de éstas, entre otras.

3.7.1 Inicio de sesión

Como se mencionó en el flujo de trabajo, los usuarios necesitan autenticarse a través de su cuenta de Twitter® o Facebook® para iniciar el proceso de análisis y extracción de sentimientos y emociones, por lo cual, a continuación, se describe lo concerniente al desarrollo de las funcionalidades de inicio de sesión en dichas redes sociales.

3.7.1.1 Inicio de sesión en Facebook®

El inicio de sesión en Facebook® requirió la creación de un archivo en PHP con la configuración necesaria para proveer datos tales como: 1) Identificador de la aplicación; 2) Clave secreta; 3) Versión de la Graph API, y 4) Permisos asociados. Asimismo, se desarrollaron las funcionalidades en PHP para emplear la configuración descrita anteriormente para iniciar sesión, lo cual se realizó siguiendo los pasos que se describen brevemente a continuación:

1. Se crea un objeto de la Graph API mediante el identificador y la clave secreta de la aplicación.
2. Se navega a la página de inicio de sesión.
3. Se obtiene la ficha de acceso de la aplicación.
4. Se informa si existe un error en caso de que se produzca por parte de la API o por parte del usuario durante la autenticación.
5. La ficha de acceso se configura para que no expire con el tiempo y posteriormente se almacena en la sesión.
6. Se navega hacia la página establecida en caso de que el inicio de sesión sea exitoso.

El listado 3.1 muestra un fragmento de código en el que se ejemplifican los pasos mencionados anteriormente (líneas 1-26).

```

1. $fb = new Facebook\Facebook(["app_id" => "ID-APP", "app_secret" => "APP-SECRET",
2.                               "default_graph_version" => "VERSION"]);
3. $helper = $fb->getRedirectLoginHelper();
4. try {
5.     $accessToken = $helper->getAccessToken();
6. } catch(Exception $e) {
7.     echo "La API devolvió el error: " . $e->getMessage();
8. }
9. if (!isset($accessToken)) {
10.    if ($helper->getError()) {
11.        header("Location: http://app.sentimentanalysis.jvmhost.net");
12.    } else {
13.        header("HTTP/1.0 400 Bad Request");
14.        echo "Bad request";
15.    }
16. }
17. $tokenMetadata->validateExpiration();
18. if (!$accessToken->isLongLived()) {
19.    try {
20.        $accessToken = $oAuth2Client->getLongLivedAccessToken($accessToken);
21.    } catch (Facebook\Exceptions\FacebookSDKException $e) {
22.        echo "Error convirtiendo la ficha de acceso: " . $helper->getMessage() . "\n\n";
23.    }
24. }
25. $_SESSION["fb_access_token"] = (string) $accessToken;
26. header("Location: http://app.sentimentanalysis.jvmhost.net/Segunda.php");

```

Listado 3.1 Fragmento de código de inicio de sesión en Facebook®.

3.7.1.2 Inicio de sesión en Twitter®

Para llevar a cabo el inicio de sesión en Twitter® se creó la clase “Config” mediante PHP, en donde se desarrolló un método encargado de mantener y proporcionar los parámetros necesarios para la conexión y la autenticación con Twitter®, los cuales son: 1) Clave de la API; 2) Clave secreta; 3) Ficha de acceso, y 4) Ficha de acceso secreta. Por otro lado, se desarrollaron las funcionalidades en PHP para utilizar los parámetros mencionados anteriormente, lo cual se detalla brevemente en los siguientes pasos:

1. Se crea un objeto de TwitterOAuth mediante la clave de la API y la clave secreta de la API.
2. Se obtiene la ficha de acceso de la aplicación mediante el objeto antes creado.
3. Se almacena en la sesión la ficha de acceso y la ficha de acceso secreta.
4. Se realiza la petición de autorización mediante la ficha de acceso obtenida.

5. Se navega hacia la página adecuada en caso de que el inicio de sesión sea exitoso o se informa en caso de que se genere algún error.

El fragmento de código (líneas 3-19) que se aprecia en el listado 3.2 ejemplifica lo concerniente al inicio de sesión en Twitter®.

```

1.  $co = new Config();
2.  $config = $co->getConfig();
3.  $twitteroauth = new TwitterOAuth($config["consumer_key"], $config["consumer_secret"]);
4.  $request_token = $twitteroauth->oauth(
5.      "oauth/request_token", [
6.          "oauth_callback" => $config["url_callback"]
7.      ]
8.  );
9.  if($twitteroauth->getLastHttpCode() != 200) {
10.     throw new \Exception("Hubo un problema con la petición");
11. }
12. $_SESSION["oauth_token"] = $request_token["oauth_token"];
13. $_SESSION["oauth_token_secret"] = $request_token["oauth_token_secret"];
14. $url = $twitteroauth->url(
15.     "oauth/authorize", [
16.         "oauth_token" => $request_token["oauth_token"]
17.     ]
18. );
19. header("Location: ". $url);

```

Listado 3.2 Fragmento de código de inicio de sesión en Twitter®.

3.7.2 Extracción de información de redes sociales

Una de las funcionalidades más importantes del módulo corresponde a la extracción de información de redes sociales, ya que a partir de dicha información se realiza el análisis y obtención de sentimientos y emociones. Dicho proceso se realiza de forma distinta entre Twitter® y Facebook®, sin embargo, a pesar de las diferencias, básicamente se siguen los siguientes pasos en ambos casos:

1. Conectar con la red social mediante la API de ésta.
2. Extraer comentarios o tuits a partir del enlace o del identificador de alguna publicación realizada en Twitter® o Facebook®.
3. Eliminar enlaces encontrados entre los comentarios.
4. Extraer los datos generales de cada usuario que efectuó cada comentario.

A continuación, se describe con mayor detalle el proceso de extracción de información para cada red social.

3.7.2.1 Extracción de datos en Facebook®

Para extraer información de Facebook® se requiere establecer una conexión con dicha red social a través de la Graph API, el identificador de la aplicación y la ficha de acceso. El listado 3.3 muestra la función para establecer la conexión con Facebook®.

```

1. public function connect() {
2.     $this->fb = new Facebook\Facebook([
3.         "app_id" => "app_id",
4.         "app_secret" => " app_secret ",
5.         "default_graph_version" => " default_graph_version "
6.     ]);
7. }

```

Listado 3.3 Función para establecer la conexión con Facebook®.

Una vez establecida la conexión es posible llevar a cabo la extracción de comentarios de una publicación empleando para ello el enlace o el identificador de ésta. En caso de que se emplee el enlace, lo que se realiza es extraer el identificador mediante él, mientras que si se emplea el identificador se evita dicho paso. De cada comentario se almacena el texto que lo conforma y el identificador del usuario que lo escribió. La petición mediante la Graph API para extraer comentarios a partir del identificador de una publicación se aprecia en el fragmento de código localizado en el listado 3.4 (líneas 2-9).

```

1. public function readPostByID($id) {
2.     $this->idUsers = array();
3.     $this->data = array();
4.     $response = $this->fb->get($id . "?fields=comments"); //Petición a la API
5.     $result = $response->getGraphObject();
6.     $ban = true;
7.     foreach ($result as $obj) { //Extracción de los resultados
8.         $ban = true;
9.         $res = json_decode($obj, true);
10.        ...
11.    }
12.    ....
13. }

```

Listado 3.4 Petición para extraer comentarios a partir de un identificador en Facebook®.

Al tener los comentarios necesarios, los pasos siguientes corresponden a eliminar enlaces encontrados en éstos para reducir la cantidad de palabras irrelevantes cuando se analicen en etapas posteriores, así como a extraer los datos generales (nombre, género, fecha de nacimiento y ubicación) de los usuarios que realizaron los comentarios, para lo cual, se emplea el identificador almacenado previamente. El listado 3.5 permite visualizar un fragmento de código relacionado con la petición para obtener datos de los usuarios mediante la Graph API (líneas 3-11).

```

1.  public function readUserData() {
2.      $batch = array();
3.      foreach ($this->idUsers as $id) {
4.          $batch[] = $requestUserName = $this->fb->request("GET", "/" . $id . "?fields=
5.                                  name,id,gender,birthday,location");
6.      }
7.      try {
8.          $responses = $this->fb->sendBatchRequest($batch);
9.      } catch (Exception $e) {
10.         echo "La API regresó un error: " . $e->getMessage();
11.     }
12.     $this->saveUsers($responses);
13. }

```

Listado 3.5 Petición para extraer datos de usuarios en Facebook®.

3.7.2.2 Extracción de datos en Twitter®

Para extraer información de Twitter® se requirió establecer una conexión con dicha red social empleando TwitterOAuth, la clave de la aplicación, la clave secreta de la aplicación, la ficha y la ficha secreta de acceso. El listado 3.6 muestra la función para establecer la conexión con Twitter®.

```

1.  public function connect() {
2.      $this->tw = new TwitterOAuth("consumer_key", "consumer_secret",
3.                                  "Access Token", "Access Token Secret");
4.  }

```

Listado 3.6 Función para establecer la conexión con Twitter®.

Tras establecer la conexión, se requiere el enlace o el identificador de un tuit en específico para realizar la extracción de las respuestas de éste. De cada tuit se almacena el texto que lo conforma y el nombre de usuario de la persona que lo escribió. La extracción

de tuits a partir del identificador se aprecia en el fragmento de código localizado en el listado 3.7 (líneas 3-28).

```

1.  public function searchByID($pag, $id) {
2.      if (!empty($pag) && !empty($id)) {
3.          $prim_id = 0;
4.          $ult_id = 0;
5.          $this->data = array();
6.          $param = ["q" => "to:" . $pag, "lang" => "en", "since_id" => $id, "count" => 100];
7.          do {
8.              $statuses = $this->tw->get("search/tweets", $param);
9.              if ($statuses != null) {
10.                 foreach ($statuses as $elem) {
11.                     if (is_array($elem)) {
12.                         foreach ($elem as $e) {
13.                             $prim_id = $ult_id;
14.                             $ult_id = $e->id;
15.                             if ($e->in_reply_to_status_id_str == $id && !(strcasecmp($e->text, "") == 0)) {
16.                                 $user = $e->user;
17.                                 $ud = new UserData();
18.                                 $ud->setComment($e->text);
19.                                 $ud->setId($user->screen_name);
20.                                 $this->data[] = $ud;
21.                             }
22.                         }
23.                     }
24.                 }
25.             }
26.             $param = ["q" => "to:" . $pag, "lang" => "en", "since_id" => $id, "max_id" => $ult_id,
27.                 "count" => 100];
28.         } while ($prim_id < $ult_id);
29.         $this->deleteURL();
30.         $this->searchUsers();
31.     }
32. }

```

Listado 3.7 Extracción de tuits a partir de un identificador en Twitter®.

De manera similar a como ocurre con Facebook®, al tener los tuits escritos como respuesta que se requieren para trabajar, se continúa con la eliminación de enlaces encontrados en éstos y con la extracción de los datos generales (nombre y ubicación) de los usuarios que escribieron los tuits, para lo cual se emplea el nombre de usuario almacenado previamente. En el listado 3.8 se visualiza un fragmento de código relacionado con la obtención de datos de los usuarios (líneas 2-15).

```

1. public function searchUsers() {
2.     $par = "";
3.     if (!empty($this->data)) {
4.         foreach ($this->data as $aux) {
5.             $par .= $aux->getId() . ",";
6.         }
7.         $i = 0;
8.         $par2 = substr($par, 0, -1);
9.         $statuses = $this->tw->get("users/lookup", ["screen_name" => $par2]);
10.        foreach ($statuses as $e) {
11.            $this->data[$i]->setName($e->name);
12.            $this->data[$i]->setLocation($e->location);
13.            $i++;
14.        }
15.    }
16. }

```

Listado 3.8 Extracción de datos de usuarios en Twitter®.

3.7.3 Extracción de sentimientos y emociones

La extracción de sentimientos y emociones requirió el desarrollo de un servicio Web que analizara un conjunto de textos mediante Natural Language Understanding®, así como el desarrollo de las funcionalidades necesarias para que el módulo consuma dicho servicio Web. A continuación, se describe lo concerniente al servicio Web y a las demás funcionalidades desarrolladas.

3.7.3.1 Servicio Web para extraer sentimientos y emociones

El servicio Web necesario para realizar la obtención de sentimientos y emociones se desarrolló mediante Java. Dicho servicio Web proporciona una operación llamada “*analyze*”, la cual recibe el conjunto de comentarios o textos a analizar. En dicha operación se obtiene para cada comentario, la polaridad sentimental y el porcentaje de enojo, desagrado, miedo, alegría y tristeza que expresa. Cabe mencionar que el análisis se lleva a cabo a nivel de comentario y a nivel de entidad.

En el análisis a nivel de comentario se obtiene la polaridad sentimental y el porcentaje de cada emoción que expresa un comentario completo, mientras que en el análisis a nivel de entidad se evalúan las mismas características para cada una de las entidades que forman parte de un comentario, en donde por entidad se entiende a cualquier

persona, lugar u objeto sobre el que se esté expresando alguna opinión, crítica o sentimiento. La Tabla 3.7 ejemplifica los valores obtenidos tras realizar el análisis a nivel de comentario y a nivel de entidad para una oración dada.

Análisis a nivel de comentario y entidad de una oración				
Oración: “I love sports but I can’t stand golf”				
Nivel de comentario		Nivel de entidad		
Aspecto		Aspecto	Entidades	
			Sports	Golf
Polaridad	0	Polaridad	0.6358	0
Enojo	0.0312	Enojo	0.0155	0.1487
Desagrado	0.0515	Desagrado	0.0091	0.3034
Miedo	0.0078	Miedo	0.0022	0.1171
Alegría	0.8708	Alegría	0.9497	0.2501
Tristeza	0.0387	Tristeza	0.0236	0.1807

Tabla 3.7 Valores obtenidos tras analizar un comentario en diferentes niveles.

Como se observa en la Tabla anterior, el análisis a nivel de comentario permite comprender cuál es la reacción que las personas expresan a través de sus comentarios de manera general, mientras que el análisis a nivel de entidad permite conocer la reacción de las personas con un mayor nivel de profundidad, pues se evalúan los sentimientos y emociones que se expresan internamente en cada comentario. Además, el análisis a nivel de entidad es bastante útil porque en muchos de los casos las personas expresan distintos sentimientos refiriéndose a diferentes personas, lugares u objetos mediante un mismo comentario. El listado 3.9 ejemplifica la petición para obtener sentimientos, emociones y entidades mediante un fragmento de código (líneas 4, 6-25).

```

1. public UserData[] analyze(@WebParam(name = "comments") String[] comments) {
2.     UserData[] ud;
3.     if (service==null)
4.         service = new NaturalLanguageUnderstanding("Version", "ClaveAPI", "Access-Token");
5.     if (comments.length > 0) {
6.         ud = new UserData[comments.length];
7.         EntitiesOptions entities = new EntitiesOptions.Builder().emotion(true).
8.             sentiment(true).build();
9.         SentimentOptions sentimentOp = new SentimentOptions.Builder().build();
10.        EmotionOptions emotionOp = new EmotionOptions.Builder().build();
11.        Features features = new Features.Builder().sentiment(sentimentOp).
12.            emotion(emotionOp).entities(entities).build();
13.    }

```

```

14.     for (int i = 0; i < comments.length; i++) {
15.         AnalyzeOptions parameters = new AnalyzeOptions.Builder().text(comments[i]).
16.             features(features).build();
17.         try{
18.             AnalysisResults response = service.analyze(parameters).execute();
19.             ud[i] = new UserData();
20.             saveResults(response, ud[i], comments[i]);
21.         }catch(Exception e){
22.             ud[i] = new UserData();
23.             saveResults(null, ud[i], comments[i]);
24.         }
25.     }
26. } else {
27.     ud = new UserData[0];
28. }
29. return ud;
30. }

```

Listado 3.9 Fragmento de código para extraer sentimientos, emociones y entidades.

3.7.3.2 Consumo de servicio Web para extraer sentimientos y emociones

Tras desarrollar el servicio Web que obtiene los sentimientos y emociones en diferentes niveles de análisis mediante el uso de Natural Language Understanding®, se desarrollaron las funcionalidades necesarias en PHP para consumir dicho servicio Web. El listado 3.10 muestra un fragmento de código en el que se ejemplifica el consumo del servicio Web desde el módulo (líneas 3-17).

```

1.     public function processSA($n, $val) {
2.         if ($this->param != null) {
3.             try {
4.                 $client = new
5.                     SoapClient("http://sentimentanalysis.jvmhost.net/ProvSA/EmoWS?wsdl");
6.                 $oResult = $client->analyze($this->param);
7.                 $this->ban = "";
8.                 if (is_array($oResult->return)) {
9.                     $this->calc->calclArray($oResult, $this->polarity, $this->emotions, $this->param);
10.                } else {
11.                    $oUd = $oResult->return;
12.                    $this->calc->calclNotArray($oUd, $this->polarity, $this->emotions);
13.                }
14.                ...
15.            } catch (Exception $ex) {
16.                echo " ERROR: " . $ex . "<br/>";
17.            }
18.        } else {

```

```

19.     $msjerror = "";
20.     if (!$val) {
21.         $msjerror = "Introduce a valid URL";
22.     } else {
23.         $msjerror = "No comments found!";
24.     }
25.     }
26.     }

```

Listado 3.10 Método que consume el servicio Web para extraer sentimientos y emociones.

Tal como se aprecia en el fragmento de código anterior, una vez que se obtiene el resultado del servicio Web, se invoca la función “calcIsArray” o la función “calcIsNotArray” dependiendo de si el resultado es un arreglo de objetos o un objeto. En cualquiera de los casos, en dichos métodos se procesa el resultado para extraer los valores de cada una de las emociones y de la polaridad sentimental, para posteriormente obtener promedios de ellos, cuidando aspectos como la división entre cero, entre algunos otros. El listado 3.11 muestra un fragmento de código que ejemplifica la función “calcIsArray”.

```

1.     public function calcIsArray($oResult, &$polarity, &$emotions, $param) {
2.         $ceros = array(0, 0, 0, 0, 0, 0);
3.         $i = 1;
4.         foreach ($oResult->return as $oUd) {
5.             $sum = $oUd->anger + $oUd->disgust + $oUd->fear + $oUd->joy + $oUd->sadness;
6.             $sum = ($sum > 0) ? $sum : 1;
7.
8.             $polarity += $oUd->polarity;
9.             $emotions[0] += ((1 / $sum) * $oUd->anger);
10.            $emotions[1] += ((1 / $sum) * $oUd->disgust);
11.            $emotions[2] += ((1 / $sum) * $oUd->fear);
12.            $emotions[3] += ((1 / $sum) * $oUd->joy);
13.            $emotions[4] += ((1 / $sum) * $oUd->sadness);
14.            $oUd->comment = "";
15.            $this->data[$i - 1] = $oUd;
16.            $this->verifyZero($ceros, $oUd);
17.            $i++;
18.        }
19.        ...
20.    }

```

Listado 3.11 Función que procesa el resultado del análisis cuando se trata de un arreglo de objetos.

3.7.4 Representación gráfica del análisis de sentimientos y emociones

Otro aspecto de gran interés en el desarrollo del módulo es la representación gráfica de los resultados obtenidos tras el análisis de textos o comentarios llevado a cabo. La representación gráfica permite que los resultados sean más fáciles de comprender, analizar y utilizar por los diferentes interesados.

Para el caso de Facebook[®] se optó por generar las siguientes gráficas: 1) Polaridad, gráfica de pastel que muestra el porcentaje de positividad y negatividad expresada hacia la publicación; 2) Emociones, gráfica de pastel que muestra el porcentaje de cada emoción (enojo, desagrado, miedo, alegría, tristeza) expresado; 3) Emociones por género, gráfica de barras que muestra la distribución de emociones por género; 4) Emociones por edad, gráfica de barras que muestra la distribución de emociones por rango de edad; 5) Emociones por ubicación, gráfica de barras que muestra la distribución de emociones por ubicación, y 6) Entidades, mapa de calor que muestra la variación en la polaridad de las diferentes entidades. Para el caso de Twitter[®] las gráficas generadas son las mismas con la excepción de las gráficas de emociones por género y edad, ya que no es posible obtener tales datos de dicha red social.

En cuanto al desarrollo de las funcionalidades requeridas para la representación gráfica se inició por calcular los promedios de la polaridad sentimental y de cada una de las emociones (enojo, desagrado, miedo, alegría y tristeza), de manera que se agruparan por género, rango de edad y ubicación para Facebook[®] y únicamente por ubicación para Twitter[®]. Para el género se consideraron tres grupos: 1) *Men*, para agrupar los valores expresados por hombres; 2) *Women*, para agrupar los valores expresados por mujeres, y 3) *No Data*, para agrupar los valores expresados por personas cuyo género es privado. Asimismo, para la edad se establecieron cuatro rangos: 1) 13-17 años; 2) 18-24 años; 3) Más de 24 años, y 4) *Unknown*, para cuando los usuarios no permitan conocer su edad ni fecha de nacimiento. Por otro lado, para el caso de la ubicación no se establecieron rangos fijos, sino que el agrupamiento se realiza tomando como base las ubicaciones extraídas. El listado 3.12 ejemplifica el agrupamiento de sentimientos y emociones por género (líneas 4-25).

```

1.  public function emotionGender() {
2.      $res = array(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0);
3.      $m = $w = $nd = 0;
4.      foreach ($this->data as $ud) {
5.          if (strcasecmp($ud->gender, "m") == 0) {
6.              $res[0] += $ud->anger;
7.              $res[1] += $ud->disgust;
8.              $res[2] += $ud->fear;
9.              $res[3] += $ud->joy;
10.             $res[4] += $ud->sadness;
11.             $m++;
12.         } else {
13.             if (strcasecmp($ud->gender, "f") == 0) {
14.                 $res[5] += $ud->anger;
15.                 $res[6] += $ud->disgust;
16.                 ...
17.                 $w++;
18.             } else {
19.                 $res[10] += $ud->anger;
20.                 $res[11] += $ud->disgust;
21.                 ...
22.                 $nd++;
23.             }
24.         }
25.     }
26.     ...
27. }

```

Listado 3.12 Agrupamiento de polaridad y emociones por género.

Una vez que se acumularon los valores de polaridad y emociones obtenidos para cada género, se procedió a calcular el promedio de éstos. Al contar con dichos valores, se desarrollaron las funcionalidades necesarias para mostrarlos mediante alguna gráfica, para lo cual se empleó Plotly.js. El listado 3.13 muestra la función desarrollada para generar una gráfica de barras que muestre la distribución de emociones por género.

```

1.  function showEmoGender(udata) {
2.      var data = [];
3.      var colors = ["rgb(129, 190, 247)", "rgb(247, 129, 159)", "rgb(245, 246, 206)"];
4.      var names = ["Men", "Women", "No data"];
5.      var k = 0;
6.      for (i = 0; i < 15; i = i + 5) {
7.          var trace1 = {
8.              x: ['Anger', 'Disgust', 'Fear', 'Joy', 'Sadness'],
9.              y: [ud[i] * 100, ud[(i + 1)] * 100, ud[(i + 2)] * 100, ud[(i + 3)] * 100, ud[(i + 4)] * 100],
10.             name: names[k],

```

```

11.     type: 'bar',
12.     marker: {
13.         color: colors[k]
14.     }
15.     };
16.     data[k] = trace1;
17.     k++;
18. }
19. var layout = {title: 'Emotions by Gender', barmode: 'group'};
20. Plotly.newPlot("dpol", data, layout);
21. }

```

Listado 3.13 Función que permite desarrollar la gráfica de emociones por género.

3.7.5 Interpretación de la información

Como se mencionó anteriormente, el resultado del análisis que lleva a cabo el módulo se representa a través de distintas gráficas con la finalidad de facilitar la comprensión e interpretación de los resultados. Entre las gráficas que se seleccionaron para mostrar los resultados se encuentran principalmente: 1) Gráficas de pastel o circulares; 2) Gráficas de barras, y 3) Mapas de calor. Dichas gráficas se seleccionaron porque tienen importantes usos desde el punto de vista de la estadística que resultan de gran utilidad en este proyecto, por lo cual algunos de sus principales usos se explican a continuación con más detalle.

1. **Gráficas de pastel:** La principal utilidad dentro de la estadística que tienen las gráficas de pastel o circulares es la representación de porcentajes y proporciones, por lo cual resultaron ser bastante adecuadas en el caso de la representación de polaridad y emociones, ya que, en ambos casos se pretendía representar el porcentaje o la proporción obtenidos para cada tipo de emoción y para cada tipo de polaridad.
2. **Gráficas de barras:** Las gráficas de barra se caracterizan por permitir la representación de un conjunto de datos mediante barras horizontales o verticales que son proporcionales a dichos datos. Asimismo, tienen la capacidad de representar porcentajes; sin embargo, la característica primordial que se consideró de interés para el módulo es la facilidad que brindan para comparar dos o más valores que corresponden a la misma categoría, por lo cual, constituyen una buena opción para visualizar la distribución de emociones por edad, género y ubicación, ya que en cada

una de éstas se requieren comparar los valores de distintos rangos o categorías para cada emoción. Un ejemplo de esto es cuando se comparan los diferentes valores de alegría que mostraron las personas de acuerdo con su género.

3. **Mapa de calor:** La utilidad de los mapas de calor es permitir la visualización de las zonas de mayor y menor interés para los usuarios desde una perspectiva global empleando una jerarquía de dos polos y distintos colores para identificar los valores. Estas últimas características resultan convenientes para representar la polaridad sentimental, por lo cual, se determinó emplear un mapa de calor para visualizar la polaridad sentimental relacionada con cada una de las entidades. De esta forma, se facilita y agiliza la identificación de cuáles son las entidades con la polaridad más positiva, y cuáles son las entidades con la polaridad más negativa.

Por otro lado, es importante mencionar que las gráficas se diseñaron de manera que resulten intuitivas para quienes intenten interpretarlas, sin embargo, algunas de ellas poseen aspectos que es necesario aclarar para que no exista ninguna confusión con la interpretación de la información. Por ejemplo, la gráfica de emociones por género (Figura 3.4) representa a través de sus distintas barras el porcentaje de cada emoción que expresaron los hombres y las mujeres en promedio. No obstante, es importante aclarar que existe una barra denominada “No Data” que expresa el porcentaje de cada emoción que expresaron aquellas personas cuyo género se desconoce a causa de que su configuración de privacidad en Facebook® impide obtenerlo.

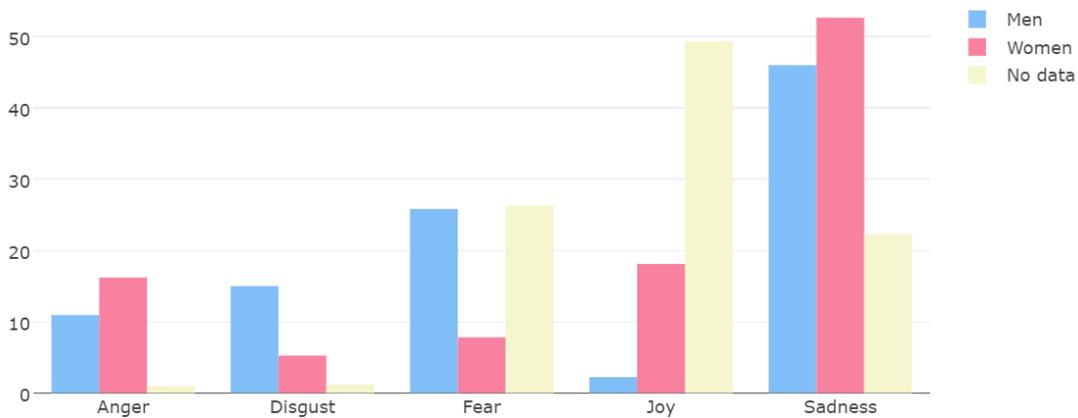


Figura 3.4 Gráfica de emociones por género.

Por otro lado, en la Figura 3.5 es posible apreciar la gráfica de emociones por edad. Las barras de esta gráfica representan el porcentaje de cada emoción que expresaron las personas cuya edad se encuentra entre los rangos: 13-17 años, 18-24 años y mayor a 24 años. Sin embargo, hay que aclarar que también existe una barra denominada “*Unknown*” que expresa el porcentaje de cada emoción expresada por aquellas personas cuya edad se desconoce porque su configuración de privacidad en Facebook® impide obtenerla.

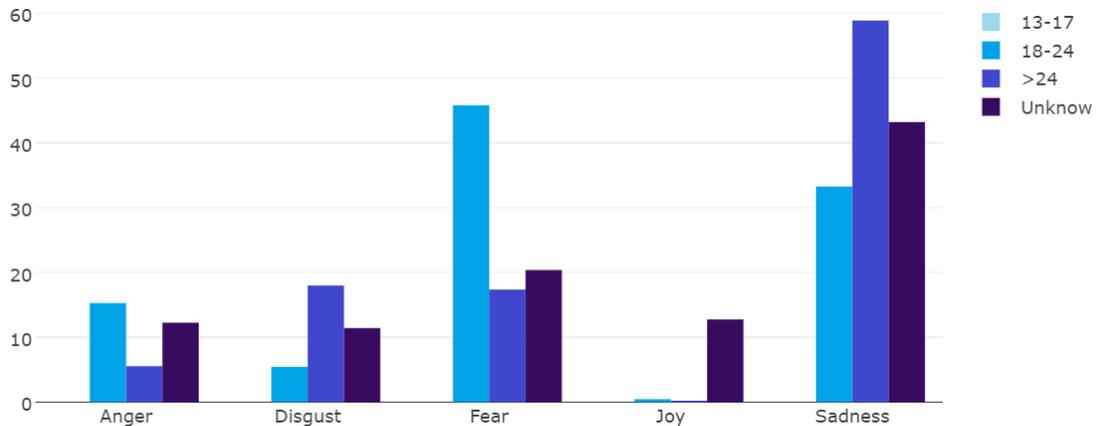


Figura 3.5 Gráfica de emociones por edad.

Finalmente, la gráfica de entidades que se muestra en la Figura 3.6 consiste en un mapa de calor compuesto por distintos recuadros de colores. Cada recuadro representa la polaridad sentimental expresada hacia una entidad en particular. El color de los recuadros corresponde proporcionalmente con el valor de polaridad que representa. De esta manera, entre más azul es el color del recuadro de una entidad, más positivo es su valor, mientras que entre más rojo es el color de un recuadro, más negativo es su valor.

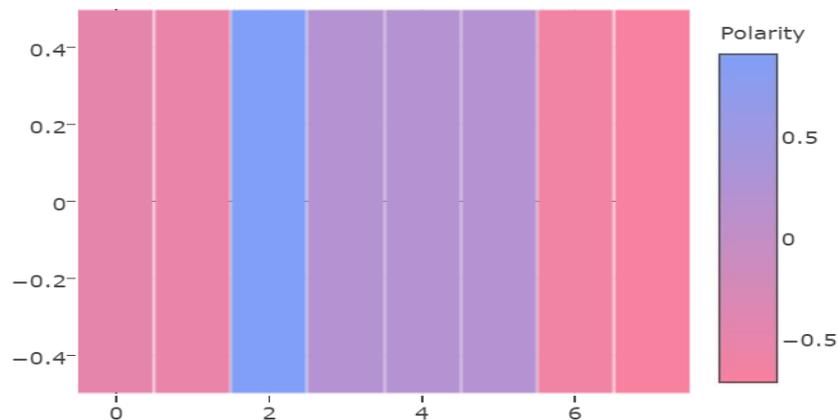


Figura 3.6 Gráfica que muestra la polaridad de entidades.

Capítulo IV. Resultados

En este capítulo se describen los resultados obtenidos con la realización del presente proyecto, para lo cual, se presentan dos casos de estudio relacionados con el ámbito de la salud en donde se visualiza el funcionamiento del módulo de obtención de emociones. Las principales funcionalidades de dicho módulo son: Extraer información de Facebook®, extraer información de Twitter®, obtener sentimientos, obtener emociones, extraer entidades y mostrar gráficas con los resultados de los distintos cálculos.

4.1 Casos de estudio

Como resultado de este proyecto se desarrolló un módulo mediante PHP que tiene la capacidad de proporcionar diferentes funcionalidades a las aplicaciones Web en las que se integre. La finalidad del módulo es analizar un conjunto de textos provenientes de comentarios hacia publicaciones de Facebook® o hacia un tuit en Twitter®, y mostrar mediante una ventana modal la información concerniente al análisis, la cual consiste en: 1) Comentarios obtenidos de la publicación; 2) Cálculo del porcentaje de polaridad sentimental y del porcentaje de cada emoción (enojo, desagrado, miedo, alegría, tristeza) identificados para cada texto; 3) Entidades encontradas, es decir, los objetos, personas, lugares, entre otros, hacia los cuales se expresó alguna opinión en los textos analizados, y 4) Gráficas que representan y resumen los resultados del análisis.

Es importante mencionar que las gráficas que el módulo presenta se detallaron en el capítulo anterior y que el beneficio principal que aporta el módulo hacia los diferentes interesados, es el de facilitar la obtención y comprensión de información relacionada con los sentimientos y emociones que las personas expresan mediante sus comentarios hacia diferente tópicos, de manera que éstos obtengan otros tipos de beneficios, por ejemplo, tomar mejores decisiones, cambiar estrategias de negocios, planear mejores campañas publicitarias, entre otros, con base en la información proporcionada. Para ilustrar el funcionamiento del módulo a continuación se describen dos casos de estudio.

4.1.1 Caso de estudio: Análisis sentimental sobre comentarios de médicos especialistas en diabetes para seleccionar el más adecuado para atender dicha enfermedad

- Supóngase que se le diagnosticó recientemente diabetes tipo 1 a una mujer de 30 años y que ésta utiliza una aplicación Web que le permite obtener información y consejos sobre su enfermedad, así como llevar un control sobre sus citas y sus pruebas clínicas.
- La aplicación además ofrece información sobre distintos médicos especialistas con los cuales la persona tiene la capacidad de atender su enfermedad; sin embargo, dicha persona no está segura acerca de qué médico es el más adecuado para atenderla porque no los conoce y por lo tanto requiere más información.
- Mediante la aplicación es posible visualizar un perfil para cada médico, en donde se visualiza un sistema de reputación y diferentes comentarios realizados por otros usuarios hacia dicho médico. No obstante, esta información no es suficiente, pues el sistema de reputación no permite conocer la razón por la cual el médico tiene dicha reputación, mientras que la lectura e interpretación de cada uno de los comentarios hacia cada médico resulta una tarea que consume mucho tiempo, es cansada y tediosa. Por otro lado, la aplicación no cuenta con algún mecanismo que facilite la tarea antes mencionada y desarrollar las funcionalidades necesarias para resolver el problema consumiría mucho tiempo y esfuerzo. Como consecuencia de esto surgen las siguientes interrogantes:
 1. ¿De qué forma el usuario tendría la posibilidad de obtener información sobre el sentir expresado hacia cada médico, que sea fácil de comprender y que no consuma mucho de su tiempo?
 2. ¿De qué manera la aplicación será capaz de integrar las funcionalidades necesarias para proporcionar información, como la que requiere la mujer en cuestión, sin que implique un gran esfuerzo en el desarrollo?

Una alternativa de solución es la integración e incorporación de las funcionalidades del módulo de obtención de emociones a la aplicación Web. A continuación, se describe la forma en la que la persona o usuario interactúa con las funcionalidades del módulo, partiendo de que ésta se encuentra observando el perfil de un médico en donde existen comentarios escritos por diferentes usuarios mediante su cuenta de Facebook®, y un botón a

través del cual la aplicación Web integra y utiliza las funcionalidades del módulo para analizar dichos comentarios.

1.- El usuario da clic sobre el botón “Analizar Opiniones” localizado dentro de la página que muestra el perfil del médico para hacer uso de la funcionalidad del módulo de extracción de textos y cálculo de sentimientos, emociones y entidades (Figura 4.1).

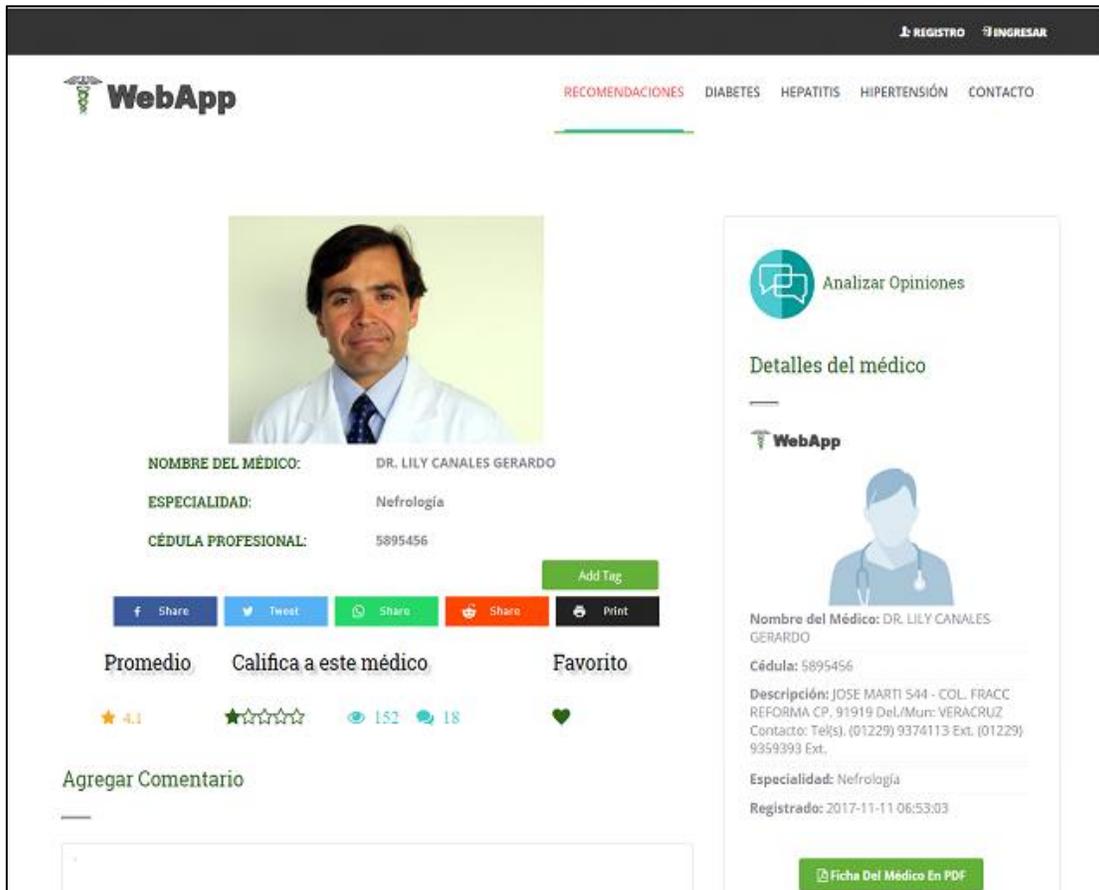


Figura 4.1 Perfil de un médico en donde es posible emplear las funcionalidades del módulo.

2.- Tras finalizar el análisis se despliegan los resultados mediante una ventana modal. Dicha ventana contiene la información resultante del análisis distribuida mediante cuatro pestañas que el usuario tiene la posibilidad de seleccionar en cualquier momento y cualquier orden, siendo la pestaña de comentarios la que aparece activa por defecto. En esta pestaña el usuario tiene la posibilidad de apreciar los comentarios que se analizaron, lo cual se muestra en la Figura 4.2.

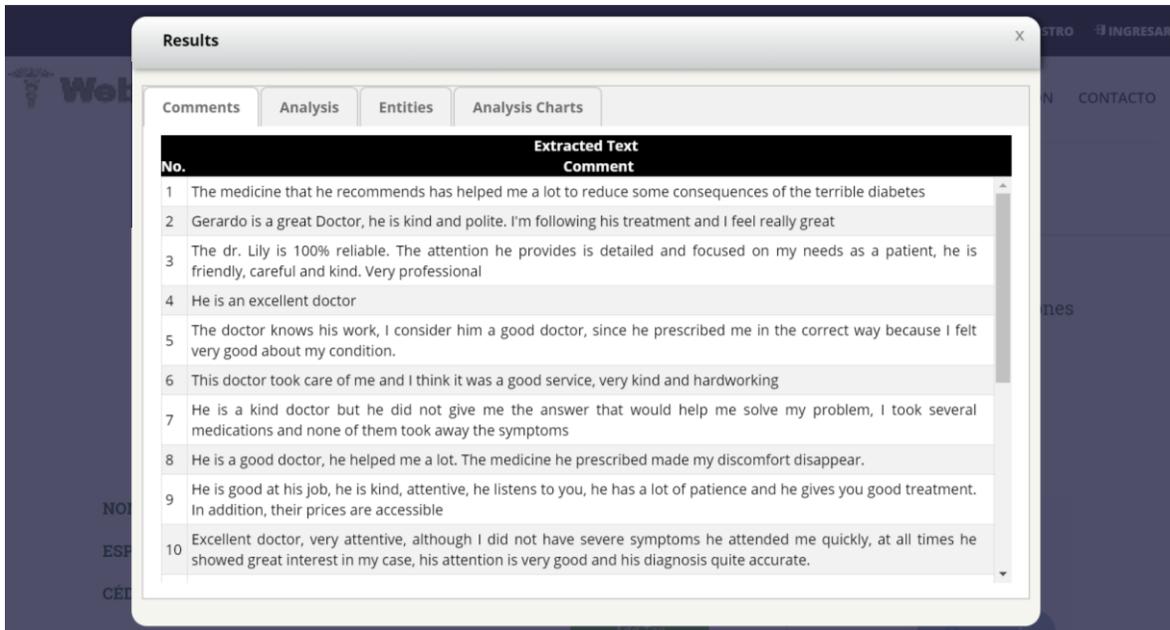


Figura 4.2 Ventana emergente que muestra los comentarios analizados.

3.- El usuario da clic a la segunda pestaña, la cual muestra el resultado del análisis de manera tabular. Mediante esta pestaña se aprecia el porcentaje de polaridad y de cada emoción asociado a cada uno de los comentarios analizados, con lo cual se sabe si el médico tiene una aceptación buena o mala de manera general (Figura 4.3).

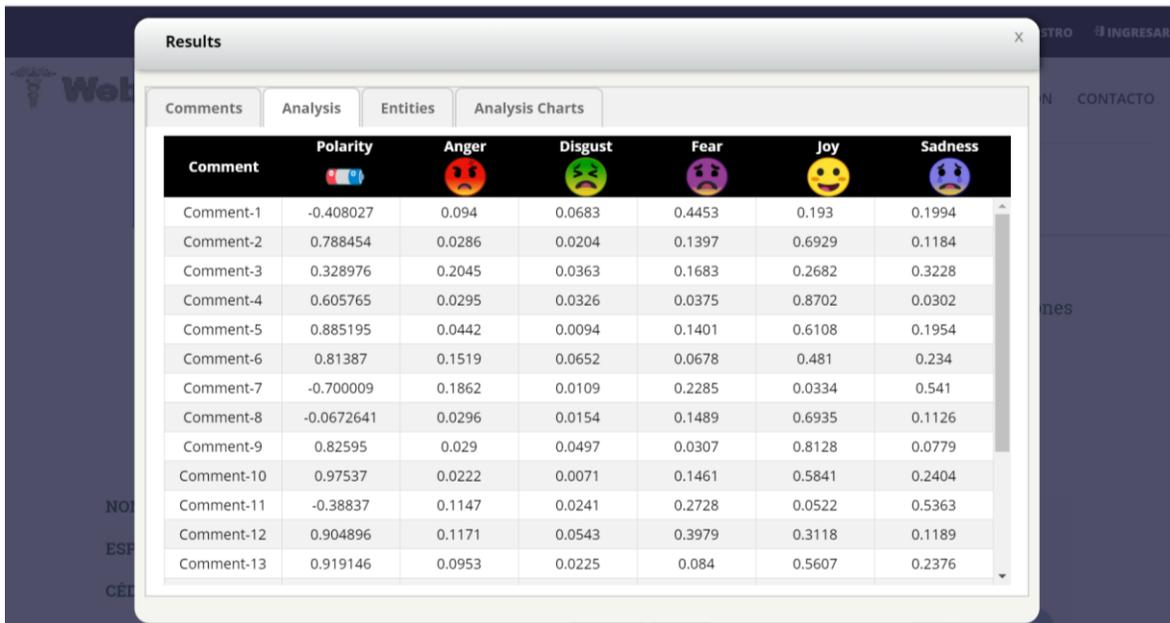


Figura 4.3 Pestaña que muestra el porcentaje de polaridad y emociones para cada comentario.

4.-Tras dar clic en la tercera pestaña el usuario visualiza de manera tabular el conjunto de entidades identificadas en cada comentario, junto con el tipo de entidad del que se trata, el porcentaje de polaridad y de emociones expresado hacia dichas entidades. Esta pestaña es muy importante porque no solo permite identificar si el sentir hacia el médico es positivo o negativo, sino que además es posible conocer cuáles aspectos generan que ese sentir sea positivo o negativo, por ejemplo, quizás las personas están expresando un sentimiento negativo hacia un médico a causa del costo de sus consultas y un sentimiento positivo por la atención que brinda a los pacientes. Esta pestaña se aprecia en la Figura 4.4.

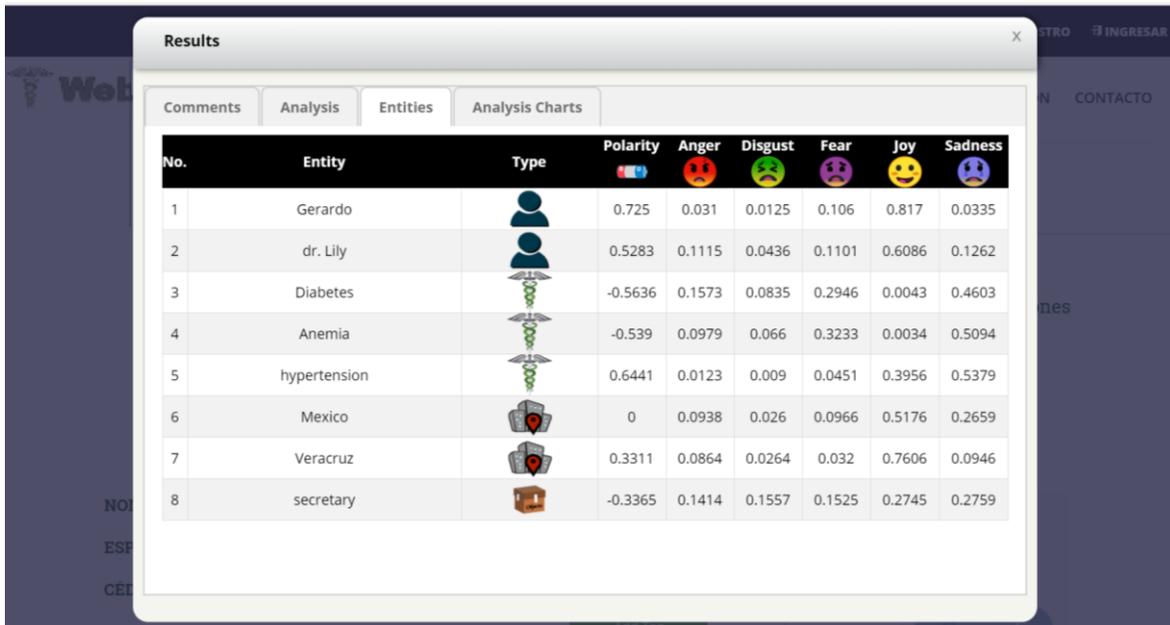


Figura 4.4 Pestaña que muestra las entidades encontradas junto con su polaridad y emociones.

5.-El usuario da clic en la última pestaña, la cual muestra el conjunto de gráficas en donde se resume la información resultante del análisis para facilitar la comprensión de ésta. En dicha pestaña el usuario tiene la posibilidad de seleccionar la gráfica que desea visualizar dando clic sobre cualquiera de las opciones que aparecen en la parte superior. Al dar clic sobre la opción “polarity” el usuario visualizará mediante una gráfica de pastel, el porcentaje de positividad y negatividad expresado en promedio hacia el médico. La positividad se muestra de color azul y la negatividad de color rojo. Dicha gráfica se aprecia en la Figura 4.5.

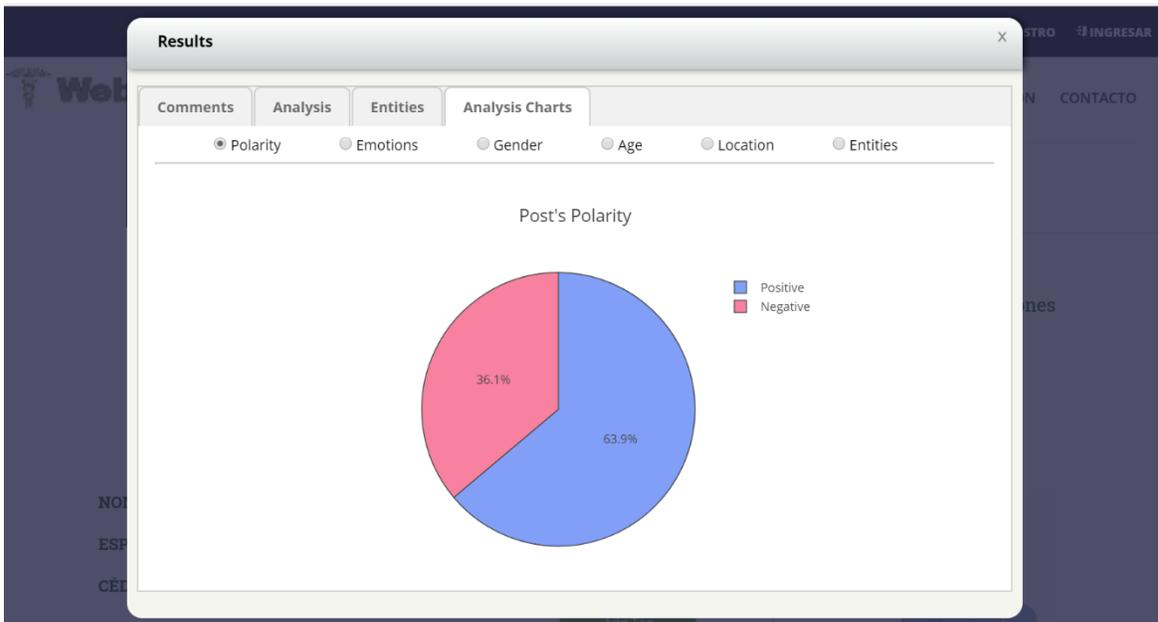


Figura 4.5 Gráfica que muestra la polaridad expresada hacia el médico.

Al seleccionar la opción “*Emotions*” se muestra una gráfica circular que representa el porcentaje de cada emoción que los usuarios manifestaron en promedio hacia el médico. Las emociones que se muestra corresponden a enojo en color rojo, desagrado en color verde, miedo en color morado, alegría en color amarillo y tristeza en color azul, lo cual se visualiza en la Figura 4.6.

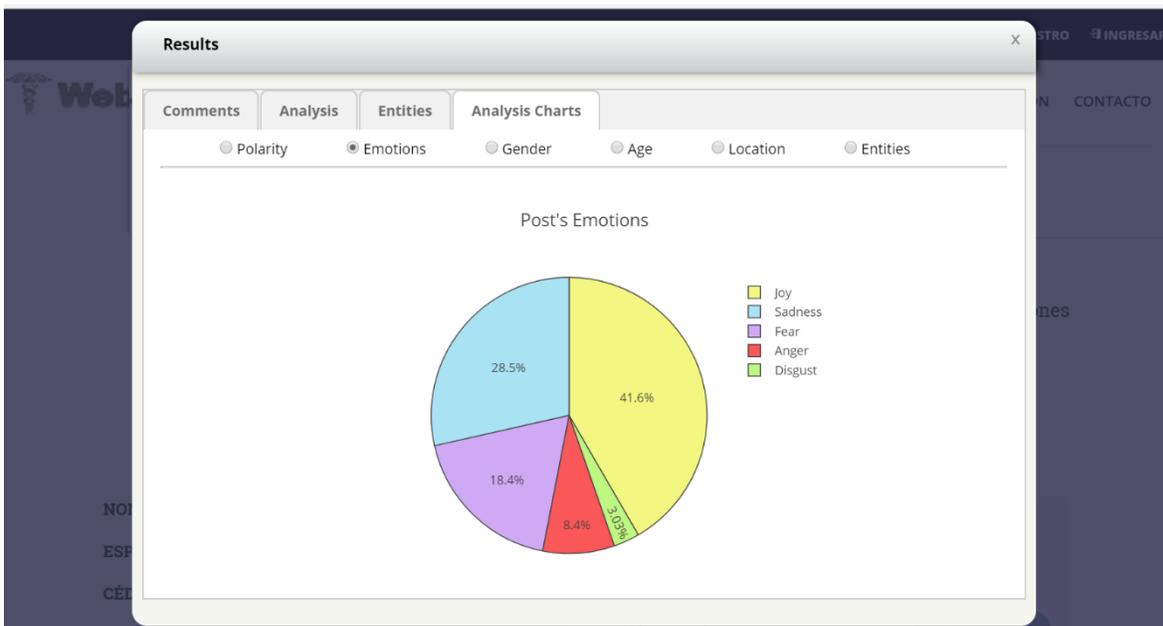


Figura 4.6 Gráfica que muestra los porcentajes de cada emoción expresados hacia el médico.

Mediante la opción “*Gender*” el usuario tiene la posibilidad de observar una gráfica de barras que muestra las emociones que manifestaron con más fuerza personas de su mismo género o de otro género. En color azul se representan los porcentajes de emociones manifestados por hombres, en rojo por mujeres y en amarillo por personas cuya configuración de privacidad en Facebook® impide conocer su género. Cabe mencionar que el rango de cada emoción por género va de 0 a 100 y que la suma de todas las emociones mostradas por género corresponde al 100%. La Figura 4.7 muestra lo descrito anteriormente.

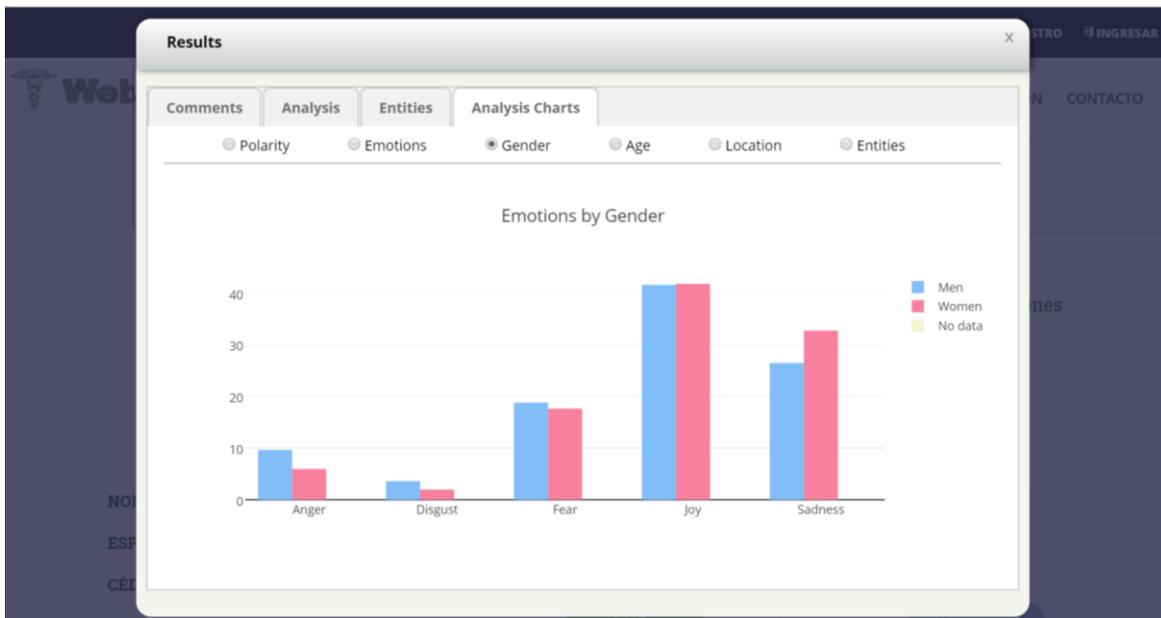


Figura 4.7 Gráfica que muestra la distribución de emociones por género.

Al dar clic en la opción “*Age*” el usuario tiene la posibilidad de observar la distribución de emociones por rangos de edad a través de una gráfica de barras, con lo cual se identifican cuáles son las emociones que manifestaron más fuerte las personas con edad similar a la suya. La gráfica cuenta con una leyenda que permite identificar la barra que corresponde a cada rango. Cabe mencionar que al igual que en la gráfica anterior, la suma de los porcentajes de cada emoción por rango es el 100% (Figura 4.8).

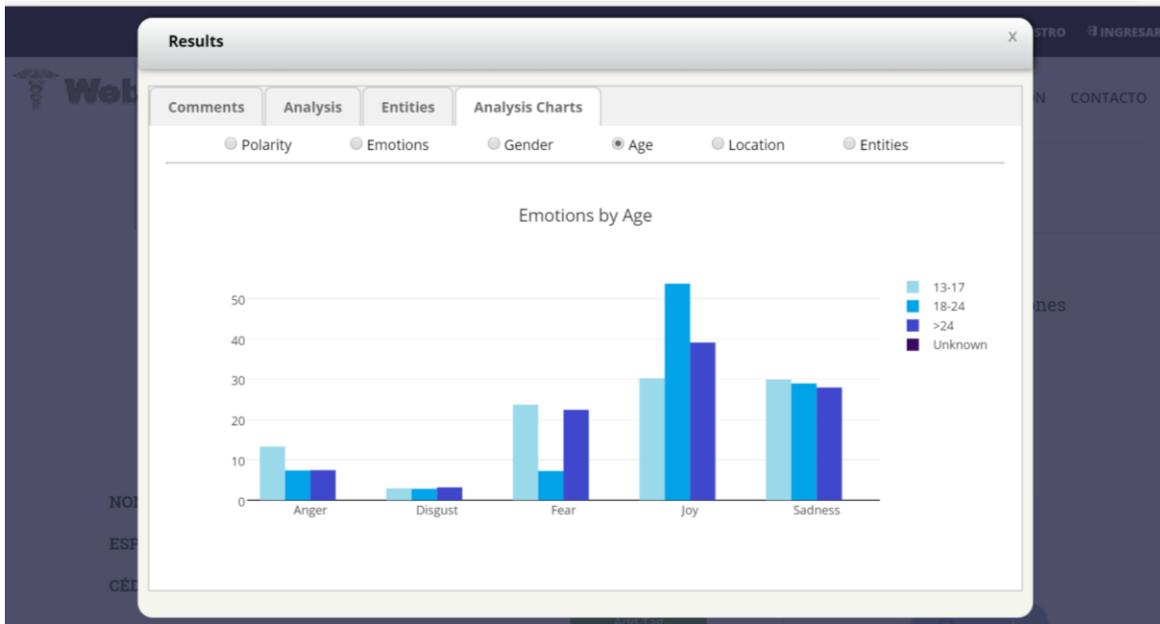


Figura 4.8 Gráfica que muestra la distribución de emociones por edad.

Asimismo, mediante la opción “*Location*” se observa la distribución de emociones por ubicación a través de una gráfica de barras. Dicha gráfica muestra cómo se distribuyen las emociones entre las ciudades a las cuales pertenecen los usuarios que comentaron. En este caso en particular los usuarios no proporcionaron información sobre su ubicación, por lo cual se muestra como desconocida (Figura 4.9). El rango de valores para cada barra va del 0 al 100 y la suma de los porcentajes de cada emoción por ciudad es del 100%.

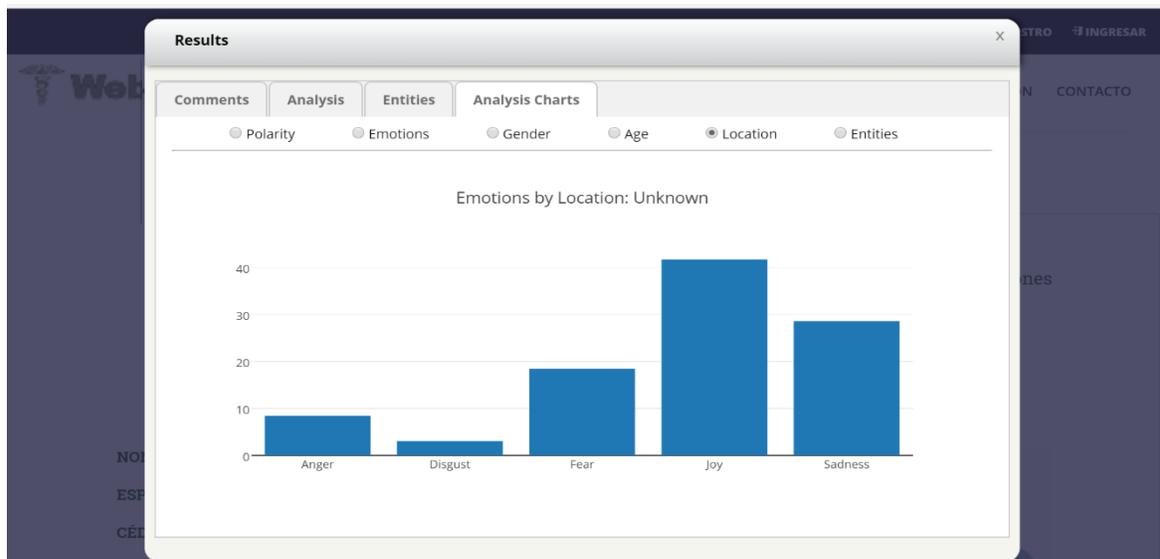


Figura 4.9 Gráfica que muestra la distribución de emociones por ubicación.

Finalmente, mediante la opción “Entities” el usuario tiene la posibilidad de observar un mapa de calor en donde se aprecia cuáles son las entidades más positivas y cuáles son las más negativas entre las entidades identificadas en los comentarios analizados sobre el médico. La positividad o negatividad de cada entidad se representa a través de una coloración que va de azul a rojo, de manera que entre más positiva es la entidad más azul es el recuadro que lo representa y entre más negativa es la entidad, más rojizo es el recuadro. Además, es importante mencionar que el rango de valores que se grafican para cada entidad se encuentra entre -1 y 1, donde -1 es muy negativo y 1 muy positivo. En la Figura 4.10 es posible observar que la entidad “Diabetes” es muy negativa, pues el recuadro que la representa tiene una coloración rojiza, lo cual es posible verificar al observar su valor, el cual es de -0.5636.

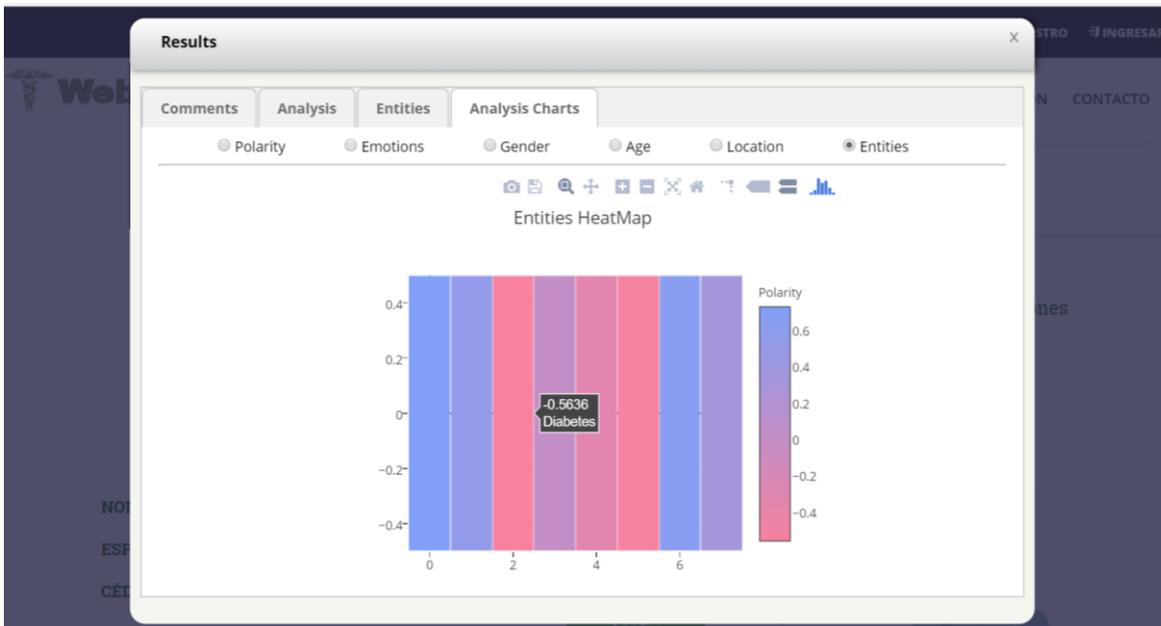


Figura 4.10 Mapa de calor que muestra la polaridad de entidades.

De esta manera, para el usuario resulta más rápido y sencillo apreciar cuáles médicos tienen mejores opiniones y conocer en qué se basan éstas, saber lo que opinan personas de su mismo género, ciudad o edad, lo cual en conjunto facilita la elección del médico para tratar y dar seguimiento a su enfermedad.

4.1.2 Caso de estudio: Análisis sentimental sobre comentarios de medicamentos para tratar la hipertensión

- Supóngase que un hombre de 40 años que padece hipertensión utiliza una aplicación Web que le permite consultar información sobre distintos tipos de medicamentos que son útiles para tratar su enfermedad. Dicho hombre, desea conocer la opinión que tienen otras personas que padecen su misma enfermedad sobre medicamentos que él está consumiendo como parte de su tratamiento.
- A través de la aplicación es posible visualizar los detalles de cada medicamento, en donde se presenta un sistema de reputación basado en una escala del 0 al 5 y el conjunto de comentarios realizados por otros usuarios hacia dicho medicamento. Sin embargo, la lectura e interpretación de cada uno de los comentarios manualmente es una tarea cansada, tediosa y que consume mucho tiempo. Además, el sistema de reputación no brinda detalles sobre los motivos por los cuáles se está mostrando cierta reputación hacia el medicamento. Por otro lado, la utilización de análisis sentimental y obtención de emociones permitiría que la aplicación otorgara información detallada sobre el sentir de otros usuarios respecto a los medicamentos. No obstante, incorporar análisis sentimental requiere una cantidad de tiempo y esfuerzo considerables. Por ello, surgen las siguientes interrogantes:

1. ¿De qué manera el usuario en cuestión obtendrá información que detalle el sentir de las personas hacia distintos medicamentos sin que involucre la inversión de una gran cantidad de tiempo y esfuerzo?
2. ¿Cómo llevar a cabo análisis sentimental y obtención de emociones en comentarios de medicamentos en la aplicación Web sin que involucre un gran esfuerzo en el desarrollo?

Mediante la integración e incorporación del módulo de obtención de emociones a la aplicación Web es posible responder las interrogantes descritas anteriormente, por lo cual, a continuación, se describe cómo es la interacción de la persona con el módulo. Para esto, supóngase que la persona que está utilizando la aplicación está visualizando el detalle de un medicamento donde existen comentarios escritos por diferentes usuarios mediante su cuenta de Twitter[®], y un botón a través del cual la aplicación Web integra las funcionalidades del módulo para analizar dichos comentarios.

1.- El usuario da clic sobre el botón “Analizar Opiniones” localizado dentro de la página que muestra el detalle de un medicamento para iniciar el proceso de análisis sentimental y extracción de emociones que ofrece el módulo (Figura 4.11).

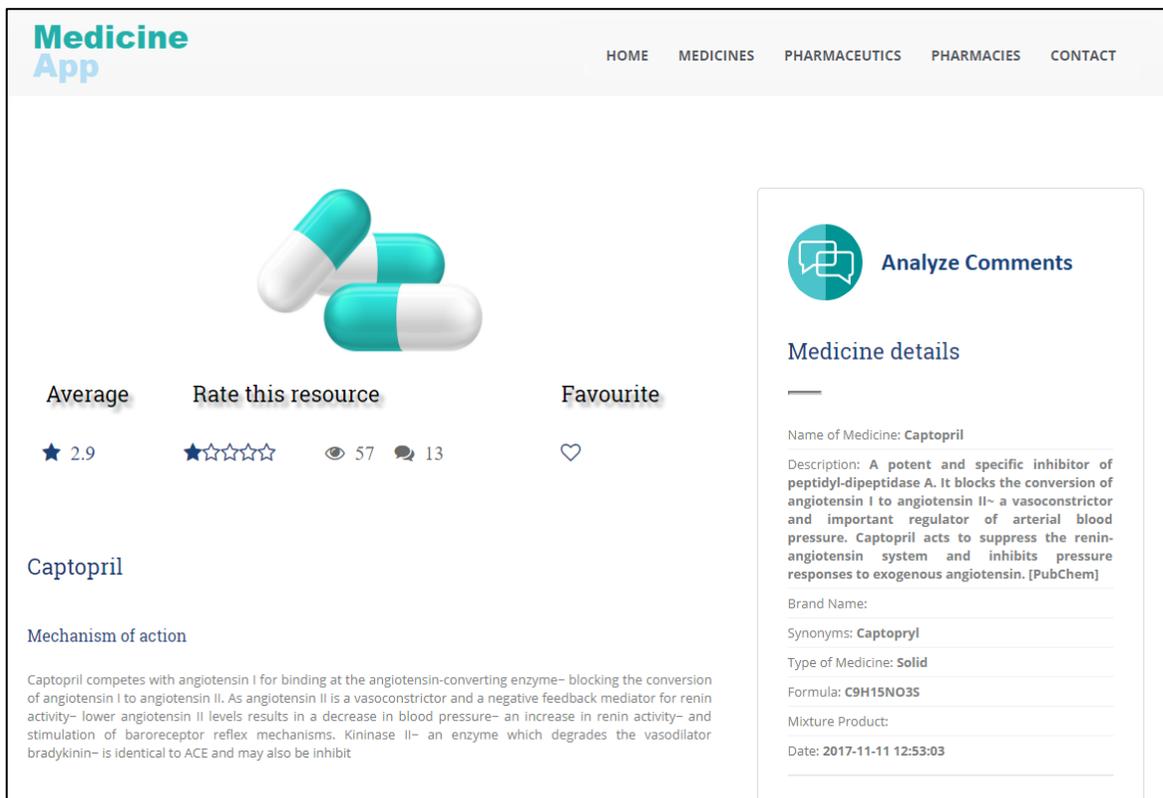


Figura 4.11 Detalle de un medicamento en donde integraron las funcionalidades del módulo.

2.- Tras presionar el botón mencionado anteriormente se extraen los comentarios correspondientes al medicamento y se lleva a cabo el análisis de éstos. Posteriormente se despliegan los resultados mediante una ventana modal. Dicha ventana es similar a la que se mostró en el caso de estudio descrito anteriormente, por lo cual no se describirá con tanto detalle. La ventana se compone por cuatro pestañas que contienen los resultados del análisis. La primera de ellas es la pestaña “Comments” en donde se muestran de forma tabular los distintos comentarios analizados sobre el medicamento, lo cual se aprecia en la Figura 4.12.

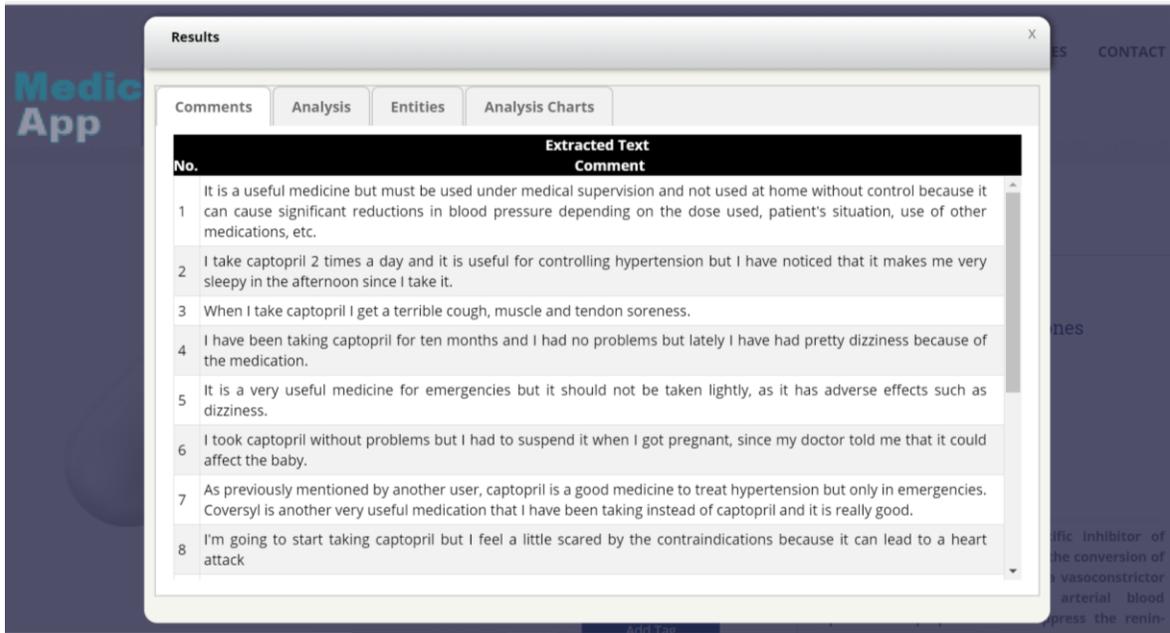


Figura 4.12 Pestaña de una ventana modal que muestra los comentarios analizados.

3.- Posteriormente, al seleccionar la pestaña “*Analysis*” se presentan de forma tabular los valores de polaridad y de cada una de las emociones (enojo, desagrado, miedo, alegría, tristeza) obtenidas tanto para cada comentario, como en promedio para el medicamento (Figura 4.13).

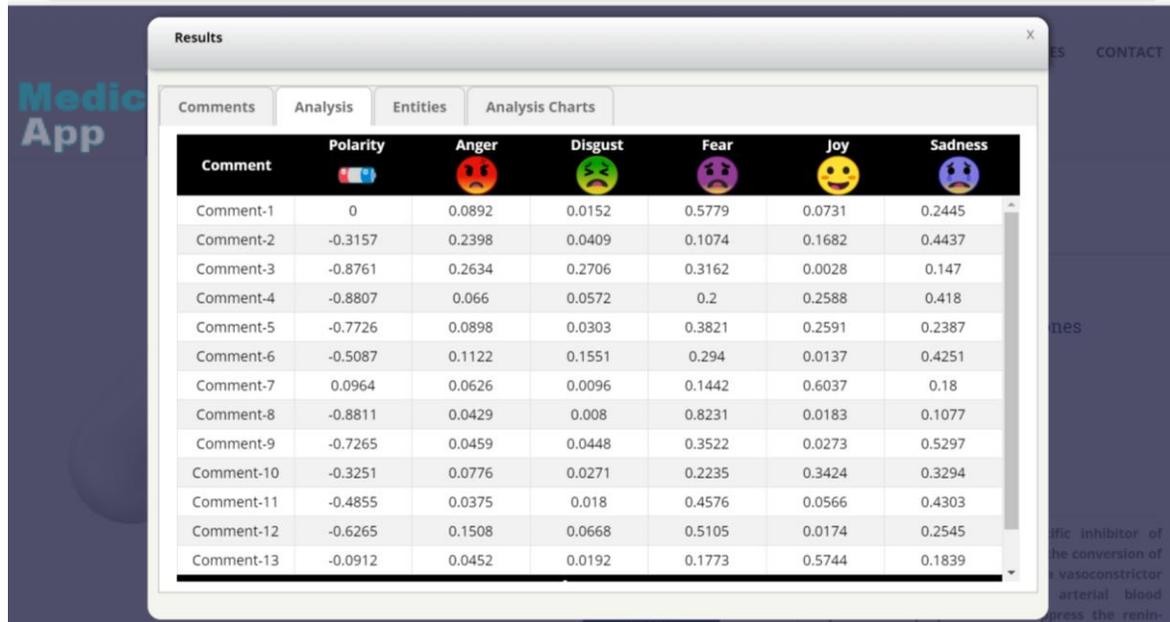


Figura 4.13 Pestaña que muestra el porcentaje de polaridad y emociones para cada comentario.

4.-La tercera pestaña, denominada “Entities”, permite visualizar de manera tabular las diferentes entidades identificadas entre los distintos comentarios, así como los valores de polaridad y de cada emoción (enojo, desagrado, miedo, alegría, tristeza) para cada una de ellas. Cabe mencionar que dichos valores se muestran en promedio, pues existe la posibilidad de que una misma entidad aparezca en distintos comentarios y en unos se muestren más fuertemente algunas emociones, mientras que en otros se muestren más fuertemente otras. Esta pestaña se aprecia en la Figura 4.14.

No.	Entity	Type	Polarity	Anger	Disgust	Fear	Joy	Sadness
1	captopril		-0.3481	0.1601	0.0965	0.265	0.1947	0.2838
2	nadolol		-0.8083	0.1639	0.0242	0.5886	0.0097	0.2135
3	cough		-0.8154	0.2905	0.3285	0.2237	0.0012	0.1561
4	dizziness		-0.814	0.0417	0.061	0.2102	0.3241	0.3629
5	hypertension		-0.3985	0.0481	0.0251	0.0291	0.8215	0.0762
6	Thyroid Gland		-0.5115	0.1095	0.0348	0.71	0.0359	0.1098
7	Coversyl		0.6232	0.1245	0.0131	0.0968	0.5744	0.1911
8	Doc		-0.7186	0.0831	0.0212	0.3422	0.1421	0.4114

Figura 4.14 Pestaña que muestra las entidades encontradas junto con su polaridad y emociones.

5.-Al dar clic en la pestaña llamada “Analysis Charts” el usuario tiene la posibilidad de seleccionar entre un conjunto de gráficas en donde se resume la información resultante del análisis. Las gráficas que se muestran son distintas al analizar publicaciones de Facebook® y de Twitter®, por lo cual para este caso de estudio las gráficas que se muestran son: polaridad, emociones, emociones por ubicación y entidades. La gráfica de polaridad permite observar mediante una gráfica de pastel el porcentaje de positividad y negatividad expresado en promedio hacia el medicamento. La positividad se muestra de color azul y la negatividad de color rojo. Dicha gráfica se aprecia en la Figura 4.15.

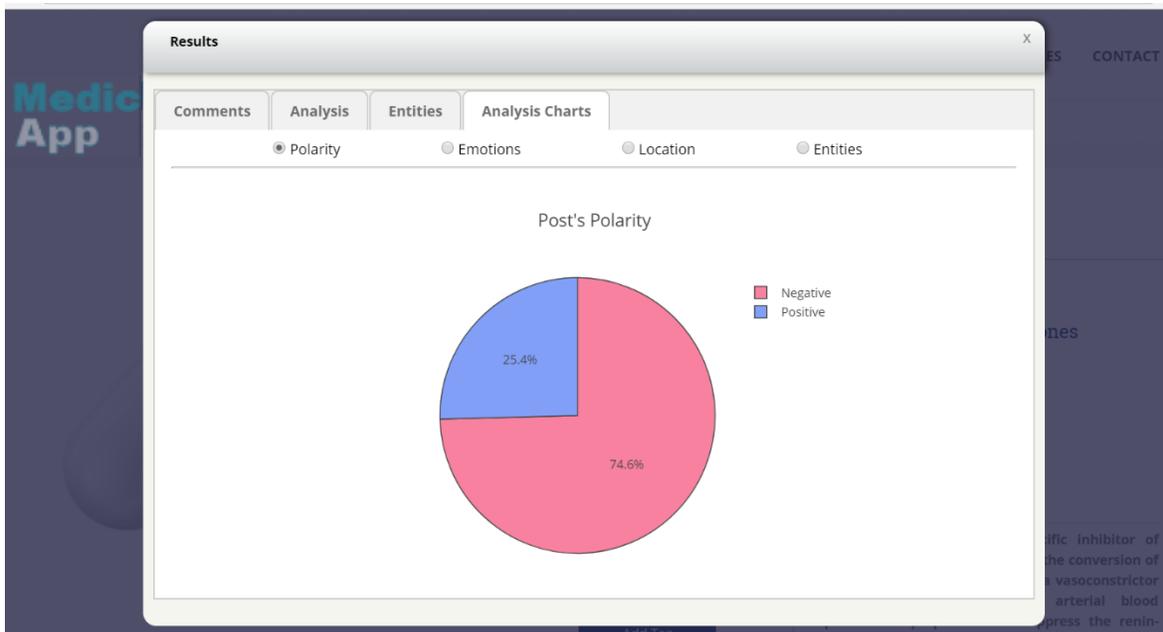


Figura 4.15 Gráfica que muestra la polaridad expresada hacia el medicamento.

La gráfica de emociones consiste en una gráfica circular que muestra el porcentaje de cada emoción que los usuarios manifestaron en promedio hacia el medicamento. Las emociones que se muestra corresponden a enojo en color rojo, desagrado en color verde, miedo en color morado, alegría en color amarillo y tristeza en color azul, lo cual se visualiza en la Figura 4.16.

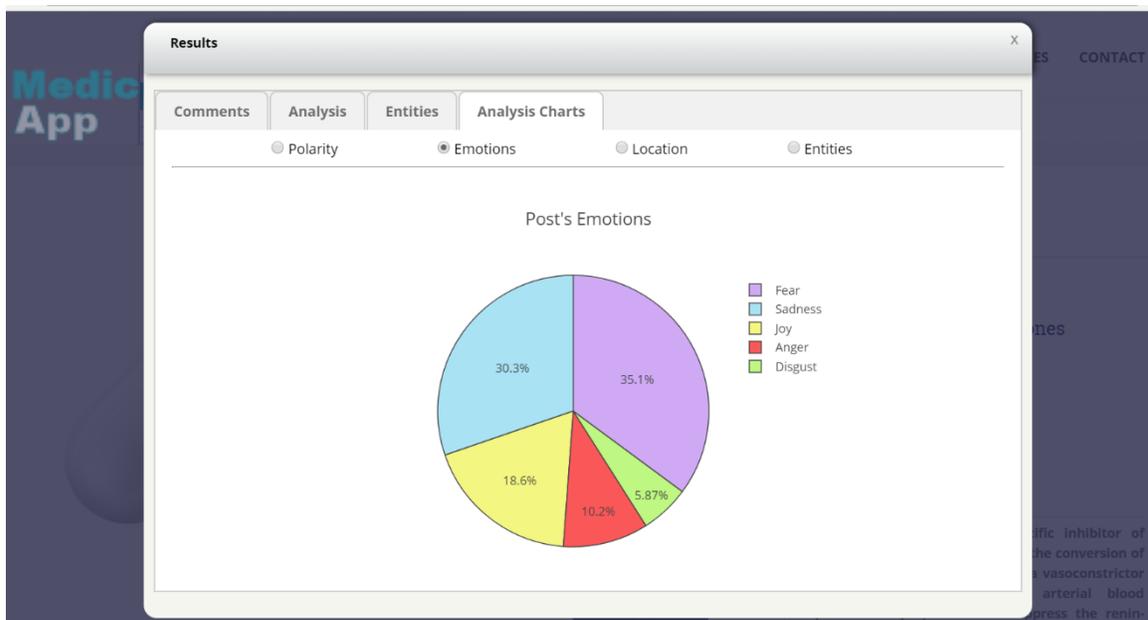


Figura 4.16 Gráfica que muestra los porcentajes de cada emoción hacia el medicamento.

La gráfica de emociones por ubicación consiste en una gráfica de barras que permite observar cómo se distribuyen las emociones entre las diferentes ciudades a las cuales pertenecen los usuarios que comentaron sobre el medicamento (Figura 4.17). El rango de valores para cada emoción va del 0 al 100 y la suma de los porcentajes de cada una de ellas por ciudad es del 100%.

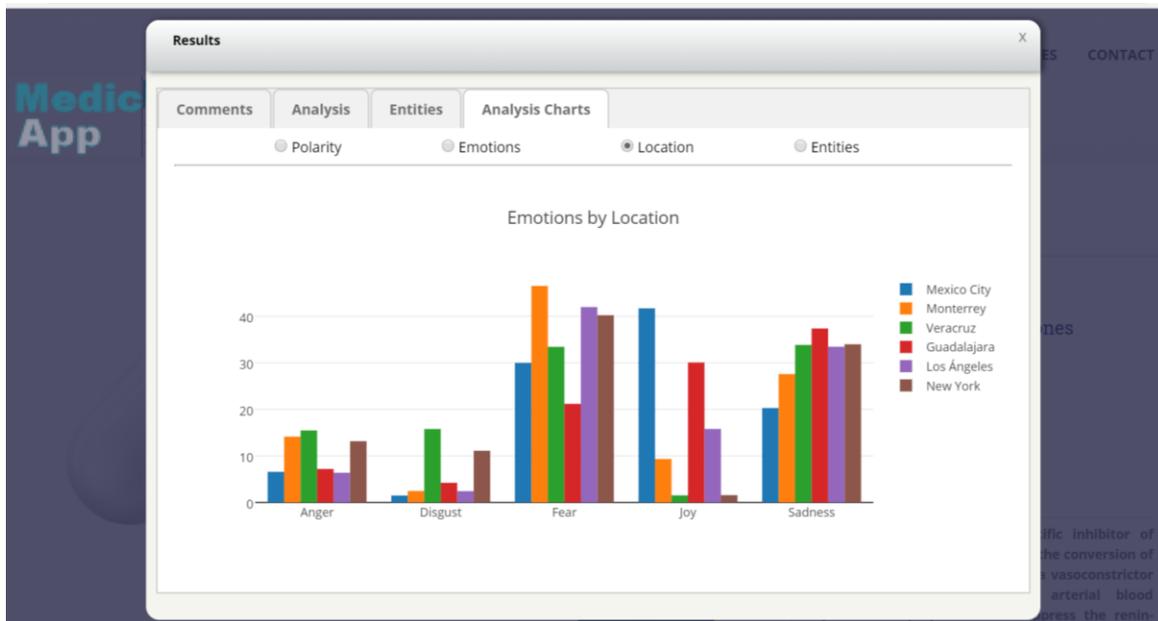


Figura 4.17 Gráfica que muestra la distribución de emociones por ubicación.

Finalmente, mediante la opción “*Entities*” el usuario tiene la posibilidad de observar un mapa de calor en donde se representa el valor de polaridad sentimental de cada entidad. Entre más cercano esté el color de cada entidad al color azul, más positiva es su polaridad, mientras que entre más cercano al color rojo sea la entidad, más negativa es su polaridad, lo cual es posible observar en la Figura 4.18.

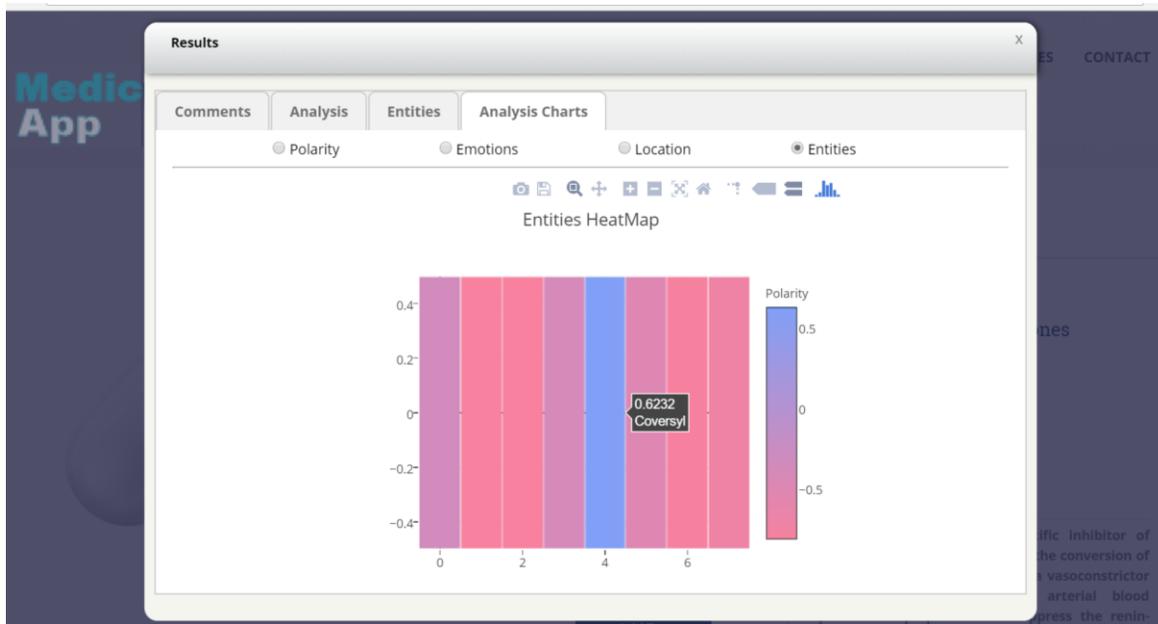


Figura 4.18 Mapa de calor que muestra la polaridad de entidades.

A través de los dos casos de estudio presentados anteriormente se validó la capacidad del módulo para integrarse con aplicaciones Web y analizar textos provenientes de redes sociales para obtener sentimientos, emociones y entidades, así como para representar gráficamente los resultados del análisis. Con ello, es posible que los diferentes interesados como lo son los pacientes en ambos casos de estudio, obtengan información valiosa y comprensible que les permita tomar decisiones y obtener distintos beneficios sin implicar una gran cantidad de tiempo o esfuerzo. Es importante mencionar que a pesar de que se presentan únicamente dos casos de estudio para validar el módulo y que éstos se enfocan en el dominio de la salud, es posible establecer algunos otros casos de estudio que involucren distintas situaciones y distintos dominios.

Capítulo V. Conclusiones

En este capítulo se describen las principales conclusiones obtenidas tras el desarrollo del presente trabajo, así como las recomendaciones o trabajo a futuro propuesto.

Mediante la realización de este trabajo se logró cumplir con el objetivo general y con los objetivos específicos del proyecto de tesis a través del desarrollo de un módulo que emplea técnicas híbridas de análisis sentimental para obtener sentimientos y emociones de textos provenientes de redes sociales. Para el desarrollo del módulo se analizaron diferentes APIs de análisis sentimental, se diseñaron y desarrollaron las diferentes funcionalidades requeridas para la obtención de emociones y se validó mediante dos casos de estudio en el ámbito de la salud. Con este proyecto se proporciona una alternativa de solución para obtener sentimientos y emociones de comentarios escritos en publicaciones de Facebook® y Twitter®, así como representar los resultados de forma que éstos sean fáciles de comprender y analizar.

La principal aportación del módulo es permitir la integración de análisis sentimental en aplicaciones Web sin implicar un gran esfuerzo en el desarrollo. Con ello, diferentes interesados tienen la capacidad de obtener información sobre las opiniones y el sentir de las personas hacia los productos o servicios que ofrecen, lo cual les permite llevar a cabo acciones para obtener beneficios, ya sea económicos o de otra índole.

Por otro lado, para la realización de este trabajo se requirió la adquisición y aplicación de nuevos conocimientos y experiencias relacionados con las diferentes tecnologías utilizadas. También, se logró adquirir y mejorar diferentes habilidades técnicas, no solo relacionadas con el análisis, diseño y desarrollo del proyecto, sino también relacionadas con la investigación y redacción. Particularmente, el conocimiento adquirido sobre análisis sentimental es muy valioso, pues es un campo de estudio muy interesante que crece y evoluciona año con año, con el cual es posible obtener diversos beneficios que se espera sean todavía más numerosos en un futuro. Precisamente, a causa de que dicho campo tiene un gran potencial es que vale la pena continuar investigando y trabajando con él.

5.1 Recomendaciones

Como parte de las recomendaciones para el módulo de obtenciones de emociones que se desarrolló, se encuentran: 1) Desarrollar funciones de análisis sentimental para analizar textos en idiomas diferentes al Inglés; 2) Obtener emociones distintas a las que ya se consideraron en este trabajo, y 3) Integrar funcionalidades que identifiquen distintos tipos de características en los comentarios analizados, tales como, sarcasmo, ironía, sátira, entre otros. A continuación, se describen con más detalle las recomendaciones anteriores.

1. Aunque en la actualidad el idioma Inglés es el que presenta un mayor grado de madurez en lo que respecta a análisis sentimental, dicho análisis está obteniendo avances considerables en otros idiomas, tales como el Español y Francés, por lo cual, vale la pena analizar e integrar distintas APIs que permitan emplear análisis sentimental sobre diferentes idiomas, para así tener la posibilidad de llevar a cabo un análisis más amplio y diverso, lo cual es de gran importancia en un mundo cada vez más globalizado.
2. La obtención de más emociones permitirá ampliar y profundizar el conocimiento sobre el sentir de las personas hacia las publicaciones, pues los resultados serán más específicos. Para lograr incorporar esta funcionalidad es necesario realizar una investigación y análisis de APIs de análisis sentimental sobre texto que brinden la capacidad de obtener emociones diferentes a las emociones con las que ya se cuenta, y que además empleen un enfoque híbrido.
3. La identificación de características como ironía, sarcasmo o sátira, permite obtener más información sobre el sentir de las personas y que ésta sea más cercana a la realidad. Para ello, se consideran dos alternativas: 1) Analizar APIs que permitan la identificación de este tipo de características empleando un enfoque híbrido, y 2) Emplear técnicas lingüísticas y técnicas de aprendizaje automático, de manera que se emplee un conjunto de prueba para crear modelos que permitan analizar comentarios nuevos y determinar si éstos poseen características como las mencionadas anteriormente.

Productos académicos

En esta sección se brinda información acerca de los productos académicos que se realizaron con base en el proyecto de tesis.

Artículos de congresos internacionales

- Ramírez-Tinoco F.J., Alor-Hernández G., Sánchez-Cervantes J.L., Olivares-Zepahua B.A., Rodríguez-Mazahua L. (2018) A Brief Review on the Use of Sentiment Analysis Approaches in Social Networks. In: Mejia J., Muñoz M., Rocha Á., Quiñonez Y., Calvo-Manzano J. (eds) Trends and Applications in Software Engineering. CIMPS 2017. Advances in Intelligent Systems and Computing, vol 688. Springer, Cham.



Registro de derechos de autor

- Alor-Hernández, G., Sánchez-Cervantes J. L., Ramírez-Tinoco F. J., “SAMOF: Módulo de análisis sentimental para comentarios de publicaciones en Facebook®”. INDAUTOR.
- Alor-Hernández, G., Sánchez-Cervantes J. L., Ramírez-Tinoco F. J., “SAMOT: Módulo de análisis sentimental para comentarios de publicaciones en Twitter®”. INDAUTOR.



Referencias

- [1] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Syst.*, vol. 89, pp. 14–46, 2015.
- [2] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Inf. Sci. (Ny)*, vol. 311, pp. 18–38, 2015.
- [3] C. Toxtli-Hernández, "Computación sentimental," *Software Guru*, 2015. [Online]. Available: <https://sg.com.mx/sgvirtual/sesion/computacion-sentimental#.WJLk5RvhA2w>.
- [4] IBM, "Natural Language Understanding," 2017. [Online]. Available: <https://www.ibm.com/watson/services/natural-language-understanding/>. [Accessed: 27-Jan-2017].
- [5] "Bitext," 2016. [Online]. Available: <https://www.bitext.com/>. [Accessed: 27-Jan-2017].
- [6] U. Krcadinac, "Synesketch," 2016. [Online]. Available: <http://krcadinac.com/synesketch/>. [Accessed: 27-Jan-2017].
- [7] "Toneapi." [Online]. Available: <http://toneapi.com/>. [Accessed: 27-Jan-2017].
- [8] "Repustate," 2015. [Online]. Available: <https://www.repustate.com/>. [Accessed: 27-Jan-2017].
- [9] "Aylien," 2017. [Online]. Available: <http://aylien.com/>. [Accessed: 27-Feb-2017].
- [10] "Text-processing." [Online]. Available: <http://text-processing.com/>. [Accessed: 27-Jan-2017].
- [11] "PreCeive," 2016. [Online]. Available: <http://www.theysay.io/product/preceive/>. [Accessed: 27-Jan-2017].
- [12] "Semantria," 2017. [Online]. Available: <https://www.lexalytics.com/>. [Accessed: 27-Jan-2017].
- [13] "Skyttle," 2017. [Online]. Available: <http://www.skyttle.com/>. [Accessed: 27-Jan-2017].
- [14] "DatumBox Machine Learning," 2017. [Online]. Available: <http://www.datumbox.com/>. [Accessed: 27-Jan-2017].
- [15] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [16] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," *Egypt. Informatics J.*, vol. 17, no. 2, pp. 199–216, 2016.

- [17] R. Benítez, G. Escudero, and S. Kanaan, *Inteligencia Artificial*. Universitat Oberta de Catalunya, 2013.
- [18] G. S. Alberto, *Inteligencia artificial: fundamentos, práctica y aplicaciones*. Madrid: RC, 2012.
- [19] The PHP Group, “PHP,” 2017. [Online]. Available: <http://php.net/>. [Accessed: 11-Nov-2017].
- [20] P. Garrido-Abenza, *Comenzando a programar con JAVA*. Elche: Universitas Miguel Hernández, 2015.
- [21] H. Ren, “Database Management System,” *Electron. Telecommun. Res. Semin. Ser.*, vol. 11, pp. 152–155, 2012.
- [22] Oracle, “MySQL,” 2017. [Online]. Available: <https://www.mysql.com/>. [Accessed: 12-Nov-2017].
- [23] Erlin, T. A. Fitri, and Susandri, “Using Social Networks: Facebook Usage at the Riau College Students,” *Procedia Comput. Sci.*, vol. 59, pp. 559–566, 2015.
- [24] H. Bicen and N. Cavus, “Social network sites usage habits of undergraduate students: case study of Facebook,” *Procedia - Soc. Behav. Sci.*, vol. 28, pp. 943–947, 2011.
- [25] O. B. Da’ar, F. Yunus, N. Md. Hossain, and M. Househ, “Impact of Twitter intensity, time, and location on message lapse of bluebird’s pursuit of fleas in Madagascar,” *J. Infect. Public Health*, 2016.
- [26] Y. Zhou, S. De, and K. Moessner, “Real World City Event Extraction from Twitter Data Streams,” *Procedia Comput. Sci.*, vol. 98, pp. 443–448, 2016.
- [27] Facebook, “Graph API,” 2017. [Online]. Available: <https://developers.facebook.com/docs/graph-api>. [Accessed: 11-Nov-2017].
- [28] A. Williams, “TwitterOAuth.” [Online]. Available: <https://twitteroauth.com/>. [Accessed: 11-Nov-2017].
- [29] Plotly, “Plotly.js,” 2015. [Online]. Available: <https://plot.ly/javascript/>. [Accessed: 11-Nov-2017].
- [30] A. Tripathy, A. Agrawal, and S. K. Rath, “Classification of Sentimental Reviews Using Machine Learning Techniques,” *Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015.
- [31] A. Ceron, L. Curini, and S. M. Iacus, “ISA: A fast, scalable and accurate algorithm for sentiment analysis of social media content,” *Inf. Sci. (Ny)*, vol. 367–368, pp. 105–124, 2016.
- [32] F. H. Khan, U. Qamar, and S. Bashir, “eSAP: A decision support framework for enhanced sentiment analysis and polarity classification,” *Inf. Sci. (Ny)*, vol. 367–

368, pp. 862–873, 2016.

- [33] “SentiWordNet,” 2010. [Online]. Available: <http://sentiwordnet.isti.cnr.it/>. [Accessed: 24-Jan-2017].
- [34] F. H. Khan, U. Qamar, and S. Bashir, “SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis,” *Knowledge-Based Syst.*, vol. 100, pp. 97–111, 2015.
- [35] M. Fernández-Gavilanes, T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. Javier González-Castaño, “Unsupervised method for sentiment analysis in online texts,” *Expert Syst. Appl.*, vol. 58, pp. 57–75, 2016.
- [36] O. Appel, F. Chiclana, J. Carter, and H. Fujita, “A hybrid approach to the sentiment analysis problem at the sentence level,” *Knowledge-Based Syst.*, vol. 108, pp. 110–124, 2016.
- [37] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, and K. Filippova, “Multi-lingual opinion mining on YouTube,” *Inf. Process. Manag.*, vol. 52, no. 1, pp. 46–60, 2016.
- [38] H. Saif, Y. He, M. Fernandez, and H. Alani, “Contextual semantics for sentiment analysis of Twitter,” *Inf. Process. Manag.*, vol. 52, no. 1, pp. 5–19, Jan. 2016.
- [39] A. Muhammad, N. Wiratunga, and R. Lothian, “Contextual sentiment analysis for social media genres,” *Knowledge-Based Syst.*, vol. 108, pp. 92–101, 2016.
- [40] “SentiStrength.” [Online]. Available: <http://sentistrength.wlv.ac.uk/>. [Accessed: 24-Jan-2017].
- [41] G. Katz, N. Ofek, and B. Shapira, “ConSent: Context-based sentiment analysis,” *Knowledge-Based Syst.*, vol. 84, pp. 162–178, 2015.
- [42] C. Bucur, “Using Opinion Mining Techniques in Tourism,” *Procedia Econ. Financ.*, vol. 23, pp. 1666–1673, 2015.
- [43] A. Ortigosa, J. M. Martín, and R. M. Carro, “Sentiment analysis in Facebook and its application to e-learning,” *Comput. Human Behav.*, vol. 31, pp. 527–541, 2014.
- [44] R. Gull, U. Shoaib, S. Rasheed, W. Abid, and B. Zahoor, “Pre Processing of Twitter’s Data for Opinion Mining in Political Context,” *Procedia Comput. Sci.*, vol. 96, pp. 1560–1570, 2016.
- [45] M. Eirinaki, S. Pisal, and J. Singh, “Feature-based opinion mining and ranking,” *J. Comput. Syst. Sci.*, vol. 78, no. 4, pp. 1175–1184, 2012.
- [46] J. F. Sánchez-Rada and C. A. Iglesias, “Onyx: A Linked Data approach to emotion representation,” *Inf. Process. Manag.*, vol. 52, no. 1, pp. 99–114, 2016.
- [47] A. Chandra Pandey, D. Singh Rajpoot, and M. Saraswat, “Twitter sentiment analysis using hybrid cuckoo search method,” *Inf. Process. Manag.*, vol. 53, no. 4, pp. 764–779, Jul. 2017.

- [48] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, “Supervised sentiment analysis in multilingual environments,” *Inf. Process. Manag.*, vol. 53, no. 3, pp. 595–607, May 2017.
- [49] V. K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, and S. Stehle, “Sentiment analysis during Hurricane Sandy in emergency response,” *Int. J. Disaster Risk Reduct.*, vol. 21, pp. 213–222, Mar. 2017.
- [50] TIOBE software BV, “TIOBE,” 2017. [Online]. Available: <https://www.tiobe.com/tiobe-index/>. [Accessed: 11-Nov-2017].