



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA**  
**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Diseño de una red social para el análisis de texto y relacionar usuarios  
por niveles de descubrimiento de lexemas mediante técnicas de  
Inteligencia Artificial

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:**

Ingeniero de Sistemas

**AUTOR:**

Torres Dávila, David Ernesto (ORCID: 0000-0003-4957-7444)

**ASESOR:**

Mg. More Valencia, Rubén Alexander (ORCID: 0000-0002-7496-3702)

**LÍNEA DE INVESTIGACIÓN:**

Sistemas de información y comunicación

PIURA - PERÚ

2022

## **Dedicatoria**

Esta tesis va dedicada para mi familia que siempre está presente en los diversos momentos de la vida, me enseñaron a afrontar cada meta con responsabilidad y sobre todo me brindaron estímulo y apoyo incondicional en todo momento además de mucha orientación.

Para mis padres Juver y Libia, Por su comprensión y ayuda en aquellos momentos difíciles, me enseñaron a encarar las adversidades sin perder nunca la dignidad ni desfallecer en el intento. Me han regalado todo lo que soy como persona, mis valores, mis principios, mi perseverancia y mi empeño siempre con una gran dosis de amor y paciencia sin pedir nada a cambio.

## **Agradecimiento**

### **A Dios**

Por darme la sabiduría y fuerza para culminar esta etapa académica

### **Al Asesor de Tesis MG. Ing. More Valencia, Rubén Alexander**

Por su guía, comprensión, paciencia, entrega y valiosos consejos a lo largo del proceso de investigación.

### **A Toda La Plana Docente De La Universidad Cesar Vallejo-Piura**

Por todas la enseñanzas y asesorías brindadas

### **A Mis Padres, Amigos y Demás Familiares**

Por el apoyo diverso en diversas situación y brindar el acompañamiento necesario.

## Índice de contenido

<b>Dedicatoria</b> .....	II
<b>Agradecimiento</b> .....	III
<b>Resumen</b> .....	IV
<b>Abstract</b> .....	V
<b>Índice de contenido</b> .....	VI
<b>Índice de Tablas</b> .....	VIII
<b>I. INTRODUCCIÓN</b> .....	1
<b>II. MARCO TEÓRICO</b> .....	4
<b>III. METODOLOGÍA</b> .....	12
3.1 Tipo y diseño de investigación .....	12
3.2 Variables y Operacionalización .....	13
3.3 Población, muestra y muestreo .....	13
3.4 Técnicas e instrumentos de recolección de datos .....	14
3.5 Procedimientos .....	15
3.6 Método de análisis de datos.....	15
3.7 Aspectos éticos .....	16
<b>IV. RESULTADOS</b> .....	17
<b>V. DISCUSIÓN</b> .....	21
<b>VI. CONCLUSIONES</b> .....	24
<b>VII. RECOMENDACIONES</b> .....	25
<b>REFERENCIAS</b> .....	26
<b>ANEXOS</b> .....	30
1.1 Tabla de Operacionalización .....	30
1.2 Matriz de consistencia.....	31
1.3 Ficha de registro N°1 .....	32
1.4 Ficha de registro N°2 .....	33

1.5	Documentos para validar los instrumentos de medición a través de juicio de expertos.....	1
1.6	Documento de metodología y desarrollo de prototipo de red social para análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de inteligencia artificial. ....	1

## **Índice de Tablas**

Tabla 1 Componentes del Corpus logístico .....	9
Tabla 2: Población de estudio .....	13
Tabla 3 Técnicas e Instrumentos de variable dependiente .....	14

## **Índice de Ilustraciones**

<b>Ilustración 1</b> Diagrama de un agente inteligente .....	11
<b>Ilustración 2</b> Histogramas de Volumen de unidades léxicas (Kb) .....	18
<b>Ilustración 3</b> Histograma de error de volumen no léxico .....	19
<b>Ilustración 4</b> Histograma de nivel de error de clasificación .....	20

## Resumen

En nuestra sociedad, se están dando cambios en los procesos comunicativos que de alguna manera reflejan en cierta medida los cambios sociales, económicos y políticos como engranaje de la transformación de la forma de educar, escribir y de expresarse, por esto el contenido digital es la tendencia de la revolución de las comunicaciones, considerando como insumo para el análisis de los textos producidos en estas nuevas formas de expresión en redes sociales de diversa índole. Ante este escenario donde las personas pasan gran parte de su tiempo y se expresan sin temores, lo que podría aprovecharse para descubrir patrones que puedan utilizarse para caracterizar perfiles de usuarios, categorizando de acuerdo a la temática o mensajería expresada en las redes sociales.

La presente investigación tiene como finalidad el diseño una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial. El tipo de investigación es de tipo aplicada y un diseño no experimental, se consideró como unidad de análisis por los textos producidos por los usuarios de una red social durante el periodo de la investigación.

En los resultados se resalta en lo que respecta al corpus textual de las unidades léxicas básicas en la red social, el volumen promedio que se almacena en cada publicación es de 0.2057 Kilobytes, el análisis del componente morfológico y el lexicográfico del lexema de los textos por usuario en la red social, es determinante el manejo del error de volumen no léxico, obteniéndose una media de 01693 Kilobytes, en cada publicación y los patrones de lexemas en los textos lo determina el nivel del error de clasificación obteniéndose una media de este error de 98.33%. Se concluye que se logra diseñar una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante el algoritmo de Inteligencia Artificial.

**Palabras Clave:** Análisis de texto, descubrimiento de lexemas, minería de texto, niveles de lexemas, inteligencia artificial

## **Abstract**

In our society, changes are taking place in communication processes that in some way reflect to a certain extent social, economic and political changes as a gear in the transformation of the way of educating, writing and expressing oneself, for this reason digital content is the trend of the communications revolution, considering as input for the analysis of the texts produced in these new forms of expression in social networks of various kinds. Faced with this scenario where people spend a large part of their time and express themselves without fear, which could be used to discover patterns that can be used to characterize user profiles, categorizing according to the theme or messaging expressed on social networks.

The present research aims to design a social network for text analysis and relate users by levels of lexeme discovery through Artificial Intelligence techniques. The type of research is of an applicative type and a non-experimental design, it was considered as a unit of analysis by the texts produced by the users of a social network during the research period.

In the results, the textual corpus of the basic lexical units in the social network is highlighted, the average volume that is stored in each publication is 0.2057 Kilobytes, the analysis of the morphological component and the lexicographic of the lexeme of the texts by user in the social network, the handling of the non-lexical volume error is decisive, obtaining an average of 01693 Kilobytes, in each publication and the lexeme patterns in the texts are determined by the level of classification error, obtaining an average of this error of 98.33%. It is concluded that it is possible to design a social network for text analysis and relate users by levels of lexeme discovery through the Artificial Intelligence algorithm.

**Keywords:** Text analysis, lexeme discovery, text mining, lexeme levels, artificial intelligence



## I. INTRODUCCIÓN

En nuestra sociedad, se están dando cambios en los procesos comunicativos que de alguna manera reflejan en cierta medida los cambios sociales, económicos y políticos como engranaje de la transformación de la forma de educar, escribir y de expresarse, por esto el contenido digital es la tendencia de la revolución de la comunicaciones, tomándolo como insumo para el análisis de los textos producidos en estas nuevas formas de expresión en redes sociales de diverso índole (Dzyaloshinsky et al., 2020).

En el mercado existen una gran diversidad de servicios y soluciones que aplican análisis de sentimientos, siendo relevante para las organizaciones, debido a que se puede aplicar para realizar seguimiento de opiniones en las diversas redes sociales, específicamente para conocer opiniones sobre una variedad de productos, marcas conociendo la subjetividad del usuario, por ello investigaciones actuales se enfocan en minería de opiniones en áreas de marketing y otras que estén directamente involucradas con el cliente.(E. Estrada, J. Hernandez, J. Morales, 2020).

En las redes sociales existe una gran variedad de opiniones que generan diferentes problemas con el lenguaje natural, debido a la forma ambigua de expresarse del lenguaje y sus niveles, lo cual se debe al desconocimiento del usuario (González García et al., 2019). La importancia del procesamiento del lenguaje natural es relevante para la clasificación sistemática y automatizada de los textos no estructurados, que se encuentran en redes sociales y para la realización del análisis de sentimientos en los textos (Mengliyev et al., 2021). El análisis de sentimientos o minerías de opiniones se basan en una variedad de técnicas para el procesamiento de lenguaje, que extrae de redes sociales por ejemplo información subjetiva a partir de contenido generado por los usuarios, posteriormente se clasifica los textos no estructurados de manera sistemática y automatizada en positivos, negativos y neutros, y encontrar relación a los usuarios considerando los niveles de descubrimiento de lexemas (Marquis et al., 2020). Esta detección ha generado una problemática para su clasificación, por tanto para evaluar clima de la opinión, los sistemas de análisis de sentimientos han convertido en atribuir un tipo de valoración cuantitativa, considerando el sentimiento (Rincón Acuña, 2016).

Actualmente las redes sociales son el escenario de expresión de las personas y pasan gran parte de su tiempo donde se expresan sin temor. Además, se

comparten gustos, preferencias, sus inquietudes, anhelos, deseos, sueños, esto en las plataformas no se clasifica de tal manera que se tengan niveles en cuanto al uso de Lexemas. Estas expresiones sociales, se vierten espontáneamente y se evidencia significativamente en los mensajes que en cada momento se están generando, y esto puede aprovecharse para descubrir patrones que puedan utilizarse para caracterizar perfiles de usuarios, categorizando de acuerdo a la temática o mensajería expresada en las redes sociales (Alvarado Atocha, 2019).

Cada usuario produce un gran volumen de información, y hoy en día existen muchas maneras del tratamiento de textos, tal como la extracción de información, la clasificación de textos y la búsqueda de respuestas entre otras; mediante esta clasificación facilitaría a la organización y determinar la categoría de un texto de entre una gama de categorías ya definidas, considerando características o tokens identificables en un texto. El propósito es el análisis de este texto dentro de una red social, mediante técnicas de inteligencia artificial, descubrir lexemas de acorde a un corpus textual predefinido y poder relacionarlos por cada usuario, diseñando agentes de clasificación automática para tal fin (Quej Cosgaya Héctor Manuel, Durán Lugo Juan Miguel, 2016).

Las redes sociales serán el contexto que permita empotrar un corpus textual, y mediante el diseño de una solución con inteligencia artificial, clasifique en forma automática tomando cadenas de texto para su análisis y categorizarlos por tipo de lexemas. Las técnicas de lógica difusa compensatoria (LDC) y sistemas expertos, podrían sincronizar con algoritmos de cálculos de la distancia de edición (DE) para descubrir lexemas, cuyo propósito sea la entrega de información y pueda emitir la categorización por niveles de clasificación, pero en relación a lo comentado por cada usuario de dicha red social (Campos García, 2020).

Ante esta realidad se enuncia el problema de investigación: ¿Cómo se analiza texto y relaciona usuarios por niveles de descubrimiento de lexemas mediante técnicas de inteligencia artificial en una red social? y como preguntas de investigación: ¿Cuál es el corpus textual de las unidades léxicas básicas que se estudiarán en la red social?, ¿Cuáles son los componentes morfológicos y el lexicográfico del lexema de los textos por usuario en la red social?, ¿Qué patrones de lexemas en los textos por cada usuario hay en la red social?

El estudio se justifica en lo social, debido a que se centra en redes donde interactúan las personas masivamente, y ante este gran volumen de datos se pretende clasificar los lexemas de acuerdo a cada usuario con técnicas de inteligencia artificial para obtener información útil para caracterizar perfiles a nivel organizacional y social (Ouhiainen, 2018).

El presente estudio se justifica teóricamente, porque para el análisis de los lexemas y su categorización por usuarios se hará uso de técnicas de inteligencia artificial, algoritmos de clasificación de datos, modelos para precisión de clasificaciones, por tal servirán como soporte para comprobar mediante una propuesta de un modelo la aplicabilidad de tales fundamentaciones teóricas en un contexto real (Faridy Bermeo, y otros, 2016).

En cuanto al aspecto metodológico se justifica debido que para la clasificación de la información se ha realizado una serie de pasos, para poder gestionar los antecedentes tanto para el desarrollo del marco teórico como para el planteamiento de la problemática contextualizándolo con la realidad existente, asimismo para el planteamiento de los indicadores de la variable en estudio según la similitud con la temática de otros autores (Vila Ponte, 2018).

Como objetivo del presente trabajo su objetivo general es: Diseñar una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial.

Como objetivos específicos tenemos: Establecer un corpus textual de las unidades léxicas básicas que se estudiarán en la red social, analizar el componente morfológico y el lexicográfico del lexema de los textos por usuario en la red social y determinar patrones de lexemas en los textos por cada usuario en la red social.

## II. MARCO TEÓRICO

En lo que respecta a los antecedentes que sustentan el presente proyecto de investigación tenemos referentes internacionales, nacionales y locales.

En el contexto internacional en Ecuador Castillo Mantuano (2020) con la investigación “Twitter, un medio social que diseña políticos: análisis de la imagen política que se construye en el perfil de Paola Pabón durante el período de campaña política seccional 2019”, cuyo objetivo fue el análisis de las construcciones acerca de la imagen política y comunicacional del perfil de Paola Pabón en la mencionada red social con miras a la campaña electoral 2019. El Tipo de investigación es aplicada y el diseño pre experimental, la población está compuesta por 2.074.703 personas, que se constituye en su público potencial, y la muestra menos del 40% de la población pichinchana. Entre sus conclusiones se tiene que, por cada 100 menciones, 41 se clasificaron como positivas, 35 menciones negativas y 24 se catalogaron neutras. Por tal se considera que el perfil de Paola Pabón alcanza aproximadamente un 50% de aceptación entre los usuarios que la siguen en la red social Twitter (Castillo Mantuano, 2020).

En México, Meza Flores y otros (2020), realizaron la investigación “Diseño de un prototipo de sistema experto que permita identificar patrones basado en las aptitudes y habilidades académicas de los alumnos de nivel superior”, cuyo objetivo general fue diseñar un sistema experto para identificar patrones basados en aptitudes y habilidades de los estudiantes de la mencionada institución, para apoyar la titulación o término de créditos en tiempo y forma. El Tipo de investigación es aplicada, el diseño pre experimental, la población y la muestra estuvo conformada por todos los estudiantes. En sus conclusiones se tiene que ofreciendo unas herramientas tecnológicas que permitan la posibilidad de que el docente diseñe o implemente planes de acción por medio de estimadores que ayuden a incrementar el tiempo de permanencia dentro de los primeros semestres detonando con ello el continuar hasta el término del mapa curricular, formando de manera paralela egresados integrales que sean capaces de usar sus habilidades y aptitudes de manera enfocada así como saber la forma de compensar sus debilidades en su formación académica y que con ello no se tenga alguna problemática grave en el desarrollo profesional o laboral (Flores, 2020).

En España, Ortega de los Ríos (2020), realizó la investigación denominada “Análisis de sentimiento en las redes sociales producido por las Fuerzas Armadas Españolas”, cuyo objetivo general fue realizar un estudio de análisis de sentimientos considerando como corpus los tweets de las fuerzas armadas. El Tipo de investigación es descriptiva, de diseño pre experimental, la población está constituida por un corpus de 3000 tweets y la muestra es toda la población. Entre los resultados se apreció un cierto sesgo en la clase de los tweets negativos, aproximadamente un 2%, por ello fue necesaria la validación cruzada. Se logra deducir poca fiabilidad de GaussNB y KNN respecto a los otros modelos considerados, ello se explica dado que se consideró que el conjunto de los datos presentaba una distribución normal. En las conclusiones se tiene la poca fiabilidad del modelo de aprendizaje sin supervisión para la predicción de la polaridad en un conjunto de datos, debido a la limitación del texto o limitaciones de la librería. Se generó modelos de aprendizaje con supervisión fiables, siendo F-1 superior al 80%. Las palabras más representativas de los tweets considerados fueron (rey, ejército, trabajo, unidad, tropa entre otros) y fueron agrupadas en cuatro categorías distintas; de los tweets considerados negativos eran a contenidos ajenas a temáticas de las fuerzas armadas (Ortega de los Ríos & Milagros, 2020).

En España, Campos García (2020) realizó la investigación diseño e implementación de un sistema de comparación de perfiles en redes sociales, cuyo objetivo general fue diseñar un sistema informático que compare dado un perfil de una red social, con una lista de perfiles de acuerdo a criterios de similitud. El Tipo de investigación es aplicada, con un diseño pre experimental, La población estuvo conformada por un corpus de 43 perfiles y la muestra es toda la población. En los resultados se tuvo que encontró valores de similitud siendo el promedio de un 10% de coincidencia, además el modelo categorizo bien los perfiles de acuerdo al tema, pero no se pudo establecer conexiones entre perfiles que son llevados por la misma persona. Asimismo, en el modelo TF-IDF, la precisión de similitud pasando de 10% a casi el 90%; a pesar de ello se obtuvo muchas clasificaciones erróneas (Campos García, 2020).

En España García del Saz (2019) realizó la investigación “Herramienta de descarga y análisis de datos de Twitter sobre participación ciudadana”, cuyo objetivo general

fue el desarrollar una herramienta altamente para la extracción de datos de Twitter. El Tipo de investigación es aplicada, y su diseño pre experimental, la población la conformó la población de Madrid, y la muestra a toda la población dentro de las redes sociales. Entre sus conclusiones se tiene que el modelo de datos que se diseñó facilita las métricas realizadas desde SQL, además permite exportar a diversos formatos estándares como CSV, XML y otros. Como limitante se tiene que el volumen de tweets que se logró extraer es bajo, tan solo 5000 tweets durante 07 meses (González García et al., 2019).

En Ecuador Alvarado Atocha (2019), realizó la investigación para determinar la incidencia de la utilización de la red social Facebook, en la comunicación de los estudiantes de primero de bachillerato del colegio industrial Febres Cordero Guayaquil. La investigación es de nivel descriptivo, tipo aplicada, el diseño no experimental, La población fueron de 280 estudiantes, y la muestra por 100 estudiantes. En cuanto a sus conclusiones, se tuvo que, en función de los tipos de redes sociales, los estudiantes usan con mayor frecuencia Facebook, además de Instagram y Spotify en menor frecuencia.(Alvarado Atocha, 2019).

En Ecuador, Gamarra Contreras (2018) desarrolló la “Investigación y análisis del lenguaje subjetivo en sistemas de minerías de opiniones existentes en el mercado”, cuyo objetivo general fue realizar un estudio de técnicas y aplicaciones concerniente al uso del procesamiento de lenguaje natural y subjetivo en los sistemas de minería de opiniones. La investigación es de tipo documental, de diseño no experimental, se consideró una población de 1000 estudiantes de los semestres de VI, VII, VII del programa de Ingeniería en Networking y Telecomunicaciones, con una muestra de 199 estudiantes. En sus conclusiones se tiene que de los encuestados, el 40% afirma que la interpretación del contenido de un texto puede dificultar el análisis del contenido en un sistema de minería de opiniones, el al 25% demostró la utilización de redes sociales para generar una determinada opinión a través de comentarios, el 40% lo genera mediante la composición de frases cortas, uso de expresiones abreviadas, uso de emoticonos o combinaciones de niveles, lo que afecta a los sistemas de minerías de opiniones al momento de interpretar el contenido del texto. Finalmente, comparando los diferentes métodos para la clasificación de sentimientos se estableció que el mejor

método con mejor precisión, es aquel basado en diccionarios léxicos, puesto que es la mejor técnica que se basa en recursos lingüísticos que se van incorporando dependiendo la necesidad de la implementación utilizando para esto la herramienta llamada SentiStrength (Gamarra Contreras, 2018).

En España Piñón Ferrer (2018) desarrollo la investigación “Minería de datos aplicada a Twitter y análisis de sentimientos mediante algoritmos de inteligencia artificial”, cuyo objetivo general es el análisis de los sentimientos utilizando técnicas de inteligencia artificial. El tipo de investigación es aplicada, de diseño pre experimental, la población está constituida por las publicaciones de Twitter, y la muestra es toda la población. Entre sus conclusiones se analizó asociaciones entre los conjuntos de datos y descubrir patrones frecuentes, encontrando ítem sets frecuentes y reglas de asociación fuertes. Además, se aplicó los algoritmos de clustering para agrupar; bien sea particionando, basándonos en la densidad o generando dendrograma de agrupación jerárquicos de forma aglomerativo y divisiva. Finalmente, algoritmos de clasificación y predicción; desde Naive Bayes y SVM hasta las Estructuras en Forma de Árbol con sus algoritmos y optimizaciones, cada uno de los modelos se utilizó para procesar texto procedente de los tuits, que extraemos de un hashtag, para obtener información sobre un evento o una entidad (Piñón Ferrer, 2018)

En Colombia Rincón Acuña (2016) realizó la investigación “Memoria y formación: Una aproximación desde la fenomenología y la inteligencia artificial”, el objetivo general de la investigación es validar si el desarrollo lexicográfico asistido por un dispositivo computacional contribuye de alguna forma al proyecto de formación del ser humano. El tipo de investigación es aplicada y el diseño pre experimental, la población está compuesta por todos aquellos posibles usuarios de RS2, es decir, cualquier persona con necesidad de estudiar textos; el tamaño de la muestra es equivalente a 18 individuos escogidos al azar. Se concluye que RS2 le permite al ser humano elaborar un producto académico con mejor respaldo de citas bibliográficas, labor mecánica, pero le quita la función de almacenamiento de contenido fáctico representado en las citas empleadas. Asimismo, a pesar de su complejidad, las descripciones fenomenológicas pueden ser modeladas computacionalmente y plasmadas en algoritmos aptos para el aprendizaje de

máquinas y el aprendizaje de máquinas. Los modelos generados presentan un enfoque alterno para la solución de problemas clásicos en los cuales la ambigüedad y la falta de contexto son un factor relevante. Además, se ha mostrado que se puede transformar un texto en un grafo que represente conceptos semánticos relevantes para el ser humano a partir de métodos no supervisados y auto contenidos esto es, sin recurrir a la manipulación del ser humano o a fuentes externas de conocimiento, sistemas operando de modo auto contenido pueden ser usados en desarrollo de cerebros robóticos (Rincón Acuña, 2016).

Asimismo, en Argentina tenemos a Ramos & Vélez (2016) desarrolló su investigación “Integración de técnicas de procesamiento de lenguaje natural a través de Servicios Web”. Siendo su objetivo la implementación de la herramienta denominada TeXTracT, para el procesamiento NLP. El tipo de investigación es aplicada y el diseño pre experimental, la población, así como la muestra estuvo conformada por los servicios web. En las conclusiones se tuvo que, para poder desarrollar la herramienta propuesta, el diseño se basó en el patrón Bróker, implementado con un módulo para independizar los servicios NLP de las aplicaciones que soliciten procesamiento. Se evaluó la herramienta, en un primer caso mediante un aplicativo que requiera servicios NLP, se demostró que la arquitectura de TeXTracT creo aplicaciones de forma rápida y sencilla, con soporte de evolución en un determinado tiempo(Ramos & Velez, 2016).

Como soporte teórico relacionado con la investigación, se desarrolla las definiciones de la variable en estudio. En el caso del análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas:

En lo que concierne a Lexema, son las representaciones textuales latentes, que también se denominan incrustaciones de palabras o representaciones de palabras distribuidas. Creado con métodos no supervisados, solo se basan en un gran corpus de texto como entrada (Thoma et al., 2017). Asimismo, se considera información sobre la estructura de información básica y adicional de la raíz y el sufijo necesario para formar un analizador morfológico. La raíz, el sufijo y los morfotactos están incluidos en el léxico (Mengliyev et al., 2021)



La Inteligencia artificial, se define como una máquina o un sistema que posee la capacidad de ejecutar acciones y asumir el comportamiento de un ser humano a mediante la racionalidad (Santaella Santos, 2019). Así que, en la situación actual la Inteligencia Artificial (IA), se toma como máquinas y sistemas que simulan procesos de la inteligencia humana a través del conocimiento, razonamiento y la autocorrección (E. Estrada, J. Hernandez, J. Morales, 2020). La IA ofrece la posibilidad de crear soluciones informáticas que doten a las computadoras de aprendizaje, propósito de Mathison Turing, Von Neumann, Wiener, y otros, que intentaron equipar computadoras con programas que contenían inteligencia, autocopiado, aprendizaje y controlando su entorno, siendo su postulado modelar el cerebro humano, imitar el aprendizaje humano y simular la evolución biológica (González García et al., 2019)

Las redes neuronales constan de pesos y sesgos como factor en la fuerza de la señal o información que se propaga desde una neurona (también conocidos como nodos) al siguiente. Para determina el valor de las ponderaciones y sesgos que conducen a una óptima (Wangperawong et al., 2018).

En cuanto a los componentes del corpus lingüístico; Corpus, es una recopilación de muestras de lengua, bien sea en su expresión oral, o en su expresión escrita y que representa el uso de una lengua natural, en su extensión como en su variedad, según Vila Ponte se puede clasificar las búsquedas según la siguiente tabla(Vila Ponte, 2018).

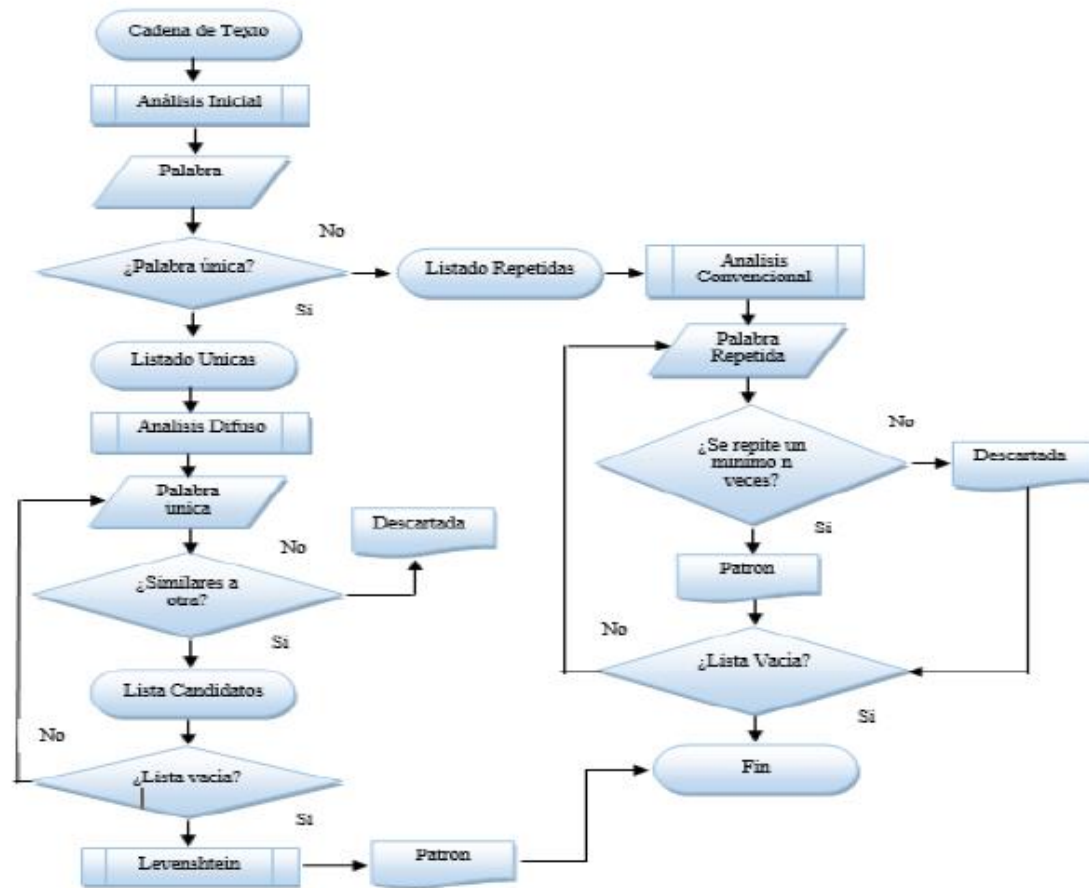
**Tabla 1** Componentes del Corpus logístico

Lema	hashtag
Número de concordancias	58 casos en 63 documentos
Distribución por zona geográfica	En España se registran 33 casos, mientras que en América 56, de los cuales 35 se registran en México y Centroamérica, 6 en el Río de la Plata, 6 en el Caribe continental, 6 en la región andina, 1 en la región chilena, 1 en las Antillas y 1 en Estados Unidos
Distribución por período	2011-2015: 62 casos; 2006-2010: 27 casos

Las unidades del análisis morfológico, en cuanto al propósito de la morfología es estudiar los componentes constitutivos de la palabra. En el análisis morfológico existen, dos unidades. La de rango superior es la palabra, y la de rango inferior sería el componente interno de la palabra, que recibe el nombre de morfema. (Marquis, y otros, 2020)

Patrones de lexemas en los textos, Es el proceso encargado del descubrimiento de conocimientos que no existían explícitamente en ningún texto de la colección, pero que surgen de relacionar el contenido de varios de ellos, mediante técnicas de minería de texto (Ortíz Morán, 2020).

Algoritmo relativo conocido como la distancia de edición de Levenshtein, consiste en el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra, se utiliza en la teoría de la información y ciencias de la computación, donde basado en la distancia de los textos es posible realizar operaciones de inserción, búsqueda, sustitución y eliminación (Quej Cosgaya, y otros, 2016).



**Ilustración 1** Diagrama de un agente inteligente

### III. METODOLOGÍA

#### 3.1 Tipo y diseño de investigación

##### 3.1.1 Tipo de investigación

La investigación es aplicada, pues se plantea una solución mediante minería de textos en una red social. El nivel de investigación es descriptivo, dado que se trata de conocer y describir cómo se analiza textos de los usuarios y relacionar por niveles de descubrimiento de lexemas basado en técnicas de inteligencia artificial (Roberto, Hernández Sampieri, Mendez Valencia Sergio, Mendoza Torres Christian Paulina, 2017).

##### 3.1.2 Diseño de investigación

El diseño, es de tipo no experimental, pues solo se describirán los resultados a partir de relacionar por niveles de descubrimiento de lexemas del análisis de los mensajes de texto.

G1: X1-----O1

Dónde:

G1: Mensajes o textos de los usuarios de la red social

X1: Corpus textual de la red social, con sugerencias de usuarios con la misma relación textual o tema de asunto...

O1: Se utilizar fichas de registro para determinar el volumen de unidades léxicas básicas.

X1: Componente morfológico y el lexicográfico de los mensajes, con sugerencias de usuarios con la misma relación textual o tema de asunto...

O1: Se utilizar fichas de registro para determinar el nivel morfológico y nivel lexicográfico.

X1: Patrones de lexemas en los mensajes, con sugerencias de usuarios con la misma relación textual o tema de asunto...

O1: Se utilizar fichas de registro para determinar el nivel de clasificación y nivel de error de clasificación.

### 3.2 Variables y Operacionalización

#### 3.2.1. Variable

Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas

Consiste en la indexación semántica para descubrir una nueva información a partir del análisis de textos estructurados, clasificación de acuerdo a niveles por los usuarios, a partir de expresiones en un lenguaje natural en el contexto de redes sociales (Cortez Vásquez Augusto, 2017).

### 3.3 Población, muestra y muestreo

#### 3.3.1 Población

Según Esteban Nieto, la población está constituida por un conjunto de individuos para ser analizados para obtener los resultados (Esteban Nieto, 2018). Mientras Arias lo define como un conjunto de casos, definido, limitado al cual se accede y será el referente para la muestra de la investigación (Arias et al., 2016).

La población estará conformada por los textos producidos por los usuarios de una red social en los meses de agosto y setiembre de 2021.

Tabla 2: Población de estudio

Unidad de análisis	Cantidad
Textos producidos por los usuarios de una red social en forma mensual	350

Fuente: Aproximación de la cantidad de usuarios y mensajes para el prototipo

#### 3.3.2 Muestra

Debido a que las técnicas requieren gran cantidad de información para poder clasificar con mayor asertividad, en la investigación se trabajará con una muestra por conveniencia, de esta manera no se requerirá realizar el muestreo.

### 3.4 Técnicas e instrumentos de recolección de datos

#### 3.4.1 Técnicas

Según Esteban Nieto, los métodos para la recopilación de información son variados tales como: Encuesta, análisis documental, la observación, entrevista, entre otros. Las técnicas que se utilizarán serán el análisis documental (Esteban Nieto, 2018).

#### Análisis documental

Esta técnica consiste en examinar los datos que se encuentran en documentos existentes en el lugar de estudio, tales como bases de datos, actas, informes, registros entre otros.

#### 3.4.2 Instrumentos

VARIABLE: Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas

Tabla 3 Técnicas e Instrumentos de variable dependiente

No	INDICADOR	TÉCNICA	INSTRUMENTO
1	Volumen de unidades léxicas básicas	Observación	Ficha de registro
2	Nivel de clasificación	Observación	Ficha de registro
3	Nivel del error de clasificación	Observación	Ficha de registro

#### 3.4.3 Validez

Según Hernández Escobar, y otros (2018), para obtener el grado en que los instrumentos cuantifican las variables, se considera la validez de constructo, contenido, criterio y el criterio de experticia de los jueces.

En la investigación se aplicará el criterio de experticia de los jueces en los instrumentos del análisis de texto y análisis de lexemas. Los Jueces expertos validarán con su firma los instrumentos que miden los indicadores que se aplicarán en la recolección de datos.

### 3.5 Procedimientos

Se realizará la búsqueda de trabajos previos afín a la temática de investigación, se identificarán y se incorporarán temáticas claves y conceptualizaciones relacionadas con el análisis de textos, niveles de descubrimiento de lexemas y técnicas de inteligencia artificial.

Para la recolección de datos se realizará mediante la técnica de observación, instrumento Guía de registro para la valoración de cada uno de los indicadores propuestos en la Tabla 2.

### 3.6 Método de análisis de datos

Para realizar el análisis de los datos recopilados, se utilizará la estadística, gráficos descriptivos, cuadros comparativos, según Hernández Sampieri, y otros (2017), se puede realizar contraste de medias de dos o más grupos.

Uso de Chi cuadrado, es una prueba no paramétrica, que evalúa la diferencia entre una distribución de frecuencias observadas y esperadas, solo considera valores positivos, debido a que la suma de sus valores se eleva al cuadrado, pero para su uso no solo considera que la muestra sea aleatoria, sino la muestra debe ser lo suficiente grande, lo que no aplica para la presente investigación por ser una muestra pequeña (Cruz, y otros, 2015).

Para hacer la prueba de hipótesis se usará el análisis paramétrico como la prueba T Student para diferencia de medias y las no paramétricas como la Prueba de Bondad de Ajuste de Kolmogorov-Smirnov (KS). Esta prueba busca determinar el nivel de acuerdo entre la distribución de una cantidad de valores y una distribución teórica determinada (Cruz, y otros, 2015).

En la estadística descriptiva, las medidas de tendencia central y de dispersión, en la escala cuantitativa son promedio, mediana y moda y en la escala cuantitativa se presentan cuadros o tablas (Alergia México, 2016). Para la presentación de los resultados de los indicadores volumen de unidades léxicas básicas, nivel morfológico, nivel Lexicográfico, nivel de clasificación y nivel del error de clasificación se utilizará promedio, medianas, varianzas y tablas de frecuencia y absoluta; asimismo cuadros comparativos y gráficos.

### 3.7 Aspectos éticos

La estructura de las citas y referencias de las fuentes consideradas para la investigación se basan en el estilo ISO 690:2010. También se considera lo expresado en el artículo N°43 del código de ética profesional del Colegio de Ingenieros del Perú, referente a la inviolabilidad del trabajo ajeno, asimismo se respetó el artículo N° 41 considerándose la idea expresada por los autores. Asimismo, lo expresado por la Universidad César Vallejo en lo que respecta a faltas de ética y sanciones. Además del artículo 1 del código de ética de la mencionada Universidad, en cuanto que su investigación se desarrolló con los máximos estándares de rigor científico, responsabilidad y honestidad en cuanto a la elaboración del informe basado estrictamente con el manejo de la información recopilada. Finalmente se considera el artículo N°15 con respecto al plagio y el artículo N°16 con respecto a la autoría del trabajo.



#### IV. RESULTADOS

A continuación, se muestran los valores de los indicadores:

<b>Número</b>	<b>I1: Volumen Unidades Lexicas(Kb)</b>	<b>I2: Error de Volumen No Lexico (KiloByte)</b>	<b>I3: Nivel del error de clasificación (Porcentaje)</b>
1	0.016	0.008	0.33
2	0.023	0.000	0.00
3	0.016	0.016	0.33
4	0.023	0.008	0.33
5	0.008	0.039	0.50
6	0.016	0.016	0.33
7	0.031	0.016	0.00
8	0.016	0.031	0.33
9	0.023	0.023	0.00
10	0.023	0.031	0.40
11	0.023	0.023	0.25
12	0.023	0.023	0.25
13	0.016	0.008	0.50
14	0.016	0.023	0.33
15	0.023	0.016	0.00
16	0.016	0.031	0.33
17	0.023	0.016	0.25
18	0.023	0.016	0.25
19	0.016	0.016	0.33
20	0.023	0.000	0.00
21	0.016	0.023	0.33
22	0.023	0.023	0.00
23	0.016	0.016	0.33
24	0.023	0.016	0.25
25	0.031	0.016	0.20
26	0.023	0.016	0.25
27	0.016	0.016	0.00
28	0.016	0.008	0.33
29	0.031	0.008	0.20
30	0.023	0.008	0.25

## Análisis de resultados

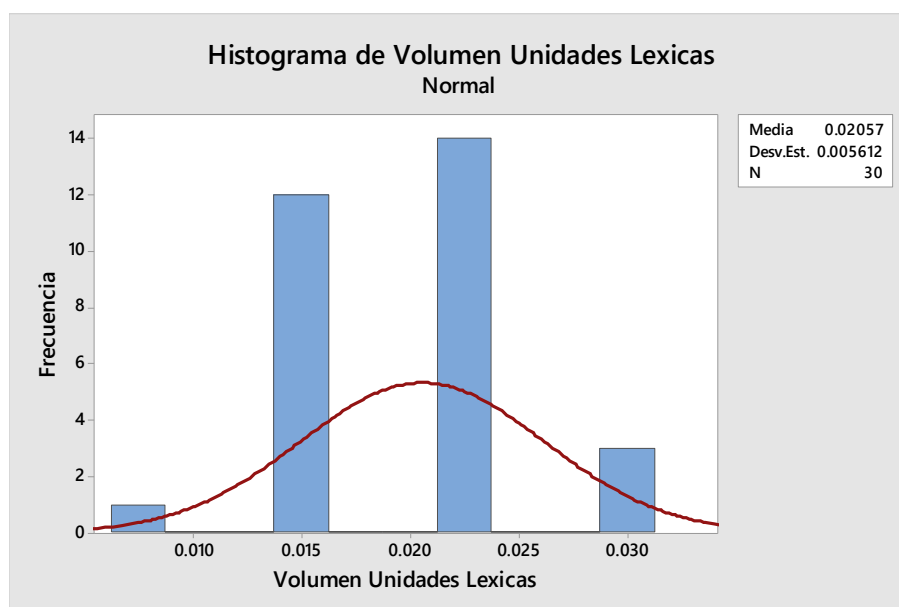
### Volumen de unidades léxicas básicas

El tamaño del vector inferido, es de acorde a la publicación realizada por el usuario; y cada uno de los caracteres de las publicaciones de estos, está representado por un byte. Por ello el volumen del vector inferido se calcula mediante la siguiente fórmula

$$\text{Volumen Vector Inferido (Kb)} = \frac{\text{Longitud de la publicación} * 8}{1024}$$

Para el caso del volumen de unidades léxicas básicas, el algoritmo selecciona el número de unidades léxicas y aplica la fórmula anterior, resultando de la siguiente manera:

$$\text{Volumen de unidades léxicas básicas (Kb)} = \frac{\text{Unidades léxicas} * 8}{1024}$$



**Ilustración 2** Histogramas de Volumen de unidades léxicas (Kb)

Como se aprecia en la ilustración 2, la media del volumen de unidades léxicas básicas es de 0.2057 Kilobytes, para 30 publicaciones de diferentes usuarios de la red social. Por tanto, la importancia de considerar el factor almacenamiento a medida que se incrementa la cantidad de usuarios.

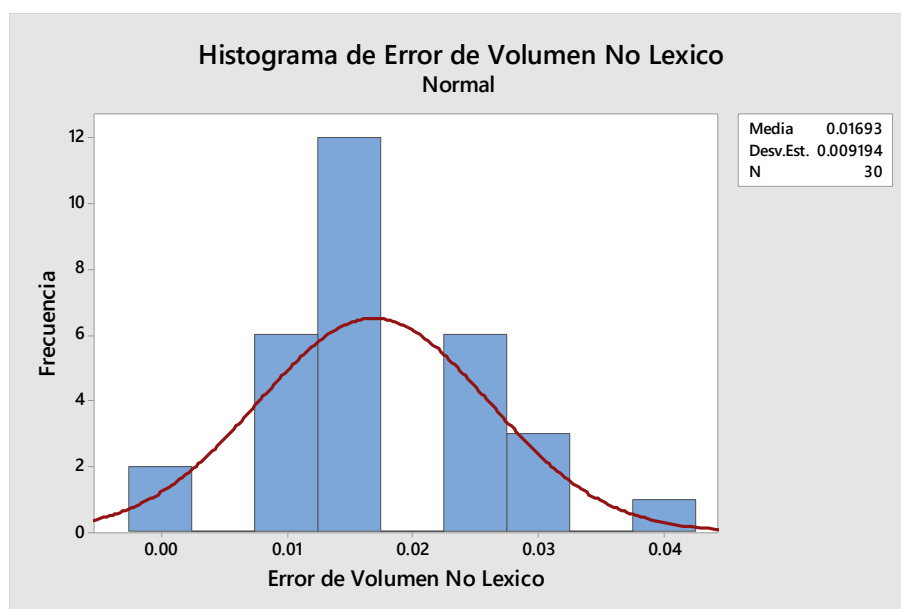
## Error de Volumen No Léxico

En primer lugar, el error, se produce debido a que el algoritmo no logra realizar una óptima segmentación, es decir se almacenan unidades léxicas que no deberían ser consideradas generando un almacenamiento falso en el corpus, lo que implica un desperdicio en Kilobytes por cada uno de estos errores generados. El cálculo del error es:

$$\text{Error} = \text{Unidades léxicas erróneas} - \text{Unidades segmentación}$$

Asimismo, para el cálculo del error de volumen no léxico, se considera el error en unidades léxicas, y como cada uno de estos se representa por un byte, se realiza mediante la siguiente formula:

$$\text{Error de Volumen No Léxico (Kb)} = \frac{\text{Error} * 8}{1024}$$



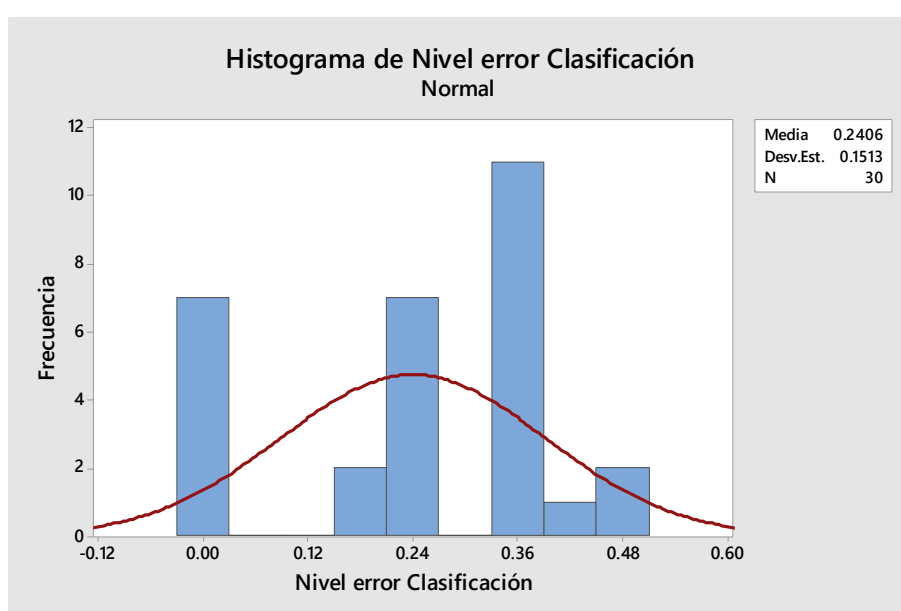
**Ilustración 3** Histograma de error de volumen no léxico

Como se detalla en la ilustración 3, la media del error de volumen de unidades no léxicas es de 0.01693 Kilobytes, para 30 publicaciones de diferentes usuarios de la red social. Por tanto, la importancia de que el algoritmo tienda a un error cero, pues a medida que se incrementa la cantidad de usuarios y publicaciones se acumula en el corpus información intrascendente.

## Nivel del error de clasificación

En lo que respecta al nivel del error de clasificación, se considera la razón entre las unidades léxicas erróneas, y se sustraerán las unidades léxicas correctamente segmentadas con las unidades léxicas correctamente segmentadas. El nivel óptimo esperado es del 100% por tal el mayor error está dado por la distancia de los valores independientes de cada publicación del valor referenciado.

$$\text{Nivel de error de clasificación} = \frac{\text{Unidades léxicas} - \text{segmentación}}{\text{Unidades léxicas}}$$



**Ilustración 4** Histograma de nivel de error de clasificación

Como se detalla en la ilustración 4, la media del nivel del error de clasificación de unidades léxicas es de 0.2406, para 30 publicaciones de diferentes usuarios de la red social. El nivel de error medio de 24.06% implica que existen unidades léxicas que no se están considerando, por tanto, es necesario un ajuste en el algoritmo para reducir el error a 0% y con ello pueda realizar las clasificaciones de los lexemas con los usuarios, de acuerdo al patrón de concordancia. Este ajuste implica una correcta clasificación de los lexemas para que formen o sugieran correctamente los grupos de acorde a la similitud de los lexemas de la publicación con los patrones almacenados en el corpus y relacionarlo por un nivel de usuario.

## V. DISCUSIÓN

Debido a los cambios en los procesos comunicativos que reflejan en cierta medida los cambios sociales, económicos y políticos como transformación de la forma de educar, escribir y de expresarse; por esto el contenido digital es la tendencia de la revolución de las comunicaciones, tomándolo como insumo para el análisis de los textos producidos en estas nuevas formas de expresión en redes sociales de diversa índole. Las redes sociales se han convertido en el escenario de expresión de las personas, estando gran parte de su tiempo, compartiendo gustos, preferencias, inquietudes, anhelos, deseos, sueños, esto en las plataformas no se clasifica de tal manera que se tengan niveles en cuanto al uso de Lexemas. Cabe resaltar el desarrollo del algoritmo de edición Levenshtein para la transformación de cadenas, implementando agente de inteligencia para optimizar el analizador léxico. Ante ello se implementó una solución que analice los textos de las publicaciones y relaciones a los usuarios por niveles de descubrimiento de lexemas, mediante la valoración de los siguientes indicadores:

### Volumen de unidades léxicas básicas

En cuanto al volumen de unidades léxicas básicas, los valores individuales del volumen de unidades léxicas básicas tienen una media de 0.020573 Kilobytes. Asimismo, el 95% de los valores del volumen de unidades léxicas básicas se encuentran entre dos desviaciones estándar de la media, es decir, entre 0.004470 y 0.007545 Kilobytes.

En ese sentido se corresponde a lo obtenido por Gamarra Contreras (2018) mediante el análisis del lenguaje subjetivo en sistemas de minerías de opiniones; pues en sus resultados tiene que el 40% confirma la dificultad de la interpretación del contenido de un texto; por la incorrecta composición del comentario generado y que la mejor técnica es ir incorporando recursos lingüísticos dependiendo las necesidades de la implementación; pues el autor confirma lo difícil que implica para los algoritmos la extracción de lexemas, interpretarlos y luego relacionarlos con los patrones existentes en el corpus, además de solo almacenar el volumen de unidades léxicas básicas. Otra similitud con Rincón Acuña (2016) quien trabaja con la fenomenología, modelando mediante enfoques computacionales para poder

descripciones con enfoque basados en la ambigüedad y falta de contexto; como concluye el autor en estos contextos se generan algoritmos más óptimos que almacenar solo las unidades de los lexemas apropiadas y almacenar solo lo adecuado al corpus.

#### Error de Volumen No Léxico

Los valores individuales del error de volumen no léxico tienen una media de 0.016927 Kilobytes. Asimismo, el 95% del error de volumen no léxico se encuentran entre dos desviaciones estándar de la media, es decir, entre 0.007322 y 0.012360 Kilobytes.

Estos resultados son semejantes a los resultados de Campos García (2020) quien realiza comparaciones de 43 perfiles, encontrándose 10% de coincidencia pero categorizó los niveles de acuerdo a los temas pero faltó establecer conexiones entre perfiles de una misma persona y obteniéndose errores en las clasificaciones a pesar de que su modelo posee una precisión de similitud de 90%; es coherente en cuanto la existencia de errores de clasificación, pues a pesar de clasificar niveles en un 95% estos errores conllevan a un error de almacenamiento de los lexemas que no debieron ser clasificados por el algoritmo y el costo en la búsqueda con los patrones almacenados en el corpus.

#### Nivel del error de clasificación

Asimismo, los valores individuales del nivel del error de clasificación tienen una media de 0.24056. Asimismo, el 95% de los valores nivel del error de clasificación se encuentran entre dos desviaciones estándar de la media, es decir, entre 0.12048 y 0.20337.

Estos resultados son coherentes en parte con lo planteado por Ortega de los Ríos (2020), quien debido a la poca fiabilidad conseguida por GaussNB y KNN por las limitaciones de las librerías, necesito conseguir modelos fiables superiores al 80%, para tener tweets más representativos, ello se explica dado que se consideró que el conjunto de los datos presentaba una distribución normal. La discordancia con el autor mencionado es considerar los datos con distribución normal, pues como se muestra en la ilustración 5, los datos no se asemejan a una distribución normal pues  $p < 0.05$ . Como el nivel del error de clasificación está en 24.06%, la coincidencia de

los lexemas de las publicaciones de cada usuario no es óptima debido a que el algoritmo de clasificación está en aprendizaje y los pocos usuarios pueden dar una falsa conclusión de resultados óptimos. En ese sentido se corresponde a lo obtenido por Piñón Ferrer (2018), que mediante algoritmos de clasificación y predicción logra optimizaciones de modelos para procesar texto de los tuits, hashtag; pues de la misma manera con algoritmos de aprendizaje como el de la distancia de edición de Levenshtein.

## **VI. CONCLUSIONES**

- a) Se establece que el corpus textual de las unidades léxicas básicas en la red social, se debe considerar el volumen del vector inferido, determinado por las unidades léxicas básicas que se almacenan de cada publicación que en promedio es de 0.2057 Kilobytes por cada publicación de usuario.
- b) Se comprueba que en el análisis del componente morfológico y el lexicográfico del lexema de los textos por usuario en la red social, es determinante el manejo del error de volumen no léxico, debido a la consideración de lexemas en el corpus por imprecisiones del algoritmo utilizado obteniéndose una media de 01693 Kilobytes, por cada una de las publicaciones
- c) Se evidencia que, al determinar patrones de lexemas en los textos por cada usuario en la red social, un factor determinante es el nivel del error de clasificación, y está ligado directamente a las unidades léxicas que se clasifican de manera errónea, además de influir en búsqueda innecesaria de estos en los patrones del corpus, obteniéndose una media de este error de 24.06%
- d) Se comprueba que se logra diseñar una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante el algoritmo de Inteligencia Artificial de la distancia de edición de Levenshtein



## **VII. RECOMENDACIONES**

- a) Se recomienda para futuras investigaciones considerar nuevos modelos implementados con algoritmos que puedan optimizar el almacenamiento de las unidades léxicas en el corpus textual
- b) Se sugiere en otras investigaciones técnicas diversas de análisis morfológico y lexicográfico de los lexemas de los textos aplicables para redes sociales temáticas
- c) Se recomienda en investigaciones posteriores el manejo de una caché de patrones de lexemas de las publicaciones para agilizar las búsquedas en contextos con grandes cantidades de usuarios.
- d) Se sugiere considerar en futuras investigaciones técnicas de almacenamiento dinámico para sobrescribir en los datos y depurar errores que se puedan generar en la clasificación temática en estos tipos de redes sociales

## REFERENCIAS

1. Alvarado Atocha, M. J. (2019). *GUAYAQUIL-ECUADOR Septiembre 2019*. *GUAYAQUIL-ECUADOR Septiembre 2019*.
2. Alergia México. 2016. Metodología de la investigación. México Alergia México : Revista , 2016.
3. Arias, J., Villasís, A., Keever, M., Guadalupe, M., & Novales, M. (2016). Metodología de la investigación. *Alergia México*, 63, 201–206. [www.nietoeditores.com.mx](http://www.nietoeditores.com.mx)
4. Campos García, J. F. (2020). DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE COMPARACIÓN DE PERFILES EN REDES SOCIALES. *Zaguan.Unizar.Es*, 70.
5. Castillo Mantuano, M. A. (2020). *Twitter, un medio social que diseña políticos: análisis de la imagen política que se construye en el perfil de Paola Pabón durante el período de campaña política seccional 2019*. <http://www.dspace.uce.edu.ec/bitstream/25000/21888/1/T-UCE-0009-CSO-288.pdf>
6. <CIP, Colegio de Ingenieros del Perú. 1987. CIP. [En línea] 1987. [http://www.cip.org.pe/publicaciones/reglamentosCNCD2018/codigo\\_de\\_etica\\_del\\_cip.pdf](http://www.cip.org.pe/publicaciones/reglamentosCNCD2018/codigo_de_etica_del_cip.pdf).
7. Cortez Vásquez Augusto. (2017). *Clasificación de textos informáticos mediante indexación semántica latente*. 109–121. [https://planetadelibrosar0.cdnstatics.com/libros\\_contenido\\_extra/40/39307\\_Inteligencia\\_artificial.pdf](https://planetadelibrosar0.cdnstatics.com/libros_contenido_extra/40/39307_Inteligencia_artificial.pdf)
8. Cortez Vásquez Augusto. (2018). Inteligencia artificial 101. *Alenta Editorial*, 352. [https://planetadelibrosar0.cdnstatics.com/libros\\_contenido\\_extra/40/39307\\_Inteligencia\\_artificial.pdf](https://planetadelibrosar0.cdnstatics.com/libros_contenido_extra/40/39307_Inteligencia_artificial.pdf)
9. Cruz, C., Olivares, S. y González, M. 2015. Metodología de la investigación. . México D.F. : Grupo Editorial Patria, 2015.
10. Dzyaloshinsky, I., Pilgún, M., & Annenkova, I. (2020). Análisis del contenido de las redes sociales: las peculiaridades y la problemática de la situación en Rusia. *Estudios Sobre El Mensaje Periodístico*, 26(1), 95–102. <https://doi.org/10.5209/esmp.67289>

11. E. Estrada, J. Hernandez, J. Morales, M. N. (2020). Introducción a la inteligencia artificial. 23, 12(32), 1. [https://d1wqtxts1xzle7.cloudfront.net/62082081/articulo-v220200212-81699-1198il1.pdf?1581568380=&response-content-disposition=inline%3B+filename%3DIntroduccion\\_a\\_la\\_inteligencia\\_artificia.pdf&Expires=1621021181&Signature=gjUlfgeDA6U~V71hGasjbUZmOiXA2751ri](https://d1wqtxts1xzle7.cloudfront.net/62082081/articulo-v220200212-81699-1198il1.pdf?1581568380=&response-content-disposition=inline%3B+filename%3DIntroduccion_a_la_inteligencia_artificia.pdf&Expires=1621021181&Signature=gjUlfgeDA6U~V71hGasjbUZmOiXA2751ri)
12. Esteban Nieto, T. N. (2018). *Tipos de investigación*. 1–4.
13. Faridy Bermeo, Yaffar, Hernández-Mosqueda, José Silvano y Tobón-Tobón, Sergio. 2016. ANÁLISIS DOCUMENTAL DE LA V HEURÍSTICA MEDIANTE LA CARTOGRAFÍA CONCEPTUAL. México : Ra Ximhai, 2016. ISSN: 1665-0441.
14. Fidias G., Arias. 2016. El proyecto de investigación. 2016. 7ma Edición
15. Flores, M. (2020). *Diseño de un prototipo de sistema experto que permita identificar patrones basado en las aptitudes y habilidades académicas de los alumnos de nivel superior*. 8, 77–83.
16. Gamarra Contreras, G. B. (2019). *INVESTIGACIÓN Y ANÁLISIS DEL LENGUAJE SUBJETIVO EN SISTEMAS DE MINERIAS DE OPINIONES EXISTENTES EN EL MERCADO*.
17. García del Saz, José Antonio. 2019. Herramienta de descarga y análisis de datos de Twitter sobre participación ciudadana. España : Universidad Autónoma de Madrid, 2019.
18. González García, C., Núñez-Valdez, E., García-Díaz, V., Pelayo G-Bustelo, C., & Cueva-Lovelle, J. M. (2019). A Review of Artificial Intelligence in the Internet of Things. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(4), 9. <https://doi.org/10.9781/ijimai.2018.03.004>
19. Hernández Escobar, Arturo Andrés, y otros. 2018. Metodología de la Investigación Científica. 2018. 978-84-948257-0-5.
20. Hernández-Sampieri, R., Fernández-Collado, R. y Baptista-Lucio, P.,. 2017. Selección de la muestra. 2017.
21. Hernández Sampieri, Roberto, Fernández Collado, Carlos y Pilar, Baptista Lucio. 2010. Metodología de la Investigación. México : McGRAW-HILL, 2010.
22. Herrera Castrillo, Cliffor Jerry. 2019. Estadística y probabilidades. s.l. : Universidad Nacional Autónoma de Nicaragua, 2019.

23. Joyanes Aguilar, Luis. 2016. *Sistemas de Información en la empresa: El impacto de la nube, la movilidad y los medios sociales*. Colombia : Alfa Omega, 2016. 2da edición.
24. Marquis, P., Papini, O., & Prade, H. (2020). *A Guided Tour of Artificial Intelligence Research*.
25. Mengliyev, B., Shahabitdinova, S., Khamroeva, S., Gulyamova, S., & Botirova, A. (2021). The morphological analysis and synthesis of word forms in the linguistic analyzer. *Journal of Language and Linguistic Studies*, 17(1), 558–564.
26. Meza Flores, Julio C. y Rodríguez Páez, Liliana. 2020. Diseño de un prototipo de sistema experto que permita identificar patrones basado en las aptitudes y habilidades académicas de los alumnos de nivel superior. s.l. : Revista Aristas: Investigación Básica y Aplicada., 2020. ISSN 2007-9478.
27. Ordaz Sanz, José Antonio, Melgar Hiraldo, María del Carmen y Rubio Castaño, Carmen María. 2016. *Métodos estadísticos y econométricos en la empresa y para finanzas*,. s.l. : Universidad Pablo de Olavide, 2016.
28. Ouhainen, LASSER. 2018. *Inteligencia Artificial 101 cosas que debes saber hoy sobre nuestro futuro*. España : Alienta Editorial, 2018. ISBN: 978-84-17568-08-5.
29. Ortega de los Ríos, & Milagros. (2020). *Centro Universitario de la Defensa en la Escuela Naval Militar Análisis de sentimiento en las redes sociales producido por las Fuerzas Armadas Españolas*. 2019–2020.
30. Ortíz Morán, Yorly Melissa. 2020. ANÁLISIS DEL COMPORTAMIENTO DEL CONSUMIDOR DIGITAL EN REDES SOCIALES MEDIANTE TÉCNICA DE MINERÍA DE TEXTO. Ecuador : Universidad de Guayaquil, 2020.
31. Piñon Ferrer, L. (2018). *Minería de datos aplicada a Twitter y análisis de sentimientos mediante algoritmos de inteligencia artificial*. 1–76. [http://repositori.uji.es/xmlui/bitstream/handle/10234/177643/TFG\\_Piñon Ferrer%2C Luis.pdf?sequence=1&isAllowed=y](http://repositori.uji.es/xmlui/bitstream/handle/10234/177643/TFG_Piñon_Ferrer%2C%20Luis.pdf?sequence=1&isAllowed=y)
32. Ramos Martín, Alicia y Ramos Martín, Jesús. 2015. *Aplicaciones Web*. España : Ediciones Nobel, 2015.
33. Santaella Santos, Daniel. 2019. *Procesamiento del Lenguaje Natural para la extracción en tiempo real de neologismos del español en redes sociales*. España : Universidad Autónoma de Madrid, 2019.

34. Quej Cosgaya Héctor Manuel, Durán Lugo Juan Miguel, E. S. G. M. (2016). *Multidisciplinas de la Ingeniería. 4*.
35. Ramos Martín, Alicia y Ramos Martín, Jesús. 2015. *Aplicaciones Web*. España : Ediciones Nobel, 2015.
36. Ramos, F. M., & Velez, J. I. (2016). *Integración de técnicas de procesamiento de lenguaje natural a través de servicios web*. 1–107. <http://www.alejandrorage.com.ar/files/advising/2016-thesis-velez&ramos.pdf>
37. Rincón Acuña, J. C. (2016). *Memoria y formación: Una aproximación desde la fenomenología y la inteligencia artificial*. 206.
38. Roberto, Hernández Sampieri, Mendez Valencia Sergio, Mendoza Torres Christian Paulina, C. R. A. (2017). *Fundamentos de Investigación - Hernández Sampieri.pdf*.
39. Thoma, S., Rettinger, A., & Both, F. (2017). Knowledge fusion via embeddings from text, knowledge graphs, and images. *ArXiv*.
40. UCV, Universidad César Vallejo. 2020. [ucv. ucv.edu.pe](https://www.ucv.edu.pe/). [En línea] 28 de Agosto de 2020. <https://www.ucv.edu.pe/wp-content/uploads/2020/11/RCUN%C2%B00262-2020-UCV-Aprueba-Actualizaci%C3%B3n-del-C%C3%B3digo-%C3%89tica-en-Investigaci%C3%B3n-1-1.pdf>.
41. Vila Ponte, Juan José. 2018. *La terminología de las redes sociales digitales: estudio morfológico-semántico y lexicográfico*. España : Universidad de Coruña, 2018.
42. Wangperawong, A., Kriangchaivech, K., Lanari, A., Lam, S., & Wangperawong, P. (2018). Comparing heterogeneous entities using artificial neural networks of trainable weighted structural components and machine-learned activation functions. *ArXiv*, 1–9.

## ANEXOS

Tabla de Operacionalización

Variable	Definición conceptual	Dimensión	Indicador	Nivel de medición
Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas (descubrimiento de lexemas)	Son las representaciones textuales latentes, que también se denominan incrustaciones de palabras o representaciones de palabras distribuidas. Creado con métodos no supervisados, solo se basan en un gran corpus de texto como entrada (Thomas et al., 2017).	Corpus textual	Volumen de unidades léxicas básicas	Continua/Razón
		Patrones de lexemas en los textos	Error de Volumen no Léxico	Continua/Razón
			Nivel del error de clasificación	Continua/Razón

## Matriz de consistencia

Título: Diseño de una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial							
Autor: David Ernesto Torres Dávila							
PROBLEMA	HIPÓTESIS	OBJETIVOS	VARIABLE	DIMENSIONES	INDICADORES	INSTRUMENTO	METODOLOGÍA
<p><b>Problema General:</b>  <b>PA:</b> ¿Cómo se analiza texto y relaciona usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial en una red social?</p> <p><b>Problemas Específicos:</b>  <b>PE1:</b> ¿Cuál es el corpus textual de las unidades léxicas básicas que se estudiarán en la red social?  <b>PE2:</b> ¿Cuáles son los componentes morfológicos y el lexicográfico del lexema de los textos por usuario en la red social?  <b>PE3:</b> ¿Qué patrones de lexemas en los textos por cada usuario hay en la red social?</p>	<p><b>General:</b>            HA: Mediante técnicas de Inteligencia Artificial en una red social se analiza texto y relaciona usuarios por niveles de descubrimiento de lexemas</p>	<p><b>General:</b>  <b>OA:</b> Diseñar una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial</p> <p><b>Específicos</b>  <b>OA1:</b> Establecer un corpus textual de las unidades léxicas básicas que se estudiarán en la red social.  <b>OA2:</b> Analizar el componente morfológico y el lexicográfico del lexema de los textos por usuario en la red social  <b>OA3:</b> Determinar patrones de lexemas en los textos por cada usuario en la red social</p>	<p>Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas</p>	<p>Corpus textual</p> <p>Patrones de lexemas en los textos</p>	<p>Volumen de unidades léxicas básicas</p> <p>Error de Volumen No Léxico</p> <p>Nivel del error de clasificación</p>	<p>Ficha de registro</p> <p>Ficha de registro</p> <p>Ficha de registro</p>	<p>METODOLOGÍA APLICADA: Técnicas de inteligencia artificial</p> <p>TIPO DE INVESTIGACION: APLICADA</p> <p>DISEÑO DE INVESTIGACION: NO-EXPERIMENTAL</p> <p>POBLACIÓN, MUESTRA            350 textos producidos por los usuarios de una red social en forma mensual</p>

Ficha de registro N°1

Ficha de registro - Volumen de unidades léxicas básicas							
<b>Investigador</b>		David Ernesto Torres Dávila			<b>Tipo de prueba</b>	DESCRIPTIVO	
<b>Institución</b>					Universidad Privada César Vallejo		
<b>Dimensión de estudio para Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas</b>					Corpus textual		
<b>Fecha de Inicio</b>			<b>Fecha final</b>				
<b>Variable</b>		<b>Indicador</b>	<b>Medida</b>		<b>Fórmula</b>		
Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas		Volumen de unidades léxicas básicas	Kb		(Solo si hay un indicador con formulación)		
<b>#</b>	<b>Publicación</b>	<b>Tamaño del vector inferido</b>	<b>Volumen Vector Inferido(bits)</b>	<b>Número de unidades léxicas</b>	<b>Volumen Unidades Léxicas (Kb)</b>	<b>Número mínimo de apariciones requeridas por termino</b>	<b>Volumen de Unidades No Léxicas</b>
1							
2							
3							
4							
5							

Donde:

**Tamaño del vector inferido:** Es el reconocimiento del lexema el cual se almacena en una estructura que se maneja como un vector

**Número mínimo de apariciones requeridas por termino:** Es el mínimo de veces que un lexema aparece dentro de un texto.



Ficha de registro N°2

Ficha de registro – Error de Volumen No Léxico y error de clasificación						
Investigador	David Ernesto Torres Dávila			Tipo de prueba	DESCRIPTIVO	
Institución	Universidad Privada César Vallejo					
Dimensión de estudio para Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas	Patrones de lexemas en los textos					
Fecha de Inicio			Fecha final			
Variable	Indicador		Medida	Fórmula		
Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas	Error de Volumen No Léxico Nivel del error de clasificación			(Solo si hay un indicador con formulación)		
#	Texto	Segmentación	Error de Volumen No Léxico (Mb)	Unidades Léxicas Erróneas	Error	Nivel del error de clasificación
1						
2						
3						
4						
5						

Donde:

**Segmentación:** Es la estructura que compone al texto

**Error:** Es la desviación de algunas de sus estructuras del texto

**Nivel del error de clasificación:** Es el error por el grado de clasificación del usuario relacionado al texto generado

Documentos para validar los instrumentos de medición a través de juicio de expertos

## CARTA DE PRESENTACIÓN

Mg. Correa Calle, Teófilo Roberto.

Presente

Asunto: VALIDACIÓN DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTO.

Me es muy grato comunicarme con usted para expresarle mi saludo y así mismo, hacer de su conocimiento que, siendo estudiante de la Escuela profesional de Ingeniería de Sistemas de la Universidad César Vallejo, en la sede de Piura, requiero validar los instrumentos con los cuales recogeré la información necesaria para poder desarrollar mi investigación.

El título de nuestro proyecto de investigación es: “**Diseño de una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial**” y siendo imprescindible contar con la aprobación de docentes especializados para poder aplicar los instrumentos en mención, he considerado conveniente recurrir a usted, ante su connotada experiencia en temas de investigación educativa.

El expediente de validación, que le hacemos llegar contiene:

- Carta de presentación.
- Definiciones conceptuales de las variables y dimensiones.
- Matriz de consistencia de las variables.
- Certificado de validez de contenido de los instrumentos.
- Instrumento de validación de cada indicador

Expresándole mi sentimiento de respeto y consideración me despido de usted, no sin antes agradecerle por la atención que dispense a la presente.

Atentamente.



---

Firma

David Ernesto Torres Dávila  
D.N.I: 74593109

## DEFINICIÓN CONCEPTUAL DE LAS VARIABLES Y DIMENSIONES

### **Variable:**

#### **Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas**

Consiste en la indexación semántica para descubrir una nueva información a partir del análisis de textos estructurados, clasificación de acuerdo a niveles por los usuarios, a partir de expresiones en un lenguaje natural en el contexto de redes sociales (Cortez Vásquez Augusto, 2017)

### **Dimensiones:**

#### **Corpus textual**

Corpus, es una recopilación de muestras de lengua, bien sea en su expresión oral, o en su expresión escrita y que representa el uso de una lengua natural, en su extensión como en su variedad, según Vila Ponte se puede clasificar las búsquedas según la siguiente tabla(Vila Ponte, 2018).

#### **Patrones de lexemas en los textos**

Es el proceso encargado del descubrimiento de conocimientos que no existían explícitamente en ningún texto de la colección, pero que surgen de relacionar el contenido de varios de ellos, mediante técnicas de minería de texto.

**MATRIZ DE CONSISTENCIA DE LAS VARIABLES**

Título: Diseño de una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial							
Autor: David Ernesto Torres Dávila							
PROBLEMA	HIPÓTESIS	OBJETIVOS	VARIABLE	DIMENSIONES	INDICADORES	INSTRUMENTO	METODOLOGÍA
<p><b>Problema General:</b>  <b>PA:</b> ¿Cómo se analiza texto y relaciona usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial en una red social?</p> <p><b>Problemas Específicos:</b>  <b>PE1:</b> ¿Cuál es el corpus textual de las unidades léxicas básicas que se estudiarán en la red social?  <b>PE2:</b> ¿Cuáles son los componentes morfológicos y el lexicográfico del lexema de los textos por usuario en la red social?  <b>PE3:</b> ¿Qué patrones de lexemas en los textos por cada usuario hay en la red social?</p>	<p><b>General:</b>            HA: Mediante técnicas de Inteligencia Artificial en una red social se analiza texto y relaciona usuarios por niveles de descubrimiento de lexemas</p>	<p><b>General:</b>  <b>OA:</b> Diseñar una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial</p> <p><b>Específicos</b>  <b>OA1:</b> Establecer un corpus textual de las unidades léxicas básicas que se estudiarán en la red social.  <b>OA2:</b> Analizar el componente morfológico y el lexicográfico del lexema de los textos por usuario en la red social  <b>OA3:</b> Determinar patrones de lexemas en los textos por cada usuario en la red social</p>	<p>Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas</p>	<p>Corpus textual</p> <p>Patrones de lexemas en los textos</p>	<p>Volumen de unidades léxicas básicas</p> <p>Error de Volumen No Léxico</p> <p>Nivel del error de clasificación</p>	<p>Ficha de registro</p> <p>Ficha de registro</p> <p>Ficha de registro</p>	<p><b>METODOLOGÍA APLICADA:</b>            Técnicas de inteligencia artificial</p> <p><b>TIPO DE INVESTIGACION:</b>            APLICADA</p> <p><b>DISEÑO DE INVESTIGACION:</b>            NO-EXPERIMENTAL</p> <p><b>POBLACIÓN, MUESTRA</b>            350 textos producidos por los usuarios de una red social en forma mensual</p>

**CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:**
**Variable: Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas**

Nº	DIMENSIONES / ítems	Pertinencia <sup>1</sup>		Relevancia <sup>2</sup>		Claridad <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
	<b>DIMENSION 1: Corpus textual</b>							
<b>1</b>	<b>INDICADOR 1: Volumen de unidades léxicas básicas</b>							
<b>a</b>	Es formulado con lenguaje apropiado.	X		X		X		
<b>b</b>	Esta expresado en conducta observable.	X		X		X		
<b>c</b>	Es adecuado el avance, la ciencia y tecnología.	X		X		X		
<b>d</b>	Existe una organización lógica.	X		X		X		
<b>e</b>	Comprende los aspectos de cantidad y calidad.	X		X		X		
<b>f</b>	Adecuado para valorar los aspectos del sistema metodológico y científico.	X		X		X		
<b>g</b>	Está basado en aspectos teóricos y científicos.	X		X		X		
<b>h</b>	En los datos respecto al indicador.	X		X		X		
<b>i</b>	Responde al propósito de investigación.	X		X		X		
<b>j</b>	El instrumento es adecuado al tipo de investigación.	X		X		X		
	<b>DIMENSION 2: Patrones de lexemas en los textos</b>	Si	No	Si	No	Si	No	Sugerencias
<b>1</b>	<b>INDICADOR 1: Error de Volumen No Léxico</b>							
<b>a</b>	Es formulado con lenguaje apropiado.	X		X		X		
<b>b</b>	Esta expresado en conducta observable.	X		X		X		
<b>c</b>	Es adecuado el avance, la ciencia y tecnología.	X		X		X		
<b>d</b>	Existe una organización lógica.	X		X		X		
<b>e</b>	Comprende los aspectos de cantidad y calidad.	X		X		X		
<b>f</b>	Adecuado para valorar los aspectos del sistema metodológico y científico.	X		X		X		
<b>g</b>	Está basado en aspectos teóricos y científicos.	X		X		X		
<b>h</b>	En los datos respecto al indicador.	X		X		X		
<b>i</b>	Responde al propósito de investigación.	X		X		X		
<b>j</b>	El instrumento es adecuado al tipo de investigación.	X		X		X		
<b>2</b>	<b>INDICADOR 2: Nivel del error de clasificación</b>	Si	No	Si	No	Si	No	Sugerencias
<b>a</b>	Es formulado con lenguaje apropiado.	X		X		X		
<b>b</b>	Esta expresado en conducta observable.	X		X		X		
<b>c</b>	Es adecuado el avance, la ciencia y tecnología.	X		X		X		
<b>d</b>	Existe una organización lógica.	X		X		X		
<b>e</b>	Comprende los aspectos de cantidad y calidad.	X		X		X		
<b>f</b>	Adecuado para valorar los aspectos del sistema metodológico y científico.	X		X		X		
<b>g</b>	Está basado en aspectos teóricos y científicos.	X		X		X		
<b>h</b>	En los datos respecto al indicador.	X		X		X		
<b>i</b>	Responde al propósito de investigación.	X		X		X		
<b>j</b>	El instrumento es adecuado al tipo de investigación.	X		X		X		

Observaciones (precisar si hay suficiencia): \_\_\_\_\_

Opinión de aplicabilidad:    **Aplicable [ X ]**            **Aplicable después de corregir [ ]**            **No aplicable [ ]**

Apellidos y nombres del juez validador.    **Mg. Correa Calle, Teófilo Roberto**

DNI: 02820231

**Especialidad del validador: Maestría en Dirección y Gestión de Tecnología de la Información y Comunicación e Ingeniero Informático.**

<sup>1</sup>**Pertinencia:** El ítem corresponde al concepto teórico formulado.

<sup>2</sup>**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo

<sup>3</sup>**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

**Nota:** Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

**02 de diciembre del 2021**



TEOFILO ROBERTO  
CORREA CALLE

-----  
**Firma del Experto Informante.**

**TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Volumen de unidades léxicas básicas**
**I. DATOS GENERALES**

Apellidos y Nombres del Experto: **Correa Calle, Teófilo Roberto**  
 Título y/o Grado Académico: **Ingeniero en Informática Mg.**

Doctor ()    Magister ()    Ingeniero ()    Licenciado ()    Otro () .....

Universidad que labora: **Universidad César Vallejo Piura**  
 Fecha: **02-12-2021**

**TESIS: "Diseño de una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial"**

**Autor: David Ernesto Torres Dávila**

**Deficiente (0-20%)    Regular (21-50%)    Bueno (51-70%)    Muy Bueno (71-80%)    Excelente (81-100%)**

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

**II. ASPECTOS DE VALIDACIÓN**

INDICADOR	CRITERIO	VALORACION				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				75	
OBJETIVIDAD	Esta expresado en conducta observable.				75	
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.				80	
ORGANIZACION	Existe una organización lógica.				80	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				75	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				75	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80	
COHERENCIA	En los datos respecto al indicador.				80	
METODOLOGIA	Responde al propósito de investigación.				80	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80	
<b>TOTAL PROMEDIO</b>					78	

**III. PROMEDIO DE VALIDACIÓN**

**78**

**IV. OPCIÓN DE APLICABILIDAD**

- (X) El instrumento puede ser aplicado, tal como está elaborado.  
 () El instrumento debe ser mejorado antes de ser aplicado.

**FIRMA DEL EXPERTO**



TEOFILO ROBERTO  
CORREA CALLE



Ficha de registro - Volumen de unidades léxicas básicas							
Investigador		David Ernesto Torres Dávila			Tipo de prueba	DESCRIPTIVO	
Institución					Universidad Privada César Vallejo		
Dimensión de estudio para Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas					Corpus textual		
Fecha de Inicio			Fecha final				
Variable		Indicador	Medida		Fórmula		
Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas		Volumen de unidades léxicas básicas	Kb		(Solo si hay un indicador con formulación)		
#	Publicación	Tamaño del vector inferido	Volumen Vector Inferido(bits)	Número de unidades léxicas	Volumen Unidades Léxicas (Kb)	Número mínimo de apariciones requeridas por termino	Volumen de Unidades No Léxicas
1							
2							
3							
4							
5							

Donde:

**Tamaño del vector inferido:** Es el reconocimiento del lexema el cual se almacena en una estructura que se maneja como un vector

**Número mínimo de apariciones requeridas por termino:** Es el mínimo de veces que un lexema aparece dentro de un texto.

**TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Error de Volumen No Léxico**
**I. DATOS GENERALES**

Apellidos y Nombres del Experto: **Correa Calle, Teófilo Roberto**  
 Título y/o Grado Académico: **Ingeniero en Informática Mg.**

**Doctor ( )    Magister ( )    Ingeniero ( )    Licenciado ( )    Otro ( ).....**

Universidad que labora: **Universidad César Vallejo Piura**  
 Fecha: **02-12-2021**

**TESIS: “Diseño de una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial”**

**Autor: David Ernesto Torres Dávila**

**Deficiente (0-20%)    Regular (21-50%)    Bueno (51-70%)    Muy Bueno (71-80%)    Excelente (81-100%)**

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

**II. ASPECTOS DE VALIDACIÓN**

INDICADOR	CRITERIO	VALORACION				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				75	
OBJETIVIDAD	Esta expresado en conducta observable.				75	
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.				80	
ORGANIZACIÓN	Existe una organización lógica.				80	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				75	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				75	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80	
COHERENCIA	En los datos respecto al indicador.				80	
METODOLOGIA	Responde al propósito de investigación.				80	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80	
<b>TOTAL PROMEDIO</b>					78	

**III. PROMEDIO DE VALIDACIÓN**

**78**

**IV. OPCIÓN DE APLICABILIDAD**

- (X) El instrumento puede ser aplicado, tal como está elaborado.  
 ( ) El instrumento debe ser mejorado antes de ser aplicado.

**FIRMA DEL EXPERTO**



TEÓFILO ROBERTO  
CORREA CALLE

Ficha de registro – Error de Volumen No Léxico y error de clasificación						
<b>Investigador</b>	David Ernesto Torres Dávila			<b>Tipo de prueba</b>	DESCRIPTIVO	
<b>Institución</b>	Universidad Privada César Vallejo					
<b>Dimensión de estudio para Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas</b>	Patrones de lexemas en los textos					
<b>Fecha de Inicio</b>			<b>Fecha final</b>			
<b>Variable</b>	<b>Indicador</b>		<b>Medida</b>	<b>Fórmula</b>		
Análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas	Error de Volumen No Léxico Nivel del error de clasificación			(Solo si hay un indicador con formulación)		
<b>#</b>	<b>Texto</b>	<b>Segmentación</b>	<b>Error de Volumen No Léxico (Mb)</b>	<b>Unidades Léxicas Erróneas</b>	<b>Error</b>	<b>Nivel del error de clasificación</b>
1						
2						
3						
4						
5						

Donde:

Segmentación: Es la estructura que compone al texto

Error: Es la desviación de algunas de sus estructuras del texto

Nivel del error de clasificación: Es el error por el grado de clasificación del usuario relacionado al texto generado

**TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Error de clasificación**
**I. DATOS GENERALES**

Apellidos y Nombres del Experto: **Correa Calle, Teófilo Roberto**  
 Título y/o Grado Académico: **Ingeniero en Informática Mg.**

Doctor ( )    Magister ( )    Ingeniero ( )    Licenciado ( )    Otro ( ).....

Universidad que labora: **Universidad César Vallejo Piura**  
 Fecha: **02-12-2021**

**TESIS: “Diseño de una red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial”**

**Autor: David Ernesto Torres Dávila**

**Deficiente (0-20%)    Regular (21-50%)    Bueno (51-70%)    Muy Bueno (71-80%)    Excelente (81-100%)**

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

**II. ASPECTOS DE VALIDACIÓN**

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				75	
OBJETIVIDAD	Esta expresado en conducta observable.				75	
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.				80	
ORGANIZACION	Existe una organización lógica.				80	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				75	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				75	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80	
COHERENCIA	En los datos respecto al indicador.				80	
METODOLOGIA	Responde al propósito de investigación.				80	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80	
<b>TOTAL PROMEDIO</b>					<b>78</b>	

**III. PROMEDIO DE VALIDACIÓN**

78

**IV. OPCIÓN DE APLICABILIDAD**

- (X) El instrumento puede ser aplicado, tal como está elaborado.  
 ( ) El instrumento debe ser mejorado antes de ser aplicado.



TEOFILO ROBERTO  
CORREA CALLE

**FIRMA DEL EXPERTO**



Documento de metodología y desarrollo de prototipo de red social para análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de inteligencia artificial.

# DOCUMENTO DE DESARROLLO

**Red social para el análisis de texto y relacionar usuarios por niveles de descubrimiento de lexemas mediante técnicas de Inteligencia Artificial.**



**Autor:**

TORRES DAVILA DAVID ERNESTO



2021

Piura – Perú

## METODOLOGÍA

La tendencia de los desarrollos de tecnología de información, el desarrollo de plataformas basadas en principios de Cloud Computing (Computadora en la nube) que nos ofrecen muchas posibilidades para administrar, mantener, escalar y adaptar a la hora de crear nuevos servicios. el desarrollo e implementación de las plataformas se centró en la aplicación de la metodología de desarrollo de software XP el cual ayudara al desarrollo del sistema en la investigación, esto implica implementar, de forma conjunta y al extremo, prácticas ya conocidas en el ámbito del desarrollo de software, el cual se agruparan las variables involucradas en un proyecto llevado a cabo con XP, que permite analizar el efecto de la Implementación conjunta de las prácticas de dicha metodología y ayuda en la gestión de este tipo de proyectos. El modelo presentado en este trabajo corresponde al subsistema de desarrollo de tareas, uno de los más importantes. Este modelo se ha validado con datos de la entidad que ha ofrecido para poder lograr con cada objetivo, además se ha diseñado y ejecutado una serie de experiencias sobre el mismo y se ha realizado el análisis de sensibilidad de sus variables.

### FASE I: EXPLORACIÓN

#### ROLES DE USUARIO:



Los **usuarios de redes sociales** las utilizan para muchos propósitos: buscar información relacionada con sus intereses, entretenerse, conectar con otras personas y difundir mensajes. ... Como consecuencia de su influencia, han cambiado el comportamiento de las personas y sus hábitos de consumo.

## FASE II: PLANIFICACIÓN

### DIAGRAMA DE VIDA DEL PROYECTO

Ítem	Nombre de Tarea	Duración	Comienzo	Fin	Predecesoras
	<b><u>Proyecto de Prototipo de Red Social de Lexemas</u></b>	57 días	05/08/2021	30/09/2021	1
1	<b><u>Fase I:</u></b> Exploración	10 días	05/08/2021	15/08/2021	2
2	<b><u>Fase II:</u></b> Planificación y Programación de Entrega	15 días	15/08/2021	30/08/2021	3
3	<b><u>Fase III:</u></b> Iteraciones	5 días	30/08/2021	04/09/2021	4
4	<b><u>Fase IV:</u></b> Producción	10 días	04/09/2021	14/09/2021	5
5	<b><u>Fase V:</u></b> Mantenimiento	10 días	14/09/2021	24/09/2021	6
6	<b><u>Fase VI:</u></b> Integración de Proyecto	6 días	24/09/2021	30/09/2021	7

## FASE III: ANÁLISIS

### REQUERIMIENTOS FUNCIONALES

#### Usuario

N°. Requerimiento:	RF1	Interfaz :	Plataforma.
Descripción Corta:	Identificación de Módulos.		
Categoría:	Funcional.	Tabla :	Usuario.
Descripción Detallada	Niveles		
	1. Usuario. a. Buscar usuario registrado e interactuar en la plataforma.		
Términos:	Registros, Usuarios. Segmentación		
Prioridad:	Alta.		

N°. Requerimiento:	RF2	Interfaz:	Acceso
--------------------	-----	-----------	--------



Descripción Corta:	Ingreso al sistema.		
Categoría:	Funcional.	Ubicación:	Usuario.
Descripción Detallada	Proceso		
	1. Ingresa al sistema con un usuario y una clave, las credenciales permiten identificar el tipo de usuario al que pertenece dicho perfil.		
Términos:	Perfil, Credenciales, Usuario y Clave		
Prioridad:	Alta.		

N°. Requerimiento:	RF3	Interfaz:	Acceso
Descripción Corta:	Validar Ingreso al Sistema.		
Categoría:	Funcional.	Tabla :	Usuario.
Descripción Detallada	Mensajes		
	<ol style="list-style-type: none"> <li>1. Validación de Usuarios Vacío.</li> <li>2. Validación de Clave Vacía.</li> <li>3. Validación de Usuario y Clave no Validos.</li> </ol>		
Términos:	Acceso, Usuario y Clave		
Prioridad:	Alta.		

## Publicación

---

N°. Requerimiento:	RF4	Interfaz:	Acceso
Descripción Corta:	Publicación.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Es el artículo que se publicara en el blog de la web y generalmente es mostrado cronológicamente junto al resto de artículos.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RF5	Interfaz:	Acceso
--------------------	-----	-----------	--------

Descripción Corta:	Buscar Publicación.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Se buscaran la publicación según interés del usuario por filtrar.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RF6	Interfaz:	Acceso
Descripción Corta:	Notificación.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Las notificaciones pueden estar caracterizadas como alertas de nuevos mensajes en la plataforma.		
Términos:	Publicación, Segmentación, Red Social, Plataforma, Notificación, Mensajes		
Prioridad:	Alta.		

N°. Requerimiento:	RF7	Interfaz:	Acceso
Descripción Corta:	Segmentación de Intereses del usuario.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Se notificaran el interés relacionados a la segmentación registrada por el usuario.		
Términos:	Publicación, Segmentación, Red Social, Plataforma, Notificación, Mensajes, Segmentación		
Prioridad:	Alta.		

## Comentario

---

N°. Requerimiento:	RF8	Interfaz:	Acceso
Descripción Corta:	Comentar.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Son las opiniones y comentarios de la audiencia en la plataforma son de suma importancia para la imagen de tu marca, ya que te permiten recibir feedback sobre tus productos y servicios, son una fuente de inspiración para nuevos contenidos y ayudan a conocer más a fondo a los usuarios		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RF9	Interfaz:	Acceso
Descripción Corta:	Publicar y Comentar.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Las publicaciones están relacionadas a los comentarios que se registraran en estas para obtener información de los usuarios que interactúan sobre un mismo interés.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

## Amigo

---

N°. Requerimiento:	RF9	Interfaz:	Acceso
Descripción Corta:	Solicitud de amistad.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Es la solicitud que se envía para poder visualizar información del otro contacto e interactuar sobre sus intereses.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RF10	Interfaz:	Acceso
Descripción Corta:	Amigo.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Es sinónimo de contacto, es con quien se interactúa en una publicación.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

## Segmentación

---

N°. Requerimiento:	RF11	Interfaz:	Acceso
Descripción Corta:	Intereses.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Es el registro de interés que se registraran para poder ser notificado sobre estos.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RF12	Interfaz:	Acceso
Descripción Corta:	Notificación de interés.		
Categoría:	Funcional.	Tabla :	Publicación.
Descripción Detallada	Es la alerta que se envía para avisar sobre una publicación registrada con un mismo interés de los usuarios.		
Términos:	Publicación, Segmentación, Red Social, Plataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RF13	Interfaz:	Grupos
Descripción Corta:	Grupos de interés.		
Categoría:	Funcional.	Tabla :	Gustos.
Descripción Detallada	Permitirá crear grupos de interés según publicaciones de lexemas relacionados con los usuarios.		
Términos:	Grupos, lexemas, usuarios		
Prioridad:	Alta.		

## REQUERIMIENTOS NO FUNCIONALES

N°. Requerimiento:	RNF1	Sistema O:	Multiplataforma
Descripción Corta:	MySQL		
Categoría:	No Funcional.	Versión:	
Descripción Detallada	MySQL es un sistema de gestión de bases de datos relacional desarrollado bajo licencia dual: Licencia pública general/Licencia comercial por Oracle Corporation y está considerada como la base de datos de código abierto más popular del mundo, <sup>12</sup> y una de las más populares en general junto a Oracle y Microsoft SQL Server, sobre todo para entornos de desarrollo web.		
Términos:	Gestión, MySQL, Licencia, Multiplataforma		
Prioridad:	Alta.		

N°. Requerimiento:	RNF2	Sistema O:	Multiplataforma
Descripción Corta:	MySQL		
Categoría:	No Funcional.	Versión:	7.3.9
Descripción Detallada	XAMPP es una distribución de Apache completamente gratuita y fácil de instalar que contiene MariaDB, PHP y Perl. El paquete de instalación de XAMPP ha sido diseñado para ser increíblemente fácil de instalar y usar.		
Términos:	MariaDB, PHP y Perl,		
Prioridad:	Alta.		

N°. Requerimiento:	RNF3	Sistema O:	Multiplataforma
Descripción Corta:	phpMyAdmin		
Categoría:	No Funcional.	Versión:	5.6.44
Descripción Detallada	phpMyAdmin es una herramienta escrita en PHP con la intención de manejar la administración de MySQL a través de páginas web, utilizando un navegador web.		
Términos:	MySQL, páginas web y navegador		
Prioridad:	Alta.		

N°. Requerimiento:	RNF4	Sistema O:	Multiplataforma
Descripción Corta:	PHP		
Categoría:	No Funcional.	Versión:	7.2.7
Descripción Detallada	PHP, acrónimo recursivo en inglés de PHP: Hypertext Pre proceso (preprocesador de hipertexto), es un lenguaje de programación de propósito general de código del lado del servidor originalmente diseñado para el pre procesamiento de texto plano en UTF-8. Posteriormente se aplicó al desarrollo web de contenido dinámico, dando un paso evolutivo en el concepto de aplicación en línea, por su carácter de servicio.		
Términos:	Recursivo, servicio, evolutivo, programación y pre procesado		
Prioridad:	Alta.		

N°. Requerimiento:	RNF5	Sistema O:	Multiplataforma
Descripción Corta:	JavaScript		
Categoría:	No Funcional.	Versión:	
Descripción Detallada	JavaScript (abreviado comúnmente JS) es un lenguaje de programación interpretado, dialecto del estándar ECMAScript. Se define como orientado a objetos, <sup>3</sup> basado en prototipos, imperativo, débilmente tipado y dinámico.		
Términos:	Prototipos, imperativo, débilmente tipado y dinámico.		
Prioridad:	Alta.		

N°. Requerimiento:	RNF6	Sistema O:	Multiplataforma
Descripción Corta:	HTML		
Categoría:	No Funcional.	Versión:	
Descripción Detallada	HTML es un lenguaje de marcado que se utiliza para el desarrollo de páginas de Internet. Se trata de las siglas que corresponden a HyperText Markup Lenguaje, es decir, Lenguaje de Marcas de Hipertexto.		
Términos:	HyperText Markup y Internet.		
Prioridad:	Alta.		

N°. Requerimiento:	RNF7	Sistema O:	Multiplataforma
Descripción Corta:	AJAX		
Categoría:	No Funcional.	Versión:	
Descripción Detallada	AJAX, acrónimo de Asynchronous JavaScript And XML (JavaScript asíncrono y XML), es una técnica de desarrollo web para crear aplicaciones interactivas o RIA (Rich Internet Applications). Estas aplicaciones se ejecutan en el cliente, es decir, en el navegador de los usuarios mientras se mantiene la comunicación asíncrona con el servidor en segundo plano.		
Términos:	Acrónimo, interactividad, RIA y aplicaciones.		
Prioridad:	Alta.		

## DICCIONARIO DE DATOS

Estructura de tabla para la tabla amigo

Columna	Tipo	Nulo	Predeterminado	Comentarios
IdAmigo	int(11)	No		
IdUsuario1	int(11)	No		
IdUsuario2	int(11)	No		
Estado	int(11)	No		

Estructura de tabla para la tabla comentario

Columna	Tipo	Nulo	Predeterminado	Comentarios
IdComentario	int(11)	No		
IdPublicacion	int(11)	No		
IdUsuario	int(11)	No		
Comentario	text	No		
FechaHoraRegistro	timestamp	No	CURRENT_TIMESTAMP	

Estructura de tabla para la tabla publicación

Columna	Tipo	Nulo	Predeterminado	Comentarios
IdPublicacion	int(11)	No		
Titulo	text	No		
Hora	timestamp	No	CURRENT_TIMESTAMP	
Estado	char(1)	No		
IdUsuario	int(11)	No		

Estructura de tabla para la tabla segmentación

Columna	Tipo	Nulo	Predeterminado	Comentarios
IdSegmentacion	int(11)	No		
IdUsuario	int(11)	No		
Descripción	varchar(100)	No		

Estructura de tabla para la tabla usuario

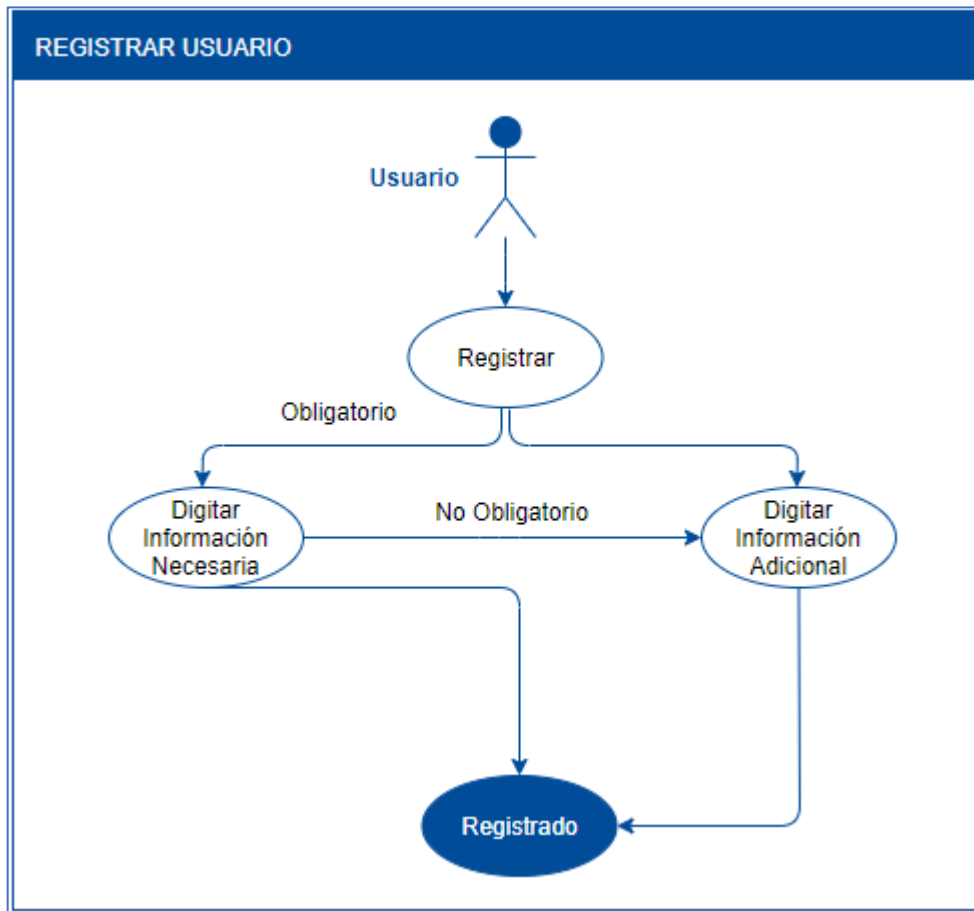
Columna	Tipo	Nulo	Predeterminado	Comentarios
IdUsuario	int(11)	No		
Nombre	varchar(20)	No		
Apellido	varchar(20)	No		
Alias	varchar(20)	Sí	NULL	
Clave	varchar(255)	No		
Correo	varchar(255)	No		
Genero	char(1)	No		
FechaNacido	date	No		
Estado	char(1)	Sí	NULL	
Detalle	text	Sí	NULL	
PuebloNatal	varchar(255)	Sí	NULL	

Estructura de tabla para la tabla gustos

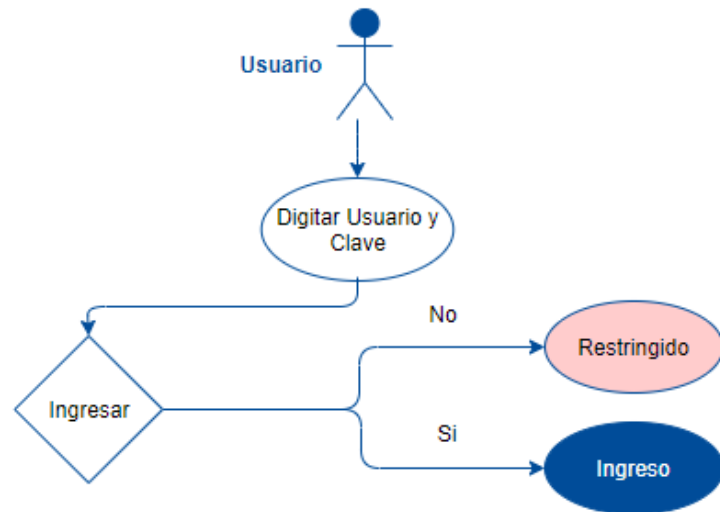
Columna	Tipo	Nulo	Predeterminado	Comentarios
IdPalabra	int(11)	No		
IdUsuario	int(11)	No		
Descripcion	varchar(100)	No		
Estado	int(11)	Sí	NULL	
FechaHoraRegistro	timestamp	No		
Ítem	varchar(15)	Si	NULL	
Contador	int(11)	No		



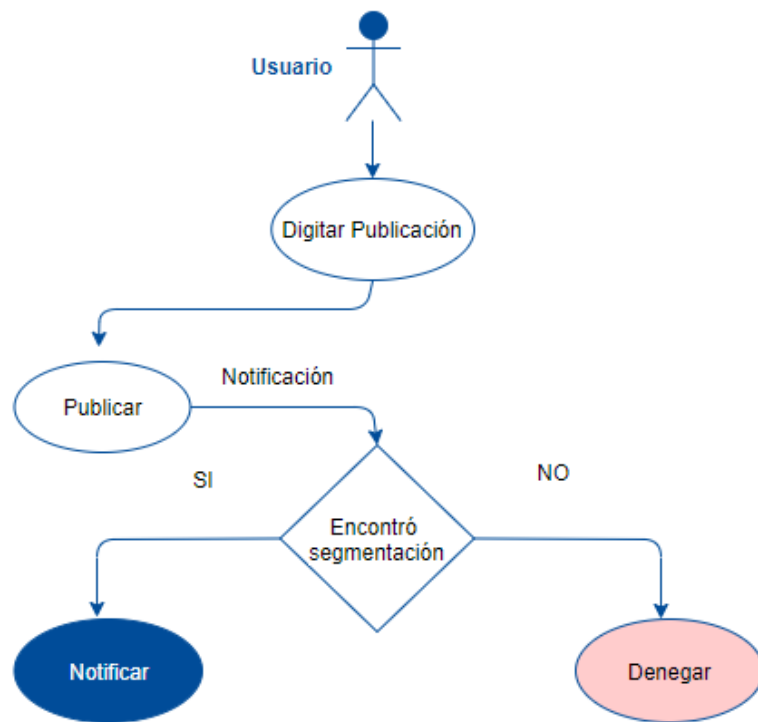
## CASOS DE PROCESOS

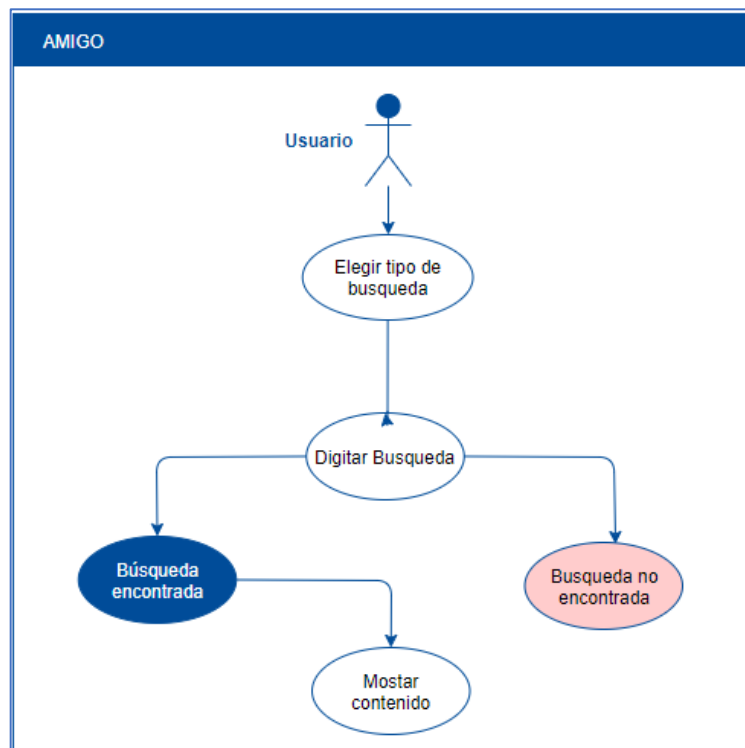
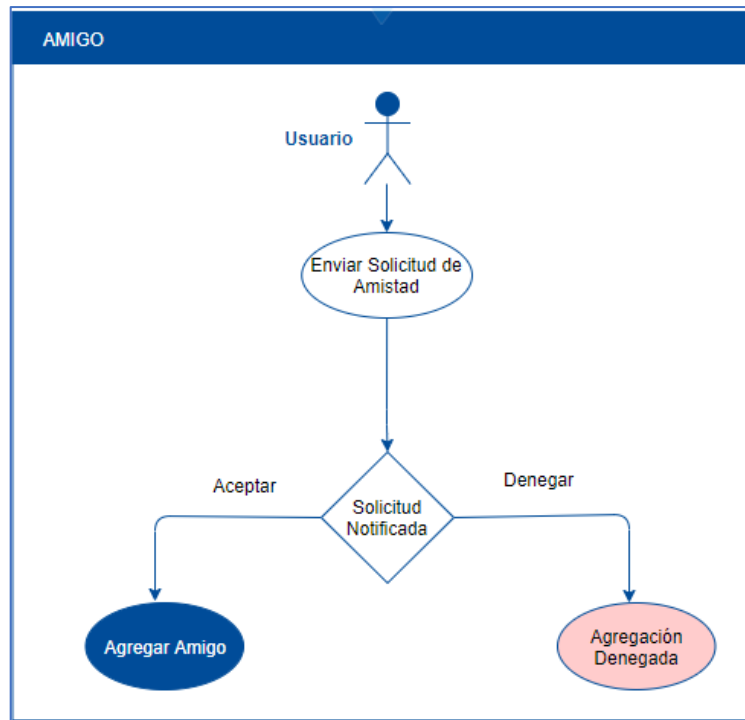


## INGRESAR

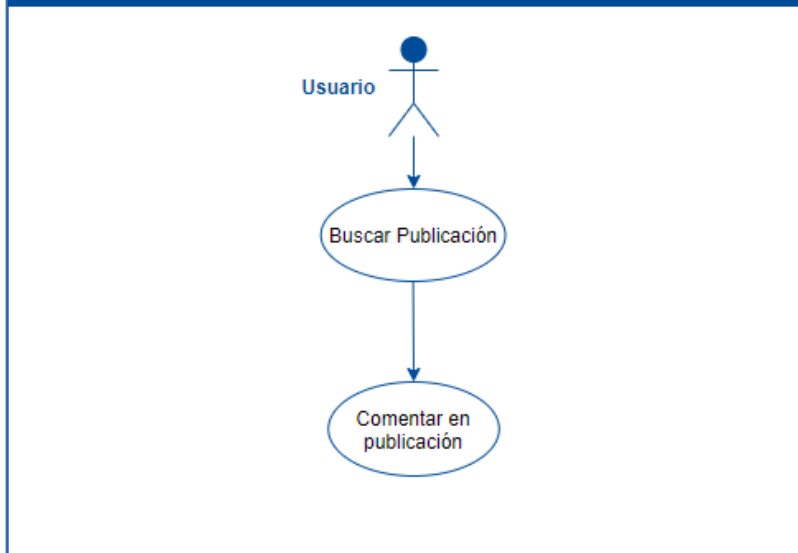


## PUBLICACIÓN

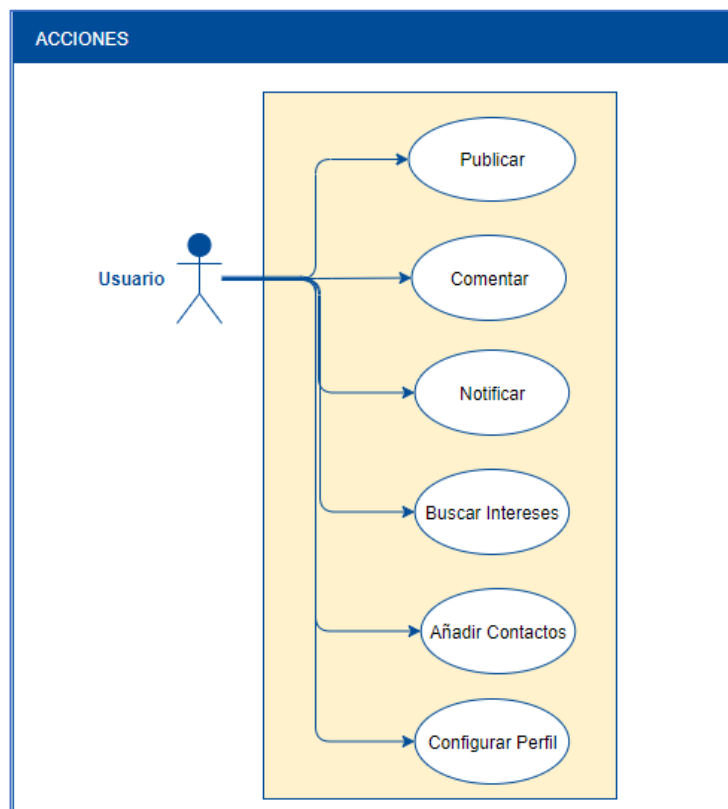




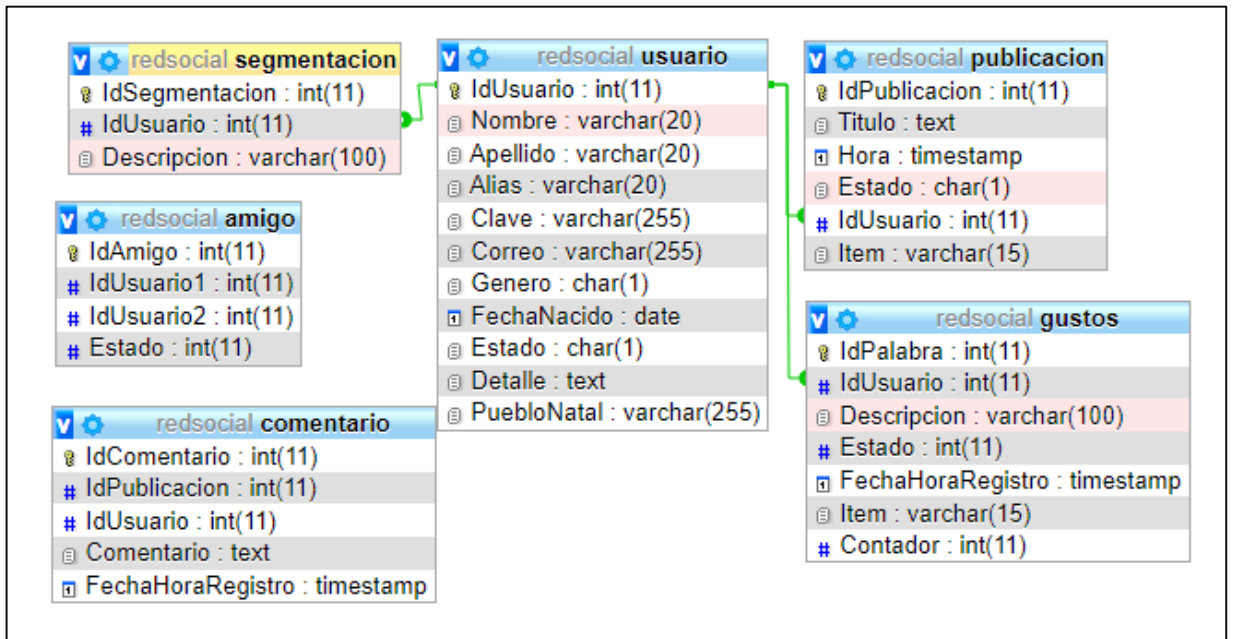
## COMENTARIO



## CASOS DE USO



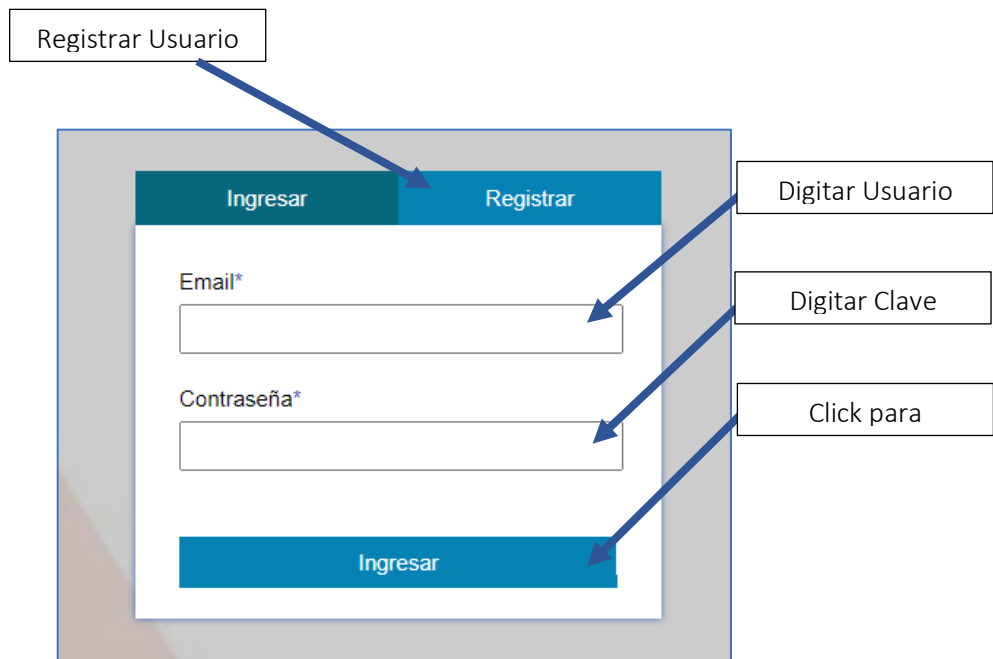
## DIAGRAMA DE FLUJO



### FASE IV: INTERFACES

#### ACCESO

Ingreso del usuario a los sistemas.



## REGISTRAR USUARIO

Registrar usuario en el sistema.

The form is divided into two tabs: 'Ingresar' and 'Registrar'. The 'Registrar' tab is selected. The form is organized into two main sections: 'Información Necesaria' and 'Información Adicional'.

**Información Necesaria:**

- Nombre\*
- Apellidos\*
- Apodos
- Clave\*
- Confirmar Clave\*

**Información Adicional:**

- Correo\*
- Fecha Nacimiento\* (1 | Enero | 1996)
- Hombre  Mujer
- Pueblo Natal
- Solter@  Comprometid@  Casad@
- Sobre mi

**Crear Cuenta**

A callout box labeled 'Crear Cuenta' points to the 'Crear Cuenta' button at the bottom right of the form.

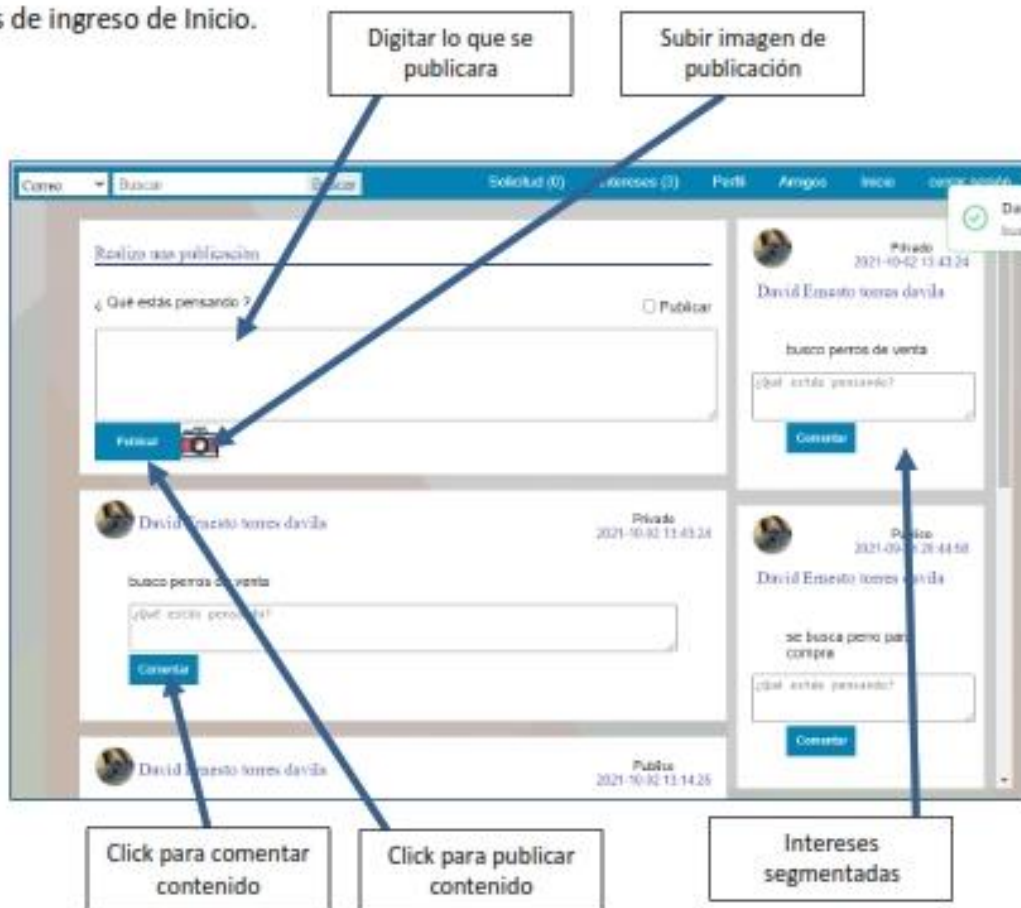
### 4.3 OPCIONES DE USUARIO

Opciones de ingreso de Usuario.



### 4.4 INICIO

Opciones de ingreso de Inicio.



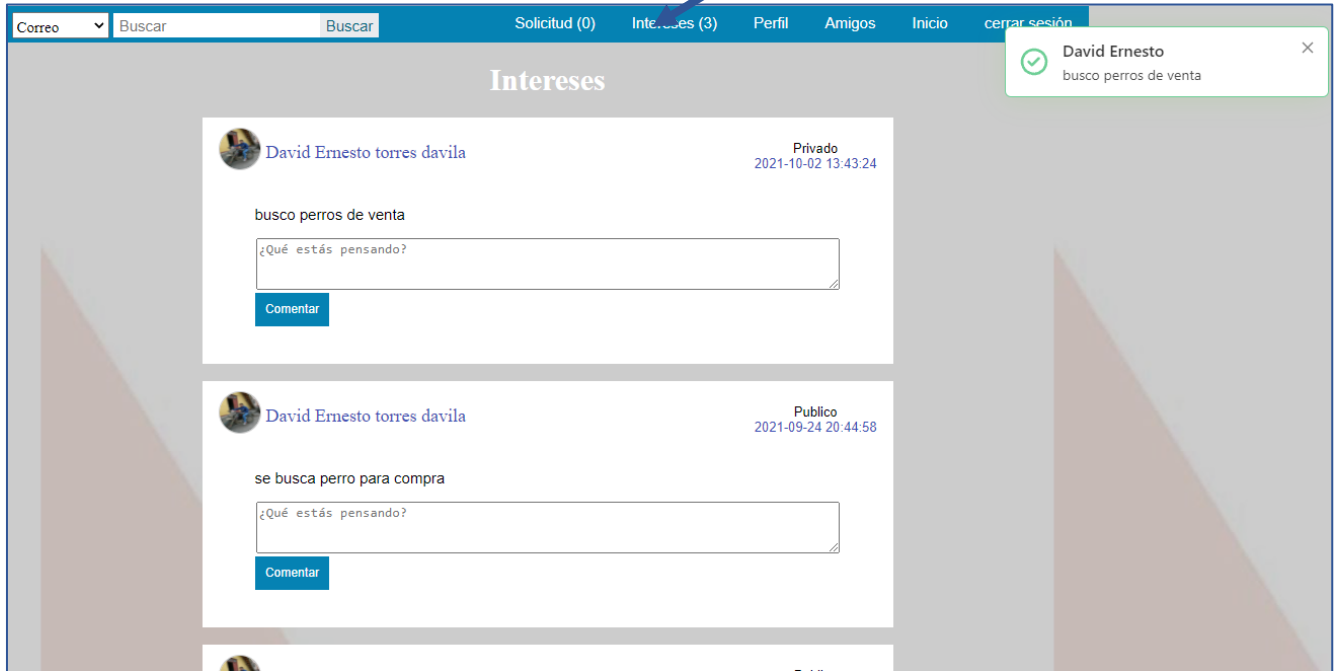




# INTERESES

Opciones de contenido de intereses.

Ventana de  
interese



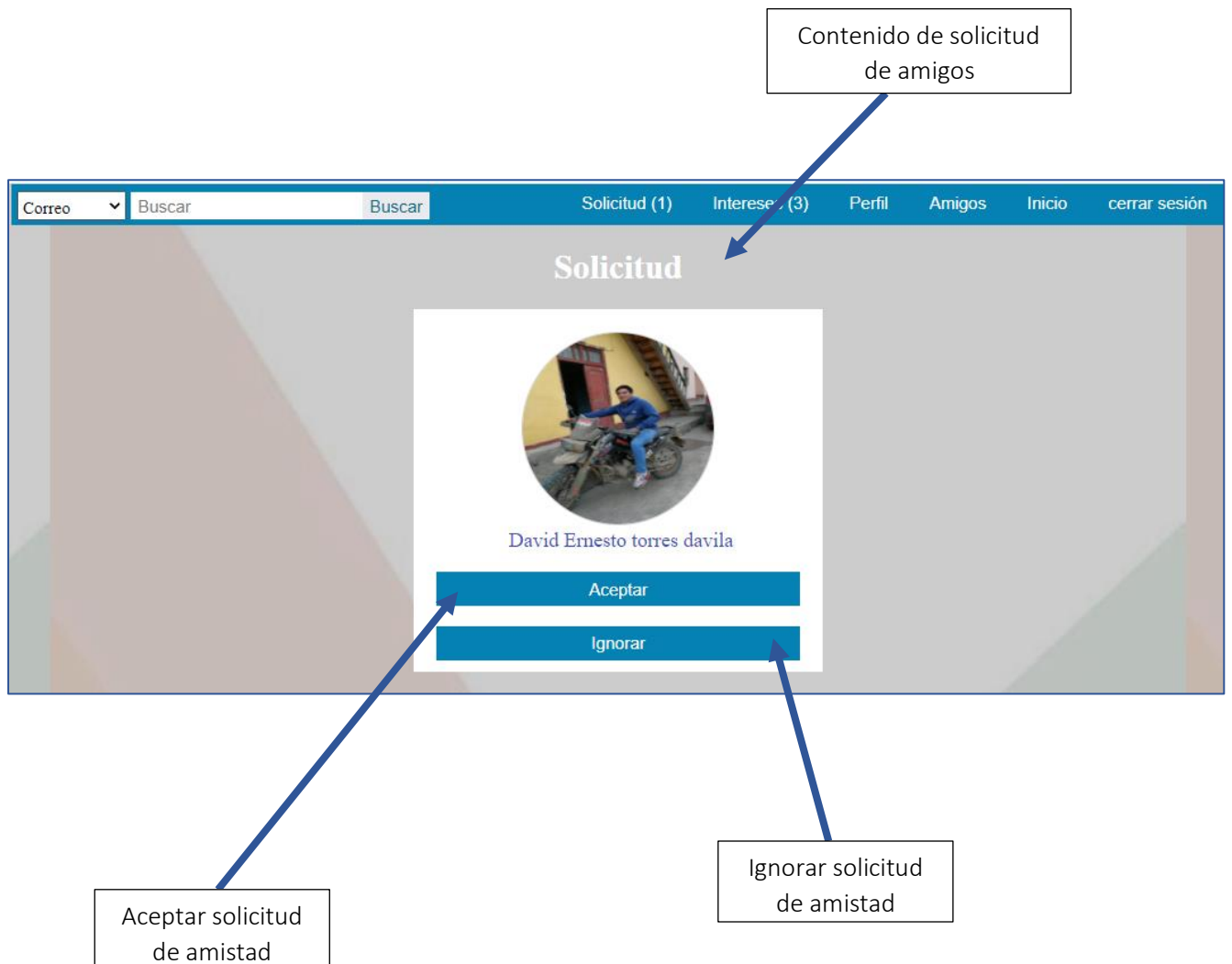
## AMIGOS

Opciones de contenido de amigos.



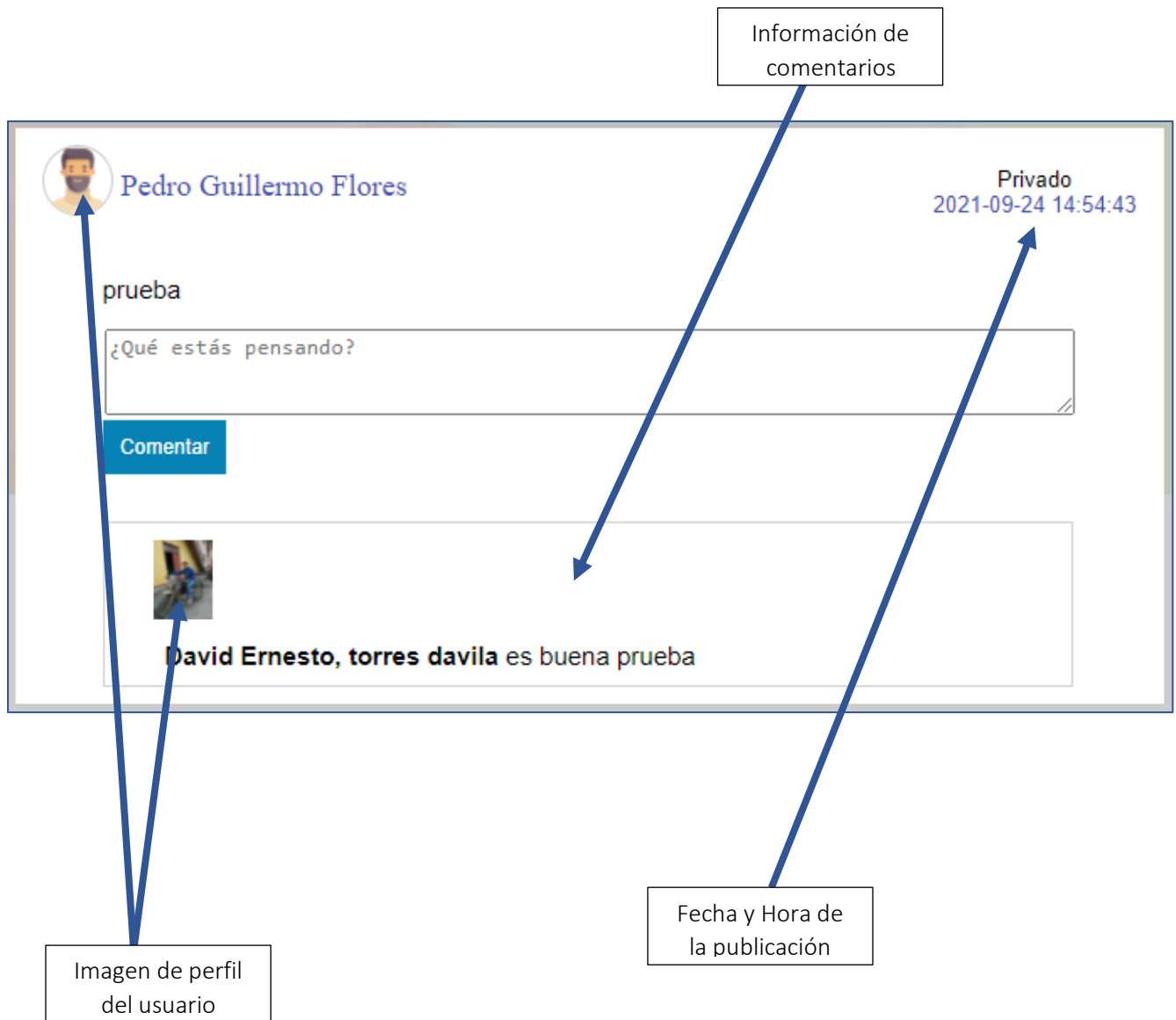
## ACEPTAR SOLICITUD DE AMIGO

Opciones de aceptar solicitud de amigos.



## COMENTARIOS

Información de comentarios de las publicaciones.



## INTERESES

Son los intereses que gustan del usuario.

Segmentación de gustos

The screenshot displays a social media profile page with the following elements:

- Gustos Asociados:** A section titled "Gustos Asociados" containing two blue buttons: "TIBURONES" and "GRUPOS". A blue arrow points from the "Segmentación de gustos" box to the "TIBURONES" button.
- Publicaciones:** A bracket on the left side groups two posts under the label "Publicaciones".
- Post 1:** By Pedro Guillermo Flores, dated 2021-11-30 11:58:38 (Privado). The text is "los tiburones son interesantes". Below the text is a red button labeled "TIBURONES".
- Post 2:** By Kelly Guarizo, dated 2021-11-30 11:58:25 (Privado). The text is "los tiburones son agresivos en ambiente hostil". Below the text are four red buttons: "TIBURONES", "AGRESIVOS", "AMBIENTE", and "HOSTIL". A blue arrow points from the "Segmentación de publicación" box to the "AGRESIVOS" button.

Publicaciones

Segmentación de publicación

## GRUPOS

Son los grupos formados por los usuarios según gustos preferenciales publicadas.

GRUPOS DE INTERESES FORMADOS

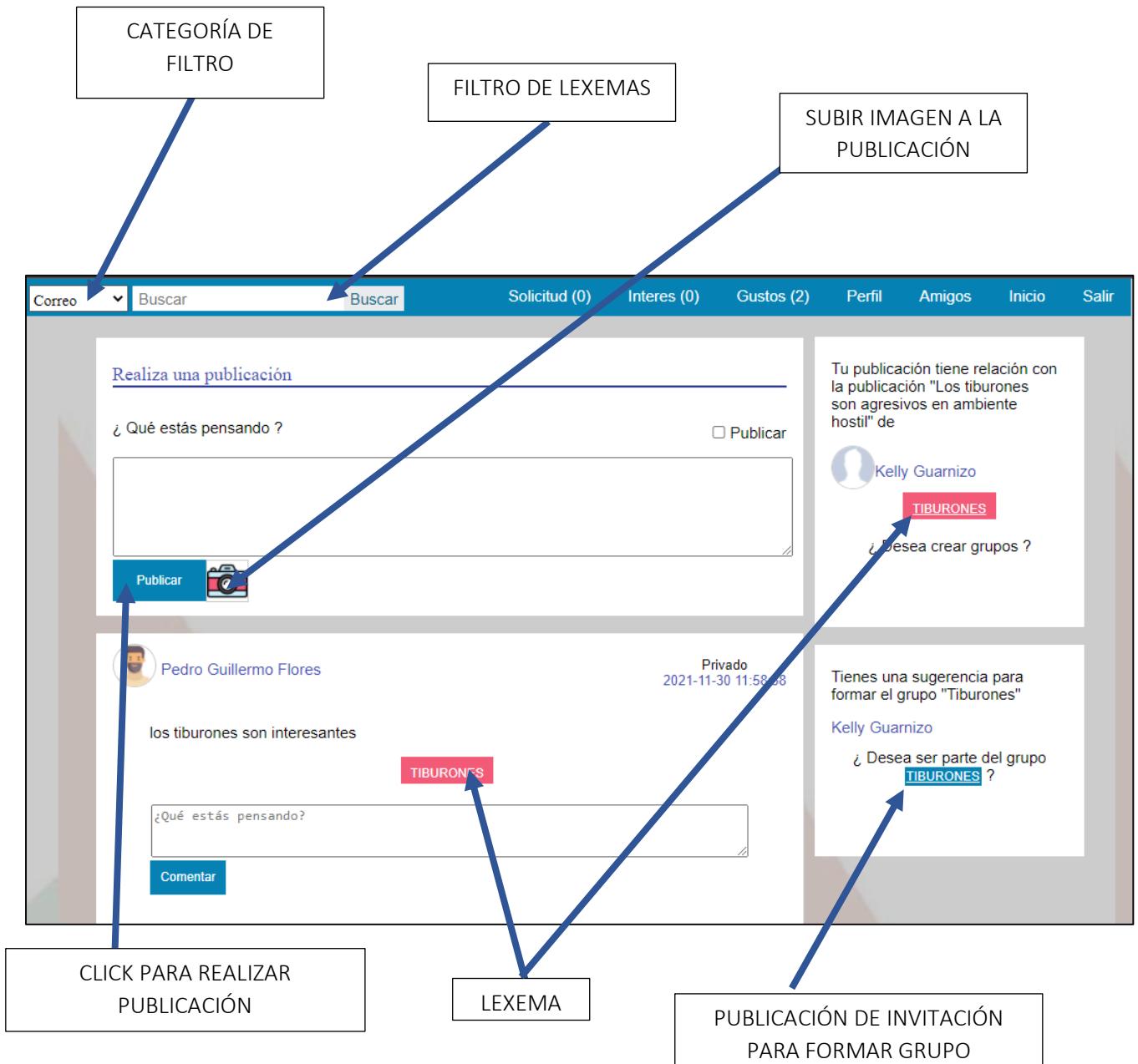
The screenshot shows a social media interface with a blue header bar containing navigation options: 'Correo', 'Buscar', 'Solicitud (0)', 'Interes (0)', 'Gustos (2)', 'Perfil', 'Amigos', 'Inicio', and 'Salir'. The main content area is titled 'Grupos de Interes' and features a group named 'TIBURONES'. Below the group name, it lists members: 'FLORES PEDRO GUILLERMO / PEDRO@GMAIL.COM' and 'GUARNIZO KELLY / KELLY\_100@GMAIL.COM'. The 'Publicaciones de Grupo' section shows a post by 'Pedro Guillermo Flores' with the text 'los tiburones son interesantes' and a comment input field with a 'Comentar' button. The post is marked as 'Privado' and dated '2021-11-30 11:58:38'.

LISTA DE INTEGRANTES DEL GRUPO

PUBLICACIÓN DE INTEGRANTE DEL

# INICIO

Muestra la publicación con los lexemas encontrados en cada publicación y sus intereses al lado de las publicaciones.



## COMENTARIO

Son los comentarios que se le realizan a las publicaciones.

LEXEMA

PUBLICACIÓN

Pedro Guillermo Flores

Privado  
2021-11-30 19:09:04

PEDRO TIENE 2 PROPUESTAS DE NEGOCIO PARA MEJORAR LA ECONOMIA DEL PAIS

PEDRO PROPUESTAS NEGOCIO ECONOMIA PAIS

¿Qué estás pensando?

Comentar

Kelly, Guarnizo / 2021-11-30 19:14:30  
genial

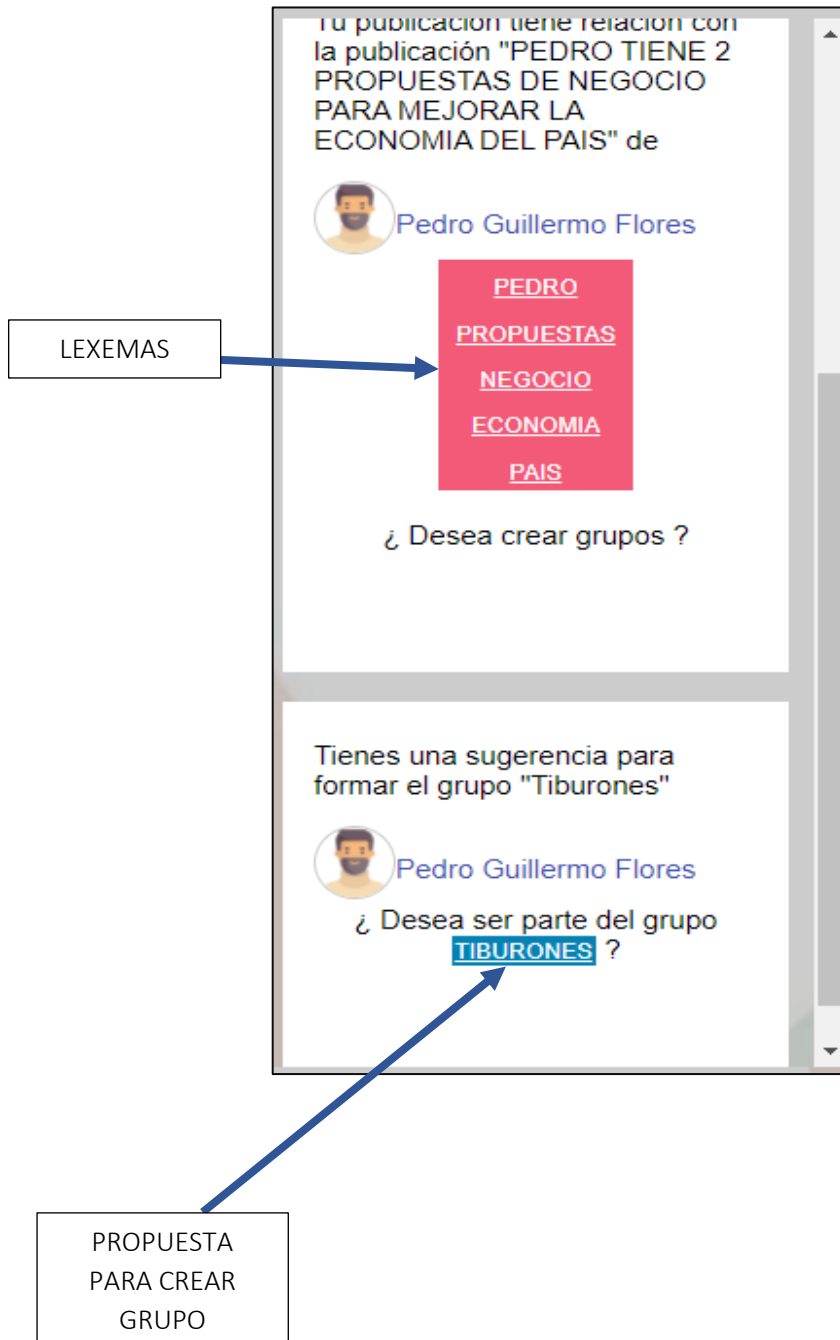
Kelly, Guarnizo / 2021-11-30 19:14:38  
me parece bien

COMENTARIO

CLICK PARA COMENTAR PUBLICACIÓN

## PUBLICACIONES DE INTERES Y NOTIFICACIONES DE UNIÓN A GRUPOS

Son las publicaciones relacionadas al lexema del usuario que publica un mismo interés.





LEXEMAS  
RELACIONAD

The screenshot displays a social media interface with two main sections. The top section is a post by Pedro Guillermo Flores, titled "Realiza una publicación". It contains a text input field with the placeholder "¿Qué estás pensando?", a "Publicar" button, and a camera icon. Below the post, there is a comment section with a profile picture of Pedro Guillermo Flores, the name "Pedro Guillermo Flores", a "Privado" status, and a timestamp "2021-11-30 19:09:04". The comment text reads "PEDRO TIENE 2 PROPUESTAS DE NEGOCIO PARA MEJORAR LA ECONOMIA DEL PAIS". Below the comment, there are five red tags: "PEDRO", "PROPUESTAS", "NEGOCIO", "ECONOMIA", and "PAIS". A text input field with the placeholder "¿Qué estás pensando?" and a "Comentar" button are also present.

The bottom section of the screenshot shows a group invitation. It features a profile picture of Pedro Guillermo Flores, the name "Pedro Guillermo Flores", and the text "¿Desea ser parte del grupo TIBURONES?". Above this, there is a notification: "Tienes una sugerencia para formar el grupo 'Tiburones'". To the left of this section, there is a notification from Kelly, Guarnizo: "Kelly, Guarnizo me parece bien".

INVITACIÓN A  
UNIRSE AL  
GRUPO