



**UNIVERSIDAD NACIONAL DEL ALTIPLANO**  
**ESCUELA DE POSGRADO**  
**DOCTORADO EN ESTADÍSTICA APLICADA**



**TESIS**

**TEXT MINING EN LA REPRESENTACIÓN SOCIAL SOBRE LA  
CONTAMINACIÓN DEL LAGO TITICACA, PUNO 2020**

**PRESENTADA POR:**

**RENZO APAZA CUTIPA**

**PARA OPTAR EL GRADO ACADÉMICO DE:**

**DOCTOR EN ESTADÍSTICA APLICADA**

**PUNO, PERÚ**

**2020**

## Reporte de similitud

NOMBRE DEL TRABAJO

**TEXT MINING EN LA REPRESENTACIÓN  
SOCIAL SOBRE LA CONTAMINACIÓN DE  
L LAGO TITICACA, PUNO 2020**

AUTOR

**RENZO APAZA CUTIPA**

RECUENTO DE PALABRAS

**23845 Words**

RECUENTO DE CARACTERES

**132135 Characters**

RECUENTO DE PÁGINAS

**96 Pages**

TAMAÑO DEL ARCHIVO

**1.5MB**

FECHA DE ENTREGA

**Feb 8, 2024 12:02 PM GMT-5**

FECHA DEL INFORME

**Feb 8, 2024 12:03 PM GMT-5**


### ● 13% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

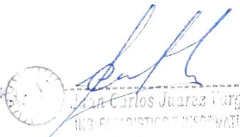
- 13% Base de datos de Internet
- Base de datos de Crossref
- 5% Base de datos de trabajos entregados
- 2% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

### ● Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Coincidencia baja (menos de 12 palabras)

  
-----  
**José P. Tito Lipa**  
Ing. Estadístico e Informático D.Sc.  
CIP. 159645



  
-----  
Carlos Juárez Vargas  
ING. ESTADÍSTICO E INFORMÁTICO  
CIP. 159645

Resumen



# UNIVERSIDAD NACIONAL DEL ALTIPLANO

## ESCUELA DE POSGRADO

### DOCTORADO EN ESTADÍSTICA APLICADA

#### TESIS

TEXT MINING EN LA REPRESENTACIÓN SOCIAL SOBRE LA  
CONTAMINACIÓN DEL LAGO TITICACA, PUNO 2020

PRESENTADA POR:

RENZO APAZA CUTIPA

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTOR EN ESTADÍSTICA APLICADA



APROBADA POR EL JURADO SIGUIENTE:

PRESIDENTE

.....  
Dr. EDMUNDO GERARDO MORENO TERRAZAS

PRIMER MIEMBRO

.....  
Dr. BERNABE CANQUI FLORES

SEGUNDO MIEMBRO

.....  
Dr. JUAN REYNALDO PAREDES QUISPE

ASESOR DE TESIS

.....  
D.Sc. JOSÉ PANFILO TITO LIPA

Puno, 20 de diciembre de 2022

**ÁREA:** Minería de Datos.

**TEMA:** Text Mining en la Representación Social Sobre la Contaminación del Lago Titicaca, Puno 2020

**LÍNEA:** Base de Datos.



## DEDICATORIA

A los docentes y compañeros del programa de doctorado en Estadística Aplicada de la Escuela de Posgrado de la Universidad Nacional del Altiplano por compartir sus experiencias y conocimientos para brindar aportes científicos que ayuden a la sociedad.



## AGRADECIMIENTOS

Agradezco de forma muy especial a la Universidad nacional del altiplano por financiar esta tesis de investigación doctoral como parte de la ejecución de los megaproyectos de investigación científica multidisciplinaria con el proyecto ganador “REPRESENTACIONES SOCIALES SOBRE CONTAMINACIÓN DE LA BAHÍA DEL LAGO TITICACA” del “CONCURSO DE PROYECTOS DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA MULTIDISCIPLINARIA – 2016, II FASE UNA PUNO” dando conformidad de lo dispuesto mediante Resolución Rectoral N° 2410-2016-R-UNA y N° 0960-2017-R-UNA.



## ÍNDICE GENERAL

	<b>Pág.</b>
DEDICATORIA	i
AGRADECIMIENTOS	ii
ÍNDICE GENERAL	iii
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS	viii
ÍNDICE DE ANEXOS	x
RESUMEN	xi
ABSTRACT	xii
INTRODUCCIÓN	1

### CAPÍTULO I

#### REVISIÓN DE LITERATURA

1.1 Marco teórico	2
1.1.1 La Cuenca del Lago Titicaca	2
1.1.2 El lago Mayor o Chucuito	4
1.1.3 Bahía Mayor de Puno	4
1.1.4 Bahía Interior de Puno	4
1.1.5 El lago menor o Huiñaymarca	5
1.1.6 Fuentes contaminantes de la cuenca del lago Titicaca	5
1.1.7 Aguas residuales municipales.	5
1.1.8 Aguas residuales y contaminantes emergentes.	7
1.1.9 Residuos sólidos peligrosos.	7
1.1.10 Residuos sólidos hospitalarios	9
1.1.11 Las letrinas y la contaminación del agua	9
1.1.12 Teoría de las Representaciones Colectivas	9
1.1.13 Teoría de las Representaciones Sociales	10
1.1.14 Proceso de conformación de una Representación Social	11
1.1.15 Proceso de objetivación	11
1.1.16 El proceso de anclaje	12
1.1.17 Tipos de Representaciones sociales	12
1.1.18 Estructura de las Representaciones Sociales	13
1.1.19 Social media	14



1.1.20	Plataforma Social media de Twitter	16
1.1.21	Minería de Social media	18
1.1.22	Minería en Grafos	19
1.1.23	Minería de Textos	20
1.1.24	Proceso General de Minería de textos en social media	22
1.1.25	Extracción Rápida y Automática de Palabras Clave - RAKE	22
1.2	Antecedentes	23
1.2.1	Estudios sobre representaciones sociales y medio ambiente	23
1.2.2	Estudios sobre la contaminación del Lago Titicaca	26
1.2.3	Estudios sobre Social media y Medio Ambiente	27

## **CAPÍTULO II**

### **PLANTEAMIENTO DEL PROBLEMA**

2.1	Identificación del problema	29
2.2	Enunciados del problema	30
2.3	Justificación	31
2.4	Objetivos	31
2.4.1	Objetivo general	31
2.4.2	Objetivos específicos	31
2.5	Hipótesis	32
2.5.1	Hipótesis general	32
2.5.2	Hipótesis específicas	32

## **CAPÍTULO III**

### **MATERIALES Y MÉTODOS**

3.1	Lugar de estudio	33
3.2	Población	34
3.3	Muestra	34
3.4	Método de investigación	34
3.5	Descripción de métodos por objetivos	34
3.5.1	Representaciones sociales expresadas en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca	34
3.5.2	Conocimientos evidenciados en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.	36



3.5.3 Actitudes y prácticas expresadas sobre la plataforma de microbloggings Twitter sobre la contaminación del Lago Titicaca.	36
---	----

## **CAPÍTULO IV**

### **RESULTADOS Y DISCUSIÓN**

4.1. Representaciones sociales expresadas en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.	38
4.1.1 Autenticación mediante protocolo OAuth 2.0 en Twitter.	38
4.1.2 Adquisición de tweets	39
4.1.3 Construcción de base de datos	39
4.1.4 Procesamiento del Lenguaje Natural de la representación social Twitter acerca de la contaminación del Lago Titicaca.	40
4.1.5 Modelado de la representación social sobre la Plataforma Twitter.	40
4.1.6 Modelo Obtenido sobre sustantivos	41
4.1.7 Modelo Obtenido sobre adjetivos y sustantivos	41
4.1.8 Modelo Obtenido Sobre sustantivos, adjetivos y Palabras Clave	42
4.1.9 Discusión sobre las representaciones sociales obtenidas	49
4.2. Conocimientos evidenciados en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.	51
4.3. Actitudes y prácticas expresadas sobre la plataforma de microbloggings Twitter sobre la contaminación del Lago Titicaca.	56
<b>CONCLUSIONES</b>	<b>63</b>
<b>RECOMENDACIONES</b>	<b>64</b>
<b>BIBLIOGRAFÍA</b>	<b>65</b>
<b>ANEXOS</b>	<b>72</b>





## ÍNDICE DE TABLAS

	<b>Pág.</b>
1. Tamaño del sistema TDPS (adaptado de World Water Assessment Programme United Nations (2003))	3
2. Características Físico Químicas de las aguas residuales municipales y sus efectos en un cuerpo de agua	6
3. Clasificación de los residuos peligrosos domésticos urbanos	8
4. Clasificación de social media por presencia social / riqueza de medios y autopresentación / autodivulgación.	15
5. Terminología de Twitter (adaptado de Mollett et al. (2011))	17
6. Ejemplo de tweets	18
7. Etiquetas de búsqueda sobre la plataforma Twitter en idioma inglés y español	39
8. Nombre y tipo de variable de la base de datos	40
9. Agrupación de temas obtenidos mediante el algoritmo LDA modelando los sustantivos.	41
10. Agrupación de temas obtenidos mediante el algoritmo LDA modelando adjetivos y sustantivos.	42
11. Agrupación de temas obtenidos mediante el algoritmo LDA modelando adjetivos, sustantivos y palabras clave de tweets acerca del lago Titicaca parte 1.	43
12. Agrupación de temas obtenidos mediante el algoritmo LDA modelando adjetivos, sustantivos y palabras clave de tweets acerca del lago Titicaca parte 2	43
13. Sustantivos más utilizados ordenados por mayor frecuencia	52
14. puntaje de palabras clave identificadas con el algoritmo Rake.	54
15. Adjetivos con más ocurrencia ordenados por frecuencia	57



<b>16.</b> Frecuencia de coocurrencias de sustantivos y adjetivos utilizados en la misma oración	58
<b>17.</b> Correlaciones de Pearson para las palabras clave sobre matrices dispersas	60
<b>18.</b> Agrupación de temas obtenidos mediante el algoritmo LDA modelado basado en los verbos de los tweets	61



## ÍNDICE DE FIGURAS

	<b>Pág.</b>
1. Proceso de conformación de una representación social	12
2. Tipos de Representaciones Sociales	13
3. Grafo de la red palabras sobre los tweets usando el hashtag climate change	20
4. Proceso General de minería de textos en social media.	22
5. Lago Titicaca.	33
6. El proceso de autenticación implementado mediante el protocolo OAuth 2.0	38
7. Grafo de coocurrencias sobre el tópico/tema1 construido sobre sustantivos adjetivos y palabras clave.	44
8. Grafo de coocurrencias sobre el tópico/tema2 construido sobre sustantivos adjetivos y palabras clave.	45
9. Grafo de coocurrencias sobre el tópico/tema 3 construido sobre sustantivos adjetivos y palabras clave	46
10. Grafo de coocurrencias sobre el tópico/tema 4 construido sobre sustantivos adjetivos y palabras clave	47
11. Grafo de coocurrencias sobre el tópico/tema 5 construido sobre sustantivos adjetivos y palabras clave	48
12. Grafo de coocurrencias sobre el tópico/tema 6 construido sobre sustantivos adjetivos y palabras clave	49
13. Gráfico de barras de los Sustantivos más utilizados ordenados por mayor frecuencia	53
14. Gráfico de barras para la identificación de palabras clave con el algoritmo Rake.	55
15. Nube de palabras de las palabras clave referente a la contaminación del Lago Titicaca con el algoritmo Rake.	56



**16. Grafo de frecuencia de coocurrencias de sustantivos y adjetivos utilizados en la misma oración**

59



## ÍNDICE DE ANEXOS

	<b>Pág.</b>
<b>1. Código fuente de Análisis Estadístico</b>	<b>72</b>



## RESUMEN

Esta investigación, basada en la teoría de las representaciones sociales, tuvo como objetivo identificar el contenido textual que refleja las representaciones sociales sobre la contaminación del Lago Titicaca, el lago más grande del Perú. Utilizando técnicas de minería de texto (Text Mining), se enfocó en las etapas de recolección de datos, extracción de características y clasificación de representaciones sociales, las cuales se expresaron en términos de conocimientos, actitudes y prácticas relacionadas con la contaminación del lago. Se creó una base de datos utilizando la plataforma de microblogging de Twitter, aplicando métodos de text mining para la colección, extracción de características y clasificación. El estudio, de diseño no experimental, transversal, explicativo y correlacional, reveló que las representaciones sociales sobre la contaminación del Lago Titicaca pueden clasificarse en tipos antropocéntrico, globalizante y antropocéntrico cultural. Además, se encontró que los conocimientos acerca de la contaminación del lago varían considerablemente. Las actitudes registradas fueron mayoritariamente positivas en cuanto a la responsabilidad de recuperar y conservar el lago, mientras que las prácticas expresadas resultaron ser heterogéneas, evidenciadas por una amplia variedad de verbos en los conjuntos de datos.

**Palabras clave:** Actitudes y prácticas, conocimientos, contaminación, Lago Titicaca, Minería de textos, procesamiento natural del lenguaje, representaciones sociales



## ABSTRACT

The research, grounded in social representation theory, aimed to identify textual content reflecting social representations of pollution in Lake Titicaca, Peru's largest lake. Using text mining techniques, it focused on data collection, feature extraction, and classification stages of social representations, expressed in terms of knowledge, attitudes, and practices related to the lake's pollution. A database was created using the microblogging platform of Twitter, applying text mining methods for collection, feature extraction, and classification. The study, non-experimental, cross-sectional, explanatory, and correlational in design, revealed that social representations of Lake Titicaca's pollution can be classified into anthropocentric, globalizing, and cultural anthropocentric types. It also found that knowledge about the lake's pollution varies considerably. The recorded attitudes were predominantly positive regarding the responsibility to recover and conserve the lake, while the expressed practices turned out to be heterogeneous, evidenced by a wide variety of verbs in the data sets.

**Keywords:** Attitudes and practices, knowledge, Lake Titicaca, Natural Language Processing, pollution, Social representations, Text mining.

## INTRODUCCIÓN

La contaminación del medio ambiente es el problema más álgido que enfrenta la humanidad, provocada en gran medida por las actividades del hombre, actividades tales como la industria minera, la industria de la construcción, la industrialización y la explotación desmedida de los recursos naturales, es decir por las actividades humanas a pequeña y a gran escala, que son las principales causas de este panorama preocupante.

Los recursos hídricos no son ajenos a esta problemática, siendo estos uno de los recursos ambientales más afectados. Así, la contaminación de los cuerpos de agua dulce son una amenaza mundial para el bienestar de la humanidad debido a que estos constituyen un recurso escaso y vital, en este panorama el Lago Titicaca, el lago navegable más alto del mundo, un lago muy importante por su biodiversidad denominado por ello como sitio RAMSAR, es un recurso natural que debe ser estudiado para proponer medidas adecuadas para su conservación. Corresponde entonces a todos y en especial a la universidad asumir esta carga contribuyendo en mayor medida a proteger nuestro medio ambiente.

El presente trabajo abordó la utilización de técnicas de minería de textos en las representaciones sociales sobre la contaminación del lago Titicaca, text mining es el área de minería de datos dedicada a buscar significado mediante métodos automáticos sobre una base de datos de textos, para ello utiliza algoritmos basados en estadística que le dan robustez.

En esta investigación esta técnica fue utilizada para tratar de dar significado a la información obtenida de una de las redes sociales y que es de uso global, la red social Twitter que es utilizada por su simplicidad y manejo de varios lenguajes.

El documento se presenta en el capítulo primero la revisión de la literatura correspondiente al tema de interés de esta investigación, sigue a continuación la presentación de los problemas de investigación tanto general como los problemas específicos, además también se presentan los objetivos e hipótesis a utilizar, En las siguientes secciones se presenta el estado del arte de las investigaciones relacionadas con el tema. Por último, presenta los resultados para finalmente mostrar las conclusiones y las sugerencias.



## CAPÍTULO I

### REVISIÓN DE LITERATURA

Este capítulo está dividido en dos secciones la primera, el marco teórico, dedicada a la exposición del sustento teórico de la investigación y la segunda sección los antecedentes de la investigación que expone las investigaciones realizadas referentes a la temática de la presente.

#### 1.1 Marco teórico

En esta sección se hace una descripción de los principales conceptos teóricos en los que la presente investigación se sustentó. Dado que el propósito principal de la investigación es obtener la representación social sobre la contaminación del Lago Titicaca, Utilizando Minería de Textos (Text Mining), se exponen en primera instancia las características generales de la Cuenca del Lago Titicaca y contaminación ambiental; a continuación, la teoría relacionada a las representaciones sociales para a luego ampliar con la teoría referente a la minería de textos y su aplicación en el ámbito de las ciencias sociales.

##### 1.1.1 La Cuenca del Lago Titicaca

A 14 grados al sur, la Cordillera de los andes se divide en dos ramas: las cordilleras oriental y occidental. Entre ellos se encuentra un sistema hidrológico cerrado de aproximadamente 140,000 kilómetros cuadrados (km<sup>2</sup>) ubicado entre 3,600 y 4,500 metros sobre el nivel del mar (m.s.n.m.). Dentro de ese sistema se encuentran cuatro cuencas principales (ver Tabla 1): Lago Titicaca (T), Río Desaguadero (D), Lago Poopó (P) y Lago Salar de Coipasa (S). El río Desaguadero es la única desembocadura del Titicaca y desemboca en el lago Poopó, cuyo desbordamiento a su vez da lugar al lago salar de Coipasa. Estas cuatro cuencas forman el sistema TDPS, cuyo elemento principal, el lago Titicaca, es el más grande de América del

Sur, el lago navegable más alto del mundo y, según la cosmología inca, el origen de la vida humana (World Water Assessment Programme (United Nations), 2003).

Tabla 1

*Tamaño del sistema TDPS (adaptado de World Water Assessment Programme United Nations (2003))*

Cuenca	Área (Km <sup>2</sup> )
Lago Titicaca	56, 300
Río Desaguadero	29, 800
Lago poopó	24, 800
Lago salar Coipasa	33, 000
<b>Tamaño total TDPS</b>	<b>143, 900</b>

La cuenca del Titicaca tiene un área superficial de 57, 293. Km<sup>2</sup>, una precipitación anual de 702 mm/año, una descarga anual de 281 m<sup>3</sup>/s y una evapotranspiración anual de 652 mm/año. Así mismo la altitud promedio es de 3810 m.s.n.m. una extensión máxima de 173 km, una extensión mínima de 70 km y una profundidad de 283m (World Water Assessment Programme (United Nations), 2003).

El lago Titicaca, ubicado en una zona intertropical, está sujeto a condiciones climáticas específicas de esta área. No es un cuerpo de agua dulce aislado; recibe materiales y sustancias tanto naturales como antropogénicas desde la atmósfera y los ríos circundantes. Elementos como carbono, nitrógeno y fósforo juegan un papel crítico en su calidad y en la estructura biológica del agua, impactando directamente en su composición. En el sector peruano recibe aguas de cuatro ríos principales: Ramis, que incluye los ríos Crucero-Azángaro y Santa Rosa-Ayaviri-Pucará, con un caudal promedio de 88,2 m<sup>3</sup> /s; Ilave (7791 km<sup>2</sup> ), con 40,1 m<sup>3</sup> /s; Illpa (1255,6 km<sup>2</sup> ), con 7,5 m<sup>3</sup> /s; Coata (4882,4 km<sup>2</sup> ), con 39,3 m<sup>3</sup> /s, y Huancané (3611,9 km<sup>2</sup> ), con 23,7 m<sup>3</sup> /s. Su caudal promedio total es del orden de los 198,8 m<sup>3</sup> /s, y equivale a 6269.36 Hm<sup>3</sup> de agua dulce/año. El río Suches (1154,6 km<sup>2</sup> ), con 8,40 m<sup>3</sup> /s, atraviesa parte del territorio peruano y descarga sus aguas en la parte boliviana del lago Titicaca. El lago Titicaca, situado a una altura entre 3,809 y 5,829 metros sobre el nivel del mar, es el lago navegable más alto en zonas tropicales. Destaca por su rica cultura, flora y fauna endémica, lo que llevó a su reconocimiento como sitio RAMSAR en 1997. Este lago se compone de cuatro áreas morfométricas y batimétricas bien diferenciadas: el lago Mayor o Chucuito, la bahía Mayor de Puno,

la bahía Interior de Puno y el lago Menor o Huiñaymarca (Autoridad Nacional del Agua, 2017).

### **1.1.2 El lago Mayor o Chucuito**

Tiene aproximadamente 7132km<sup>2</sup>, con zonas mayores de 200 metros de agua, profundidad media de entre 100 a 200 metros (135 metros en promedio), zonas profundas intermedias desde los 20 a 1000 metros ubicados en las bahías de Puno y Achacachi con un borde litoral con profundidad menor a 20 metros estrecha a lo largo de la bahía de Puno, Rio Ramis y de Achacachi; y con una profundidad máxima de 284m medida por BOULANGE Y AQUIZE JAEN (1981), alcanzado en el lado peruano, al noreste, en las proximidades de la isla Soto, frente a Conima. El lago mayor tiene pendientes muy pronunciadas cuando se desplaza lago adentro, además el volumen del agua que contiene es el 98.5% del total del agua en el lago (Dejoux y Iltis, 1991),

### **1.1.3 Bahía Mayor de Puno**

Con una superficie cercana a los 540 km<sup>2</sup>, conformada por las penínsulas de Capachica y Chucuito, posee una profundidad media de 35 m. En la zona norte desemboca el río Coata. Alberga 29 150 Ha de totorales que constituyen gran parte de la Reserva Nacional del Titicaca, incluye las islas de los Uros. La crianza de truchas en jaulas de forma intensiva es una actividad que se desarrolla entre las localidades de Chimu y Barco, zonas oeste y sur, a cargo de casi ochenta piscicultores entre empresas y personas naturales. Constituye la principal fuente de provisión de agua potable para la ciudad de Puno, la captación se ubica en la zona de Chimu, a 400 m del acceso del canal de navegación (Autoridad Nacional del Agua, 2017).

### **1.1.4 Bahía Interior de Puno**

Es un cuerpo natural de agua de 17,5 km<sup>2</sup>, equivalente al 0,02 % de la superficie total del lago Titicaca (8400 km<sup>2</sup>). Morfométricamente, es un cuerpo de agua casi cerrado, debido a la presencia de barreras naturales formadas por totorales que la separan de la bahía Mayor de Puno limitando el intercambio de agua. Las corrientes superficiales son muy bajas: sus velocidades fluctúan entre 0,54 cm/s y 19,31 cm/s, equivalentes a 0,018 km/h y 0,70 km/h. Este es un factor importante que limita la

dispersión de las masas de agua eutrofizadas hacia la bahía Mayor de Puno (Autoridad Nacional del Agua, 2017).

#### **1.1.5 El lago menor o Huiñaymarca**

El lago Menor, que se ubica en el extremo sur, tiene una superficie de aproximadamente 1470 km<sup>2</sup> (16 % de la superficie total del lago Titicaca) y se caracteriza por su baja profundidad, la cual oscila entre 5 m y 9 m. Su profundidad máxima es de 41 metros. La mayor superficie se encuentra en territorio boliviano (Autoridad Nacional del Agua, 2017).

#### **1.1.6 Fuentes contaminantes de la cuenca del lago Titicaca**

Existen diferentes procesos de contaminación provocados por las actividades humanas dentro de la cuenca del Lago Titicaca. La Autoridad Nacional del Agua (ANA) en base a la información brindada por las administraciones locales del Agua y también de fuentes secundarias relacionadas con aspectos demográficos, ha hecho público el inventario de fuentes de contaminación de la cuenca hidrográfica del Lago Titicaca. Este considera la producción de residuos líquidos y sólidos, pasivos ambientales mineros y de actividades productivas como la minería y la piscicultura (Autoridad Nacional del Agua, 2017).

#### **1.1.7 Aguas residuales municipales.**

Todos los cuerpos de agua poseen una capacidad natural y de auto purificación de los elementos que se incorporan a ellos, conocida como capacidad de asimilación o capacidad de carga (Rivas, 2004). Se considera la contaminación de un cuerpo de agua como la presencia de diversas formas de energía, elementos, compuestos orgánicos o inorgánicos que, disueltos, dispersos o en suspensión, que alcanzan una concentración que limita cualquier tipo de uso (consumo humano, agrícola, pecuario, industrial, recreativo, estético, de conservación de fauna y flora acuática, etcétera). La definición aclara que el uso del agua está en función directa de sus características físicas, químicas, microbiológicas y organolépticas que definen su calidad a partir del uso establecido en la normativa pertinente, en este caso, de los estándares de calidad ambiental para agua establecidos en el D.S N.º 015-2015-MINAM.

Dentro de la cuenca del lago Titicaca, se ha identificado a las aguas residuales domésticas y municipales como una de las principales fuentes contaminantes de las aguas superficiales, ríos, lagos y lagunas. Están compuestas por elementos físicos, químicos y biológicos, es decir, son una mezcla de materiales orgánicos e inorgánicos, suspendidos o disueltos en el agua, las características físico-químicas así como las microbiológicas correspondiente a las aguas residuales municipales son presentadas en la la Tabla 2.

Tabla 2

*Características Físico Químicas de las aguas residuales municipales y sus efectos en un cuerpo de agua*

Contaminantes	Descripción del efecto
Patógenos	En los cuerpos de agua receptores de aguas residuales domésticas o municipales, existe un riesgo de transmisión de enfermedades gastrointestinales y contagiosas. Esto se debe a la presencia de organismos patógenos como bacterias, virus y parásitos contenidos en dichas aguas residuales.
Contaminantes prioritarios	Estos compuestos, ya sean orgánicos o inorgánicos, se clasifican según su capacidad carcinogénica, mutagénica, teratogénica o toxicidad aguda, tanto conocida como sospechada. Una gran cantidad de estos compuestos se encuentra presente en las aguas residuales. Esta categoría de contaminantes incluye residuos de detergentes, productos farmacéuticos y de uso personal, que se conocen como contaminantes emergentes.
Sólidos en suspensión	abarcaban materiales en distintas formas, desde partículas hasta elementos disueltos, afectando la vida acuática y la cadena alimentaria. Se clasifican en suspendidos y disueltos, con los suspendidos potencialmente sedimentables y los disueltos clasificables en orgánicos (volátiles) o inorgánicos (fijos). Su presencia en cuerpos de agua puede generar lodos y alterar el ecosistema acuático.
Materia orgánica biodegradable	Las proteínas, carbohidratos, grasas de origen animal y vegetal son su principal composición, que al ser vertidas sin el tratamiento adecuado, producen el agotamiento del oxígeno, generando condiciones sépticas (malos olores) debido a la descomposición en fase anaeróbica, que genera sulfuro de hidrógeno en altas concentraciones, consecuencia de la reducción de los sulfatos por acción microbiana.
Nutrientes	El nitrógeno, el fósforo y el carbono son elementos vitales que se encuentran en las aguas residuales y, cuando son liberados en cuerpos de agua, promueven un crecimiento excesivo de ciertos organismos acuáticos, como el fitoplancton, así como de diversas plantas acuáticas.
Sólidos inorgánicos disueltos	El calcio, sodio y los sulfatos, son los que constituyen en su mayoría.
Sustancias tóxicas: cianuros, metales, fenoles, etcétera	Su presencia en las aguas residuales en altas concentraciones, tendrán como efecto la extinción de la vida acuática, eliminarán bacterias y frustrarán el debido proceso de autodepuración.

Fuente: adaptado de Autoridad Nacional del Agua (2017)

### **1.1.8 Aguas residuales y contaminantes emergentes.**

Entre los más relevantes se encuentran surfactantes, medicamentos, productos para el cuidado personal, aditivos de las gasolinas, retardantes del fuego, antisépticos, aditivos industriales, esteroides, hormonas, subproductos de la desinfección del agua, plaguicidas, productos para diagnósticos médicos y aditivos alimentarios. Estos productos no causan efectos negativos sin estar permanentemente en el ambiente, dado que sus altas tasas de transformación/remoción se pueden compensar por su introducción continua en el ambiente. La mayor parte de este tipo de contaminantes no son ni acumulativos ni volátiles, a diferencia de los perfumes, que son tóxicos, bioacumulativos, persistentes y volátiles. Los efectos en el hombre y en la biota no son conocidos, pero al ser introducidos continuamente en el medio se convierten en contaminantes persistentes, aún si su semivida es corta (Daughton, 2001).

Los detergentes de tipo alquilfenol etoxilado y algunos fármacos son disruptores endocrinos, y una exposición a ellos podría causar alteraciones en el crecimiento, desarrollo, reproducción y comportamiento de los organismos, y una de las más mejor documentadas es la feminización en organismos acuáticos superiores (Barceló y López, 2008). El mayor porcentaje procede de las medicinas que son excretadas a través de la orina y las heces, vertidas en las aguas por hospitales y domicilios. Otros a partir del lixiviado de botaderos de residuos sólidos urbanos o agroindustriales; este es un problema común en el ámbito de los centros urbanos de la cuenca del lago Titicaca, donde no existen sistemas de rellenos sanitarios (Autoridad Nacional del Agua, 2017).

### **1.1.9 Residuos sólidos peligrosos.**

Denominados a los que provienen de actividades humanas y animales y que son desechados como inútiles o superfluos. Comprenden la masa heterogénea de los desechos de la comunidad urbana y la acumulación más homogénea de los residuos agrícolas, industriales y mineros (Tchobanoglous et al., 1994).

Están compuestos por elementos que, por su composición química y en algunos casos biológica, son peligrosos por ser fuentes de contaminación de los cuerpos de agua superficiales y subterráneos; entre ellos se tienen: baterías, tubos fluorescentes, focos ahorradores, aceites y grasas inorgánicas procedentes de talleres de mecánica, filtros

de aceite, etcétera. A pesar de constituir una fracción pequeña de la composición de los residuos domésticos, son importantes cualitativamente debido a las sustancias químicas que los constituyen. A este tipo de residuos se les denomina residuos peligrosos urbanos (RPU) (Autoridad Nacional del Agua, 2017). La tabla 3 muestra una clasificación de este tipo de residuos.

Tabla 3

*Clasificación de los residuos peligrosos domésticos urbanos*

<b>Tipo</b>	<b>Descripción</b>
Envases de aerosoles	Los envases de aerosoles pueden ser extremadamente inflamables debido a su gas propelente. Además, frecuentemente contienen sustancias tóxicas y peligrosas..
Residuales de productos farmacéuticos	pastillas, sobres, jarabes, aerosoles, pomadas, jeringuillas y objetos punzantes, ampollas y radiografías, productos caducados domiciliarios o de consultas y centros de salud, productos veterinarios.
Productos de limpieza de cocina, sanitarios, suelos, paredes y cristales	Comprenden desengrasantes, desincrustantes, desinfectantes, desatascadores, ceras, limpiacristales, productos para la limpieza de alfombras y moquetas, productos amoniacales
Residuales de detergentes (vajilla y ropa)	Jabones, detergentes, lavavajillas, cuyos envases con algo de producto van a los botaderos al suelo y a los cuerpos de agua.
Aceites vegetales	A través de los desagües, también se desechan en los recipientes de basura, terminando finalmente en los botaderos.
Pinturas y disolventes	Pigmentos (5 % a 25 %) y disolventes (75 % a 95 %). Pinturas basadas en agua o látex y esmaltes o pinturas basadas en aceite, disolventes, diluyentes y decapantes. Los envases contienen restos o se desechan sin estar terminados y acaban en la basura
Productos para la conservación de la madera	Limpiadores, decapantes y barnices. Son productos con pesticidas para proteger de plagas como la polilla y productos para proteger de las inclemencias del tiempo.
Productos de bricolaje	Sellantes, disolventes y colas.
Productos para el revelado fotográfico	Negativos, baños inhibidores, reveladores y fijadores.
Productos eléctricos y electrónicos	Engloban pilas, acumuladores, tubos fluorescentes, etcétera. Pilas de formato normal (zinc-carbono); pilas de botón con óxido de mercurio son las más contaminantes (30 % de mercurio).
Productos para automóviles	Aceites de motores, baterías, líquido de frenos, de transmisión, filtros y neumáticos. Las baterías contienen plomo (65 %) y ácido sulfúrico (27 %); consumen dos terceras partes del plomo total producido y, por tanto, son una fuente de plomo secundario importante.

Fuente: adaptado de López y Dorado (2008)



### **1.1.10 Residuos sólidos hospitalarios**

Generados en los centros de salud ubicados en el ámbito de la cuenca del lago Titicaca como hospitales, clínicas, policlínicos, centros médicos, laboratorios clínicos y consultorios, producidos durante la prestación de servicios de hospitalización, diagnóstico, prevención, tratamiento y curación, incluyendo los generados en los laboratorios.

Se caracterizan por estar contaminados con agentes infecciosos que pueden contener altas concentraciones de microorganismos que son de potencial peligro, tales como agujas, hipodérmicas, gasas, algodones, medios de cultivo, órganos patológicos, restos de comida, papeles de embalaje, material de laboratorio, entre otros (Autoridad Nacional del Agua, 2017).

### **1.1.11 Las letrinas y la contaminación del agua**

En el ámbito rural circunlacustre los habitantes del lugar no cuentan con sistemas de alcantarillado o saneamiento, razón por la que los hoyos y los pozos sépticos son utilizados, que son fuentes de contaminación de las aguas subterráneas, puesto que estos son construidas sobre el suelo natural, sin el acondicionamiento que la apertura de un pozo, que tiene una profundidad de aproximadamente 2 m y que en muchos casos puede estar a escasos metros por encima del nivel freático, que dentro del área circunlacustre puede variar entre 3 m y 8 m. (Huajuapan et al., 2007).

### **1.1.12 Teoría de las Representaciones Colectivas**

Vera (2002), interpreta a Durkheim, quien sostiene que las categorías del pensamiento y la razón son producto de la sociedad. Durkheim argumenta que los conceptos, siendo elementos del pensamiento, emergen de la experiencia colectiva y el intercambio lingüístico en la sociedad. Según él, las representaciones colectivas como la religión, la moral y la economía, son sistemas de valores sociales que influyen en la formación de estos conceptos, siendo el lenguaje un reflejo de cómo la sociedad conceptualiza los objetos. Durkheim enfatiza que el pensamiento lógico es impersonal y está arraigado en estas representaciones colectivas, que a su vez dependen del conjunto de individuos asociados. Las representaciones colectivas se producen por el intercambio de acciones que realizan los individuos como colectividad, en el seno de la vida social y constituyen, por lo tanto, hechos sociales



que sobrepasan y se imponen al individuo, pues las propiedades individuales, al sumarse en la colectividad, pierden su especificidad y se constituyen en fenómenos eminentemente sociales (Piñero, 2012).

### **1.1.13 Teoría de las Representaciones Sociales**

Esta teoría sostiene que los individuos se forman e interactúan en un entorno social y material. Esta teoría se enfoca en cómo las personas internalizan y contribuyen a crear representaciones en su vida diaria. Explora la diferencia entre el pensamiento científico, basado en procesos reflexivos, y el pensamiento de sentido común, arraigado en creencias populares. Su objetivo es comprender cómo estas representaciones influyen en la percepción y comportamiento cotidiano de las personas. (Moscovici, 1979). Indica que los esquemas mentales constituyen las representaciones que los individuos elaboran para comprender, dar sentido común al mundo y comunicarse con los demás, sin que el colectivo se la imponga. Esto permite conocer y explicar los fenómenos donde los humanos son protagonistas tomando en consideración la interacción colectiva y la participación del sujeto (Jodelet, 2008, 1986).

Las representaciones sociales son un sistema que interpreta la realidad, son sistemas cognitivos que poseen lenguaje y lógica propia, representan ramas del conocimiento, con derechos propios para el descubrimiento y organización de la realidad, permite a las personas orientarse en lo material y social para dominarlo; y en segundo lugar permitir establecer comunicación con otros miembros de la comunidad con un código que se les es proporcionado para intercambiar socialmente y para clasificar y nombrar claramente aspectos diversos de su mundo y de su historia en grupo e individual (Farr, 1983), mediante la construcción de mecanismos y fenómenos necesarios en el desarrollo de la vida de los individuos en la sociedad (Ibáñez, 1988).

Es función de las representaciones sociales la elaboración del conocimiento, la elaboración de los comportamientos y la comunicación entre los individuos. La representación es un corpus organizado de conocimientos y una de las actividades psíquicas gracias a las cuales los hombres hacen inteligible la realidad física y social, se integran en un grupo o en una relación cotidiana de intercambios, liberan los poderes de su imaginación. Dicho en términos más fáciles es el conocimiento del sentido común que tiene como objetivos, comunicar, estar al día y sentirse dentro del

ámbito social, y que se origina en el intercambio de comunicaciones del grupo social (Mora, 2002).

Es importante destacar lo que rescata Piñero (2012) donde expresa que la sociología de Pierre Bourdieu representa una aproximación que nos permite entender la forma en que se llevan a cabo los procesos de reproducción y diferenciación social dentro de una sociedad determinada. En este sentido, emplea las nociones de sistemas simbólicos y relaciones de poder para explicar la forma en que son construidas las clases sociales, a las que él denomina espacios sociales.

#### **1.1.14 Proceso de conformación de una Representación Social**

Las representaciones se conforman a través de dos procesos: la objetivación y el anclaje (ver Figura 1) procesos que articulan las funciones cognitivas, de integrar lo nuevo, interpretar la realidad, el comportamiento, las relaciones, la postura ante alguna situación, la transformación del conocimiento y el simbolizar la realidad (Pérez, 2003).

#### **1.1.15 Proceso de objetivación**

En este proceso los entes abstractos, conceptuales se convierten en elementos icónicos, imágenes lo abstracto se convierte en concreto y familiar, mediante tres fases (Pérez, 2003):

- a) *Construcción selectiva*: Esta fase se encarga de la apropiación de los conocimientos referidos al objeto de representación, se seleccionan y descontextualizan los elementos significativos para el sujeto.
- b) *Esquematización estructurante*: Brinda a los sujetos la visión de una realidad que permite una organización coherente de la imagen simbólica del objeto. Se estructuran y organizan en el núcleo figurativo o esquema, columna vertebral de la representación.
- c) *Naturalización*: En esta fase el núcleo adquirido cobra su propia existencia, y deja los niveles de abstracción y funcionan como categorías sociales del lenguaje.

### 1.1.16 El proceso de anclaje

Posterior al proceso de objetivación se integran las informaciones al sistema de pensamiento ya configurado, esto permitirá integrar las nuevas representaciones dentro del sistema de representaciones ya existentes, de esta manera la visión de la realidad se reconstruye permanentemente.

Lo nuevo es incorporado de forma autónoma y creativa al mismo tiempo que se familiariza ante lo desconocido, llevando lo nuevo a un plano conocido donde se puede clasificar a partir de un sustrato cognitivo y emocional previo.



*Figura 1.* Proceso de conformación de una representación social

### 1.1.17 Tipos de Representaciones sociales

Moscovici (1988) describe tres tipos de representaciones sociales: la hegemónica, la emancipada y la polémica (Pérez, 2003), como se aprecia en la Figura 2.

#### a) Representaciones Hegemónicas

Donde los individuos miembros del grupo llegan a un algo grado de consenso, estas se corresponden más con las representaciones colectivas de Durkheim.

#### b) Representaciones emancipadas

Son representaciones propias de determinados subgrupos, que poseen nuevas formas de pensamiento social creando versiones propias de conocimiento.

#### c) Representaciones polémicas

Surgen ante hechos relevantes socialmente, donde grupos específicos atraviesan situaciones conflictivas con foco a acontecimientos u objetos sociales.

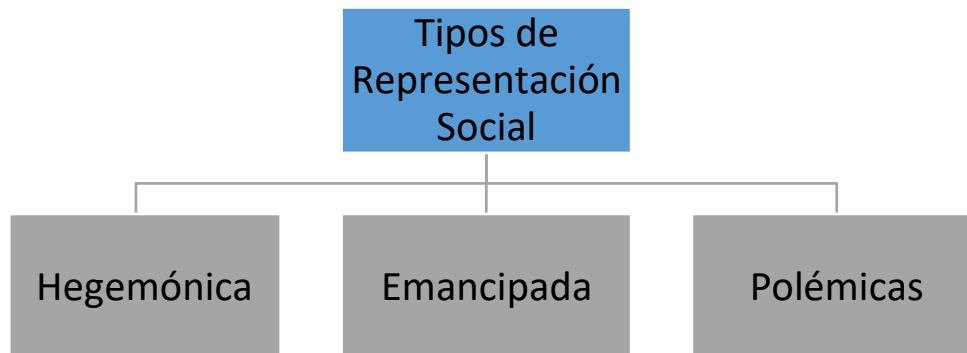


Figura 2. Tipos de Representaciones Sociales

### 1.1.18 Estructura de las Representaciones Sociales

#### a) La dimensión Afectiva

Es la dimensión de la actitud que la orienta en forma favorable o desfavorable frente al objeto de representación, es decir orienta a la actitud y comportamiento de diversas formas, intensidad y dirección

#### b) La dimensión de los conocimientos

Esta dimensión conocida también como la dimensión de la información, da cuenta de los conocimientos en torno al objeto de representación, la calidad es determinada por varios factores destacando la pertenencia grupal y la inserción social.

#### c) La dimensión del campo de la representación

Esta se esquematiza alrededor de un sistema central o núcleo y otro periférico; el núcleo es un grupo de componentes que estructuran y dan significado a las representaciones; tienen una función generadora. Los periféricos son un conjunto que interactúan con el núcleo para dar estructura o modificarla (Abric, 2004).

Sugiere un modelo referido a la jerarquía y orden de los contenidos de las representaciones, se estructura en torno del núcleo y es utilizada por el

investigador para construir en función de la dimensión del conocimiento y la afectiva (Pérez, 2003).

### **1.1.19 Social media**

Social media (medios sociales de comunicación en la internet), es una forma de comunicación que utiliza herramientas en línea como por ejemplo Facebook, Twitter, Instagram, entre otros, Un grupo de aplicaciones basadas en Internet que construidos sobre los fundamentos ideológicos y tecnológicos de la Web 2.0 y que permiten la creación e intercambio de contenido generado por el usuario (Kaplan y Haenlein, 2010).

Existen varios tipos de social media, para realizar un esquema de clasificación y hacerlo de manera sistemática confiamos en un conjunto de teorías en el campo de la investigación de medios (presencia social, riqueza de medios) y procesos sociales (auto -presentación, auto - divulgación), los dos elementos clave de las redes sociales. Con respecto al componente de redes sociales relacionado con los medios, la teoría de la presencia social establece que los medios difieren en el grado de "presencia social", definido como el contacto acústico, visual y físico que se puede lograr, que permiten emerger entre dos compañeros de comunicación (Parker et al., 1978). La presencia social está influenciada por la intimidad (interpersonal frente a mediada) y la inmediatez (asíncrona frente a síncrona) del medio, y se puede esperar que sea inferior para mediada (p. Ej., Conversación telefónica) que interpersonal (p. Ej., Debate cara a cara) y para comunicaciones asíncronas (p. ej., correo electrónico) que síncronas (p. ej., chat en vivo).

Cuanto mayor es la presencia social, mayor es la influencia social que los socios de comunicación tienen en el comportamiento de los demás. Estrechamente relacionado con la idea de presencia social está el concepto de riqueza mediática. La teoría de la riqueza de los medios indica Daft y Lengel (1986) se basa en el supuesto de que el objetivo de cualquier comunicación es la resolución de la ambigüedad y la reducción de la incertidumbre. Establece que los medios difieren en el grado de riqueza que poseen, es decir, la cantidad de información que permiten transmitir en un intervalo de tiempo determinado, y que, por lo tanto, algunos medios son más efectivos que otros para resolver la ambigüedad y la incertidumbre. Aplicado al contexto de las redes sociales, se supone que se puede hacer una primera clasificación basada en la

riqueza del medio y el grado de presencia social que permite. Con respecto a la dimensión social de las redes sociales, el concepto de autopresentación establece que en cualquier tipo de interacción social las personas tienen el deseo de controlar las impresiones que otras personas forman de ellas (Goffman, 1959). Esto es hecho con el objetivo de influir en otras personas para obtener algo a favor (recompensas), se puede mencionar como ejemplo, causar una impresión positiva en una entrevista de trabajo; Por otro lado, está impulsado por el deseo de crear una imagen que sea consistente con la identidad personal de uno, por ejemplo, usar un atuendo de moda para ser percibido como joven o moderno. Trasladando estos conceptos a social media, la razón clave por la cual las personas deciden crear una página web personal es, por ejemplo, el deseo de presentarse o representarse en el ciberespacio (Schau y Gilly, 2003). Generalmente, la presentación es efectuada por intermedio de la divulgación personal; es decir, tanto consciente o inconsciente se revela información personal (pensamientos, sentimientos, gustos, disgustos) consistente con la imagen que uno desea dar. Aplicado al contexto de las redes sociales, suponemos que se puede hacer una segunda clasificación en función del grado de revelación requerido y el tipo de presentación que permitido.

En la Tabla 4 se visualiza la clasificación de social media mediante la combinación de la dimensión autopresentación con la dimensión presencia social.

Tabla 4

*Clasificación de social media por presencia social/riqueza de medios y autopresentación/autodivulgación.*

Autopresentación n /Autodivulgación	Presencia Social / Riqueza de medios		
	Bajo	Medio	Alto
Alto	Blogs	Social networking sites (Facebook)	Virtual social worlds (second life)
Bajo	Collaborative projects (Wikipedia)	Content communities	Virtual game worlds (World of warcraft)

Las redes sociales abarcan muchas plataformas desarrolladas sobre internet que ayudan a la divulgación de las emociones humanas, como son:

- Networking, por ejemplo, Facebook, LinkedIn
- Micro blogging, por ejemplo, Twitter Tumblr
- Photo Sharing, por ejemplo, Instagram, Flickr
- Video sharing, por ejemplo, YouTube.
- Stack exchanging, por ejemplo, Stack Overflow, Github.
- Instant messaging, por ejemplo, Whatsapp, Hike

El uso de redes sociales en marketing ha evolucionado, pasando de ser solo una plataforma para anuncios a una herramienta crucial para analizar la opinión del consumidor. La minería de datos en redes sociales proporciona a las empresas una comprensión profunda de las perspectivas de clientes actuales y potenciales. Esta información es clave para tomar decisiones empresariales informadas que pueden influir significativamente en el futuro del negocio, más allá del simple impacto publicitario. (Garg y Ravindran, 2015).

Los sistemas de gestión de relaciones con clientes (CRM) actuales utilizan perfiles detallados de consumidores, combinando datos demográficos y patrones de compra anteriores para mejorar las estrategias de marketing. Estos sistemas permiten un seguimiento exhaustivo de los consumidores, aprovechando la información detallada disponible en redes como LinkedIn y Facebook. Esta información, incluyendo los intereses de los clientes, es invaluable para desarrollar planes de marketing efectivos. Las comunidades en línea como Facebook y Twitter generan constantemente grandes volúmenes de datos, que si se analizan adecuadamente, pueden revelar tendencias emergentes y oportunidades en tiempo real. (Garg y Ravindran, 2015).

### **1.1.20 Plataforma Social media de Twitter**

Twitter es una herramienta gratuita de redes sociales ampliamente utilizada que permite a las personas compartir información, en una fuente de noticias en tiempo real (Mistry, 2011) mediante la publicación de breves comentarios sobre sus experiencias y pensamientos (Bristol, 2010). Los mensajes públicos enviados y recibidos a través de Twitter, o "tweets", están limitados a no más de 140 caracteres y pueden incluir enlaces a blogs, páginas web, imágenes, videos y cualquier otro material en línea. A pesar de la brevedad impuesta por esta herramienta de medios, el uso de Twitter se usa ampliamente en una amplia variedad de circunstancias y, de acuerdo con Mollett et al. (2011), miles de académicos e investigadores en todos los

niveles de experiencia y en todas las disciplinas ya usan Twitter diariamente ". Después de configurar una cuenta de Twitter ([www.twitter.com](http://www.twitter.com)), los usuarios establecen un perfil y un "nombre" de Twitter, por ejemplo, @OTprofile, y luego puede enviar y recibir tweets, a los que se accede a través de cualquier computadora o dispositivo móvil en red. Una vez que se envía un tweet, aparece en el "bandeja de entrada" de Twitter del usuario y en la bandeja de entrada de cualquier persona que lo siga. La Tabla 5 proporciona un resumen de la terminología común de Twitter. La búsqueda también se puede utilizar para encontrar tweets relevantes. Esto puede ser por palabras clave, a menudo identificadas por hashtags definidos por el usuario, identificadas por un símbolo inicial '#' (por ejemplo, #contaminación o #revista).

Tabla 5

*Terminología de Twitter (adaptado de Mollett et al. (2011))*

<b>Término</b>	<b>Definición</b>
<b>Seguir</b>	Seguir a otro usuario significa que todos sus tweets aparecerán en su bandeja de entrada.
<b>A quién seguir</b>	Esta es una lista de las sugerencias de Twitter de personas u organizaciones que puede seguir, en función de puntos de similitud con su perfil.
<b>Dejar de seguir</b>	Para dejar de ver los tweets de otra persona.
<b>Bloquear</b>	De vez en cuando puede aparecer un spammer u otro personaje desagradable en su lista de seguidores y se eliminarán de la lista de Seguidores.
<b>Retweet o RT</b>	Para compartir el tweet de otra persona que has visto en tu Bandeja de entrada.
<b>Responder</b>	Para responder al tweet de otra persona.
<b>@</b>	Para responder al tweet de otra persona, se usa en tweets cuándo quieres mencionar a otro usuario. También la primera parte de cada nombre de usuario de Twitter, por ejemplo @BAOTCOT.
<b>Menciones</b>	Para ver cuando otros lo mencionan en un tweet por su nombre de usuario. Si menciona a otros, esto aparecerá en su pestaña de Mención
<b>#</b>	Hashtag: se usa para clasificar los tweets. Los temas populares se denominan temas de tendencia y a veces van acompañados de hashtags, como # london2012, #OTuesday u #Occhat.
<b>Mensaje directo o DM</b>	Estos son mensajes privados que puede enviar a otros usuarios de Twitter
<b>URL acortadas</b>	Dado que una dirección web típica es bastante larga y torpe, los sitios gratuitos de acortamiento de URL como <a href="http://bitly.com">bitly.com</a> y <a href="http://tinyurl.com">tinyurl.com</a> proporcionan enlaces que puede pegar en tweets.



Los hashtags ayudan a localizar áreas particulares de discusión (Bristol, 2010) y algunos hashtags que se referentes a la contaminación del Lago Titicaca pueden verse en los tweets de ejemplo en la Tabla 6.

Como herramienta de comunicación, Twitter permite el libre intercambio de ideas globalmente, entre personas interesadas en áreas similares de especialización, así como brindar la oportunidad de participar en debates críticos (Maclean et al., 2013).

Tabla 6

*Ejemplo de tweets*

<b>Tweets sobre el lago Titicaca.</b>
<b>Hospital General Monte Sinaí @HGMonteSinai, 6min</b> Personal de Atención al Usuario enseña a reconocer los colores para la clasificación de residuos. El reciclaje ayuda a combatir la contaminación del planeta y fomenta entornos saludables. #MSPdeTodosyParaTodos
<b>Carolina Perdomo @Carolpg12,4h</b> PERÚ Los andes cobijan bajo sus montañas y escarpes un mundo de mitos y tradiciones ancestrales que permanecen vivas . Machu Picchu y el Lago Titicaca enmarcan este asombroso viaje.
<b>Hernán Green @GreeHernn · 8 ene.</b> PERÚ Los andes cobijan bajo sus montañas y escarpes un mundo de mitos y tradiciones ancestrales que permanecen vivas . Machu Picchu y el Lago Titicaca enmarcan este asombroso viaje.

### 1.1.21 Minería de Social media

La minería de redes sociales implica un análisis sistemático de grandes volúmenes de información generada en estas plataformas, conocido como Big Data. Aproximadamente el 20% de los datos de redes sociales son relevantes, presentando el desafío de filtrar y procesar eficientemente esta información masiva. La tecnología juega un papel crucial en la gestión y análisis de estos conjuntos de datos, que a menudo son de gran tamaño, para extraer información útil y significativa, considerando aspectos como la relación señal/ruido (Garg y Ravindran, 2015).

La minería de redes sociales (SMM) utiliza un conjunto de herramientas y técnicas para analizar datos de redes sociales. Esta técnica permite a las empresas comprender la influencia de los usuarios en la web y detectar las preocupaciones de los clientes en tiempo real, facilitando la planificación proactiva. Además, SMM es útil para identificar clientes potenciales a través del análisis de sus actividades en línea y de

las de sus conexiones, proporcionando una solución eficiente al antiguo desafío empresarial de localizar prospectos (Garg y Ravindran, 2015). Existe mucha investigación en múltiples disciplinas de las redes sociales:

- ¿Por qué es importante la minería en las redes sociales?
- Si puede medirlo, puede mejorarlo
- Comportamiento de modelado
- Análisis predictivo
- Recomendar contenido

La minería de las redes sociales tiene sus raíces en muchos campos, como las estadísticas, el aprendizaje automático, la recuperación de información, el reconocimiento de patrones y la bioinformática. Los campos principales no están exentos de desafíos. La gran cantidad de datos que se generan a diario es asombrosa, pero las técnicas actuales permiten nuevas soluciones de minería de datos y modelos computacionales escalables con la ayuda de los conceptos y teorías y algoritmos fundamentales.

### **1.1.22 Minería en Grafos**

Normalmente, una red social está representada como un grafo. El como explotar los patrones en un grafo para tareas como detección de comunidades, clasificación de redes, detección de valores atípicos entre otros, es un tópico de mucho interés gracias a la disponibilidad de gran cantidad de datos de las redes sociales (Tang y Liu, 2010).

Típicamente, las comunidades de usuarios constituyen un grupo de nodos en grafos de red donde los nodos dentro de la misma comunidad o clúster tienden a compartir características comunes.

La minería de grafos se puede describir como el proceso de extracción de conocimiento útil (patrones, valores atípicos, etc.) de una relación social entre los miembros de la comunidad, esta relación se puede representar como un grafo. El ejemplo más influyente de la minería de grafos es Facebook Graph Search. La Figura 3, muestra un grafo de red de palabras de los tweets usando el hashtag Climate Change.



efectivamente esta avalancha de texto en una amplia variedad de aplicaciones (Allahyari et al., 2017).

Los enfoques de minería de texto están relacionados con la minería de datos tradicional, y métodos de descubrimiento de conocimiento, con algunas especificidades. Existen varias definiciones para el descubrimiento de conocimiento o el descubrimiento de conocimiento en bases de datos (KDD) y la minería de datos en la literatura. Lo definimos de la siguiente manera: el descubrimiento de conocimiento en bases de datos está extrayendo información implícita válida, nueva y potencialmente útil de los datos, que no es trivial (Frawley et al., 1992); (Fayyad et al., 1996). Data Mining es una aplicación de algoritmos particulares para extraer patrones de datos. KDD tiene como objetivo descubrir patrones ocultos y conexiones en los datos. Basado en lo anterior la definición de KDD se refiere al proceso entero de descubrir conocimiento útil de los datos mientras que minería de datos se refiere a un paso específico en este proceso donde la data puede ser estructurada como una base de datos o no estructurada como datos dentro de un archivo de texto común.

La minería de textos o descubrimiento de conocimiento desde el texto (KDT) (Feldman y Dagan, 1995), se refiera al proceso de extraer información de calidad desde textos como por ejemplo datos desde Sistemas de Administración de base de datos relacionales (RDBMS) (Džeroski, 2009); (Chen et al., 1996), semi estructurados como XML y JSON (Doroodchi et al., 2009; Pouriyeh y Doroodchi, 2009) y fuentes de texto no estructurados como documentos de procesadores de texto, videos e imágenes. Cubre un largo conjunto de tópicos relacionados y algoritmos para analizar textos, involucra:

- recuperación de información
- Procesamiento del Lenguaje Natural,
- Extracción de información desde texto
- Resumen de textos
- Métodos de aprendizaje no supervisados
- Métodos de aprendizaje supervisado
- Métodos probabilísticos para la minería de textos
- Flujos de texto y minería de medios sociales
- Minería de Opinión y Análisis sentimental

- Minería de texto en ayunas

#### 1.1.24 Proceso General de Minería de textos en social media

Cualquier actividad de minería de datos sigue o involucra en su protocolo algunos pasos genéricos para obtener ideas útiles de los datos. Dentro de los pasos, tomando datos de la plataforma de Twitter tenemos lo siguiente: autenticación del sitio web social, visualización de datos, limpieza y preprocesamiento, modelado de datos utilizando algoritmos estándar como minería de opinión, agrupación, detección de anomalías / spam, correlaciones y segmentaciones, recomendaciones y visualización de resultados este proceso se resume en la Figura 4.

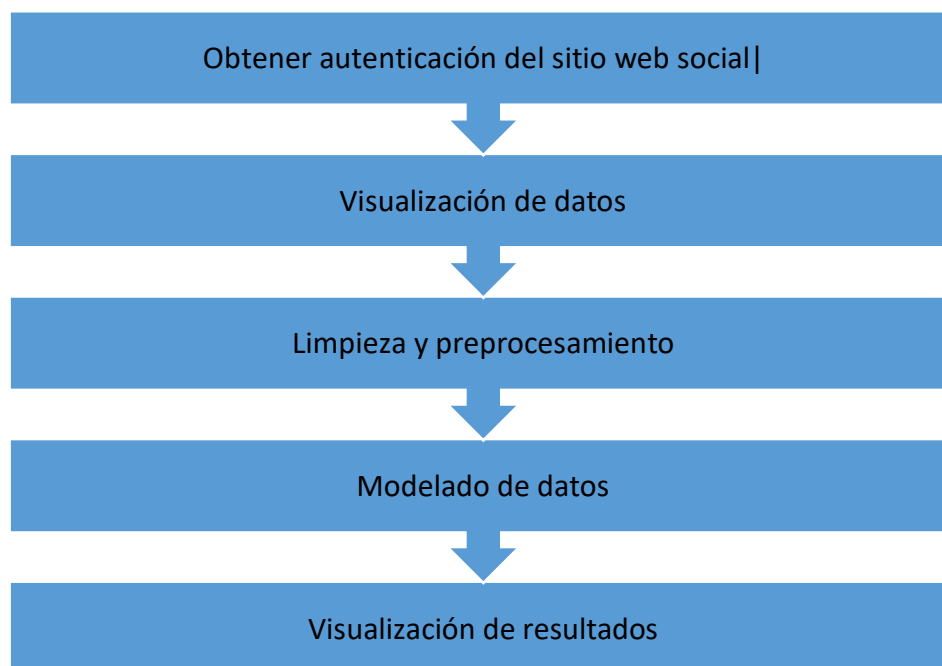


Figura 4. Proceso General de minería de textos en social media.

#### 1.1.25 Extracción Rápida y Automática de Palabras Clave - RAKE

RAKE es una técnica no supervisada para obtener palabras clave de un texto. Las palabras clave tienen varias palabras de contenido que son informativas. RAKE puede ser relevante para cualquier tipo de documento independiente de los dominios, especialmente aquellos que no persiguen tradición lingüística explícita. Los parámetros de entrada para RAKE son una lista de palabras vacías, delimitadores de frases. Las palabras que no parecen estar en la lista de palabras vacías se toman en cuenta como palabras clave candidatas (Thushara et al., 2019).

El algoritmo RAKE incluye los siguientes pasos:

- 1) **Selección de candidatos:** elimina todas las palabras vacías y delimitadores de frases y luego extrae todas las posibles frases clave.
- 2) **Cálculo de características:** El valor de la característica en RAKE es la matriz de puntuación de palabras clave  
Las estimaciones distintivas utilizadas para la construcción de matriz de puntaje ponderados.  
**Frecuencia de palabras:** Muestra el recuento de ocurrencias de una palabra en el artículo.  
**Grado de palabra:** el grado de una palabra representa cuán a menudo co-ocurren con diferentes palabras dentro de la andidades a palabras clave.  
**Razón entre el grado y la frecuencia:** la matriz de puntuaciones ponderadas elaborada es utilizada en la selección de palabras clave del artículo.
- 3) **Selección de palabras clave:** las palabras mejor clasificadas son consideradas como las palabras clave entre las frases candidatas extraídas.

## 1.2 Antecedentes

A continuación, detallaremos los antecedentes que la investigación consideró para la elaboración del presente.

### 1.2.1 Estudios sobre representaciones sociales y medio ambiente

Este apartado está dirigido a las investigaciones orientadas al análisis de las representaciones sociales y su relación con el medio ambiente:

Andrade et al. (2004), destaca cómo la globalización impulsa la discusión sobre temas globales como la educación ambiental y para la salud. Investigando 204 estudiantes universitarios en Río de Janeiro, el estudio reveló tres categorías en educación ambiental y cuatro en educación sanitaria, mostrando que el medio ambiente y la salud se perciben aún como temas separados. Esta separación afecta la eficacia de las acciones tanto de la población como de las entidades gubernamentales y no gubernamentales en estos campos.

Navarro (2004) enfatiza la importancia global del agua en términos políticos y geoestratégicos, señalando cómo Colombia, a pesar de su abundancia hídrica, solo provee agua a la mitad de su población. El estudio expone los problemas de

degradación de recursos y ecosistemas, contaminación, despilfarro y desigualdad en el acceso al agua, obstaculizando el desarrollo sostenible. Investigando representaciones sociales del agua, Navarro entrevistó a 150 estudiantes de ciencias sociales en Bogotá, empleando análisis prototípico y categorial para entender cómo este grupo percibe el agua.

Albero (2005) Este enfoque renueva la comprensión pública de la ciencia y la tecnología, desafiando la percepción social tradicional de estos campos. Utiliza el concepto de representaciones sociales y revisa críticamente argumentos y evidencias sobre la apreciación pública, el interés y la relación entre educación y aceptación de la tecnociencia. Finalmente, propone un modelo alternativo que relaciona diversas representaciones sociales de ciencia y tecnología con diferentes contextos sociales, alejándose del enfoque lineal y educativo tradicional..

Flores (2008) analiza las representaciones sociales (RS) del medio ambiente en estudiantes de la Benemérita Escuela Nacional de Maestros en la Ciudad de México. Se enfoca en cómo las RS funcionan como un tipo de conocimiento en sociedades modernas, donde las personas interpretan ideas científicas preestablecidas. Los hallazgos revelan cinco categorías de RS sobre el medio ambiente: antropocéntricas utilitaristas, pactuadas y culturales, naturalistas y globalizantes, mostrando la diversidad en la percepción y comprensión del medio ambiente entre los estudiantes..

Tochetto (2009) examina la asimilación de la gestión ambiental en empresas del noreste, destacando tres fuerzas impulsoras: mercado, estado y sociedad. Se enfoca en la influencia de la sociedad local, utilizando la Nueva Escala del Paradigma Ecológico. Se aplicaron cuestionarios a estudiantes de universidades en Rio Grande do Sul y Ceará, buscando datos comparativos sobre la percepción ambiental. Los resultados muestran diferencias significativas en los paradigmas ecológicos entre las sociedades de ambos estados, validando la influencia de la sociedad como fuerza impulsora en la gestión ambiental.

Flores (2010) describen los aspectos centrales de la investigación Representaciones del medio ambiente de los estudiantes de la licenciatura en educación primaria. En función de las representaciones los estudiantes discriminan espacio temporalmente los elementos del medio ambiente, consolidando correspondencias con las prácticas

sociales. Entre los resultados destacan la existencia de una relación entre las representaciones del medio ambiente y las corrientes de la educación ambiental.

González y Valdez (2012) Los estudios sobre educación ambiental en España, Brasil y México han utilizado principalmente categorías naturalistas, antropocéntricas y globalizantes para clasificar representaciones sociales. Sin embargo, la mayoría se centra en un solo grupo, como profesores o estudiantes, sin analizar su interacción. Se sugiere una mayor exploración en este campo, incorporando diferentes sujetos y aplicando los hallazgos para mejorar los procesos educativos y promover hábitos y valores proambientales.

Navarro (2013) indica, Las representaciones sociales, especialmente en el ámbito ambiental, ayudan a comprender comportamientos y posturas relacionadas con temas como la polución. Un estudio involucró a 305 individuos en Medellín, comparando los residentes de zonas con alta y baja contaminación del aire. Se evidenció que la percepción del "riesgo" es un componente central en la representación social del medio ambiente. La exposición a la polución influye significativamente en cómo las personas atribuyen causas a la contaminación del aire.

Rodríguez y Paba (2013) muestran que la representación social de la crisis ambiental entre estudiantes de educación media y superior en Santa Marta, Colombia, con 292 participantes. Utilizando un enfoque metodológico de complementariedad que combina análisis descriptivo y teoría de facetas, se descubrió que las representaciones se centran en enfermedades y contaminación como los problemas ambientales más comunes, vinculados a factores ambientales, humanos y climáticos. Se enfocó en identificar conocimientos, actitudes y percepciones sobre planes, programas, proyectos y docencia en educación ambiental..

Piermattéo et al. (2016) evaluó cómo la invalidación de expectativas influye en el compromiso con programas proambientales. Se examinó la jerarquía de expectativas, distinguiendo entre cogniciones centrales y periféricas en el marco de la teoría de las representaciones sociales. Se halló que invalidar una cognición central aumenta significativamente la probabilidad de abandonar el programa en comparación con una cognición periférica..



Polli y Camargo (2016) Investigaron las actitudes y representaciones sociales del medio ambiente en diferentes grupos de edad utilizando la nueva escala de paradigma ambiental (NEP) y un análisis de contraste con el programa Alceste. Participaron 150 personas en tres grupos de edad, sin encontrar diferencias significativas entre ellos. Todos mostraron concordancia con la idea contemporánea de cuidado y reconocimiento de la fragilidad del medio ambiente.

Levidow y Upham (2017), Combinaron perspectivas teóricas de expectativas tecnológicas y representaciones sociales, aplicándolas al tratamiento térmico de residuos sólidos mediante gasificación. Este estudio muestra cómo las representaciones sociales influyen las expectativas tecnológicas y cómo estos dos elementos interactúan para crear una percepción más favorable de la tecnología de gasificación, considerándola como una transición hacia métodos más "avanzados".

### **1.2.2 Estudios sobre la contaminación del Lago Titicaca**

Fontúrbel (2005) evaluó las condiciones ambientales en seis sitios del Lago Titicaca, Bolivia, usando cuatro indicadores de eutrofización y análisis de pH y micro meso fauna. Se hallaron niveles variables de eutrofización debido a influencias antropogénicas y contaminación.

Beltrán (2015) evaluó la calidad acuática de la bahía interior de Puno mediante muestreos mensuales y análisis de parámetros fisicoquímicos. Se encontró contaminación significativa cerca de la salida de la laguna de estabilización de Espinar, con altos niveles de coliformes fecales..

Archundia et al. (2017) Analizaron la contaminación en la cuenca hidrográfica de Katari y su impacto en la Bahía Cohana del Lago Titicaca, utilizando un enfoque integrador que incluye datos hidrológicos, fisicoquímicos, químicos y bacterianos. La concentración de contaminantes tiene patrones que están determinados por las variaciones hidrológicas estacionales. La baja calidad de las aguas superficiales en la cuenca representa un riesgo para las poblaciones humanas y animales, así como para la calidad de los acuíferos ubicados debajo de la ciudad de El Alto.

Gutiérrez (2018) Investigó la influencia de la producción truchícola en Chucuito, Juli y Pomata en el impacto ambiental, utilizando análisis cualitativos y pruebas estadísticas para evaluar los factores óptimos de producción y sus efectos

ambientales. Los resultados de las Evaluaciones de Impacto Ambiental de cada zona de estudio, indican que es posible seguir llevando adelante esta actividad acuícola; considerando siempre, medidas correctivas y de mitigación para los impactos negativos y medidas de optimización para los impactos positivos.

Moreno et al. (2018) cuantificaron metales en sedimentos superficiales de la bahía interior de Puno, Lago Titicaca, mediante espectrometría. Los metales presentaron concentraciones dentro de los rangos recomendados por las normas ambientales. Se concluyó que, los sedimentos superficiales de la bahía interior de Puno, no representan riesgo por exposición a metales totales, ya que sus concentraciones se encontraron en el rango de los valores permisibles.

Cahui (2019) estudió la presencia de antibióticos en el agua superficial y potable en la zona sur del Lago Titicaca y en la ciudad de Puno, encontrando concentraciones significativas que podrían implicar un riesgo para la salud pública. La presencia de estos contaminantes en agua y sedimento puede ser atribuida a la constante aplicación de estos fármacos para la prevención y tratamiento de enfermedades en la producción de truchas lo cual, sin embargo, puede significar un grave riesgo para la salud pública de la población de Puno.

Escobar (2019) indica que evaluó la contaminación en el Lago Titicaca, específicamente en la Bahía de Puno, analizando características químicas y físicas del agua y niveles de metales. Aunque la calidad del agua está dentro de los estándares, las actividades de producción truchícola podrían comprometer la sostenibilidad del lago. Las tendencias de crecimiento en la actividad productiva de truchas en el lago comprometen la calidad del agua y, por lo tanto, la sostenibilidad del Lago Titicaca. Por lo tanto, se sugiere implementar medidas responsables mínimas y buenas prácticas en esta actividad económica para abordar este problema.

### **1.2.3 Estudios sobre Social media y Medio Ambiente**

Mei et al. (2014) Propusieron modelos de aprendizaje automático para estimar el Índice de Calidad del Aire (AQI) en China utilizando contenido de redes sociales, enfocándose en correlaciones espacio-temporales. Sus experimentos con datos de Sina Weibo de 108 ciudades durante un período de un mes demuestran el rendimiento preciso de predicción AQI de nuestro enfoque.



Kay et al. (2015) analizaron el impacto de la publicación de datos sobre calidad del aire por la Embajada de EE. UU. en Beijing en Twitter, destacando cómo los microblogs pueden empoderar a los ciudadanos y ser utilizados por el gobierno y empresas en China. muestran que, aunque los microblogs son capaces de capacitar a los ciudadanos para avanzar en una causa ambiental, las redes sociales también han sido empleadas cada vez más por el gobierno como una herramienta para el monitoreo y control social y por las empresas como una plataforma para beneficiarse de la contaminación del aire.

Jiang et al. (2015) desarrollaron un marco analítico para monitorear la calidad del aire en ciudades utilizando mensajes de redes sociales, encontrando una fuerte correlación entre estos mensajes y el AQI. Nuestro estudio indica que los mensajes filtrados de las redes sociales están fuertemente correlacionados con el AQI y pueden usarse para monitorear la dinámica de la calidad del aire hasta cierto punto.

## CAPÍTULO II

### PLANTEAMIENTO DEL PROBLEMA

#### 2.1 Identificación del problema

La crisis ambiental es un problema que involucra a toda a toda la población mundial, la problemática que se afronta va desde casos como la desertificación de los suelos (suelos que ya no producen vegetación), la explotación indiscriminada de los recursos naturales (por ejemplo, la pesca industrializada), el manejo inadecuado de los recursos hídricos entre otros. En este panorama se presenta el Lago Titicaca como una reserva de agua y biodiversidad importante donde se desarrollan diferentes actividades económicas y culturales, actualmente esta reserva natural es golpeada a lo largo de toda su rivera afrontando problemas como la minería ilegal, el mal manejo de residuos sólidos, la explotación inadecuada del recurso hídrico entre otros. Así mismo pese a ser una problemática que involucra a los diferentes actores de la sociedad, los ciudadanos y el estado a través de sus diferentes representantes, hacen muy poco por conservar este recurso tan importante.

El Lago Titicaca, especialmente a lo largo de su ribera que es donde se concentra la mayoría de las ciudades y población de la meseta del altiplano entre los países de Perú y Bolivia, viene siendo objeto de un proceso creciente de contaminación, producido por las actividades humanas. Dicha problemática, ha generado estudios diagnóstico y evaluaciones, enfatizando en sus fuentes y la magnitud de los efectos sobre los recursos naturales, la salud humana y el medio ambiente. Sin embargo, los estudios desde un punto de vista social, en especial sobre las representaciones sociales de la población sobre la contaminación del Lago Titicaca son casi inexistentes o poco conocidos o difundidos. La problemática abarca diferentes dimensiones desde la social donde se evidencia poco o nulo interés de la población pese a existir canales adecuados tales como el presupuesto

participativo; otra dimensión de interés es la dimensión tecnológica principalmente referente al tratamiento de aguas, la región de Puno aún no cuenta un sistema completo para el tratamiento de aguas residuales por lo que en toda la rivera del lago estos desperdicios ingresan directamente al agua sin tratamiento adecuado; finalmente existe una participación muy baja de parte de las instituciones que deberían encargarse de la descontaminación y conservación del Lago Titicaca.

Con la evolución de los canales de comunicación hasta abarcar el mundo digital de la red internet conocido como social media, y que son actualmente estos medios, los que agrupan a gran cantidad de ciudadanos no solo locales sino de todo el mundo, además la dinamicidad de estas comunicaciones permite el registro de opiniones, expresiones y emociones de forma espontáneas. Estos registros son poco tratados y analizados en nuestro medio convirtiéndose en campo atractivo desde una perspectiva social para abordar la problemática de la contaminación del Lago Titicaca. Así desde la Minería de Datos corresponde hacer aportes que permitan no solo coleccionar datos sino de ayudar a la toma de decisiones dirigidas a solucionar el problema de la contaminación del Lago Titicaca, de forma dinámica y propendiendo a un manejo en tiempo real de la información para generar conocimientos oportunos.

Dado lo mencionado en los párrafos anteriores se plantea la siguiente interrogante general de investigación:

## **2.2 Enunciados del problema**

La interrogante que buscó responder la investigación fue:

¿Qué representaciones sociales se evidencian mediante la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca?

Y como parte del proceso dar respuesta a las siguientes interrogantes específicas:

¿Qué conocimientos se evidencian en los mensajes sobre la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca?

¿Qué actitudes y prácticas se expresan sobre la plataforma de microbloggings Twitter sobre la contaminación del Lago Titicaca?

## 2.3 Justificación

El Lago Titicaca, es el lago navegable más alto del mundo, en su entorno se desarrollaron grandes civilizaciones (Pucara, Tiahuanaco, Lupaca y Colla) y es reconocida como maravilla natural de la humanidad por sus condiciones geomorfológicas, su potencialidad y alta biodiversidad y ser fuente de vida permanente.

La relevancia del estudio, está centrado en dos niveles: de un lado, el tema de la contaminación, constituye parte de la agenda no solo regional y nacional, sino de la agenda global, signado por el calentamiento global y el cambio climático. De otro lado, debido a que, para plantear propuestas o alternativas integrales en perspectiva sostenida frente a la contaminación del Lago Titicaca, previamente, se debe captar las percepciones de los actores sociales. Por tanto, la investigación contribuirá con la caracterización de los conocimientos y actitudes de los usuarios de la red social Twitter frente a la contaminación del lago.

Así enfocado desde una perspectiva de la Estadística y su área de conocimiento Minería de Datos, la importancia del presente Proyecto de Tesis, radica en que sus aportes ayudarán a la toma de decisiones dirigidas a solucionar el problema de la contaminación del Lago Titicaca.

## 2.4 Objetivos

### 2.4.1 Objetivo general

Determinar qué representaciones sociales se expresan sobre la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.

### 2.4.2 Objetivos específicos

- Determinar qué conocimientos se evidencian sobre los mensajes vertidos en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.
- Determinar qué actitudes y prácticas se expresan sobre la plataforma de microbloggings Twitter sobre la contaminación del Lago Titicaca.



## 2.5 Hipótesis

### 2.5.1 Hipótesis general

Se expresan sobre la plataforma de microblogging Twitter representaciones sociales divergentes sobre la contaminación del Lago Titicaca.

### 2.5.2 Hipótesis específicas

- Los conocimientos vertidos en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca son diferenciados.
- Las actitudes y prácticas expresadas sobre la plataforma de microbloggings Twitter sobre la contaminación son heterogéneas.



## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1 Lugar de estudio

La presente investigación ubica su ámbito de estudio sobre la plataforma digital de microblogging Twitter, acerca de las representaciones sociales de la Contaminación del Lago Titicaca, el Lago Titicaca está ubicado en América del Sur en la Cordillera de los Andes, región Altiplano andino, en la ecorregión Puna, es un área protegida como sitio Ramsar, su cuenca es de sistema endorreico Titicaca-Desaguadero-Poopó-Salar de Coipasa y se encuentra en las coordenadas  $15^{\circ}45'00''\text{S}$   $69^{\circ}25'00''\text{O}$ .



*Figura 5.* Lago Titicaca.

Fuente: Wikimedia Commons



## **3.2 Población**

### **3.2.1 En Social media Microblogging**

Sobre la totalidad de las plataformas de social media accesibles en internet dedicadas a microblogging, la población está conformada por la totalidad de los posts referentes al lago Titicaca registrados sobre el servicio de microblogging de la plataforma Twitter.

## **3.3 Muestra**

### **3.3.1 En Social media Microblogging**

La muestra está conformada por los posts vertidos sobre el servicio de microblogging de Twitter en el mes de septiembre del 2019 al mes de febrero del 2020 en idioma español.

## **3.4 Método de investigación**

La presente investigación es del tipo no experimental puesto que no existe manipulación de ningún variable o nivel de estas, es descriptivo pues se describen las representaciones sociales sobre la contaminación del Lago Titicaca sobre la plataforma de microblogging Twitter y es de corte transversal pues se ejecutó en un solo periodo de tiempo.

## **3.5 Descripción de métodos por objetivos**

### **3.5.1 Representaciones sociales expresadas en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca**

#### **a) Autenticación mediante protocolo OAuth 2.0 sobre la plataforma de microblogging Twitter**

Para poder acceder a los contenidos vertidos en la plataforma de microblogging Twitter es necesario un método de autenticación seguro, este método se denomina OAuth y es el utilizado en la presente investigación. OAuth es una tecnología de identificación digital accedida por internet mediante un Interfaz de programación de aplicaciones API, es un protocolo para establecer la administración de identidad estándar a través de un servicio web, provee una alternativa para compartir los nombre de usuarios y contraseñas (Leiba y Technologies, 2012). Una vez identificados en la

plataforma podremos realizar las operaciones autorizadas sobre la plataforma de Twitter mientras dura la sesión de autenticación utilizada es decir este proceso también servirá para poder continuar con la solución de los objetivos específicos propuestos.

#### **b) Adquisición de tweets**

Los tweets son las unidades de análisis que se utilizará, los tweets son las publicaciones realizadas sobre la plataforma de Twitter por sus usuarios, para la adquisición de los tweets relacionados se elaboró un listado de palabras clave, estas palabras clave son las que se consultan sobre la plataforma de Twitter además estas siguen el estándar propuesto por la plataforma utilizando el carácter “#”, que permite vincular todos los tweets que tienen la palabra clave de interés, la ejecución de este proceso también servirá para poder continuar con la solución de los objetivos específicos propuestos.

#### **c) Construcción de base de datos**

Para la construcción de la base de datos se agruparon los comentarios de forma legible y de fácil lectura para su análisis, se utilizó el formato Tidy data (Wickham, 2014), este formato agrupa las variables en columnas y cada una de las observaciones en filas a partir de cada uno de los comentarios obtenidos en el proceso descrito en b). La base de datos obtenida organiza la información recolectada al realizar la consulta con la interfaz de programación de la plataforma de Twitter a la base de datos, esta contiene las respuestas en idioma español.

Este proceso también es necesario para la ejecución de cada uno de los objetivos propuestos.

#### **d) Text Mining – Procesamiento del Lenguaje Natural**

Los procedimientos utilizados para obtener estadísticas que representan los contenidos textuales son:

- Post etiquetado
- lematización
- coincidencias
- Anotación de texto
- Análisis de Frecuencias
- Selección automática de palabras clave
- Análisis de coocurrencias relación sustantivos y adjetivos

- Clasificación automática de Texto utilizando Latent Dirichlet Allocation.

### **3.5.2 Conocimientos evidenciados en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.**

Para evidenciar los conocimientos sobre la contaminación del Lago Titicaca que han sido vertidos en la plataforma de microblogging Twitter mediante texto se realizó:

- Post etiquetado
- lematización
- coincidencias
- Anotación de texto
- Análisis de Frecuencias
- Selección automática de palabras clave
- Análisis de coocurrencias relación sustantivos y adjetivos
- Clasificación automática de Texto utilizando Latent Dirichlet Allocation.

Estos procesos brindarán como resultado un listado de tópicos que serán interpretados para explicar cuales son los conocimientos que han sido evidenciados.

### **3.5.3 Actitudes y prácticas expresadas sobre la plataforma de microbloggings Twitter sobre la contaminación del Lago Titicaca.**

La tarea de determinar qué actitudes y qué prácticas han sido expresadas sobre la plataforma de microblogging Twitter referente al tema de la contaminación del Lago Titicaca se realiza:

- Post etiquetado
- lematización
- coincidencias
- Anotación de texto
- Análisis de Frecuencias
- Selección automática de palabras clave



- Análisis de coocurrencias relación sustantivos y adjetivos
- Clasificación automática de Texto utilizando Latent Dirichlet Allocation.

Al igual que para el objetivo específico anterior los procedimientos descritos brindarán como resultado un listado de tópicos que serán interpretados para determinar cuáles son las actitudes y prácticas que han sido expuestas sobre la plataforma de Twitter.

## CAPÍTULO IV RESULTADOS Y DISCUSIÓN

En este capítulo se exponen los resultados obtenidos agrupados según los objetivos planteados tanto general y específicos.

### 4.1. Representaciones sociales expresadas en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.

Se exponen a continuación los resultados del proceso de obtención de información sobre la plataforma Twitter y los resultados obtenidos para la determinación de las representaciones sociales expresadas en la plataforma de microblogging Twitter.

#### 4.1.1 Autenticación mediante protocolo OAuth 2.0 en Twitter.

El proceso de autenticación implementado mediante el protocolo OAuth 2.0 se muestra en el diagrama de procesos de la Figura 6.

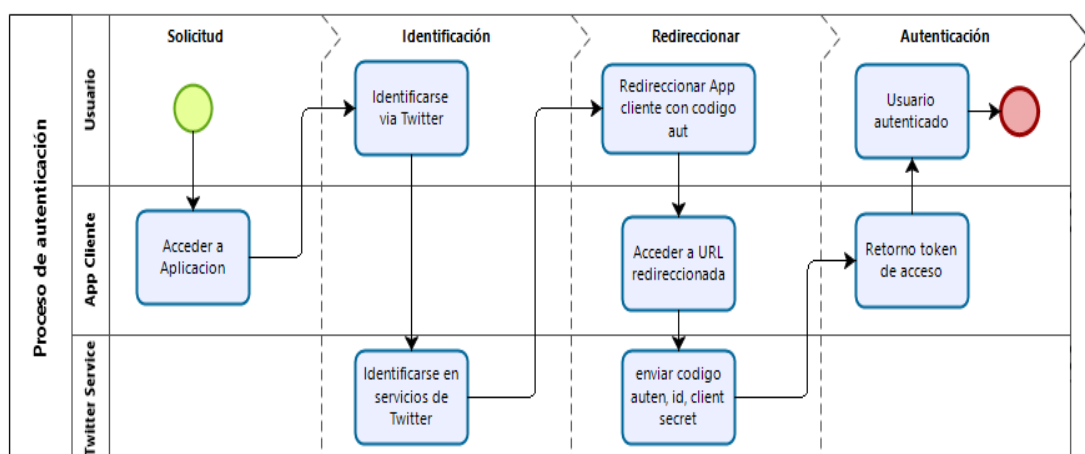


Figura 6. El proceso de autenticación implementado mediante el protocolo OAuth 2.0

En la implementación de este proceso se utilizó el lenguaje de programación “R” (R Core Team, 2019), y el paquete “twitteR” para la administración de la conexión con OAuth y la plataforma de Twitter (Gentry, 2016).

#### 4.1.2 Adquisición de tweets

Para la recolección de los tweets relacionados con el Lago Titicaca se utilizaron las etiquetas mostradas en la Tabla 7. Para realizar este proceso también se utilizó el lenguaje de programación “R”.

Tabla 7

*Etiquetas de búsqueda en Twitter en idioma inglés y español*

Idioma Inglés	Idioma Español
Titicaca	Titicaca
Lago Titicaca	Lago Titicaca
Titicaca Lake	Titikaka
#titicaca	#titicaca
#lagotiticaca	#lagotiticaca
#titicacalake	#titikaka

#### 4.1.3 Construcción de base de datos

La base de datos se construyó con la totalidad de los tweets adquiridos en función de las etiquetas de la tabla 7, la base de datos consta de 17 variables detalladas en la Tabla 8 y de 2499 observaciones.

Una vez hecha la lectura de la información se almacenó el conjunto de datos en la estructura de datos DataFrame del Lenguaje R.

Tabla 8

*Nombre y tipo de variable de la base de datos*

<b>Nombre de Variable</b>	<b>Tipo de datos en R</b>
\$ text	character
\$ favorited	logi
\$ favoriteCount	int
\$ replyToSN	Factor
\$ created	Factor
\$ truncated	logi
\$ replyToSID	num
\$ id	num
\$ replyToUID	num
\$ statusSource	Factor
\$ screenName	Factor
\$ retweetCount	int
\$ isRetweet	logi
\$ retweeted	logi
\$ longitude	num
\$ latitude	num
\$ idioma	Factor

#### **4.1.4 Procesamiento del Lenguaje Natural de la representación social Twitter acerca de la contaminación del Lago Titicaca**

La composición universal del discurso (UPOS – Universal Parts Of Speech), es utilizada para analizar el discurso mediante técnicas de Procesamiento del Lenguaje Natural, se enfoca esta tarea principalmente en el análisis de los sustantivos, adjetivos, verbos y palabras clave (combinación de sustantivos, adjetivos y verbos)

#### **4.1.5 Modelado de la representación social sobre la Plataforma Twitter**

La representación social sobre la plataforma Twitter respecto al Lago Titicaca, es analizada mediante el contenido de las Partes Universales del Discurso, por ello se realizó el modelado de tópicos o temas es decir los conocimientos de los usuarios de Twitter, el modelado se realizó por los sustantivos contenidos en el discurso, a continuación, se realizó el modelado por los adjetivos y un tercer modelo que contiene sustantivos, adjetivos y palabras clave obtenidas mediante el algoritmo RAKE.

#### 4.1.6 Modelo Obtenido sobre sustantivos

Los sustantivos tomados como entidades funcionan como núcleos, en ese sentido el modelo obtenido mediante el método Latent Dirichlet Allocation presenta la agrupación de términos es decir los sustantivos y se propuso agruparlos en 4 temas o tópicos, estos resultados se muestran en la Tabla 9.

Tabla 9

*Agrupación de temas obtenidos mediante el algoritmo LDA modelando los sustantivos.*

topic_001		topic_002		topic_003		topic_004	
term	prob	term	prob	term	prob	term	prob
lago	0.26888	lago	0.53	día	0.0830997	isla	0.152
titicaca	0.17568	agua	0.04	perú	0.06067063	puno	0.134
hombre	0.04563	recuperación	0.04	bolivia	0.05730627	lago	0.06
ano	0.04346	picchu	0.02	viaje	0.04721319	uro	0.052
fiesta	0.02287	titikaka	0.02	copacabano	0.04609173	titicacar	0.049
comunidad	0.01962	plan	0.02	foto	0.03151284	sol	0.038

Haciendo una interpretación del conocimiento de los usuarios de Twitter sobre lago Titicaca basado en esta agrupación de sustantivos se puede observar que el primer tópico hallado agrupa los sustantivos referentes al lago Titicaca y su relación con las actividades culturales del hombre, el segundo tópico agrupa sustantivos que hacen referencia al lago Titicaca y su relación con el agua y su recuperación, a continuación, el tercer tópico agrupa sustantivos referentes a la actividad turística realizada sobre el corredor Perú Bolivia, por último el cuarto tópico agrupa sustantivos también relacionados a la actividad turística esta vez relacionados directamente con el lago Titicaca y sus islas.

#### 4.1.7 Modelo Obtenido sobre adjetivos y sustantivos

El siguiente modelo realizado tomó en consideración los adjetivos y sustantivos identificados de los tweets vertidos en la plataforma Twitter, dado que un adjetivo complementa al sustantivo calificándolo y expresando características propias de los objetos de interés, el modelo hallado por el método Latent Dirichlet Allocation, se muestran en la Tabla 10.



Tabla 10

*Agrupación de temas obtenidos mediante el algoritmo LDA modelando adjetivos y sustantivos.*

topic_001		topic_002		topic_003		topic_004		topic_005		topic_006	
term	prob	term	prob	term	prob	term	Prob	term	prob	term	prob
titicacar	0.265	lago	0.479	lago	0.340	titicaca	0.225	perú	0.081	isla	0.153
lago	0.170	bolivia peruan	0.056	titicaco	0.179	agua	0.047	día	0.077	titicacar	0.118
sol copacaban o	0.065	o	0.053	gran	0.033	hombre	0.045	mejor	0.06	puno	0.117
	0.047	alto	0.026	orilla recuperac ión	0.032	solo	0.043	viaje	0.045	uro	0.077
ano	0.042	parte primer	0.021	ión	0.027	lugar	0.038	nuevo	0.036	ambiental	0.035
paraíso	0.024	o	0.021	experto	0.018	foto	0.033	picchu	0.03	flotante	0.034

De la tabla 10 propone 6 tópicos o temas que incluyen tanto a sustantivos y adjetivos, el primer grupo o tópico evidencia una apreciación positiva sobre el Lago Titicaca, el segundo tópico encontrado muestra contenido referente al corredor turístico entre los países de Perú y Bolivia, el tercer tópico evidencia contenido referente a la recuperación del Lago Titicaca y la participación de expertos en el tema, el cuarto tópico agrupa términos que hacen referencia al lago Titicaca y su ubicación, el quinto tópico encontrado muestra el contenido asociado con otros destinos turísticos del Perú (Machu Picchu) , el contenido evidencia una relación positiva entre ambos destinos turísticos por parte de los usuarios de Twitter, el último tópico hallado agrupa tanto sustantivos y adjetivos relacionados al medio ambiente del Lago además de las islas de los Uros como atractivos turísticos.

#### 4.1.8 Modelo Obtenido Sobre sustantivos, adjetivos y Palabras Clave

Si bien los sustantivos y los adjetivos muestran los contenidos vertidos por los usuarios de Twitter estos pueden perder interpretabilidad, motivo por el que, para hacerlos más interpretables, en el siguiente modelo de temas se incluirán las palabras clave compuestas en el modelo.

La Tabla 11 y 12 muestra la agrupación de temas obtenidos mediante el algoritmo LDA, el modelo de tópicos para todos los sustantivos, todas las palabras clave compuestas que consisten en sustantivos y adjetivos y en todas las frases nominales identificadas, vertidas por los usuarios de Twitter.

Tabla 11

*Agrupación de temas obtenidos mediante el algoritmo LDA modelando adjetivos, sustantivos y palabras clave de tweets acerca del lago Titicaca parte 1.*

topic_1		topic_2		topic_3	
term	prob	term	prob	term	prob
lago titicaca	0.4660	titicaca	0.3812	perú	0.1603
bolivia	0.0464	puno	0.0949	tour	0.1234
#puno	0.0424	hombre	0.0929	lima	0.0782
picchu	0.0323	lago solo	0.0849	días	0.0700
ruta #viajes	0.0223	#lagotiticaca	0.0506	machupicchu	0.0372
ríos	0.0203	bolivia	0.0445	islas flotantes	0.0351
lago navegable	0.0183	lugar	0.0224	paz	0.0351
peruano	0.0183	copacabana #bolivia	0.0204	peruano	0.0331
lake titicaca	0.0163	lugares favoritos	0.0204	fotos	0.0310
película	0.0163	@ambienteyagua	0.0163	gran parte	0.0289
taquile	0.0163	gobierno	0.0123	islas	0.0269
túneles secretos	0.0163	plan	0.0123	lago	0.0269
destinos	0.0143	salar de uyuni	0.0123	cusco	0.0248
lago	0.0143	territorio	0.0123	viaje	0.0207
orillas	0.0143	días visita	0.0103	mejores ofertas	0.0187
turistas nacionales	0.0122	seguro	0.0103	uros	0.0187
cusco	0.0102	-	-	vacaciones	0.0187
forma	0.0102	-	-	viajes	0.0187
lagos	0.0102	-	-	fin	0.0125

Tabla 12

*Agrupación de temas obtenidos mediante el algoritmo LDA modelando adjetivos, sustantivos y palabras clave de tweets acerca del lago Titicaca parte 2*

topic_4		topic_5		topic_6	
term	prob	term	prob	term	prob
lago	0.1146	lago titicaca	0.4432	lago titicaca	0.4129
sol	0.1040	titikaka	0.0582	puno	0.1203
isla	0.0955	pies	0.0333	recuperación ambiental	0.0544
uros	0.0934	copacabana	0.0292	mundo	0.0448
paraíso	0.0426	lago #titicaca	0.0292	alto	0.0331
expertos	0.0341	mauri	0.0250	anos	0.0312
mejores experiencias	0.0320	alternativas	0.0230	nuevo	0.0293
#bolivia	0.0299	cultura	0.0209	carachi	0.0215
mar	0.0299	valle sagrado	0.0168	día	0.0215
binacional	0.0256	lado	0.0147	perú	0.0215
agua	0.0235	tatora	0.0147	ciudad	0.0176
orillas	0.0235	vez	0.0147	copacabana	0.0157
especies	0.0214	ticaca	0.0126	titicaca	0.0157
#titicaca	0.0193	isla	0.0126	brisas	0.0138
manana	0.0172	lake titicaca	0.0126	objetivo	0.0138
peruano	0.0172	msnm	0.0126	llachón	0.0118
aguas residuales	0.0129	Paz	0.0126	picchu	0.0118
gente	0.0129	región puno	0.0126	poder	0.0118
nueva aventura	0.0129	Viaje	0.0126	vista insuperable	0.0118

Los 6 tópicos o temas resultados del modelamiento mostrados en las Tablas 11 y 12 propone también 6 grupos que incluyen tanto a sustantivos, adjetivos y las palabras clave, donde: el primer grupo o tópico trata del corredor turístico alrededor del Lago Titicaca; el segundo tópico también trata de turismo pero esta vez enfocado en el corredor que corresponde al lado Boliviano del Lago Titicaca; El tópico 3 también corresponde a turismo pero enfocado esta vez al corredor correspondiente al lado peruano del lago Titicaca; el grupo de términos del tópico 4 indica las islas como destino turístico (los uros y del sol) pero además, también indica aspectos referidos a la ecología del lago como el manejo de aguas residuales y las especies que habitan en él; el quinto tópico agrupa aspectos culturales y hace referencia también a especies endémicas como el mauri y la totora; en el último tópico, el tópico 6 se agrupan los términos que hacen referencia directa a temas ambientales como la recuperación del medio ambiente, especies endémicas como el carachi además de evidenciar sentimiento positivos.

Si bien las Tabla 11 y 12 resumen la agrupación de los temas es más fácil apreciar la relación entre los términos de cada tópico de forma gráfica, para ello se muestran a continuación los grafos correspondientes a cada tópico.

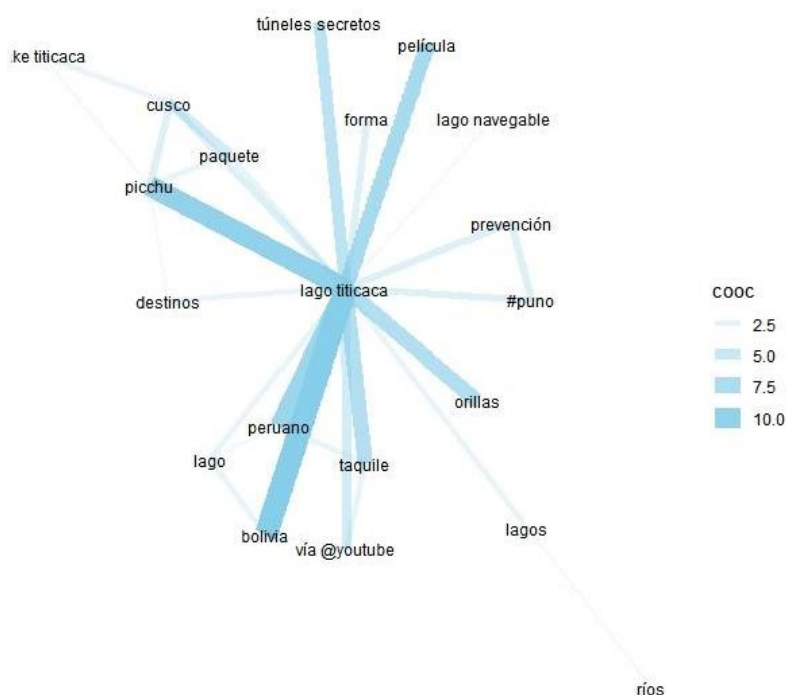


Figura 7. Grafo de coocurrencias sobre el tópico/tema 1 construido sobre sustantivos adjetivos y palabras clave.

La Figura 7 muestra el grafo de coocurrencias para el primer grupo o tópico que trata del corredor turístico alrededor del Lago Titicaca, en este gráfico se aprecia las fuertes relaciones entre el lago Titicaca con los términos que hacen referencia a su nacionalidad (peruano o boliviano), sus atractivos turísticos (la isla de Taquile, sus orillas), su relación con el mayor destino turístico del Perú Machu Picchu. También se observa relaciones con los términos que evidencian aspectos turísticos como: #puno, destinos, paquete, cusco, vía @youtube. Y otros aspectos con relaciones más débiles como: lagos, ríos, prevención, lago navegable.

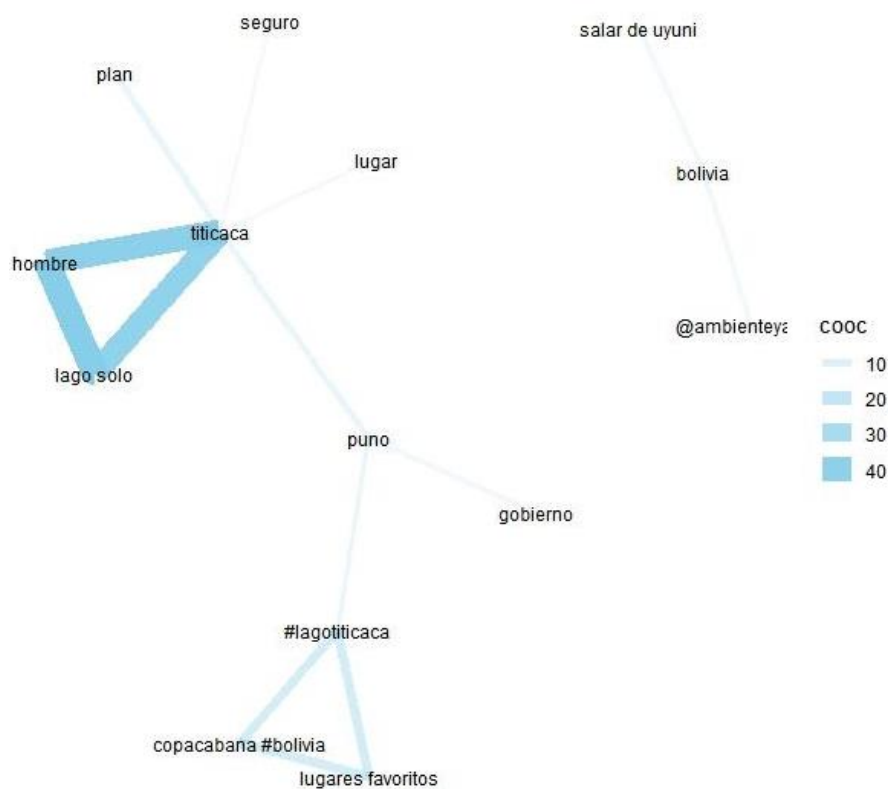


Figura 8. Grafo de coocurrencias sobre el tópico/tema2 construido sobre sustantivos adjetivos y palabras clave.

La Figura 8 el segundo tópico catalogado como el turismo enfocado en el corredor que corresponde al lado Boliviano del Lago Titicaca, el grafo mostrado en la figura muestra en primer lugar una asociación fuerte entre los términos “hombre”, “Titicaca” y “lago solo”, términos que podrían evidenciar la relación entre el Lago Titicaca y las actividades humanas a su alrededor, en este mismo grupo se muestran los términos con relaciones más débiles, donde destaca que para hacer referencia al destino turístico de “copacabana #bolivia” esta referenciada por el término “#lagotiticaca” y esta a la vez por “puno”. También en este grupo se muestra términos débilmente

referenciados por “titicaca” como “plan”, “lugar” y “seguro”. Otro aspecto importante que destacar de este grafo es la agrupación independiente de los términos “Bolivia”, “salar de Uyuni” y “@ambienteyagua”.

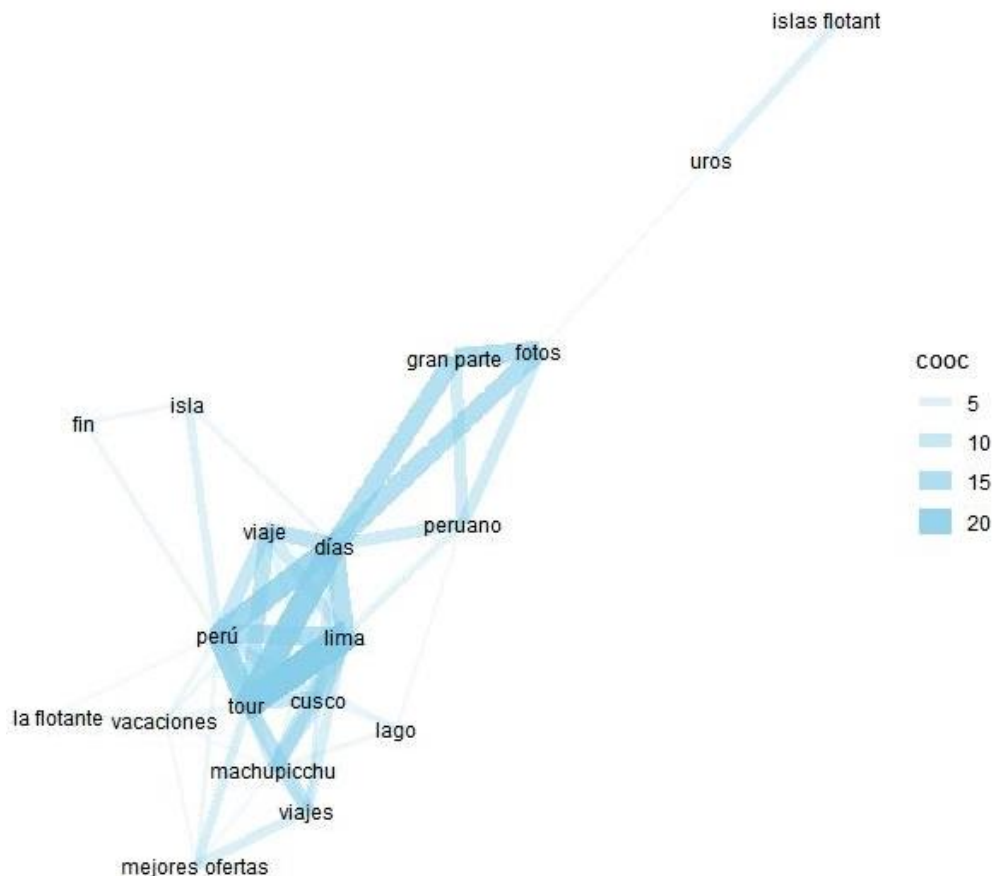


Figura 9. Grafo de coocurrencias sobre el tópico/tema 3 construido sobre sustantivos adjetivos y palabras clave

En la Figura 9 se aprecia el grafo correspondiente al tópico 3, este tópico también corresponde a turismo pero enfocado esta vez al corredor correspondiente al lado peruano del lago Titicaca, como se puede apreciar en el grafo se ve de manera más articulada las relaciones marcadas por las coocurrencias de los términos utilizados por los usuarios de Twitter estos términos son: “gran parte”, “todos”, “peruano”, “días”, “viaje”, “perú”, “tour”, “Lima”, “cusco”, “machupicchu”, “viajes”; también se puede apreciar en una escala menor la relaciones de los términos “isla flotante”, “vacaciones”, “mejores ofertas”, “lagos”, “isla”, “fin”, “uros”; que también hacen referencia al aspecto turístico del Lago Titicaca.

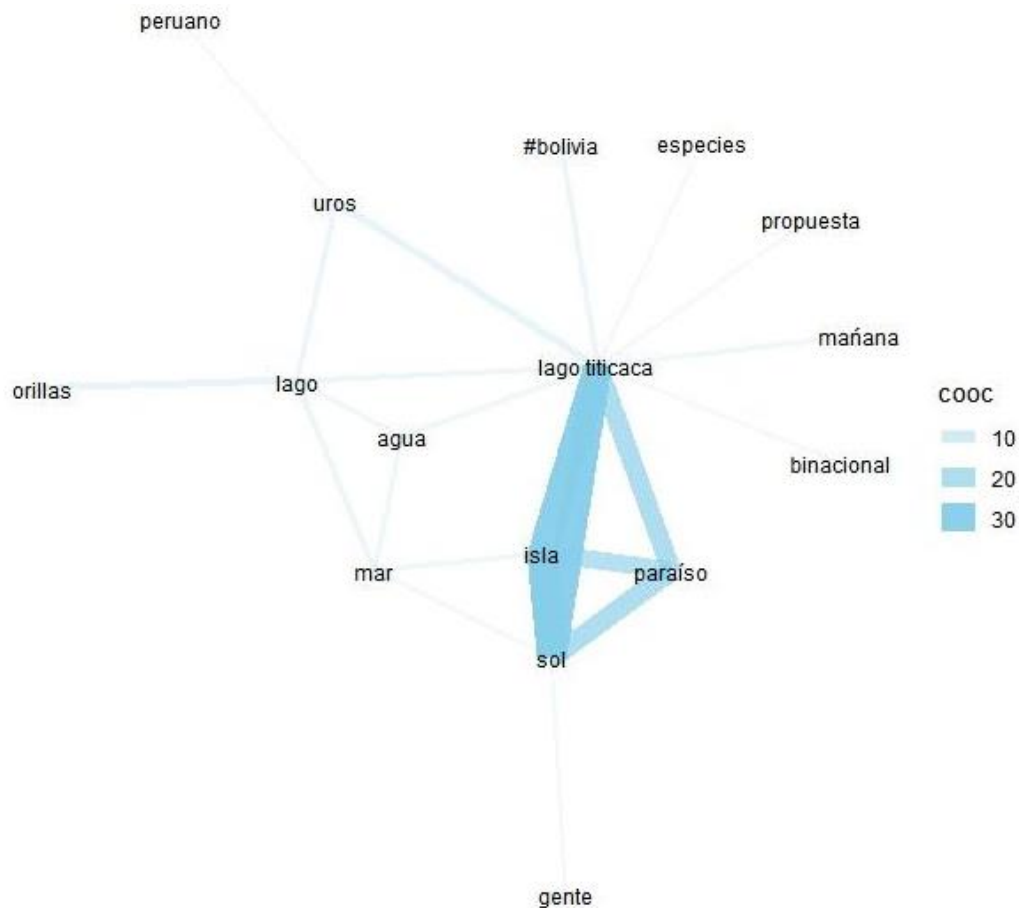


Figura 10. Grafo de coocurrencias sobre el tópicos/tema 4 construido sobre sustantivos adjetivos y palabras clave

El grafo mostrado en la Figura 10, resume el grupo de términos del tópicos 4, este indica a las islas del sol y los uros, como destino turístico e indica aspectos referidos a la ecología. Se observa del grafo que existe una relación marcada entre los términos “lago titicaca”, “paraíso”, “isla” y “sol”; también se observan relaciones más moderadas cerca a “Lago Titicaca” con los términos “agua”, “uros”, “gente”, “mar”, “lago”, “orillas”, “#bolivia”, “especies”, “propuesta”, “propuesta”, “mañana”, “peruano” y “binacional”; que se muestran de forma muy dispersa y entre estos el término “peruano” se relaciona con el término “lago Titicaca” por intermedio de la palabra uros.

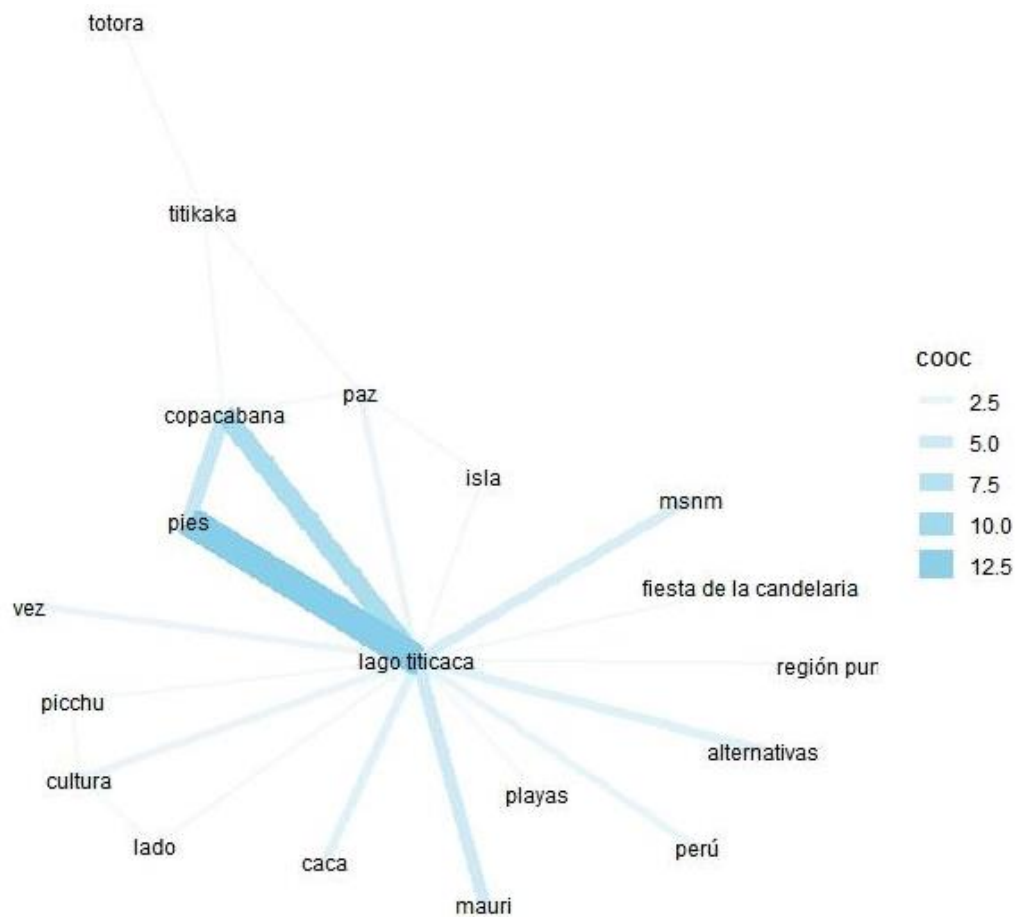


Figura 11. Grafo de coocurrencias sobre el tópico/tema 5 construido sobre sustantivos adjetivos y palabras clave

En la Figura 11 se muestra el grafo que pertenece al tópico 5, este quinto tópico agrupa aspectos culturales y hace referencia también a especies endémicas del Lago Titicaca, de este grafo se aprecia también que el centro de las relaciones o relaciones con mayor co ocurrencia son los términos “lago titicaca”, “Copacabana” y “pies”; siguen las relaciones del término “lago Titicaca” con los términos “msnm”, “mauri” y “alternativas”; también se observa relaciones más discretas con los términos “isla”, “paz”, “isla”, “fiesta de la candelaria”, “región puno”, “perú”, “playas”, “lado”, “cultura”, “picchu”, “vez”; por último se aprecia también las relaciones dispersas entre el término “titikaka” tanto con los términos la “paz” y “copacabana” ambos que hacen referencia a la ubicación geográfica por el lado boliviano.

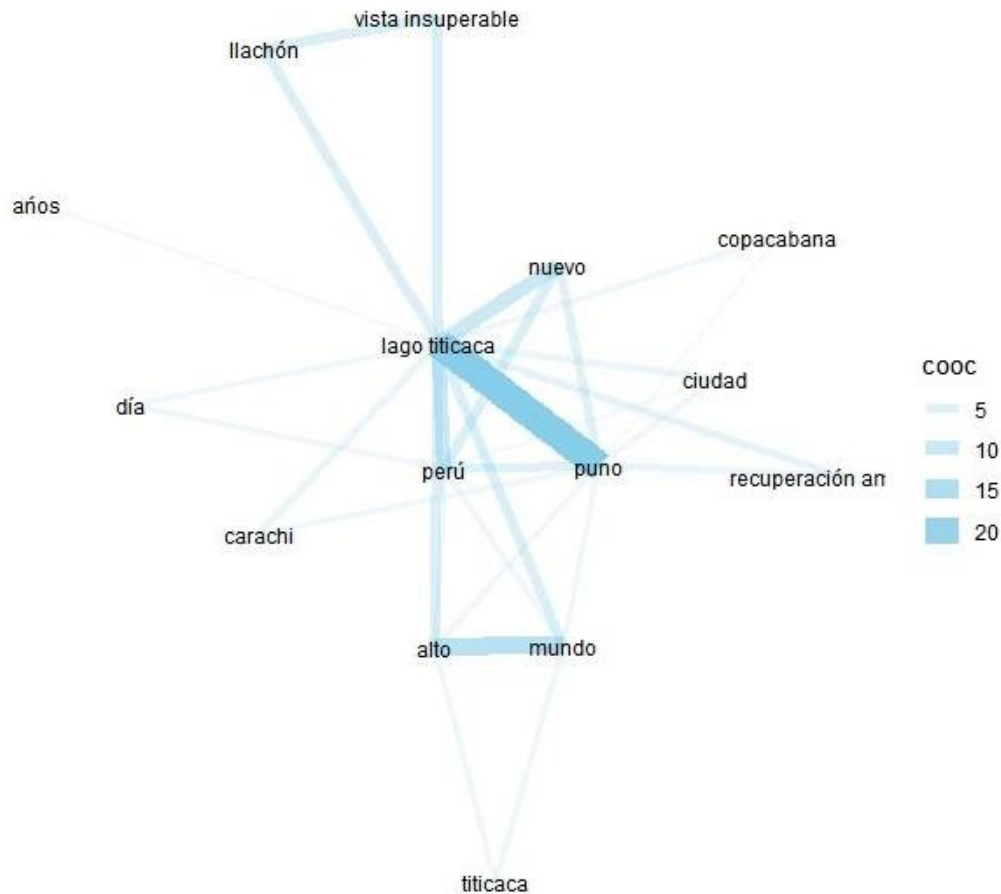


Figura 12. Grafo de coocurrencias sobre el tópico/tema 6 construido sobre sustantivos adjetivos y palabras clave

La Figura 12 corresponde al grafo del tópico 6, este tópico que agrupa los términos que hacen referencia a temas ambientales como “recuperación ambiental”, “carachi” (especie endémica del Titicaca); también los términos “llachon”, “vista insuperable”, “Copacabana”, “ciudad”, “nuevo”, “perú”, “día”, “mundo” y “alto” que están relacionados con el turismo y que evidencian una apreciación positiva de los usuarios de twitter sobre el Lago Titicaca.

#### 4.1.9 Discusión sobre las representaciones sociales obtenidas

Las cuentas de usuario de la plataforma de microblogging Twitter conserva las representaciones individuales en la percepción de los objetos y comunicación con otros individuos característica similar a lo que expone Jodelet (2008, 1986). Al ser analizados cada una de las expresiones vertidas por los usuarios mediante el algoritmo LDA, además, de analizar su representación gráfica mediante redes de grafos, estos



nos brindan evidencias producto de aspectos sociales como indica Vera (2002), Estas evidencias de las representaciones son mostrados como temas o tópicos.

Según la clasificación de Moscovici (1988) las representaciones sociales encontradas son de tipo hegemónicas, dado que en los 6 tópicos hallados existe un consenso sobre el un conocimiento en los temas ambientales y sobre todo de la actividad turística realizada sobre el Lago Titicaca actividad que contribuye a la contaminación y afecta al medio ambiente. Acerca de la estructura de las Representaciones Sociales que destaca Abric (2004), se halló en la dimensión afectiva expresiones positivas acerca del Lago Titicaca vertidas sobre la plataforma de microblogging Twitter, en la dimensión de los conocimientos destacan la ubicación geográfica los atractivos naturales propios del Lago y los que están relacionados por los corredores turísticos que involucran a los países de Perú y Bolivia, por otro lado en la dimensión del campo de la representación el núcleo también gira en torno a las actividades propias del turismo.

Respecto a las representaciones sociales medio ambientales se evidenció que los primeros tres tópicos pertenecen a una Representación Social del tipo Antropocéntrica dado tienen temática sobre las actividades humanas que impactan en el medio ambiente como el corredor turístico alrededor del Lago Titicaca, turismo que corresponde al lado Boliviano del Lago Titicaca y turismo correspondiente al lado peruano del lago Titicaca, como es mencionada por Raúl Calixto Flores (2013) y Reigota (2004) para las representaciones sociales del medio ambiente.

Los siguientes tres tópicos tienen temática referente a aspectos ecológicos, ambientales, turísticos y culturales, además, de evidenciar sentimientos positivos. Las representaciones sociales ambientales que corresponden a este tópico son la Globalizante dado que las evidencias muestran interacciones entre los aspectos sociales y naturales; y que en determinados términos se adecuan también a una representación antropocéntrica cultural.

La representación Antropocéntrica hallada concuerda parcialmente con los hallazgos de: los estudios de educación ambiental de Andrade et al. (2004), las representaciones sociales del medio ambiente en estudiantes realizada por Flores (2008), con los hallazgos de Tochetto (2009) que en resumen presenta una representación social antropocéntrica utilitaria y con Navarro (2013) que trata la exposición a fuentes de

polución del aire; que si bien no tratan de forma directa el tema propuesto por la presente investigación marcan un precedente de estudios ambientales basados en la teoría de las representaciones sociales.

Referente a las representaciones sociales Globalizante y Antropocéntrica cultural que fueron encontradas se aproximan a los hallazgos de: González y Valdez, (2012) que habla de investigaciones de educación ambiental, (Flores, 2010) que evidencia la existencia de una relación entre las representaciones del medio ambiente y las corrientes de la educación ambiental, Rodríguez y Paba (2013) que relaciona la crisis ambiental con enfermedades y contaminación, Piermattéo et al. (2016) resuelve tener en cuenta las representaciones para explicar o prevenir la retirada de programas proambientales, Polli y Camargo (2016) que trata la forma de pensar que reconoce la fragilidad y necesidad de atención del medio ambiente; que pese a no tener una relación directa con la contaminación del lago Titicaca dan un marco de referencia sobre las representaciones sociales halladas sobre la problemática ambiental.

#### **4.2. Conocimientos evidenciados en la plataforma de microblogging Twitter acerca de la contaminación del Lago Titicaca.**

Los conocimientos recuperados de los datos obtenidos sobre los tweets fueron realizados mediante el análisis del discurso tomando como centro a los sustantivos y las palabras clave identificadas mediante el algoritmo RAKE.

En la Tabla 13 se resume los 20 primeros sustantivos identificados por frecuencia y su respectiva frecuencia porcentual que constituyen el 32.60% de la totalidad de sustantivos hallados en los tweets vertidos por los usuarios respecto al Lago Titicaca; de estos sustantivos “lago” ocupa el 13.73% de ellos, “Titicaca” ocupa el 3.96%, “Puno” ocupa el 2.10%, “isla” ocupa el 1.66%, “Bolivia” ocupa el 1.08%, continua “Perú” con el 1.03%, el 1.02% lo ocupa “días”, el 0.97% está ocupado por “Uros”, el 0.88% es ocupado por “sol”, el 0.78% por “hombre”, el 0.73% lo ocupa “Copacabana”, “islas” tiene una frecuencia del 0.66%, “paz” tiene una frecuencia de 0.58%, seguidas del 0.56% que corresponde a “años”, el 0.47% “mundo”, “orillas” y “picchu” con 0.46%, “#lagotiticaca” con 0.42% y “parte” con 0.41%.

Esta información también se puede apreciar en el Gráfico de barras de la Figura 13.

Tabla 13

*Sustantivos más utilizados ordenados por mayor frecuencia*

Nro	key	freq	freq_pct
1	lago	811	13.73
2	titicaca	234	3.96
3	puno	124	2.10
4	isla	98	1.66
5	bolivia	64	1.08
6	perú	61	1.03
7	días	60	1.02
8	uros	57	0.97
9	sol	52	0.88
10	hombre	46	0.78
11	copacabana	43	0.73
12	islas	39	0.66
14	paz	34	0.58
15	anos	33	0.56
16	mundo	28	0.47
17	orillas	27	0.46
18	picchu	27	0.46
19	#lagotiticaca	25	0.42
20	parte	24	0.41

Como se aprecia los sustantivos de mayor frecuencia hacen alusión directa al Lago Titicaca y rescata como su principal característica su ubicación geográfica referenciada en primera instancia por la ciudad de Puno, luego por el país de Bolivia, seguido por Perú, la ciudad de Copacabana, la ciudad de la Paz; es decir el conocimiento evidenciado en los principales destinos del corredor turístico del lago entre Puno y Bolivia, además también se podría incluir en este grupo al término picchu que hace referencia a Machu Picchu. También en menor frecuencia los términos de isla, Uros, sol, islas y orillas evidencia el conocimiento de los atractivos al interior y propios del Lago Titicaca. Y los términos mundo, parte, días, hombre, años por si solos no guardan una relación directa con el Lago, a continuación, se expone la evaluación de términos y las palabras clave.

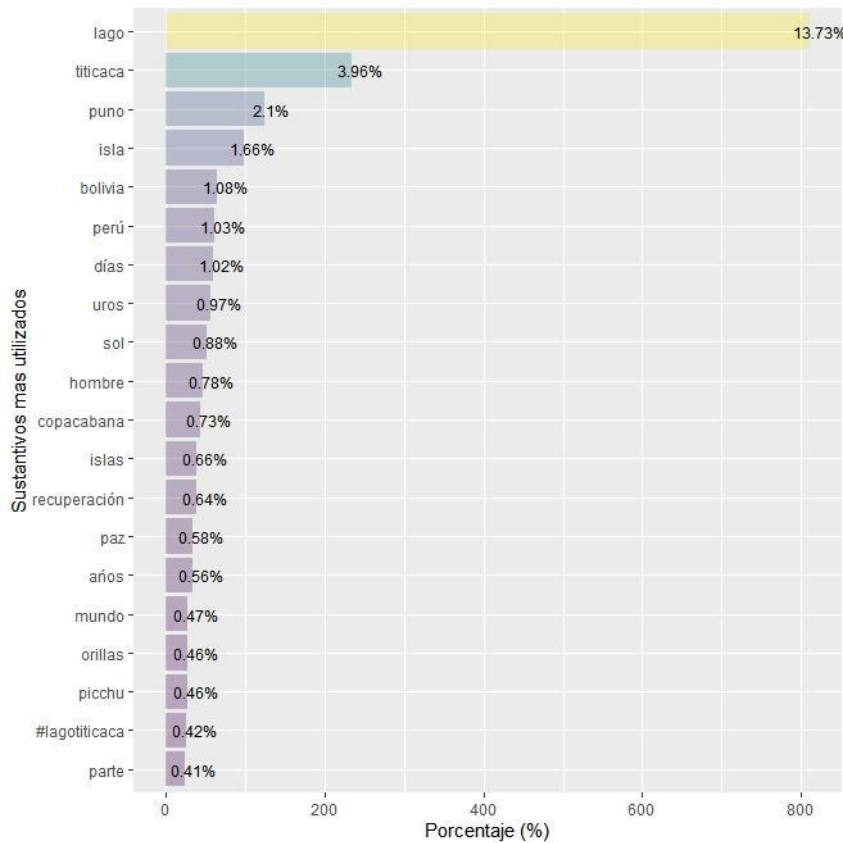


Figura 13. Gráfico de barras de los Sustantivos más utilizados ordenados por mayor frecuencia

En la Tabla 14 se pueden apreciar las palabras clave obtenidas de los tweets vertidos en la plataforma de Twitter que hacen referencia al Lago Titicaca ordenados según el puntaje obtenido por el algoritmo RAKE, de forma gráfica se pueden observar en las las Figuras 13 y 14, Ejemplo de los tweets de los que fue obtenida las palabras clave son:

*“Comisión multisectorial para la prevención y recuperación ambiental de la cuenca del lago Titicaca debe ser abierta...”.*

*“#Puno: Se inició sesión ordinaria de la Comisión Multisectorial para la prevención y recuperación ambiental del lago...”.*

*“Puno: Expertos proponen tecnología móvil para descontaminar el lago Titicaca”.*

*“Bolivia y Perú: buenas noticias para la rana gigante del Titicaca”.*

Destacan en estos resultados las palabras clave “comisión multisectorial” con un puntaje de 2.4 cuyo, “tecnología móvil” con un puntaje de 2, “recuperación ambiental” con un

puntaje de 1.8, “rana gigante” con un puntaje de 1.67, “gobierno regional” también con 1.67 y “agua residual” con 1.37; las palabras claves listadas constituyen elementos del campo **de conocimiento que hacen referencia a la contaminación del Lago Titicaca.**

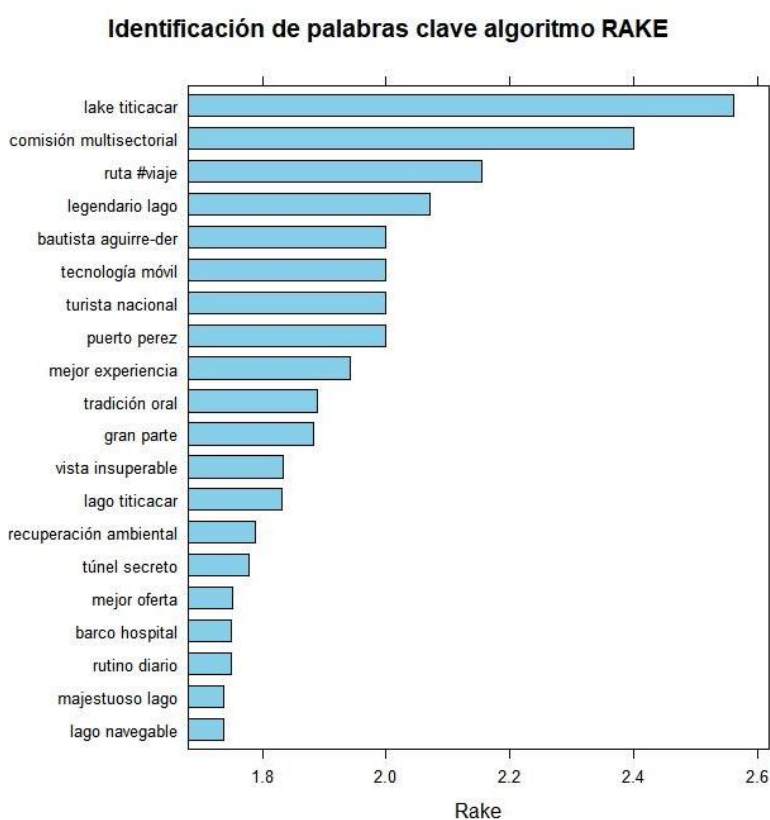
Tabla 14

*Puntaje de palabras clave identificadas con el algoritmo Rake.*

Id	keyword	fq	rake	key	Id	keyword	fq	Rake	key
9	lake titicacar	5	2.6	lake titicacar	179	lago menor	4	1.74	lago menor
17	comisión multisectorial	6	2.4	comisión multisectorial	180	lago solo	44	1.74	lago solo
28	ruta #viaje	12	2.2	ruta #viaje	189	lago titicaco	100	1.72	lago titicaco
33	legendario lago	4	2.1	legendario lago	200	rana gigante	6	1.67	rana gigante
59	bautista aguirre-der	4	2	bautista aguirre-der	201	gobierno regional	4	1.67	gobierno regional
121	tecnología móvil	4	2	tecnología móvil	208	valle sagrado	8	1.67	valle sagrado
122	turista nacional	6	2	turista nacional	210	#titicaco #isla sonar	4	1.64	#titicaco #isla
128	puerto perez mejor	4	2	puerto perez mejor	221	increíble	6	1.57	sonar increíble
135	experiencia	16	1.9	experiencia	223	rio sar	5	1.56	rio sar
141	tradición oral	6	1.9	tradición oral	225	isla flotante	22	1.53	isla flotante
143	gran parte	15	1.9	gran parte	227	primero vez	4	1.52	primero vez
144	vista insuperable	8	1.8	vista insuperable	238	especie nativo	4	1.50	especie nativo
145	lago titicacar recuperación ambiental	251	1.8	lago titicacar recuperación ambiental	248	lago #titicacar	6	1.49	lago #titicacar
149	recuperación ambiental	28	1.8	recuperación ambiental	251	lugar favorito	13	1.47	lugar favorito
152	túnel secreto	8	1.8	túnel secreto	257	magico tour	4	1.43	magico tour
154	mejor oferta	10	1.8	mejor oferta	267	lago #titicaco	5	1.38	lago #titicaco
159	barco hospital	5	1.8	barco hospital	270	isla amantaní	4	1.37	isla amantaní
165	rutino diario	6	1.8	rutino diario	274	agua residual	6	1.37	agua residual
177	majestuoso lago	4	1.7	majestuoso lago	280	fútbol peruano copacabano	4	1.33	fútbol peruano copacabano
178	lago navegable	9	1.7	lago navegable	304	#bolivio	13	1.22	#bolivio

Estos resultados muestran un conocimiento e interés de cómo son sistematizados los esfuerzos por los actores sociales y lo piden mediante una “comisión multisectorial” y la participación del “gobierno regional” para la “recuperación ambiental” proponiendo tecnologías de recuperación como la “tecnología móvil”; manejo adecuado de “aguas residuales” y recuperación de especies endémicas como la “rana gigante”. Esto ante evidencias de los estudios de Fontúrbel (2005) y Archundia et al. (2017) que hallaron contaminación de la bahía del lado boliviano del lago por influencia de las actividades humanas; de Beltrán Farfán et al. (2015) que encontró el área próxima a la isla Espinar una zona crítica de contaminación en la bahía interior de Puno; de Moreno et al. (2018),

que encontró diferentes niveles de concentración de metales en los sedimentos superficiales de la bahía interior de Puno, pese a no superar las concentraciones de la norma ambiental Interim Sediment Quality Guideline, Canadá; de Gutiérrez (2018) que evidencia los efectos del manejo de la producción de trucha y el impacto causado a nivel ambiental de la región de Puno; de Cahui (2019) que halló antibióticos de uso veterinario en el agua superficial en la zona sur del lago Titicaca y en el agua potable de la ciudad de Puno; y de Escobar (2019), que indica que algunas zonas del Lago Titicaca muestran indicios de contaminación debido al vertimiento de residuos urbanos, relaves mineros y nutrientes inorgánicos derivados de la crianza de truchas.



*Figura 14.* Gráfico de barras para la identificación de palabras clave con el algoritmo Rake.

Los conocimientos acerca de la contaminación del Lago Titicaca muestran ser diferenciados al mostrar las evidencias que los conocimientos se refieren a la contaminación, preservación y recuperación del lago y otro a las actividades turísticas sobre atractivos naturales del Lago Titicaca.





Tabla 15

*Adjetivos con más ocurrencia ordenados por frecuencia*

Nro	key	freq	freq_pct	Nro	key	freq	freq_pct
1	titicaca	440	21.39	21	multisectorial	10	0.49
2	solo	44	2.14	22	binacional	10	0.49
3	ambiental	37	1.80	23	visita	10	0.49
4	mejores	33	1.60	24	nacionales	9	0.44
5	gran	31	1.51	25	increíble	9	0.44
6	lago	28	1.36	26	secretos	9	0.44
7	peruano	22	1.07	27	navegable	9	0.44
8	mejor	20	0.97	28	taquile	9	0.44
9	alto	19	0.92	29	excelente	8	0.39
10	flotantes	19	0.92	30	buenas	8	0.39
11	nuevo	18	0.88	31	grandes	8	0.39
12	#bolivia	16	0.78	32	primer	8	0.39
13	#titicaca	15	0.73	33	flotante	8	0.39
14	uros	15	0.73	34	insuperable	8	0.39
15	nueva	14	0.68	35	boliviana	7	0.34
16	favoritos	13	0.63	36	lima	7	0.34
17	#viajes	13	0.63	37	impresionantes	7	0.34
18	sagrado	13	0.63	38	juntos	7	0.34
19	perú	13	0.63	39	primera	7	0.34
20	hermoso	12	0.58	40	san	7	0.34

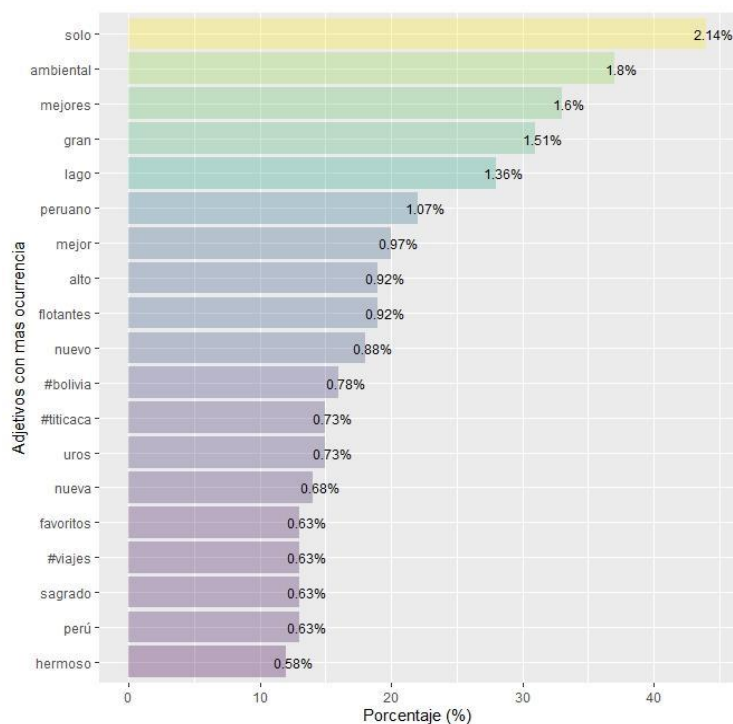


Figura 16. Gráfico de barras para los adjetivos con mayor frecuencia.



La Tabla 16 muestra las 40 primeras coocurrencias encontradas de los sustantivos y adjetivos que fueron utilizados en la misma oración por los usuarios de Twitter. Estos resultados nos permiten realizar una representación gráfica de las relaciones entre los adjetivos y sustantivos y poder determinar las relaciones que existen.

De las coocurrencias halladas y mostradas en la Tabla 16 y el grafo de la Figura 17 Se evidenciaron actitudes positivas frente a la responsabilidad de preservación y recuperación ambiental del Lago Titicaca, así mismo se evidenció también una actitud positiva ante la a las actividades turísticas que se realizan en el Lago Titicaca sin una relación directa con el tema ambiental pese a ser una fuente de contaminación.

Tabla 16

*Frecuencia de coocurrencias de sustantivos y adjetivos utilizados en la misma oración*

Nro	term1	term2	cooc	Nro	term1	term2	cooc
1	lago	titicacar	301	21	ambiental	recuperación	28
2	lago	titicaco	129	22	lago	recuperación	28
3	isla	lago	103	23	sol	titicacar	28
4	lago	titicaca	93	24	día	perú	26
5	lago	puno	88	25	día	viaje	26
6	isla	titicacar	64	26	bolivia	lago	25
7	hombre	lago	48	27	ambiental	lago	24
8	lago	solo	46	28	día	lago	24
9	isla	sol	45	29	flotante	lago	23
10	hombre	solo	44	30	copacabano	titicacar	23
11	lago	sol	37	31	día	lima	22
12	hombre	titicaca	36	32	lago	paz	22
13	lago	uro	36	33	isla	puno	22
14	lago	peruano	35	34	titicacar	uro	22
15	puno	titicacar	35	35	gran	lago	21
16	isla	uro	35	36	alto	lago	20
17	solo	titicaca	34	37	foto	lago	20
18	flotante	isla	33	38	lago	orilla	20
19	lago	perú	32	39	isla	paraíso	20
20	copacabano	lago	29	40	lago	paraíso	20

En detalle, como se aprecian en el grafo de coocurrencias de la Figura 17, las relaciones identificadas mediante las coocurrencias muestran en torno al término lago que se relaciona de forma directa con los términos “titicaca”, “#titicacar”, “uro”, “solo”, “legendario”, “navegable”, “grande”, “titicaco”, “titcacar” también mediante un término de conexión los términos los términos “alto” por intermedio de “navegable”, “Bolivia”



en la frecuencia, la correlación entre 2 términos también puede ser altas incluso si 2 términos ocurren solo una pequeña cantidad de veces, pero siempre aparecen juntos. Las correlaciones perfectas ocurren entre los términos insuperable y llachón, carachi y mauri, ecosistema y peligro, estudiante y mil, oral y sonar, internacional y turista, mil y turista; que evidencias expresiones del tipo globalizantes dado que involucran elementos de la naturaleza y su interacción con las actividades humanas que pueden interpretarse como actitudes positivas. A continuación sobre el cuerpo de expresiones vertidas se muestran las correlaciones altas entre los términos “hombre” y “solo, biorremediación y fundamento, espejo y salar, oral y tradición, sonar y tradición, diario y rutino, estrecho y panorama, #internacional y nacional, mil y nacional, nacional y turista, altura y metro, huata y Santiago, biorremediación y sar, fundamento y sar, #lagotiticaca y favorito, increíble y oral, increíble y sonar, gigante y rana, insuperable y vista, llachón y vista, ambiente y ministerio, ambiente y técnico, ministerio y técnico, secreto y túnel, biorremediación y río, increíble y tradición, fundamento y río, biorremediación y gestión; asociaciones que pueden clasificarse como naturalistas y globalizantes y que muestran una actitud positiva de los usuarios de twitter frente a los temas ambientales como recuperación y conservación del lago Titicaca.

Tabla 17

*Correlaciones de Pearson para las palabras clave sobre matrices dispersas*

id	term1	term2	Corr	id	term1	term2	Corr
36436	insuperable	llachón	1	40641	altura	metro	0.876
39323	carachi	mauri	1	59565	huata	santiago	0.874
48764	ecosistema	peligro	1	59757	biorremediacion	sar	0.874
54959	estudiante	pureza	1	59817	fundamento	sar	0.842
40891	#internacional	mil	1	23407	#lagotiticaca	favorito	0.826
62041	oral	sonar	1	45850	increíble	oral	0.816
67791	#internacional	turista	1	61990	increíble	sonar	0.816
67941	mil	turista	1	55517	gigante	rana	0.815
61714	hombre	solo	0.98079	71944	insuperable	vista	0.815
26401	biorremediacion	fundamento	0.96354	71959	llachón	vista	0.815
58992	espejo	salar	0.94864	41181	ambiente	ministerio	0.799

66345	oral	tradición	0.9370 3	62970	ambiente	técnico	0.799 48
66405	sonar	tradición	0.9370 3	63100	ministerio	técnico	0.799 48
58446	diario	rutino	0.9126 5	67474	secreto	túnel	0.790 54
46619	estrecho	panorama	0.9044 9	57336	biorremediación	rio	0.786 14
43312	#internacional	nacional	0.8845 5	66294	increíble	tradición	0.764 32
43462	mil	nacional	0.8845 5	57396	fundamento	rio	0.757 14
67950	nacional	turista	0.8845 5	27208	biorremediación	gestión	0.751 11

La identificación de las prácticas realizadas se muestra en la Tabla 18 mediante la agrupación de verbos identificados sobre los tweets vertidos por los usuarios de Twitter agrupados en 2 tópicos o temas.

Las prácticas evidenciadas son **divergentes** como muestra el primer tópico que agrupa a los verbos ir, tener, salvar, ver, saber, ser, viajar, encontrar, llegar, buscar; que son prácticas variadas y que son características de las representaciones sociales polémicas.

El segundo tópico revela una **representación social hegemónica** de las prácticas como se evidencia en el grupo conformado por los verbos hacer, conocer, estar, visitar, navegar, descontaminar, lograr, creer; dado que los verbos mencionados involucran al emisor con la interacción del Lago Titicaca, donde destaca la presencia del verbo descontaminar que en combinación con los demás verbos revela justamente la representación social hegemónica.

Tabla 18

*Agrupación de temas obtenidos mediante el algoritmo LDA modelado basado en los verbos de los tweets*

topic_001		topic_002	
term	prob	term	prob
ir	0.10690853	titicacar	0.29127745
tener	0.07989193	hacer	0.0668446
salvar	0.06831339	conocer	0.04258158
ver	0.06702689	estar	0.04258158
saber	0.05416184	visitar	0.03530268
ser	0.04772932	inca	0.02923693
viajar	0.03743728	navegar	0.02802378
encontrar	0.03100476	descontaminar	0.02559748
llegar	0.02457224	lograr	0.02074487
buscar	0.02328573	creer	0.01710542



Como se apreció el análisis reveló que las actitudes mostradas sobre la plataforma de Twitter son positivas ante la responsabilidad de recuperación y conservación del Lago Titicaca, y que las prácticas expresadas en verbo sobre la plataforma de microbloggings Twitter sobre la contaminación son heterogéneas como se aprecia al tener conjuntos de verbos claramente dispersos.

## CONCLUSIONES

Mediante las técnicas de text mining aplicadas a los contenidos de la red social twitter sobre el lago Titicaca se pudo evidenciar que con respecto a la contaminación del lago existen representaciones sociales de tipo hegemónicas sobre la temática ambiental y la actividad turística realizada sobre el Lago Titicaca actividad que contribuye a la contaminación y afecta al medio ambiente. sobre la estructura se halló posturas positivas y conocimientos de las bondades turísticas que posee y que el núcleo gira en torno a las actividades turísticas. Respecto a las representaciones sociales medio ambientales se evidenció las del tipo Antropocéntrica, Globalizante y antropocéntrica cultural.

Los conocimientos acerca de la contaminación del Lago Titicaca muestran ser diferenciados identificando la contaminación ambiental, el relacionado al turismo, preservación y recuperación del lago. Basados en conocimiento sobre los esfuerzos realizados por los actores sociales y proponiendo tecnologías para recuperar los sectores dañados del lago Titicaca.

Respecto a las actitudes, como se apreció el resultado del análisis reveló que las actitudes mostradas sobre la plataforma de Twitter son positivas ante la responsabilidad de recuperación y conservación del Lago Titicaca, y que por otro lado las prácticas expresadas son heterogéneas como se aprecia al tener conjuntos de verbos muy dispersos.

## RECOMENDACIONES

Se recomienda ampliar el horizonte del estudio para aplicar las técnicas de minería de textos a un ámbito mayor abarcando muchas más redes sociales y tiempo, esto con el fin de incrementar la precisión de los resultados y abarcar las dimensiones de estudio que influyen en los conocimientos actitudes y prácticas de las personas referentes a la recuperación y preservación del Lago Titicaca..

Dada la identificación de un conocimiento divergente entre los usuarios además de la creciente participación de la ciudadanía en las redes sociales, se sugiere adoptar políticas educativas y multisectoriales sobre medios digitales, con la finalidad de mantener a los ciudadanos mejor informados y educados, para poder usar y preservar nuestro recurso hídrico lago Titicaca de forma responsable.

Dada la identificación de actitudes positivas y de prácticas heterogéneas ante la responsabilidad de recuperación y conservación del Lago Titicaca se sugiere integrar a la sociedad en conjunto en tareas de preservación y recuperación articulando esfuerzos con los representantes de la población.

## BIBLIOGRAFÍA

- Albero, C. T. (2005). Representaciones sociales de la ciencia y la tecnología. *Reis*, *111*, 9. <https://doi.org/10.2307/40184698>
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., y Kochut, K. (2017). *A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques*. <http://arxiv.org/abs/1707.02919>
- Andrade Júnior, H. de, Souza, M. A. de, y Brochier, J. I. (2004a). Representação social da educação ambiental e da educação em saúde em universitários. *Psicologia: Reflexão e Crítica*, *17*(1), 43–50. <https://doi.org/10.1590/s0102-79722004000100007>
- Andrade Júnior, H. de, Souza, M. A. de, y Brochier, J. I. (2004b). Representação social da educação ambiental e da educação em saúde em universitários. *Psicologia: Reflexão e Crítica*, *17*(1), 43–50. <https://doi.org/10.1590/S0102-79722004000100007>
- Archundia, D., Duwig, C., Spadini, L., Uzu, G., Guédron, S., Morel, M. C., Cortez, R., Ramos Ramos, O., Chincheros, J., y Martins, J. M. F. (2017). How Uncontrolled Urban Expansion Increases the Contamination of the Titicaca Lake Basin (El Alto, La Paz, Bolivia). *Water, Air, and Soil Pollution*, *228*(1). <https://doi.org/10.1007/s11270-016-3217-0>
- Autoridad Nacional del Agua. (2017). *Un aporte al conocimiento de las causas que amenazan la calidad del agua del maravilloso lago Titicaca LAGO TITICACA*. [www.ana.gob.pe](http://www.ana.gob.pe)
- Barceló, D., y López, M. J. (2008). Contaminación y calidad química del agua: el problema de los contaminantes emergentes. *Jornadas de Presentación de Resultados: El Estado Ecológico de Las Masas de Agua. Panel Científico-Técnico de Seguimiento de La Política de Aguas, Sevilla*.
- Beltrán Farfán, D. F., Palomino Calli, R. P., Moreno Terrazas, E. G., Peralta, C. G., Montesinos-Tubée, D. B., para correspondencia, A., Científica, N., Farfán, B. D., Palomino Calli, R., Moreno Terrazas, E., y Peralta, C. (2015). Calidad de agua de la bahía interior de Puno, lago Titicaca Calidad de agua de la bahía interior de



- Puno, lago Titicaca durante el verano del 2011 Water quality of the inner Puno Bay, Titicaca Lake, during summer 2011. *Rev. Peru. Biol*, 22(223), 335–340.  
<https://doi.org/10.15381/rpb.v22i3.11440>
- Bristol, T. J. (2010). Twitter: Consider the possibilities for continuing nursing education. *Journal of Continuing Education in Nursing*.  
<https://doi.org/10.3928/00220124-20100423-09>
- Cahui Galarza, N. (2019). *Detección y cuantificación de residuos de antibióticos de uso veterinario en muestras de sedimento en zonas productoras de trucha (Oncorhynchus mykiss) y en agua potable de la zona sur de la ciudad de Puno* [Universidad Nacional del Altiplano].  
<http://repositorio.unap.edu.pe/handle/20.500.14082/12271>
- Chen, M. S., Han, J., y Yu, P. S. (1996). Data mining: An overview from a database perspective. In *IEEE Transactions on Knowledge and Data Engineering*.  
<https://doi.org/10.1109/69.553155>
- Daft, R. L., y Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*.  
<https://doi.org/10.1287/mnsc.32.5.554>
- Daughton, C. C. (2001). Pharmaceuticals and personal care products in the environment: Overarching issues and overview. *ACS Symposium Series*.  
<https://doi.org/10.1021/bk-2001-0791.ch001>
- Dejoux, C., y Iltis, A. (1991). *El Lago Titicaca: síntesis del conocimiento limnológico actual*. Orstom. <https://books.google.com.pe/books?id=l6KKzA6zQ10C>
- Doroodchi, M., Iranmehr, A., y Pouriyeh, S. A. (2009). An investigation on integrating XML-based security into Web services. *2009 5th IEEE GCC Conference and Exhibition, GCC 2009*. <https://doi.org/10.1109/IEEEGCC.2009.5734300>
- Džeroski, S. (2009). Relational data mining. In *Data Mining and Knowledge Discovery Handbook* (pp. 887–911). Springer.
- Escobar Mamani, F. (2019). *Determinación de parámetros físico-químicos y niveles de metales pesados en agua y sedimentos en la zona de crianza de truchas*



- (*Oncorhynchus mykiss*), bahía de Puno del lago Titicaca [Universidad Nacional del Altiplano]. <http://repositorio.unap.edu.pe/handle/20.500.14082/10379>
- Farr, R. M. (1983). Escuelas europeas de psicología social: la investigación de representaciones sociales en Francia. *Escuelas europeas de psicología social: la investigación de representaciones sociales en Francia. Revista Mexicana de Sociología*. <https://doi.org/10.2307/3540263>
- Fayyad, U., Piatetsky-Shapiro, G., y Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*.
- Feldman, R., y Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). *International Conference on Knowledge Discovery and Data Mining (KDD)*. <https://doi.org/10.1.1.47.7462>
- Flores, R. C. (2008). Representaciones sociales del medio ambiente. *Perfiles Educativos*. [https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-26982008000200003](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982008000200003)
- Flores, R. C. (2010). Medio ambiente y educación ambiental: *Internacio -Nal de Investigación En Educación*, 2(4), 401–414.
- Fontúrbel Rada, F. (2003). Algunos Criterios Biológicos sobre el proceso de Eutrofización a orillas de seis localidades del Lago Titikaka. *Ecología Aplicada*, 2(1). [http://www.scielo.org.pe/scielo.php?script=sci\\_arttext&pid=S1726-22162003000100011](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1726-22162003000100011)
- Frawley, W. J., Piatetsky-Shapiro, G., y Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, 13(3), 57.
- Gantz, J., Reinsel, D., y Shadows, B. D. (2012). The Digital Universe in 2020. *IDC IView "Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East."*
- Garg, V., y Ravindran, S. K. (2015). *Mastering Social Media Mining with R*. Packt. <https://www.packtpub.com/product/mastering-social-media-mining-with-r/9781784396312>
- Gentry, J. (2016). R Based Twitter Client. *Cran*. <http://lists.hexdump.org/listinfo.cgi/twitter-users-hexdump.org>

- Goffman, E. (1959). The Presentation of Self In Everyday Life (1959). In *Contemporary sociological theory*, 46-61. <https://doi.org/10.2307/258197>
- González Gaudiano, E., y Valdez, R. E. (2012). Enfoques y sujetos en los estudios sobre representaciones sociales de medio ambiente en tres países de Iberoamérica. *CPUE, Revista de Investigación Educativa*, 14. [http://www.uv.mx/cpue/num14/inves/gonzalez\\_](http://www.uv.mx/cpue/num14/inves/gonzalez_)
- Gutiérrez, S. P. (2018). *Influencia de la producción de trucha en el impacto ambiental en la región de Puno 2017*. <http://repositorio.unap.edu.pe/handle/UNAP/8523>
- Huajuapán, C. De, Gómez-anguiano, M., y Ramírez-chávez, R. J. (2007). *Fuentes de contaminación de las aguas sub-terráneas*. 23–28.
- Ibáñez, T. (1988). Representaciones sociales, teoría y método. *Ideologías de La Vida Cotidiana. Barcelona: Sendai*, 13–90.
- Jean-Claude Abric. (2004). Prácticas sociales y representaciones. *Prácticas Sociales y Representación*, 47, 140.
- Jensen Schau, H., y Gilly, M. C. (2003). We Are What We Post? Self-Presentation in Personal Web Space. *Journal of Consumer Research*, 30(3), 385–404. <https://doi.org/10.1086/378616>
- Jiang, W., Wang, Y., Tsou, M. H., y Fu, X. (2015). Using social media to detect outdoor air pollution and monitor air quality index (AQI): A geo-targeted spatiotemporal analysis framework with sina weibo (Chinese twitter). *PLoS ONE*, 10(10), 1–18. <https://doi.org/10.1371/journal.pone.0141185>
- Jodelet, D. (2008). El movimiento de retorno al sujeto y el enfoque de las representaciones sociales. *Cultura y Representaciones Sociales*.
- Jodelet, D. (1986). (1986). La representación social: fenómenos, concepto y teoría. In *Psicología social II*.
- Kaplan, A. M., y Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>

- Kay, S., Zhao, B., y Sui, D. (2015). Can Social Media Clear the Air? A Case Study of the Air Pollution Problem in Chinese Cities. *Professional Geographer*, 67(3), 351–363. <https://doi.org/10.1080/00330124.2014.970838>
- Leiba, B., y Technologies, H. (2012). *IEEE Xplore - OAuth Web Authorization Protocol*. 0–3.  
<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6123701&queryText=oauth>
- Levidow, L., y Upham, P. (2017). Linking the multi-level perspective with social representations theory: Gasifiers as a niche innovation reinforcing the energy-from-waste (EfW) regime. *Technological Forecasting and Social Change*, 120, 1–13. <https://doi.org/10.1016/j.techfore.2017.03.028>
- López, G., y Dorado, S. (2008). Residuos peligrosos domésticos en los residuos sólidos urbanos. *Castellán: Redisa*.
- Maclean, F., Jones, D., Carin-Levy, G., y Hunter, H. (2013). Understanding twitter. *British Journal of Occupational Therapy*, 76(6), 295–298.  
<https://doi.org/10.4276/030802213X13706169933021>
- Mei, S., Li, H., Fan, J., Zhu, X., y Dyer, C. R. (2014). Inferring air pollution by sniffing social media. *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 534–539.  
<https://doi.org/10.1109/ASONAM.2014.6921638>
- Mistry, V. (2011). Critical care training: Using Twitter as a teaching tool. *British Journal of Nursing*. <https://doi.org/10.12968/bjon.2011.20.20.1292>
- Mollett, A., Moran, D., y Dunleavy, P. (2011). *Using Twitter in university research, teaching and impact activities A guide for academics and researchers*.  
[www.twitter.com/signup](http://www.twitter.com/signup)
- Mora, M. (2002). La teoría de representaciones sociales de Serge Moscovici. *Man*, 25.  
<https://doi.org/10.2307/2802632>
- Moreno Terrazas, E., Argota Pérez, G., Alfaro Tapia, R., Aparicio Saavedra, M., Atencio Limachi, S., y Goyzueta Camacho, G. (2018). Cuantificación de metales

- en sedimentos superficiales de la bahía interior, lago Titicaca-Perú. *Revista de Investigaciones Altoandinas - Journal of High Andean Research*, 20(1), 09–18.  
<https://doi.org/10.18271/ria.2018.326>
- Moscovici, S. (1979). El Psicoanálisis. *Papers on Social Representations: Electronic Version*, 4(2).
- Moscovici, S. (1988). Notes towards a description of Social Representations. *European Journal of Social Psychology*, 18(3), 211–250.  
<https://doi.org/10.1002/ejsp.2420180303>
- Navarro Carrascal, O. E. (2004). Representación social del agua y de sus usos Psicológica desde el Caribe. *Psicología Desde El Caribe*, 16.  
<http://www.redalyc.org/pdf/213/21301409.pdf>
- Navarro, O. (2013). Social representation of the environment and air pollution: effect of overlap of two objects. *Revista CES Psicología*, 6(1), 104–121.  
<http://dialnet.unirioja.es/servlet/articulo?codigo=4370209&info=resumen&idioma=SPA>
- Parker, E. B., Short, J., Williams, E., y Christie, B. (1978). The Social Psychology of Telecommunications. *Contemporary Sociology*. <https://doi.org/10.2307/2065899>
- Pérez, M. P. (2003). A propósito de las representaciones sociales: apuntes teóricos, trayectoria y actualidad. *La Habana: CD Caudales. CIPS*.
- Piermattéo, A., Lo Monaco, G., y Girandola, F. (2016). When Commitment Can Be Overturned: Anticipating Recycling Program Dropouts Through Social Representations. *Environment and Behavior*, 48(10), 1270–1291.  
<https://doi.org/10.1177/0013916515597511>
- Piñero, S. (2012). La teoría de las representaciones sociales y la perspectiva de Pierre Bourdieu: Una articulación conceptual. *CPUE, Revista de Investigación Educativa*, 7, 33–51. <https://doi.org/10.25009/cpue.v0i7.96>
- Polli, G., y Camargo, B. (2016). *social representations of the environment for people of different ages*. 22, 392–406.

- Pouriyeh, S. A., y Doroodchi, M. (2009). Secure SMS Banking Based On Web Services. *SWWS*, 79–83.
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*.  
<https://www.r-project.org/>
- Reigota, M. (2004). *Meio ambiente e representação social*.
- Rivas, A. (2004). *Fundamentos de purificación del agua*. 82–104.  
<http://documentacion.ideam.gov.co/openbiblio/bvirtual/018834/MEMORIAS2004/CapituloII/2Fundamentosdelapurificaciondelagua.pdf>
- Rodríguez de Ávila, U., y Paba Barbosa, C. (2013). Representación social de la crisis ambiental. *Psicogente*, 16(29), 84–102.
- Tang, L., y Liu, H. (2010). *Graph Mining Applications to Social Network Analysis*.  
[https://doi.org/10.1007/978-1-4419-6045-0\\_16](https://doi.org/10.1007/978-1-4419-6045-0_16)
- Tchobanoglous, G., y others. (1994). *Gestión integral de residuos sólidos: volumen 1*.
- Thushara, M. G., Mownika, T., y Mangamuru, R. (2019). A comparative study on different keyword extraction algorithms. *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, 969–973. <https://doi.org/10.1109/ICCMC.2019.8819630>
- Tochetto, M. R. L. (2009). Análise Comparativa do Novo Paradigma Ecológico em dois Estados Brasileiros: A Gestão Ambiental além do Mercado e do Estado. *GESTÃO.Org : Revista Eletrônica de Gestão Organizacional*, 7(1), 84–101.
- Vera, H. (2002). Representaciones y clasificaciones colectivas. La teoría sociológica del conocimiento de Durkheim. *Sociológica*, 17(50), 103–121.
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1–23.  
<https://doi.org/10.18637/JSS.V059.I10>
- World Water Assessment Programme (United Nations). (2003). *Water for people, water for life : a joint report by the twenty three UN agencies concerned with freshwater*. UNESCO Pub.

## ANEXOS

### Anexo 1 : Código fuente de Análisis Estadístico

#Adquisición de Twits

```
datainicio <- ‘ ‘ # ejemplo 2020-02-17
```

```
datafinal <- ‘ ‘ # ejemplo 2020-02-21
```

```
setup_twitter_oauth(consumer_key ,consumer_secret,access_token ,access_secret)
#en español
tweetss <- searchTwitter("#titicaca",n = 400,lang = "es", since = datainicio, until = 'datafinal')
tweetss1 <- strip_retweets(tweetss) #strip retweets
df_titicaca_en <- twListToDF(tweetss1) #convert to dataframe
#
tweetss <- searchTwitter("#lagotiticaca",n = 400,lang = "es", since = 'datainicio', until = 'datafinal')
tweetss1 <- strip_retweets(tweetss) #strip retweets
df_lagotiticaca_en <- twListToDF(tweetss1) #convert to dataframe
#
tweetss <- searchTwitter("#titikaka",n = 400,lang = "es", since = 'datainicio', until = 'datafinal')
tweetss1 <- strip_retweets(tweetss) #strip retweets
df_titicacalake_en <- twListToDF(tweetss1) #convert to dataframe
#####
tweetss <- searchTwitter("titicaca",n = 400,lang = "es", since = 'datainicio', until = 'datafinal')
tweetss1 <- strip_retweets(tweetss) #strip retweets
df_titicaca_en1 <- twListToDF(tweetss1) #convert to dataframe
#
tweetss <- searchTwitter("lago titicaca",n = 400,lang = "es", since = 'datainicio', until = 'datafinal')
tweetss1 <- strip_retweets(tweetss) #strip retweets
df_lagotiticaca_en1 <- twListToDF(tweetss1) #convert to dataframe
#
tweetss <- searchTwitter("titikaka",n = 400,lang = "es", since = 'datainicio', until = 'datafinal')
tweetss1 <- strip_retweets(tweetss) #strip retweets
df_titicacalake_en1 <- twListToDF(tweetss1) #convert to dataframe
#
datatiticaca_espa <- rbind.data.frame(df_titicaca_en, df_lagotiticaca_en, df_titicacalake_en,
df_titicaca_en1, df_lagotiticaca_en1, df_titicacalake_en1)

#almacenar dataframe
write.csv(datatiticaca_espa, "datatiticaca_espa_9.csv", row.names=FALSE)
udmodel <- udpipe_download_model(language = "spanish")
udmodel <- udpipe_load_model(file = udmodel$file_model)
data titicaca espa <- read.csv("español1.csv", sep = ";")
#data titicaca espa <- datatiticaca_espa[1:50,]

#datatiticaca_espa <- español
udmodel <- udpipe_download_model(language = "spanish")
udmodel <- udpipe_load_model(file = udmodel$file_model)

datatiticaca_espa$text <- iconv(datatiticaca_espa$text, "ISO_8859-2", "UTF-8")
datatiticaca_espa$text1 <- tolower(datatiticaca_espa$text)
#datatiticaca_espa$text1 <- toupper(datatiticaca_espa$text)
x <- udpipe_annotate(udmodel, x = datatiticaca_espa$text1, doc_id = datatiticaca_espa$id)
#x <- udpipe_annotate(udmodel, x = respuesta$text, doc_id = respuesta$identificador)
Encoding(x$conllu)
x <- as.data.frame(x)
#x
```



```
#analizando la data
library(lattice)
library(ggplot2)

## PARTES UNIVERSALES DEL HABLA
stats <- txt_freq(x$upos)
stats$key <- factor(stats$key, levels = rev(stats$key))
#barchart(key ~ freq, data = stats, col = "cadetblue",
#  main = "UPOS (Partes universales del discurso)\n frecuencia de ocurrencia",
#  xlab = "Freq")
#stats
p <- ggplot(stats, aes(x=reorder(key, freq), y=freq, fill=freq)) +
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_fill_viridis_c(trans = "sqrt", alpha = .3) +
  geom_text(aes(label=paste(round(freq_pct, digits = 2), "%", sep = "")), just=0.3, color="black", size=3)+
  theme(legend.position="none") +
  labs(
    x="Partes universales del discurso (UPOS)\n frecuencia de ocurrencia",
    y="Porcentaje (%)")
p

stats

## NOUNS - SUSTANTIVOS
stats <- subset(x, upos %in% c("NOUN"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
#barchart(key ~ freq, data = head(stats, 20), col = "cadetblue",
#  main = "Sustantivos con mas ocurrencia", xlab = "Freq")
p <- ggplot(stats[1:20,], aes(x=reorder(key, freq), y=freq, fill=freq)) +
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_fill_viridis_c(trans = "sqrt", alpha = .3) +
  geom_text(aes(label=paste(round(freq_pct, digits = 2), "%", sep = "")), hjust=0.3, color="black",
size=3)+
  theme(legend.position="none") +
  labs(
    x="Sustantivos mas utilizados",
    y="Porcentaje (%)")
p
head(stats, n = 20)

## ADJECTIVES - ADJETIVOS
stats <- subset(x, upos %in% c("ADJ"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
#barchart(key ~ freq, data = head(stats, 20), col = "cadetblue",
#  main = "Adjetivos con mas ocurrencia", xlab = "Freq")
p <- ggplot(stats[2:20,], aes(x=reorder(key, freq), y=freq, fill=freq)) +
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_fill_viridis_c(trans = "sqrt", alpha = .3) +
  geom_text(aes(label=paste(round(freq_pct, digits = 2), "%", sep = "")), hjust=0.3, color="black",
size=3)+
  theme(legend.position="none") +
  labs(
    x="Adjetivos con mas ocurrencia",
```



```
y="Porcentaje (%)"
p
head(stats, n = 20)

p <- ggplot(stats[21:40,], aes(x=reorder(key, freq), y=freq, fill=freq)) +
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_fill_viridis_c(trans = "sqrt", alpha = .3) +
  geom_text(aes(label=paste(round(freq_pct, digits = 2), "%", sep = "")), hjust=0.3, color="black",
size=3)+
  theme(legend.position="none") +
  labs(
    x="Adjetivos con mas ocurrencia",
    y="Porcentaje (%)"
  )
p
head(stats, n = 40)

p <- ggplot(stats[41:60,], aes(x=reorder(key, freq), y=freq, fill=freq)) +
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_fill_viridis_c(trans = "sqrt", alpha = .3) +
  geom_text(aes(label=paste(round(freq_pct, digits = 2), "%", sep = "")), hjust=0.3, color="black",
size=3)+
  theme(legend.position="none") +
  labs(
    x="Adjetivos con mas ocurrencia",
    y="Porcentaje (%)"
  )
p
head(stats, n = 60)

p <- ggplot(stats[61:80,], aes(x=reorder(key, freq), y=freq, fill=freq)) +
  geom_bar(stat = "identity")+
  coord_flip()+
  scale_fill_viridis_c(trans = "sqrt", alpha = .3) +
  geom_text(aes(label=paste(round(freq_pct, digits = 2), "%", sep = "")), hjust=0.3, color="black",
size=3)+
  theme(legend.position="none") +
  labs(
    x="Adjetivos con mas ocurrencia",
    y="Porcentaje (%)"
  )
p
head(stats, n = 80)

#DETECCIÓN DE PALABRAS CLAVES
## Using RAKE
stats <- keywords_rake(x = x, term = "lemma", group = "doc_id",
  relevant = x$upos %in% c("NOUN", "ADJ"))
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ rake, data = head(subset(stats, freq > 3), 40), col = "skyblue",
  main = "Identificación de palabras clave algoritmo RAKE",
  xlab = "Rake")

#install.packages("wordcloud")
library(wordcloud)
wordcloud(words = stats$key, freq = stats$freq, min.freq = 3, max.words = 100,
  random.order = FALSE, colors = c("#1B9E77", "#D95F02", "#7570B3", "#E7298A", "#66A61E",
"#E6AB02"))
```

```
## Using Pointwise Mutual Information Collocations
x$word <- tolower(x$token)
stats <- keywords_collocation(x = x, term = "word", group = "doc_id")
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ pmi, data = head(subset(stats, freq > 3), 40), col = "skyblue",
  main = "Identificación de palabras clave algoritmo PMI Collocation",
  xlab = "PMI (Pointwise Mutual Information)")

## Using a sequence of POS tags (noun phrases / verb phrases)
x$phrase_tag <- as_phrasemachine(x$upos, type = "upos")
stats <- keywords_phrases(x = x$phrase_tag, term = tolower(x$token),
  pattern = "(A|N)*N(P+D*(A|N)*N)*",
  is_regex = TRUE, detailed = FALSE)
stats <- subset(stats, ngram > 1 & freq > 3)
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ freq, data = head(stats, 40), col = "skyblue",
  main = "Keywords - simple noun phrases", xlab = "Frequency")

cooc <- cooccurrence(x = subset(x, upos %in% c("NOUN", "ADJ")),
  term = "lemma",
  group = c("doc_id", "paragraph_id", "sentence_id"))
head(cooc, n = 40)

#install.packages("igraph")
#install.packages("ggraph")
#install.packages("ggplot2")
library(igraph)
library(ggraph)
library(ggplot2)

wordnetwork <- head(cooc, 50)
wordnetwork <- graph_from_data_frame(wordnetwork)
ggraph(wordnetwork, layout = "fr") +
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "blue") +
  geom_node_text(aes(label = name), col = "darkgreen", size = 4) +
  theme_graph(base_family = "Arial Narrow") +
  theme(legend.position = "none") +
  labs(title = "Cocurrencias dentro de la oración", subtitle = "Sustantivos y adjetivos")

.
cooc <- cooccurrence(x$lemma, relevant = x$upos %in% c("NOUN", "ADJ"), skipgram = 1)
head(cooc, n=50)
#
wordnetwork <- head(cooc, 50)
wordnetwork <- graph_from_data_frame(wordnetwork)
ggraph(wordnetwork, layout = "fr") +
  geom_edge_link(aes(width = cooc, edge_alpha = cooc)) +
  geom_node_text(aes(label = name), col = "darkgreen", size = 3) +
  theme_graph(base_family = "Arial Narrow") +
  labs(title = "Palabras que se suceden", subtitle = "Sustantivos y adjetivos")

#correlations
x$id <- unique_idenfier(x, fields = c("sentence_id", "doc_id"))
dtm <- subset(x, upos %in% c("NOUN", "ADJ"))
dtm <- document_term_frequencies(dtm, document = "id", term = "lemma")
dtm <- document_term_matrix(dtm)
```

```
dtm <- dtm_remove_lowfreq(dtm, minfreq = 5)
termcorrelations <- dtm_cor(dtm)
y <- as_cooccurrence(termcorrelations)
y <- subset(y, term1 < term2 & abs(cooc) > 0.2)
y <- y[order(abs(y$cooc), decreasing = TRUE), ]
head(y, n = 50)

termcorrelations <- dtm_chisq(dtm)
y <- as_cooccurrence(termcorrelations)
y <- subset(y, term1 < term2 & abs(cooc) > 0.2)
y <- y[order(abs(y$cooc), decreasing = TRUE), ]
head(y, n = 50)

#MODELADO

## Define the identifier at which we will build a topic model
x$topic_level_id <- unique_identifier(x, fields = c("doc_id", "paragraph_id", "sentence_id"))
## Get a data.frame with 1 row per id/lemma
dtf <- subset(x, upos %in% c("NOUN"))
dtf <- document_term_frequencies(dtf, document = "topic_level_id", term = "lemma")
head(dtf)

## Create a document/term/matrix for building a topic model
dtm <- document_term_matrix(x = dtf)
## Remove words which do not occur that much
dtm_clean <- dtm_remove_lowfreq(dtm, minfreq = 5)
head(dtm_colsums(dtm_clean))

#
library(topicmodels)
m <- LDA(dtm_clean, k = 4, method = "Gibbs", control = list(nstart = 5, burnin = 2000, best = TRUE,
seed = 1:5))

scores <- predict(m, newdata = dtm, type = "topics", labels = c("labela", "labelb", "labelc", "xyz"))
str(scores)
predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)

## Build document term matrix on nouns/adjectives only
dtf <- subset(x, upos %in% c("NOUN", "ADJ") &
!lemma %in% c("appartement", "appart", "eter", "tres"))
dtf <- document_term_frequencies(dtf, document = "topic_level_id", term = "lemma")
dtm <- document_term_matrix(x = dtf)
dtm_clean <- dtm_remove_lowfreq(dtm, minfreq = 5)

## Build topic model + get topic terminology
m <- LDA(dtm_clean, k = 4, method = "Gibbs",
control = list(nstart = 5, burnin = 2000, best = TRUE, seed = 1:5))
topicterminology <- predict(m, type = "terms", min_posterior = 0.025, min_terms = 10)
scores <- predict(m, newdata = dtm, type = "topics")

predict(m, type = "terms", min_posterior = 0.05, min_terms = 6)

## Build document term matrix on VERB only
dtf <- subset(x, upos %in% c("VERB", "NOUN") &
!lemma %in% c("appartement", "appart", "eter", "tres"))
dtf <- document_term_frequencies(dtf, document = "topic_level_id", term = "lemma")
```

```
dtm <- document_term_matrix(x = dtf)
dtm_clean <- dtm_remove_lowfreq(dtm, minfreq = 5)
## Build topic model + get topic terminology
m <- LDA(dtm_clean, k = 4, method = "Gibbs",
        control = list(nstart = 5, burnin = 2000, best = TRUE, seed = 1:5))
topicterminology <- predict(m, type = "terms", min_posterior = 0.025, min_terms = 10)
scores <- predict(m, newdata = dtm, type = "topics")

predict(m, type = "terms", min_posterior = 0.05, min_terms = 6)

## Build document term matrix on VERB only
dtf <- subset(x, upos %in% c("VERB"))
dtf <- document_term_frequencies(dtf, document = "topic_level_id", term = "lemma")
dtm <- document_term_matrix(x = dtf)
dtm_clean <- dtm_remove_lowfreq(dtm, minfreq = 5)

## Build topic model + get topic terminology
m <- LDA(dtm_clean, k = 2, method = "Gibbs",
        control = list(nstart = 5, burnin = 2000, best = TRUE, seed = 1:5))
topicterminology <- predict(m, type = "terms", min_posterior = 0.025, min_terms = 10)
scores <- predict(m, newdata = dtm, type = "topics")

predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)

x_topics <- merge(x, scores, by.x="topic_level_id", by.y="doc_id")
wordnetwork <- subset(x_topics, topic %in% 4 & lemma %in% topicterminology[[4]]$term)
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "lemma")
wordnetwork <- graph_from_data_frame(wordnetwork)
ggraph(wordnetwork, layout = "fr") +
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "pink") +
  geom_node_text(aes(label = name), col = "darkgreen", size = 3) +
  theme_graph(base_family = "Arial Narrow") +
  labs(title = "Words in topic 1 ", subtitle = "Nouns & Verbs cooccurrence")
#install.packages("qgraph")
#library(qgraph)
topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)
termcorr <- subset(x_topics, topic %in% 4 & lemma %in% topicterminology[[4]]$term)
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "lemma")
termcorr <- document_term_matrix(termcorr)
termcorr <- dtm_cor(termcorr)
termcorr[lower.tri(termcorr)] <- NA
diag(termcorr) <- NA

qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,
        borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)
## Find keywords with RAKE
keyw_rake <- keywords_rake(x,
                          term = "token", group = c("doc_id", "paragraph_id", "sentence_id"),
                          relevant = x$upos %in% c("NOUN", "ADJ"),
                          ngram_max = 3, n_min = 5)

## Find simple noun phrases
x$phrase_tag <- as_phrasemachine(x$upos, type = "upos")
keyw_nounphrases <- keywords_phrases(x$phrase_tag, term = x$token,
                                     pattern = "(A|N)*N(P+D*(A|N)*N)*", is_regex = TRUE,
                                     detailed = FALSE)
keyw_nounphrases <- subset(keyw_nounphrases, ngram > 2)

## Recode terms to keywords
x$term <- x$token
x$term <- txt_recode_ngram(x$term,
```

```
      compound = keyw_rake$keyword, ngram = keyw_rake$ngram)
x$term <- txt_recode_ngram(x$term,
      compound = keyw_nounphrases$keyword, ngram = keyw_nounphrases$ngram)
## Keep keyword or just plain nouns
x$term <- ifelse(x$upos %in% "NOUN", x$term,
      ifelse(x$term %in% c(keyw_rake$keyword, keyw_nounphrases$keyword), x$term, NA))

## Build document/term/matrix
dtm <- document_term_frequencies(x, document = "topic_level_id", term = "term")
dtm <- document_term_matrix(x = dtm)
dtm <- dtm_remove_lowfreq(dtm, minfreq = 5)

m <- LDA(dtm, k = 6, method = "Gibbs", control = list(nstart = 5, burnin = 2000, best = TRUE, seed =
1:5))

topicterminology <- predict(m, type = "terms", min_posterior = 0.010, min_terms = 6)
topicterminology

scores <- predict(m, newdata = dtm, type = "topics")
predict(m, type = "terms", min_posterior = 0.05, min_terms = 6)

#m <- LDA(dtm_clean, k = 4, method = "Gibbs", control = list(nstart = 5, burnin = 2000, best = TRUE,
seed = 1:5))
#topicterminology <- predict(m, type = "terms", min_posterior = 0.025, min_terms = 10)
#scores <- predict(m, newdata = dtm, type = "topics")
#predict(m, type = "terms", min_posterior = 0.05, min_terms = 6)

x$doc_id1 <- as.integer(x$doc_id1)
scores$doc_id2 <- as.integer(scores$doc_id)
x_topics <- merge(x, scores, by.x="topic_level_id", by.y="doc_id2")
wordnetwork <- subset(x_topics, topic %in% 1 & term %in% topicterminology[[1]]$term)
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "term")
wordnetwork <- graph_from_data_frame(wordnetwork)
ggraph(wordnetwork, layout = "fr") +
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "skyblue") +
  geom_node_text(aes(label = name), col = "black", size = 3) +
  theme_graph(base_family = "Arial Narrow") +
  labs(title = "Tema 1")

wordnetwork <- subset(x_topics, topic %in% 2 & term %in% topicterminology[[2]]$term)
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "term")
wordnetwork <- graph_from_data_frame(wordnetwork)
ggraph(wordnetwork, layout = "fr") +
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "skyblue") +
  geom_node_text(aes(label = name), col = "black", size = 3) +
  theme_graph(base_family = "Arial Narrow") +
  labs(title = "Tema 2")

wordnetwork <- subset(x_topics, topic %in% 3 & term %in% topicterminology[[3]]$term)
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "term")
wordnetwork <- graph_from_data_frame(wordnetwork)
ggraph(wordnetwork, layout = "fr") +
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "skyblue") +
  geom_node_text(aes(label = name), col = "black", size = 3) +
  theme_graph(base_family = "Arial Narrow") +
  labs(title = "Tema 3")

wordnetwork <- subset(x_topics, topic %in% 4 & term %in% topicterminology[[4]]$term)
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "term")
wordnetwork <- graph_from_data_frame(wordnetwork)
```

```
ggraph(wordnetwork, layout = "fr") +  
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "skyblue") +  
  geom_node_text(aes(label = name), col = "black", size = 3) +  
  theme_graph(base_family = "Arial Narrow") +  
  labs(title = "Tema 4")
```

```
wordnetwork <- subset(x_topics, topic %in% 5 & term %in% topicterminology[[5]]$term)  
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "term")  
wordnetwork <- graph_from_data_frame(wordnetwork)  
ggraph(wordnetwork, layout = "fr") +  
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "skyblue") +  
  geom_node_text(aes(label = name), col = "black", size = 3) +  
  theme_graph(base_family = "Arial Narrow") +  
  labs(title = "Tema 5")
```

```
wordnetwork <- subset(x_topics, topic %in% 6 & term %in% topicterminology[[6]]$term)  
wordnetwork <- cooccurrence(wordnetwork, group = c("topic_level_id"), term = "term")  
wordnetwork <- graph_from_data_frame(wordnetwork)  
ggraph(wordnetwork, layout = "fr") +  
  geom_edge_link(aes(width = cooc, edge_alpha = cooc), edge_colour = "skyblue") +  
  geom_node_text(aes(label = name), col = "black", size = 3) +  
  theme_graph(base_family = "Arial Narrow") +  
  labs(title = "Tema 6")
```

```
topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)  
termcorr <- subset(x_topics, topic %in% 1 & term %in% topicterminology[[1]]$term)  
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "term")  
termcorr <- document_term_matrix(termcorr)  
termcorr <- dtm_cor(termcorr)  
termcorr[lower.tri(termcorr)] <- NA  
diag(termcorr) <- NA  
qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,  
  borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)
```

```
topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)  
termcorr <- subset(x_topics, topic %in% 2 & term %in% topicterminology[[2]]$term)  
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "term")  
termcorr <- document_term_matrix(termcorr)  
termcorr <- dtm_cor(termcorr)  
termcorr[lower.tri(termcorr)] <- NA  
diag(termcorr) <- NA  
qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,  
  borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)
```

```
topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)  
termcorr <- subset(x_topics, topic %in% 3 & term %in% topicterminology[[3]]$term)  
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "term")  
termcorr <- document_term_matrix(termcorr)  
termcorr <- dtm_cor(termcorr)  
termcorr[lower.tri(termcorr)] <- NA  
diag(termcorr) <- NA  
qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,  
  borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)
```

```
topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)  
termcorr <- subset(x_topics, topic %in% 4 & term %in% topicterminology[[4]]$term)  
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "term")  
termcorr <- document_term_matrix(termcorr)
```



```
termcorr <- dtm_cor(termcorr)
termcorr[lower.tri(termcorr)] <- NA
diag(termcorr) <- NA
qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,
        borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)

topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)
termcorr <- subset(x_topics, topic %in% 5 & term %in% topicterminology[[5]]$term)
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "term")
termcorr <- document_term_matrix(termcorr)
termcorr <- dtm_cor(termcorr)
termcorr[lower.tri(termcorr)] <- NA
diag(termcorr) <- NA
qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,
        borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)

topicterminology <- predict(m, type = "terms", min_posterior = 0.05, min_terms = 10)
termcorr <- subset(x_topics, topic %in% 6 & term %in% topicterminology[[6]]$term)
termcorr <- document_term_frequencies(termcorr, document = "topic_level_id", term = "term")
termcorr <- document_term_matrix(termcorr)
termcorr <- dtm_cor(termcorr)
termcorr[lower.tri(termcorr)] <- NA
diag(termcorr) <- NA
qgraph(termcorr, layout = "spring", labels = colnames(termcorr), directed = FALSE,
        borders = FALSE, label.scale = FALSE, label.cex = 1, node.width = 0.5)
```





## AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Renzo Apaza Cutipa,  
identificado con DNI 42676385 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado

Estadística Aplicada,  
informo que he elaborado el/la  Tesis o  Trabajo de Investigación denominada:

“ Text mining en la representación social sobre la contaminación del Lago Titicaca, Puno 2020 ”

para la obtención de  Grado,  Título Profesional o  Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

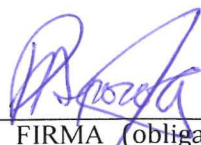
En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno \_\_\_\_\_ de \_\_\_\_\_ del 20\_\_



FIRMA (obligatoria)



Huella





## DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Renzo Apaza Cutipa,  
identificado con DNI 42676385 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado  
Estadística Aplicada

informo que he elaborado el/la  Tesis o  Trabajo de Investigación denominada:

“Text minima en la representación social  
sobre la contaminación del lago Titicaca  
PUNO 2020”

Es un tema original.

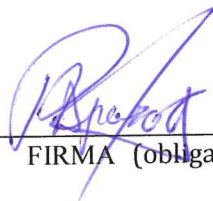
Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno \_\_\_\_\_ de \_\_\_\_\_ del 20\_\_

  
FIRMA (obligatoria)



Huella