

**APLICACIÓN DE TÉCNICAS DE ANÁLISIS DE INFORMACIÓN DE LA RED  
SOCIAL TWITTER, PARA LA VISUALIZACIÓN DE TENDENCIAS Y NECESIDADES  
LABORALES Y DE FORMACIÓN EN EL SECTOR DE T.I.**

**LUIS ALBERTO ORTIZ PALMA  
SEMILLERO DE INVESTIGACIÓN SIAVA**

**UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA - UNAD  
ESCUELAS DE CIENCIAS BÁSICAS, TECNOLOGÍA E INGENIERIA  
PROGRAMA DE INGENIERIA DE SISTEMAS  
POPAYÁN, OCTUBRE 2 DE 2019**

**APLICACIÓN DE TÉCNICAS DE ANÁLISIS DE INFORMACIÓN DE LA RED  
SOCIAL TWITTER, PARA LA VISUALIZACIÓN DE TENDENCIAS Y NECESIDADES  
LABORALES Y DE FORMACIÓN EN EL SECTOR DE T.I.**

**LUIS ALBERTO ORTIZ PALMA  
SEMILLERO DE INVESTIGACIÓN SIAVA**

Trabajo de Grado presentado como requisito parcial para optar al título de Ingeniería de Sistemas  
en la modalidad de proyecto aplicado.

**Director**

**M.Sc. HERMES MOSQUERA  
Docente Universidad Nacional Abierta y a Distancia - UNAD**

**UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA - UNAD  
ESCUELAS DE CIENCIAS BÁSICAS, TECNOLOGÍA E INGENIERIA  
PROGRAMA DE INGENIERIA DE SISTEMAS  
POPAYÁN, OCTUBRE 2 DE 2019**

**Nota de Aceptación**

-----  
-----  
-----  
-----  
-----

**Director**-----

**M.Sc. HERMES MOSQUERA**

**Jurado**-----

**Ing. HERNANDO JOSÉ PEÑA HIDALGO**

**Fecha de Sustentación: Popayán, octubre 2 de 2019**

## **AGRADECIMIENTOS**

Son muchas las personas que han contribuido al proceso y conclusión de este trabajo. En primer lugar, gracias a mi familia por todo el apoyo y por dar inicio a mi formación personal. En segundo lugar, a la Universidad Nacional Abierta y a Distancia UNAD, lugar de crecimiento académico y profesional. También, al Departamento de Ingeniería de Sistemas y a todos sus profesores, en particular al profesor Hermes Mosquera, por su disposición, asesoría, apoyo personal e institucional y la motivación constante para que concluyera este trabajo. A los Ingenieros Iván López y Hernando Peña por sus aportes y recomendaciones.

Es un momento muy especial que espero, prevalezca en el tiempo, no solo en la mente de las personas a quienes agradecí, sino también a quienes invirtieron su tiempo para leer mi proyecto de grado, a ellos asimismo les agradezco con todo mi ser.

## Tabla de Contenido

Pág.

1. Introducción.....	1
2. Definición del Problema.....	3
2.1 Descripción.....	3
2.2 Formulación pregunta de investigación .....	4
3. Justificación .....	5
4. Objetivos .....	8
4.1 Objetivo general. ....	8
4.2 Objetivos específicos.....	8
5. Marco de referencia .....	9
5.1 Antecedentes .....	9
5.2 Importancia de la tecnología en el contexto.....	11
5.3 Marco Teórico.....	13
5.4. Marco conceptual .....	17
6. Metodología .....	24
6.1. Revisión de referentes teóricos .....	26
6.2. Identificar una herramienta informática para captura de información.....	30
7.2.1. Selección de la herramienta.....	44
7.1 Captura de información.....	45
7.3.1.Esquema de conexión.....	48
7.3.4Método para captura de información .....	53
7.3.5 Codificación y parametrización.....	55
7.3.6 Validación del script.....	58
7.4. Preparación de información .....	63
7.4.1.Clasificación de tweets.....	64

7.4.2. Minería de textos .....	69
7.4.3. Script de limpieza .....	72
7.4.4. verificación de datos .....	74
7.5. Análisis de resultados .....	81
7.5.1. Análisis de sentimientos .....	81
7.5.2. Clasificación .....	84
7.5.3. Modelo de clasificación .....	91
7.5.4. Reportes y Graficas .....	106
7.6. Diseño de guías .....	113
7.6.1. Guía de usuario .....	113
7.6.2. Manual técnico .....	113
8. Conclusiones .....	114
9. Recomendaciones .....	116
10. Referencias .....	118
11. ANEXOS .....	123
Anexo A – Guía de usuario .....	123
Anexo B – Manual técnico .....	123
Anexo C – Herramienta ARS SIAVA .....	123

## LISTA DE FIGURAS

	Pág.
Figura 1. Interconexión UNAD .....	11
Figura 2. Aplicaciones UNAD.....	12
Figura 3. Mapa de procesos UNAD.....	13
Figura 4. Gestión de Redes Sociales.....	31
Figura 5. Estadísticas de actividad en Twitter .....	32
Figura 6. Interfaz de búsqueda de Socialmention.....	33
Figura 7. Interfaz de búsqueda de HowSociable .....	34
Figura 8. Interfaz de búsqueda de meltwater .....	34
Figura 9. Interfaz de Ncapture .....	36
Figura 10. Interfaz de Nvivo.....	38
Figura 11. Interfaz de Fanpagekarma .....	40
Figura 12. Interfaz de Netvizz .....	41
Figura 13. Entorno del IDE RStudio. Barra del menú principal (1), Editor (2), Consola (3), Historial de objetos almacenados en memoria (4), Directorio de trabajo (5).....	46
Figura 14. Descubrimiento de conocimiento en datos/textos .....	47
Figura 15. API de Twitter desde una cuenta de desarrollador.....	48
Figura 16. Entorno del área de estudio. ....	49
Figura 17. Permisos de Acceso API según nivel de suscripción. ....	53
Figura 18. Visualización y acceso del módulo de Recopilar datos.....	54
Figura 19. Entorno principal del módulo de Recopilación.....	54
Figura 20 Formulario de conexión.....	59

Figura 21. Botón para establecer conexión con API Twitter .....	59
Figura 22. Notificación estado de conexión .....	60
Figura 23. Formulario de búsqueda .....	61
Figura 24. Notificación de finalización de consulta a través de API de Twitter .....	61
Figura 25. Tabla generada en pantalla de Tweets consultados .....	62
Figura 26. Documento exportado a carpeta local .....	63
Figura 27. Ruta para importar datos CSV a Microsoft Excel .....	65
Figura 28. Pasos en ventana para importar documento .....	66
Figura 29. Ventana de carga y transformación de datos .....	67
Figura 30. Documento Excel con columna sentiment .....	68
Figura 31. Cinta de opciones Microsoft Excel.....	68
Figura 32. Ventana importar datos en Microsoft Excel .....	69
Figura 33. Visualización y acceso del módulo de Análisis .....	70
Figura 34. Técnicas de tratamiento de datos del Corpus .....	71
Figura 35. Frecuencia de palabras por tweet .....	78
Figura 36. Reporte de frecuencias de la matriz.....	85
Figura 37. Reporte de frecuencias de la matriz eliminado palabras de baja frecuencia .....	87
Figura 38. Comparación de matrices antes y después de filtrar por frecuencias .....	88
Figura 39. Base de datos en formato R.....	89
Figura 40. Base de datos inicial con tweets importados .....	90
Figura 41. Visualización de la representación de un Hiperplano .....	92
Figura 42. Hiperplano en espacio de características 2D .....	92
Figura 43. Vectores de soporte .....	93

Figura 44. Creación y visualización del Split en el ambiente de trabajo de R .....	96
Figura 45. Creación y visualización de las bases de datos entrenamiento y evaluación en el ambiente de trabajo de R.....	97
Figura 46. Tabla de datos de entrenamiento .....	98
Figura 47. Tabla de datos de evaluación.....	99
Figura 48. Modelo de soporte vector machine.....	101
Figura 49. Objeto con tweets y su predicción de sentimiento .....	104
Figura 50. Reporte del modelo de predicción de tweets. ....	105
Figura 51. Tabla de frecuencias con sentimiento positivo.....	106
Figura 52. Tabla de palabras y frecuencia .....	107
Figura 53. Palabras agrupadas por frecuencias.....	108
Figura 54. Nube de palabras tweets positivos.....	110
Figura 55. Histograma de frecuencias tweets positivos.....	111
Figura 56. Tabla de frecuencias análisis de sentimientos.....	113

## LISTA DE TABLAS

	Pág.
<b>Tabla 1.</b> Niveles de suscripción de API Twitter. ....	52
<b>Tabla 2.</b> Relación de tweets vs sentimientos.....	75
<b>Tabla 3.</b> Ingreso de tweets a la base de datos.....	82
<b>Tabla 4.</b> Distribución del sentimiento en base de datos de evaluación.....	100

## RESUMEN

El desarrollo de este proyecto tiene como entorno de trabajo a Twitter y contempla dos fases.

Una primera fase donde se realiza la minería de textos, la cual obtiene datos a través de una API, datos que se recopilan en forma de tweets, importados a formato CSV (Valores separados por comas). el cual es entendible para el lenguaje R, posteriormente se limpian, estandarizan y convierten los datos de acuerdo con los requerimientos de este lenguaje. La segunda parte corresponde al análisis de sentimientos, esta partirá de un pre procesado de los datos, para este proyecto se parte de un subconjunto de 218 Tweets. A este conjunto de datos se les analiza el sentimiento. Para lo cual un algoritmo recibe entrenamiento manual, por medio de Machine Learning, aprende y finalmente es capaz de predecir tendencias, en este caso tendencias laborales en TI, las cuales se proyectan en gráficos tipo Histograma y Nubes de palabras además de tablas de frecuencias y visualización datos estadísticos como media, mediana y moda, se encuentra 29 vectores de soporte y dos clases clasificadas con un porcentaje de precisión del 98%.

Para facilitar el uso al usuario, se realiza una interfaz gráfica desarrollada por medio de una aplicación web, la cual está diseñada por medio de la librería Shiny desde R. El software contiene dos entornos, el primero dedicado a la recopilación de datos desde Twitter y el segundo para análisis de textos.

**Palabras Clave:** Redes sociales, abstracción, aplicación, técnicas, herramientas de visualización, Twitter, ARS, Análisis de redes sociales.

## ABSTRACT

The development of this project has Twitter as its working environment and contemplates two phases. A first phase where text mining is carried out, which obtains data through an API, data that is collected in the form of tweets, imported into CSV format (Comma Separated Values), which is understandable for the R language, then cleaned, standardized and converted data according to the requirements of this language. The second part corresponds to the analysis of feelings, this will start from a pre-processing of the data, for this project we start from a subset of 218 Tweets. This data set is analyzed for sentiment. For which an algorithm receives manual training, through Machine Learning, learns and finally is able to predict trends, in this case working trends in IT, which are projected in graphs such as Histogram and Clouds of words in addition to frequency tables and visualization statistical data such as average, median and fashion, there are 29 support vectors and two classes classified with a percentage of accuracy of 98%.

In order to make the use easier for the user, a graphic interface is developed through a web application, which is designed through the Shiny library from R. The software contains two environments, the first dedicated to data collection from Twitter and the second for text analysis.

**Keywords:** Social networks, abstraction, application, techniques, visualization tools, Twitter, ARS, Social network analysis.

## 1. Introducción

La popularidad de los sitios de medios sociales y la facilidad con que sus datos están disponibles significa que estas plataformas son cada vez más las fuentes primarias para la investigación social. Wasim Ahmed presenta un rápido vistazo a algunas de las herramientas disponibles para los científicos sociales para analizar los datos de los medios de comunicación social y también reflexiona sobre las limitaciones de las plataformas y los métodos utilizados para este tipo de investigación. (Ahmed, 2015).

Por las razones anteriormente expuestas es importante adelantar investigaciones que permitan conocer y/o exponer el uso de redes sociales, en este caso particular con Twitter en la obtención de datos para toma de decisiones, para ello se hace necesario recopilar las principales guías y métodos que sirvan de apoyo para las personas dedicadas a esta labor. Dichas técnicas, guías y modelos que permitan realizar el análisis de los datos recopilados desde la red social Twitter, el presente proyecto muestra el proceso de consolidación apoyándose en la formulación de guías de las herramientas y técnicas de análisis más usadas, logrando clasificarlas de acuerdo a eficacia, funcionalidad y compatibilidad.

El documento está distribuido de la siguiente manera: el primer capítulo presenta una breve inducción a los medios sociales, su uso y descripción de las técnicas utilizadas. El segundo capítulo describe la metodología, diseño, análisis de datos y guías para la recopilación de datos

en Twitter. Para analizar los datos y obtener datos estadísticos, se utiliza el paquete *tm*, paquete que facilita el uso de formatos de texto heterogéneos y cuenta con soporte de base de datos integrado para minimizar las demandas de memoria. (Calvo, 2017). El tercer capítulo describe los resultados obtenidos y la discusión o argumentación a la técnica utilizada, finalmente las conclusiones sobre el estudio realizado y recomendaciones.

## 2. Definición del Problema

### 2.1 Descripción

El proceso de gestión de la oferta educativa en sus procedimientos: gestión de nuevos registros calificados y renovación de registro calificado, donde se involucra la gestión de la información y del conocimiento organizacional es complejo, dado que también interviene la normatividad planteada por el sistema de Calidad e la universidad SIC (SIC, 2015), además de cuestiones propias de la estructura de la Escuela. En este sentido para los líderes de programa es fundamental recoger las evidencias que se tengan en cada uno de los centros del país en relación a las necesidades de oferta académica en cada región.

Por lo anterior uno de los factores de gran importancia no solo para cumplir con los requerimientos del ministerio, si no para estar informados permanentemente en relación a los cambios en tendencias y necesidades de formación es el referido a la característica número 2 relacionada con el ítem de “Relevancia académica y pertinencia social del programa: El programa es relevante académicamente y responde a necesidades locales, regionales, nacionales e internacionales”, y en esta característica el Ministerio a través del CNA (consejo Nacional de Acreditación) evalúa los siguientes aspectos como: análisis sobre tendencias y líneas de desarrollo de la disciplina o profesión en el ámbito local, regional, nacional e internacional, y su incidencia en el programa así como también estudios orientados a identificar las necesidades y requerimientos del entorno laboral (local, regional y nacional) en términos productivos y de competitividad, tecnológicos y de talento humano, entre muchos otros, sin embargo de lo antes mencionado, se hace un estudio en el momento de realizar el documento para presentar al CNA,

pero no se hace un seguimiento a lo largo del tiempo para tener un análisis efectivo, identificando los aspectos predominantes que le permitan a las Escuelas (para el caso de la UNAD), tomar decisiones en cuanto a contenidos y requerimientos propios del mercado.

Ahora bien, es importante mencionar que la universidad hace presencia a nivel nacional e internacional, permitiendo tener una mirada amplia de estas tendencias desde las regiones y del entorno internacional. Lo que “dificultad” aún más tener la información actualizada y sectorizada que permita tener un currículo flexible y acorde a las diferentes regiones del país.

## **2.2 Formulación pregunta de investigación**

¿Cómo aprovechar la red social Twitter para identificar tendencias, necesidades, comportamientos y requerimientos de formación en el contexto nacional e internacional?

### 3. Justificación

En los últimos años, las Redes Sociales han experimentado un crecimiento continuo y como resultado se han convertido en uno de los principales canales, si no el principal, para las marcas que disponen de una infinidad de plataformas, como Facebook y Twitter, y otras de nuevo nicho. El potencial del Social Media se hace evidente para muchas empresas, pero también complejo para otras. La situación actual en el sector del Marketing y la Comunicación presenta un desafío para los profesionales. (Tournery, 2015). Sin embargo de acuerdo con las apreciaciones de Marianne Tournery, Directora de proyectos de Marketing de Kantar Media cita en uno de sus escritos que “la vigilancia no es suficiente, es necesario saber interpretar los datos que tienes, por ello las herramientas de evaluación se vuelven esenciales, Cuando se gestiona bien, el seguimiento y el análisis de Redes Sociales son una herramienta excepcionalmente poderosa para mejorar tu reputación de marca, así como todas las actividades de tu negocio. (Tournery, 2015).

Así mismo es importante el desarrollo de guías que permitan la extracción de datos para instituciones de educación superior, toda vez que estos productos servirán de apoyo en la organización, oferta y desarrollo de programas académicos de educación superior, ajustados a las regiones, donde tiene su campo de acción. En este sentido el presente trabajo se convierte en un apoyo para el cumplimiento de ciertos estándares de calidad y/o registros calificados del Ministerio de educación de acuerdo a lo establecido en el decreto 1295, 2010. 2010 (Ministerio de Educación Nacional de Colombia, 2010).

Partiendo de lo antes mencionado y teniendo en cuenta que “En los procesos de solicitud, renovación y /o modificación de Registro Calificado de programas académicos, las Instituciones de Educación Superior deben presentar un Documento Maestro con la información que permita verificar el cumplimiento de cada una de las 15 condiciones de calidad establecidas en el Artículo 2.5.3.2.2.1. Evaluación de las condiciones de calidad de los programas, del Decreto 1075 de 2015”. Es menester que las instituciones realicen un verdadero análisis de cada una de las condiciones y el presente proyecto aportará información actualizada para la construcción de la condición 2 en:

- Incluye un análisis de las necesidades (actuales y futuras) de la población, de la región y de los sectores productivos afines al ámbito de formación del programa.
- Identifica la oferta regional actual de formación (en especial en el nivel de formación del programa en evaluación) y establece una propuesta académica teniendo en cuenta las tendencias nacionales e internacionales de formación en el campo.

De acuerdo con lo anterior es diseñar una herramienta que permita una recolección y análisis de información generada por la red Twitter, a partir de ahí generar guías y herramientas de evaluación garantes, que brinden apoyo en la consecución, verificación y visualización de información, asimismo instruir y dar las pautas necesarias para una buena captura de información a través de la red social Twitter. Estos datos le permitirán posteriormente a la institución identificar las tendencias y necesidades del sector en cuanto a requerimiento de profesionales, capacitación en competencias, y planes de mejora que permitan ser partícipes en la formación de

las nuevas generaciones y aportar herramientas a los profesionales para enfrentar los nuevos retos del entorno laboral y social. Es de importancia tener en consideración los resultados obtenidos de estos análisis permitirán construir instrumentos necesarios para los diagnósticos requeridos en cada uno de los procesos de Autoevaluación previo a la entrega de los documentos al Ministerio para la visita de los pares académicos.

## **4. Objetivos**

### **4.1 Objetivo general.**

Aplicar técnicas de análisis de información de la red social Twitter, para identificar tendencias, necesidades, comportamientos y requerimientos de formación en el contexto nacional e internacional en el sector de T.I.

### **4.2 Objetivos específicos.**

- Identificar una herramienta para la recopilación de información de la red social Twitter.
- Determinar los pasos necesarios para la extracción y visualización de información de la red social Twitter.
- Documentar técnicas para la representación de información extraída de la red social Twitter.

## 5. Marco de referencia

### 5.1 Antecedentes

La UNAD, dentro de su estructura administrativa cuenta con el Sistema Integrado de Gestión SIG, el cual dicta los lineamientos para las funciones desarrolladas al interior de la universidad, en él se encuentra el mapa de procesos (ver **Figura 3**) que establece el proceso C8 - Gestión de la Oferta Educativa, con el cual se busca “garantizar el cumplimiento de las condiciones de calidad y el impacto de la oferta académica de la Universidad, en el marco de los lineamientos normativos, educativos y formativos, de orden institucional, local, nacional e internacional” y que a su vez incorpora 2 procedimientos relacionados con la gestión y en cada uno de los procedimientos (UNAD, 2017).

Para el caso el sistema nacional de acreditación presenta los lineamientos para la acreditación de programas de pregrado de los cuales la característica 3 “Relevancia académica y pertinencia social del programa” debe involucrar los siguientes aspectos: “

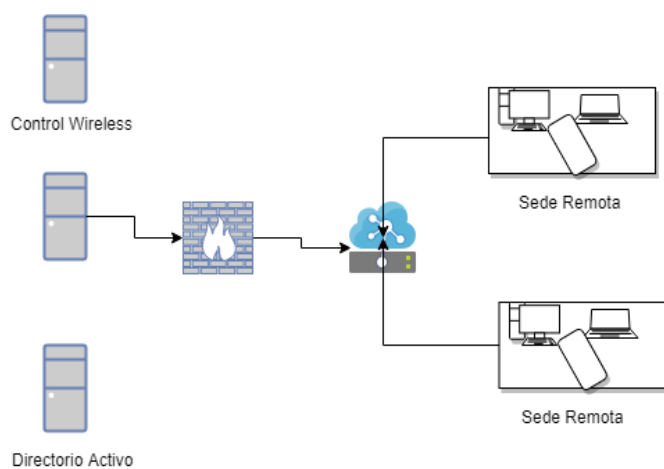
- a) Análisis realizados sobre las tendencias y líneas de desarrollo de la disciplina o profesión en el ámbito local, regional, nacional e internacional, y su incidencia en el programa.
- b) Estudios orientados a identificar las necesidades y requerimientos del entorno laboral (local, regional y nacional) en términos productivos y de competitividad, tecnológicos y de talento humano. Acciones del programa para atenderlos.

- c) Estudios que demuestren la necesidad social del programa en la metodología que se ofrece.
- d) Correspondencia entre el perfil laboral y ocupacional del sector y el perfil profesional expresado en el Proyecto Educativo del Programa.
- e) Estudios y/o proyectos formulados o en desarrollo, que propendan por la modernización, actualización y pertinencia del currículo de acuerdo con las necesidades del entorno.
- f) Estudios actualizados sobre las necesidades formativas en la región de influencia del programa.
- g) Cambios en el plan de estudios, resultantes de experiencias relativas al análisis y propuestas de solución a los problemas del contexto.
- h) Proyectos que adelanta el programa, mediante sus funciones de docencia, investigación, innovación, creación artística y cultural, y extensiones tendientes a ejercer un impacto sobre el medio, de acuerdo con el Proyecto Educativo del Programa.
- i) Estudios orientados a evaluar el impacto del programa con respecto al cumplimiento de sus propósitos y objetivos, así como la incidencia en el entorno social y su grupo de referencia disciplinario profesional.” (CNA, 2013).

Para el caso la ECBTI, no cuenta con un sistema que le brinde información de los requerimientos del mercado laboral que le permita incluirlos en sus planes de estudio y tiene que recurrir a encuestas y quizá algunas entrevistas a empresarios del sector para cumplir con los informes requeridos.

## 5.2 Importancia de la tecnología en el contexto

Para la Universidad Nacional Abierta y a Distancia el uso de la tecnología para apoyar y mejorar cada uno de sus procesos es de vital importancia y para ello la UNAD ha invertido gran esfuerzo tanto técnico como humano para lograr ser una de las pioneras en el uso de las TI, en la (Figura 1) se puede ver parte del esquema físico de conexión, específicamente para las labores de tipo académico.



**Figura 1.** Interconexión UNAD

Adaptado (UNAD, 2019)

Para el caso de los sistemas de información que puede ser uno de los que compete al proyecto propuesto, la UNAD, ha venido desarrollando una serie de sistemas que apoyan los procesos

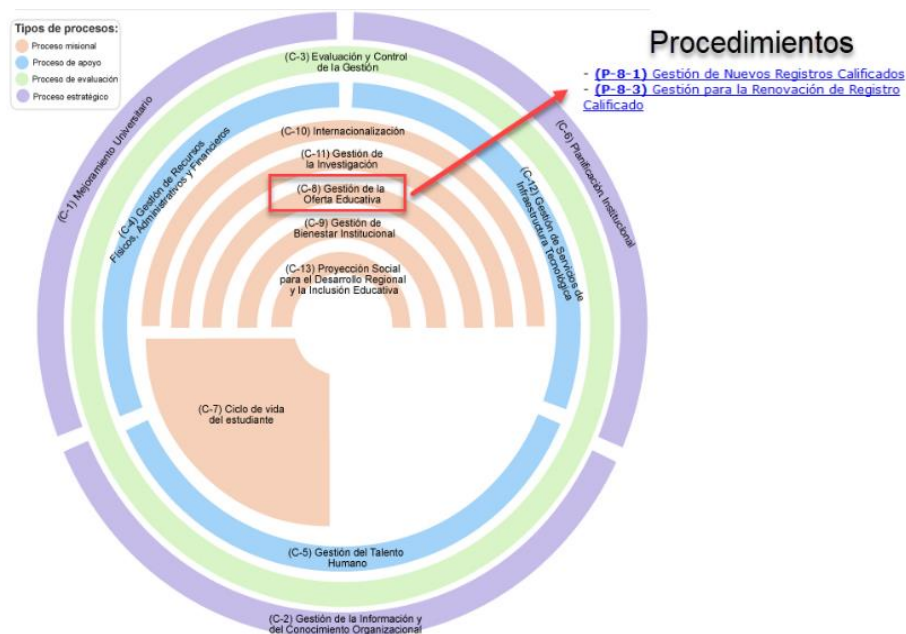
administrativos y de gestión de servicio y se han agrupado en académico, administrativo, financiero y de soporte; cómo se puede ver en la (Figura 2):



**Figura 2.** Aplicaciones UNAD

Fuente: (UNAD, 2019)

Con lo anterior el proyecto fundamenta su desarrollo como una herramienta que apoye un componente del proceso misional, referido a la oferta educativa y específicamente a la recolección de información para los procedimientos Gestión de nuevos registros calificados y Gestión para la renovación de registro calificado. (P-1-1 y P-8-3) como se muestra en la (Figura 3).



**Figura 3. Mapa de procesos UNAD**

Adaptado (UNAD, 2019)

### 5.3 Marco Teórico

Existen diferentes autores que citan la importancia del uso de las redes sociales como medio de análisis y fuente de información para el diseño herramientas que suplan necesidades en el entorno académico, en una investigación de (Rodríguez Ruibal & Santamaría Cristino, 2012), se identificó como las “Las Universidades españolas se encuentran en un momento en el que han superado su adaptación al uso de las Tecnologías de la Información y de la Comunicación (TIC), no sólo en su faceta más académica y de gestión, sino también en su uso como herramienta de comunicación y marketing”. Se puede identificar que las universidades españolas tienen estrategias particulares para su uso, de igual manera que el seguidor de estas no corresponde con

el número de estudiantes, entre otros aspectos. Asimismo (Criado & Rojas, 2013), en su texto “Las redes sociales digitales en la gestión y las políticas públicas. Avances y desafíos para un gobierno abierto”, en uno de sus apartes identifican que “los responsables públicos están identificando en las redes sociales una herramienta que les puede ayudar a mejorar su relación con la ciudadanía, así como innovar en sus procesos de gobierno y, por ello, están apostando de una manera generalizada por su uso y difusión”.

Un caso particular es el estudio de “Análisis de redes sociales aplicado a redes de investigación en ciencia y tecnología”. Realizado en la universidad de Valdivia Chile, los cuales “expone un sistema que toma los currículos de investigadores en ciencia y tecnología de la plataforma SICTI para generar archivos de redes sociales, los que sirven como entrada para software de análisis y visualización de redes sociales”. Para los procesos de análisis, diseño e implementación del sistema han tendido en cuenta, “por un lado, se tiene a SICTI que provee el módulo Currículo y con este módulo se tiene la red de investigadores; por otro lado, para realizar el análisis de redes sociales se eligió a Pajek y NetDraw. Entonces lo que falta es una herramienta que permita tomar toda la información disponible en el módulo Currículo, y la transforme de modo que pueda ser leída por los softwares para análisis de redes sociales, para luego realizar los análisis apropiados a las redes generadas aplicando el modelo de análisis creado.” (Navarro & Salazar, 2007).

Así mismo (Montoya & Carvalho, 2016), presentan un estudio “Flujo documental entre áreas administrativas de una entidad bancaria: una aproximación desde el análisis de redes sociales (ARS)”, donde “analizar doce departamentos administrativos de una entidad bancaria y sus

relaciones en términos documentales (flujos), abordados desde la unidad de investigación del Análisis de Redes Sociales (ARS).

Dichas interacciones de acuerdo con (Navarro & Salazar, 2007), se logran de acuerdo a mediciones en redes sociales a nivel de número, tamaño y conexiones entre los grupos o subgrupos, las cuales pueden explicar la conducta de la red como un todo. Se puede observar las conductas de los actores en función de su ubicación en los subgrupos. Éstos pueden actuar de puentes entre grupos, ser actores aislados o actores cosmopolitas.

A nivel de herramientas para análisis de redes sociales se encuentra el trabajo de (Rodríguez, 2007) donde se presenta el paquete *mpa* desarrollado en el software R que ejecuta el método de las palabras asociadas (MPA) y con la cual se analizan los temas presentes en los artículos publicados en la Revista Colombiana de Estadística. El análisis de estas temáticas se complementa con el análisis de correspondencias simples de las tablas de autores por año y palabras por año.

En el sector asegurador también existen trabajos como el de (Martínez, 2017) donde se realiza un análisis de sentimiento en Twitter de las principales compañías del sector asegurador español (PCSAE) mediante la utilización del software estadístico R y de léxicos de sentimiento. Es busca obtener y comprender la gran cantidad de información gratuita que circula por Internet, la cual es de mucho interés para las empresas. Por ejemplo, en la red social Twitter circulan al día millones

de opiniones sobre diversos temas, por lo que se decide utilizar esta red social para captar la opinión de los clientes sobre las PCSAE. En primer lugar, se extraen los comentarios de Twitter, denominados tuits, para posteriormente prepararlos y limpiarlos para el análisis. Resulta clave la creación de un léxico específico para el sector asegurador (Lexiseg), cuya construcción se realiza partiendo de otros léxicos en castellano ya existentes y la agregación manual de palabras que mejoran los resultados de exactitud. Una vez se dispone de la base datos y del Lexiseg se procede al análisis de sentimiento, en el que se obtienen diversos resultados para cada una de las PCSAE; polaridad, emociones, tópicos, etc. Finalmente se realiza un análisis y comparación de resultados entre las compañías y se abordan las principales conclusiones obtenidas en el trabajo.

En el área de la Minería de Datos e Inteligencia de Negocios está el trabajo de (Fernández, 2016), quién realiza un análisis de los contenidos expresados por los usuarios en las redes sociales, y más concretamente sobre Twitter. Se muestran los comportamientos sobre los temas seleccionados (BCE finanzas y cuestiones de reglamentación). Se muestran los temas, hashtags y cuestiones en torno BCE y la regulación financiera con el fin de construir toda una ontología para una clasificación efectiva. Se descubren temas, subtemas, hashtags y entidades semánticas relevantes.

A nivel de la política también existen trabajos como el de (García, 2014) en Chile, el cual desarrolla un método para extraer la opinión de los usuarios chilenos en Twitter, a través de un análisis de sentimientos, para catalogar cada tweet como positivo, negativo o neutro con respecto a los candidatos de las primarias Andrés Allamand y Pablo Longueira. Una vez extraída la

opinión por usuario, se buscaba predecir el resultado de las elecciones realizadas el año 2013. Para el análisis de los mensajes se desarrolla un programa en Perl que busca determinar la polaridad de cada tweet asociado a un usuario en específico y se determina la polaridad de cada mensaje, se utilizan herramientas estadísticas basadas en la frecuencia y el carácter de ciertas palabras, por lo cual se ocupó un diccionario con palabras positivas y negativas las cuales tienen un puntaje asociado que establece la polaridad (positivo, negativo o neutro) del tweet para cada candidato.

#### **5.4. Marco conceptual**

El análisis de sentimiento consiste en el estudio de la información contenida en un texto para identificar reacciones, actitudes, contexto y emociones, clasificándola de forma masiva y automática. Cabe resaltar que dicha clasificación se basa en relaciones estadísticas y de asociación, mas no en análisis lingüístico. (Jáuregui Fernández, 2018)

El análisis de sentimiento está basado en lexicons. Un lexicon es, en términos simples, un vocabulario (diccionario). En el presente contexto un lexicon es un conjunto de palabras puntuadas positiva y negativamente.

Las métricas más usadas de acuerdo a (Navarro & Salazar, 2007) son:

- a) **Clique:** es un subgrupo de una red en que sus actores están más cercanos y unidos que el resto de los actores de la red. Formalmente, un clique es un subgrupo de actores en el cual están presentes todos los vínculos posibles entre éstos.
- b) **N-clique:** es un subgrupo en el que todos los actores están separados a lo más por una distancia de N de los demás actores.
- c) **Componentes:** grupo en que todos sus miembros están conectados con al menos otro miembro.
- d) **Puntos de corte:** son nodos que, si se remueven, dividen la red en dos o más componentes.
- e) **Lexicon:** Es un conjunto de palabras puntuadas positiva y negativamente

Existen diferentes términos que se usan en la red social Twitter, de acuerdo al portal (webempresa, 2018) están:

**Análisis de redes sociales (ARS):** El análisis de redes sociales procede a través de dos grandes grupos de estrategias para la identificación de estructuras en una red: las medidas de cohesión y las medidas de equivalencia. En el primer grupo se encuentran todas aquellas medidas que parten de los lazos que los actores tienen entre sí. El ejemplo más sencillo de ello sería un clique o grupo de nodos conectados entre sí. Este principio se puede relajar y admitir conexiones a todos los miembros menos un número dado, a todos los miembros a una determinada distancia, etc. El segundo grupo de medidas intenta hallar posiciones equivalentes, es decir, patrones similares de relaciones con el resto de nodos. El caso extremo lo constituirían dos actores que, si

intercambiasen su posición, no afectarían al resto de propiedades de la red. (Molina, Muñoz, & Domenech, 2002)

**Twittero:** Es cada usuario registrado. Se representa con @NombreDelUsuario.

**Tweet (tuit):** Es cada uno de los mensajes que se publica. Recordar que cada uno de ellos contiene hasta 280 caracteres (se amplió de 140 caracteres en 2018) sin contar el material multimedia que se incluya en los contenidos.

**Time Line (línea de tiempo):** Es la parte de una cuenta en la que se puede ver, por orden cronológico, los mensajes de los usuarios que determinada persona sigue.

**Retweet (RT):** Es la republicación de un tweet lanzado por otro usuario.

**Follower (seguidor):** Es el usuario que sigue tu cuenta. Este puede ver todos los tweets que usted pública.

**Following (siguendo a):** Es la acción de seguir a otro usuario dentro de esa red.

**Lista:** Es un listado que se puede configurar con tus cuentas favoritas. Puede crear la cantidad de listas como desee y otorgarles un nombre.

**IM (Instant Message):** Es un mensaje directo y privado que se envía a un usuario de Twitter.

**Me gusta:** Está representado por un icono de corazón. Este se clikea si a un usuario le gusta un tweet.

**Hashtag:** El rey de Twitter. Se representa con un icono de almohadilla (#) y permite añadir tras él los términos que el usuario quiera. Se utiliza para facilitar búsquedas. Por ejemplo, usando #WordPress en el buscador, se puede encontrar un listado de los usuarios que han utilizado ese término en sus tweets.

**Trending Topic:** Son los temas más comentados del momento, es decir, las palabras con más menciones de la red social en un determinado periodo de tiempo. (webempresa, 2018)

**Lenguaje R:** Es un entorno y lenguaje de programación libre para computación estadística y gráficos. Compila y se ejecuta en una amplia variedad de plataformas UNIX, Windows y MacOS. (The R Foundation, 2019). Es un proyecto GNU. Por lo tanto, los usuarios son libres de modificarlo y extenderlo. Se trata de un lenguaje basado en comandos, en lugar de pinchar y arrastrar iconos o menús con el ratón se escriben comandos o instrucciones que son ejecutados. Una sucesión de instrucciones o comandos de R que implementa un flujo de trabajo para realizar una tarea se denomina script o guión R. Existe una amplia variedad de entornos de desarrollo para R que facilitan escribir scripts de R tales como R commander, RKWard y RStudio. (Romero, 2013)

R es un lenguaje orientado a objetos, es un lenguaje para el cálculo estadístico y la generación de gráficos, que ofrece una gran variedad de técnicas estadísticas y gráficas. Es un entorno de análisis y programación. Es un lenguaje de programación completo con el que se añaden nuevas técnicas mediante la definición de funciones. Aunque comenzar a trabajar con R es más complejo que hacerlo con programas como SPSS, Statgraphics, sin embargo, tiene muchas ventajas sobre ellos una de la más importante es que es un software libre en el que colaboran muchos usuarios para ampliar sus funciones. R se distribuye de acuerdo a GNU GENERAL PUBLIC LICENSE. (Lara & Hidalgo, 2016)

Según el libro (Santana, 2014), si este es el primer acercamiento a R, es muy probable que te cuestiones sobre la ventaja y la utilidad de R sobre otras paqueterías de estadística; pero R es más que eso. Es un software poderoso, que puede ser aplicado ampliamente en el procesamiento de datos en ciencias. Empezar diciendo que R es un lenguaje de programación interpretado, de distribución libre, bajo Licencia GNU, y se mantiene en un ambiente para el cómputo estadístico y gráfico. Este software corre en distintas plataformas Linux, Windows, MacOS, e incluso en PlayStation 3. El término ambiente pretende caracterizarlo como un sistema totalmente planificado y coherente, en lugar de una acumulación gradual de herramientas muy específicas y poco flexibles, como suele ser con otro software de análisis de datos. El hecho que R sea un lenguaje y un sistema, es porque forma parte de la filosofía de creación, como lo explica John Chambers (Chambers and Hastie [1991]), citó:

*“Buscamos que los usuarios puedan iniciar en un entorno interactivo, en el que no se vean, conscientemente, a ellos mismos como programadores.*

*Conforme sus necesidades sean más claras y su complejidad se incremente, deberían gradualmente poder profundizar en la programación, es cuando los aspectos del lenguaje y el sistema se vuelven más importantes.”*

**RStudio:** Entorno de desarrollo integrado para R. Un IDE que fue construido solo para R.

Disponible en ediciones de código abierto y comercial en el escritorio (Windows, Mac y Linux)

y desde un navegador web a un servidor Linux que ejecuta RStudio Server o RStudio Server Pro.

(RStudio, 2019)

“Características:

- Resaltado de sintaxis, finalización de código y sangría inteligente
- Ejecutar código R directamente desde el editor de fuente
- Saltar rápidamente a definiciones de funciones.
- Ayuda y documentación integradas de R
- Gestione fácilmente múltiples directorios de trabajo usando proyectos
- Navegador de espacios de trabajo y visor de datos.
- Depurador interactivo para diagnosticar y corregir errores rápidamente.
- Amplias herramientas de desarrollo de paquetes.” (RStudio, 2019)

## 6. Metodología

Para el desarrollo del proyecto es necesario realizar un estudio de enfoque mixto como lo indica (Hernández, Fernández, & Baptista, 2003) involucrando estudios cualitativos – cuantitativo de tipo descriptivo e interpretativo.

Para el logro de los resultados esperados en la investigación desde el contexto de las redes sociales se hace uso de: análisis ARS (Análisis de redes Sociales), “un avance teórico y metodológico que investiga relaciones, enlaces, contactos, pautas relacionales y estructuras, es decir, redes. Las redes están compuestas de nodos (actores) y líneas (enlaces), y la finalidad es analizar todo este conjunto de nodos y líneas: la cohesión, los subgrupos, la centralidad de los nodos, la composición de los nodos y su relevancia en la creación de enlaces, y muchos otros procedimientos.” (Cárdenas, 2016)

Se enmarca dentro de lo cuantitativo dado que el ARS impone la medición tanto del número de posibles relaciones existentes entre los involucrados, como también la dirección y profundidad de la red existente, para lo cual se utilizará instrumentos de medición y visualización existentes en el software seleccionado.

En lo cualitativo dado que se tendrá en cuenta las particularidades presentes en cada uno de los grupos objeto de estudio, lo que lleva a que en el proceso de la investigación se haga las

indagaciones e interpretaciones de la información que determinarán y se recopilarán desde el contexto descriptivo e interpretativo.

El proyecto se desarrolló en las siguientes fases:

Fase 1: **Revisión de referentes teóricos:** consulta de información relacionada con ARSV validando información relacionada del orden académico y laboral con el propósito de ampliar el marco teórico.

Fase 2: **Identificar y/o diseñar una herramienta informática para captura de información:** Identificar herramientas informáticas que permitan capturar y analizar datos de las redes sociales que se adapte a las necesidades objeto de investigación o diseñar una a la medida de las necesidades.

Fase 3: **Captura de información:** capturar la información pública en redes virtuales de las instituciones identificadas.

Fase 4: **Preparación de información:** Verificar calidad y pertinencia de la información recolectada obtenida con la herramienta propuesta.

Fase 5: **Análisis de resultados:** Presentación de los resultados de tendencias laborales y de formación obtenidos con la herramienta web, bajo el lenguaje R y el resultado del comportamiento de la misma.

Fase 6: **Diseño de guía de usuario:** Diseñar una guía con los parámetros básicos que permitan hacer ARS utilizando la herramienta informática ARS SIAVA propuesta y que sirva de soporte a los programas de la UNAD.

Fase 7: **Socialización de resultados del proyecto:** Presentación de los resultados obtenidos a la comunidad académica.

A continuación, se describen en el desarrollo del proyecto.

### **6.1.Revisión de referentes teóricos**

Para el caso del proyecto varios de los referentes teóricos se presentaron en el apartado referido al Marco Teórico, sin embargo, para el caso se hace importante fundamentar con otros referentes de estudios y se retoma que el concepto de red social ha adquirido una importancia notable en los últimos años. Se ha convertido en una expresión del lenguaje común que se asocia a nombres como Facebook o Twitter. Pero su significado es mucho más amplio y complejo. Las redes sociales son, desde hace décadas, objeto de estudio de numerosas disciplinas. Alrededor de ellas

se han generado teorías de diverso tipo que tratan de explicar su funcionamiento y han servido, además, de base para su desarrollo virtual. Con la llegada de la Web 2.0, las redes sociales en Internet ocupan un lugar relevante en el campo de las relaciones personales y son, asimismo, paradigma de las posibilidades que nos ofrece esta nueva forma de usar y entender Internet (Ponce, 2019).

Con lo anterior se pueden establecer diversos estudios que permite determinar que este tipo de investigaciones son válidas para identificar patrones, tendencias y predecir comportamientos de diversos grupos de personas al interior de una red, como es el caso del (Alcantara, 2015), en su estudio sobre el caso “El Caso De Los Representantes Sectoriales En El Consejo Territorial De Planeación Del Municipio De Arauca (2012 – 2015)”, donde hace un análisis de “las estructuras que subyacen a propósito de la participación de actores de la sociedad civil en los gobiernos locales, tomando como referencia El caso particular del Consejo Territorial de Planeación del Municipio de Arauca (2012-2015)”, donde, uno de los principales hallazgos es: Al examinar mediante Análisis Egocéntrico de Redes Sociales, las relaciones que en torno a la discusión y tratamiento de aspectos sociales, económicos y políticos, generaban los representantes de los sectores en el Consejo Territorial de Planeación del Municipio de Arauca (2012 -2015), se logró identificar estructuras pequeñas, poco densas, homófilas, con escasa participación de mujeres y jóvenes, compuestas principalmente por amigos, que interactúan frecuentemente en torno a los temas ya señalados.

Otro tipo de estudio como el de (Rodríguez Ruibal & Santamaría Cristino, 2012), donde identifica que las “Las Universidades españolas se encuentran en un momento en el que han superado su adaptación al uso de las Tecnologías de la Información y de la Comunicación (TIC), no sólo en su faceta más académica y de gestión, sino también en su uso como herramienta de comunicación y marketing”. Donde puede identificar que las universidades españolas tienen estrategias particulares para su uso, de igual manera que el seguidor de estas no corresponde con el número de estudiantes, entre otros aspectos.

(Criado & Rojas, 2013), en su texto “Las redes sociales digitales en la gestión y las políticas públicas. Avances y desafíos para un gobierno abierto”, en uno de sus apartes identifican que “los responsables públicos están identificando en las redes sociales una herramienta que les puede ayudar a mejorar su relación con la ciudadanía, así como innovar en sus procesos de gobierno y, por ello, están apostando de una manera generalizada por su uso y difusión”,

Un caso particular es el estudio de “Análisis de redes sociales aplicado a redes de investigación en ciencia y tecnología”. Realizado en la universidad de Valdivia Chile, los cuales “expone un sistema que toma los currículos de investigadores en ciencia y tecnología de la plataforma SICTI para generar archivos de redes sociales, los que sirven como entrada para software de análisis y visualización de redes sociales”. Para los procesos de análisis, diseño e implementación del sistema han tendido en cuenta, “por un lado, se tiene a SICTI que provee el módulo Currículo y con este módulo se tiene la red de investigadores; por otro lado, para realizar el análisis de redes sociales se eligió a Pajek y NetDraw. Entonces lo que falta es una herramienta que permita tomar

toda la información disponible en el módulo Currículo, y la transforme de modo que pueda ser leída por los softwares para análisis de redes sociales, para luego realizar los análisis apropiados a las redes generadas aplicando el modelo de análisis creado.” (Navarro & Salazar, 2007).

(Montoya & Carvalho, 2016), presentan un estudio “Flujo documental entre áreas administrativas de una entidad bancaria: una aproximación desde el análisis de redes sociales (ARS)”, donde “analizar doce departamentos administrativos de una entidad bancaria y sus relaciones en términos documentales (flujos), abordados desde la unidad de investigación del Análisis de Redes Sociales (ARS).”, permitió el análisis de una estructura organizacional desde un modelo alternativo, en el que se observó las relaciones de los autores dentro de un campo específico, permitiendo así con esta aproximación, dejar abierta la puerta para análisis más profundos y con diversos objetivos que puedan dar cuenta de las dificultades a las que se enfrenta una organización, especialmente cuando se trata del capital social al interior de ella. Por tanto, el ARS es una herramienta que puede ser aplicada dentro de las organizaciones para observar especialmente esos “cuellos de botella” que dificultan las interacciones entre los colaboradores y no permiten el avance deseado por la compañía.

Para lograr el cometido de la presente investigación es importante realizar una revisión de herramientas informáticas que permitan un análisis adecuado de la información, es así como (Larrosa, 2016), hace una recopilación de 30 herramientas de análisis y visualización de las redes sociales

## 6.2. Identificar una herramienta informática para captura de información

En esta etapa de exploración de herramientas para la extracción y el análisis de datos en la Redes Sociales se han identificado algunas herramientas de Software en su mayoría de uso libre que cumplen la función de monitorización de redes sociales, extracción de datos y análisis de la información.

La clave para sacarle todo el provecho a las redes sociales es escuchar lo que la audiencia dice sobre ti, analizar esos datos y finalmente alcanzar inteligencia empresarial: usando toda esta información podrás conocer mejor a tus clientes y mejorar tu estrategia de marketing.

**Hootsuite:** Es una herramienta gratuita disponibles para la administración de las redes sociales, cubre múltiples plataformas tales como Twitter, Facebook, LinkedIn, WordPress, Foursquare y Google+ (Figura 4).

Los informes de análisis semanales y su excelente aplicación de flujo de trabajo (permite la delegación de tareas y enviar mensajes privados) son muy útiles cuando hay más de una persona manejando las cuentas de las redes sociales. (Hootsuite, 2019)

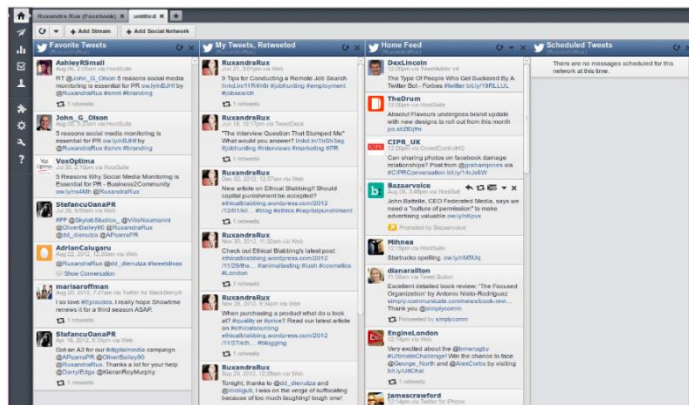


Figura 4. Gestión de Redes Sociales

Fuente: (Hootsuite, 2019)

**TweetReach:** Es una herramienta propietaria no de uso libre que permite monitorizar el alcance que tienen los tweets, Analiza la audiencia de una cuenta Twitter, de un hashtag o una palabra clave, esta herramienta mide el impacto real y las implicaciones de las conversaciones de las redes sociales (Figura 5). Es una forma de enterarse de quiénes son los seguidores con más influencia, y lo guía hacia la audiencia correcta en la que se debe centrar cuando se quiere compartir y promover contenido online. (Tweatreach, 2018)



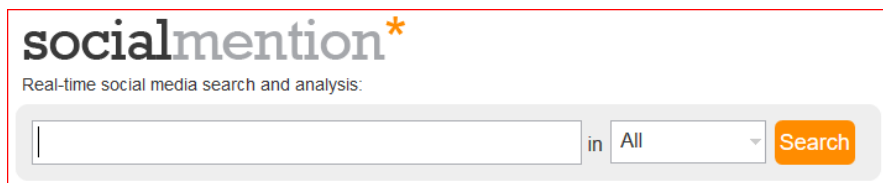
**Figura 5.** Estadísticas de actividad en Twitter

Fuente: (Tweetreach, 2018)

**Social Mention:** Es una herramienta de uso libre que permite monitorizar más de cien sitios de redes sociales es una aplicación web de búsqueda y análisis de todos los contenidos agregados por los usuarios en internet. permite monitorear quién te menciona, dónde y cuándo. Con ella puedes medir la influencia de tu marca, evento o asunto en blogs, microblogs, redes sociales, bookmarks, etc. Puede rastrear tu influencia en más de 80 medios sociales entre ellos: Twitter, Facebook, LinkedIn, YouTube, Google, Digg, etc. (Emprender Fácil, 2016)

Ingresando a la página principal de **Social Mention**, se muestra una barra de búsqueda, donde se puede insertar la palabra clave a buscar (Figura 6).

A la derecha de esta barra de búsqueda, está un menú desplegable que podrá ver y seleccionar la o las plataformas específicas que quiere que el buscador actúe. La configuración por defecto es “todo”. Al seleccionar una de ellas, la herramienta filtrará los resultados de búsqueda y te proporcionará sólo las métricas y los resultados para la plataforma que hayas seleccionado. Una vez que se haya seleccionado la palabra clave y la plataforma de búsqueda, haz clic en “buscar”, y observa los resultados.



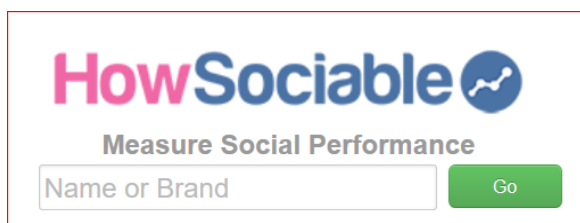
**Figura 6.** Interfaz de búsqueda de Socialmention

Fuente: (Emprender Fácil, 2016)

**HowSociable:** Es una herramienta muy útil para medir la presencia de tu marca y la de tus competidores en las redes sociales.

Una cuenta gratis te permite rastrear 12 plataformas de redes sociales, incluyendo Tumblr y WordPress. Si quieres monitorizar 24 plataformas más: Facebook, Pinterest, Twitter, etc., entonces necesitas una cuenta pro. (HowSociable, 2019)

HowSociable presenta los datos de manera un poco diferente, con diferentes puntuaciones para cada plataforma, lo que te permite comprobar qué plataforma funciona mejor para tu marca y cuáles son las que necesitan más atención. (Figura 7)

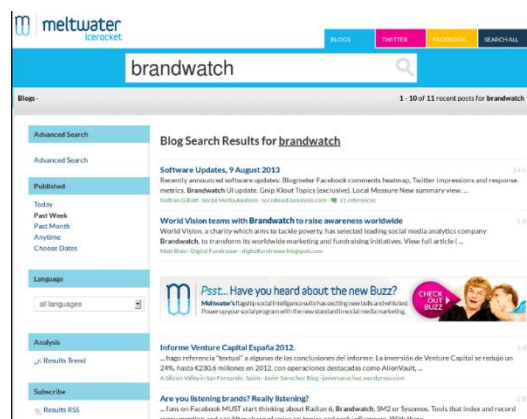


## Figura 7. Interfaz de búsqueda de HowSociable

Fuente: (HowSociable, 2019)

**IceRocket:** Esta herramienta te permite monitorizar Facebook, Twitter y blogs en 20 idiomas y te presenta también los resultados en gráficos. Permite seleccionar los periodos de tiempo que te interesa monitorear. (Meltwater, 2019)

Puede ser muy útil para estar al tanto de la actividad de tus bloggers ya que IceRocket tiene cerca de 200 millones de blogs en su base de datos, y también te da la posibilidad de encontrar las últimas tendencias relacionadas con tu búsqueda (Figura 8).



## Figura 8. Interfaz de búsqueda de meltwater

Fuente: (Meltwater, 2019)

**Ncapture:** Es una herramienta de uso libre para la captura de contenido en redes sociales tales como Facebook, YouTube, Twitter y entre otras, es fácil de adquirir y usar, tiene la opción de instalarla en Chrome como en internet Explorer (Figura 9). Es una herramienta gratuita que permite capturar contenido de la web y redes sociales como Facebook, Twitter y YouTube para importarlo y analizarlo en NVivo. (QSR International, 2019).

Ventajas:

- Gratis y resultados en PDF
- Rápida y fácil manejo
- Instalación y descarga fácil. Disponible en Buscador Chrome e internet Explorer
- Captura páginas web y PDF en línea, publicaciones en la pared de Facebook y comentarios, videos y comentarios de YouTube, discusiones en linkedIn Tweets de Twitter incluyendo palabras, frases o hashtags particulares o Tweets de un usuario en particular.
- Es en la nube

Desventajas:

- No hay muchos ejemplos o referentes de uso en español
- Idioma ingles
- No es novedoso
- Formato solo para Nvivo y no incluye captura de datos privados

- Permite solo capturar solo un artículo, publicación, comentario, etc. a la vez

Cuando se usa de Ncapture se obtiene las siguientes características:

Captura páginas web y PDF en línea, publicaciones en la pared de Facebook y comentarios, videos y comentarios de YouTube, discusiones en LinkedIn Tweets de Twitter incluyendo palabras, frases o hashtags particulares o Tweets de un usuario en particular. El formato de la información extraída es único para visualizarla y analizarla en la herramienta NVivo, la captura de texto se realiza en tipo PDF de una sección o de toda la página y en video la captura se hace en formato de video. Se encuentra disponible para Windows y Mac, no captura datos privados. Presenta la siguiente interfaz.



**Figura 9.** Interfaz de Ncapture

Fuente: (QSR International, 2019)

**Nvivo:** Es un software de uso comercial para investigaciones con datos cualitativos y métodos mixtos, ayuda a organizar, analizar y encontrar perspectivas en datos no estructurados como por ejemplo entrevistas, artículos y contenido de las redes sociales (QSR International, 2019)

Ventajas:

- Analiza Facebook, Twitter, webs, archivos desde el equipo y desde la nube
- Usuarios Windows y Mac
- Almacena y organiza información demográfica cuantitativa e información cualitativa
- Brinda herramientas para categorizar y clasificar datos: Clasifica automáticamente el sentimiento, temas y atributos
- Muestra fácilmente los datos
- Importa datos de prácticamente cualquier fuente: texto, audio, video, correos electrónicos, imágenes, hojas de cálculo, encuestas en línea, contenido social y web, y más
- Permite el manejo de la información y sus opciones de visualización
- Fácil instalación
- Versión en inglés y español

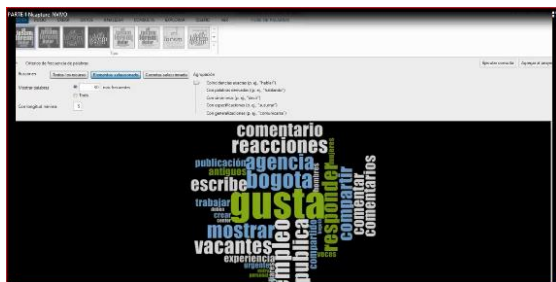
Desventajas:

- Licencia de pago
- Es compleja
- Hay poca información en español
- No tiene formas de visualización avanzada

No permite la entrada legible de la información de Twitter

Nvivo analiza de acuerdo a las necesidades, cuenta con las herramientas de análisis como categorías, búsqueda, clasificación, nodos, recursos y otras. Presenta los informes del análisis de tipo gráfico o tabla o imagen, es un programa muy completo por sus características de funcionamiento, fácil uso, múltiples idiomas y permite trabajar colaborativamente (Figura 10).

Nvivo ofrece una versión de prueba gratis y de pago. Para la instalación de la versión gratis en Windows requiere Procesador 1.2 o 1.4 GHz, Memoria RAM 2 GB o más, Espacio disco 5GB o más, Pantalla 1024x768, sistema operativo Windows 7 o superior. Presenta la siguiente interfaz.



**Figura 10.** Interfaz de Nvivo

Fuente: (QSR International, 2019)

**FangPage Karma:** Es una herramienta comercial que permite realizar recolección de datos y un análisis detallado y completo de páginas en Facebook, Twitter y otras redes sociales (Uphill GmbH, 2019)

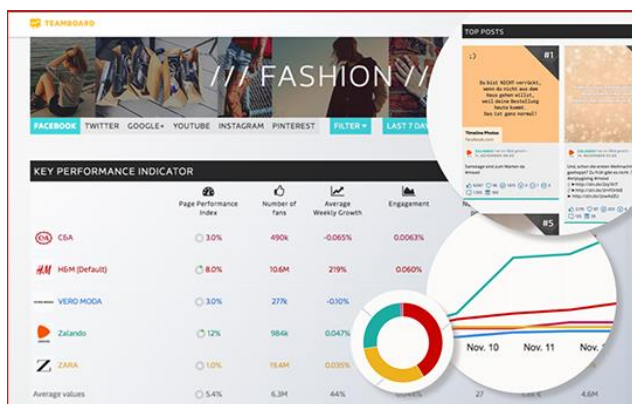
Ventajas:

- Versión de prueba
- Recolecta datos
- Idioma español y fácil de usar
- Resumen de Fan page y analítica detallada de contenido, publicaciones y otras
- Actualizaciones, categorías e informes en Excel y Power Point
- Comparación entre páginas
- Idioma español
- Buena visualización de los análisis
- Disponible para uso en la nube
- Para Facebook y Twitter y más redes sociales

Desventajas:

- Licencia de pago
- Solo da 14 días de prueba
- No permite al usuario manipular los datos y por lo tanto se limita a las necesidades del usuario
- Se vuelve lenta cuando analiza demasiada información

Es una herramienta en línea de marketing, se encuentra en varios idiomas incluido el español, ofrece una versión gratuita y de pago. FangPage Karma permite la recolección y análisis detallado y completo de páginas en Facebook, twitter y otras redes sociales y entre otras características. Se usa específicamente para observar o comparar la competencia. Para su uso puede loguearse con Facebook. Presenta informes detallado en Excel o PowerPoint (Figura 11).

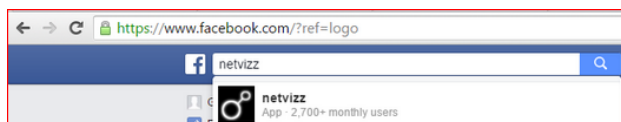


**Figura 11.** Interfaz de Fanpagekarma

Fuente: (Uphill GmbH, 2019)

**Netvizz:** es una herramienta online de uso libre para la captura de datos en Facebook, permite extraer información de secciones de Facebook, ya sea de páginas, grupos, imágenes, estadísticas, ofrece la opción de obtener un informe por alcance de fecha hasta 999 mensajes, periodo de publicaciones, estadísticas (métricas de publicaciones, estadísticas por día y fanáticos por país), estadísticas y 200 comentarios mejor clasificados por publicación o datos completos. Las salidas de datos son en formato comprimido (UP2, 2018).

Netvizz está limitada para grupos y datos privados, es sencilla y fácil de usar, disponible en lenguaje inglés. Para hacer uso de esta herramienta hay que loquearse con Facebook, agregar Netvizz app desde el buscador interno de la cuenta de Facebook, seguir los pasos de instalación y configuración (Figura 12).



**Figura 12.** Interfaz de Netvizz

Fuente: (UP2, 2018)

**Exportcoments:** Es una herramienta online muy útil para exportar los comentarios primarios públicos más recientes, desde cualquier publicación o video público de Facebook, Instagram,

Twitter, TikTok o YouTube a CSV Excel. Entre las características de Exportcoments es que puede acceder a los comentarios sin que el usuario sea propietario de cuenta de Facebook, o Twitter que produjo la publicación. Se puede acceder a publicaciones como un estado, una foto, un video o una URL, para exportar los comentarios siempre que sea público. La facilidad de uso simplemente copie en la URL de la publicación, y se exportan todos los comentarios disponibles.

Como la facilidad de uso, la disponibilidad de esta al ser una herramienta Online, el tipo de archivo que genera fruto de la exportación, es propicia para extraer información de las diferentes redes sociales, la capacidad de extracción, entre otros. Exportcoments es una herramienta online para exportar y descargar en formato Excel o CVS comentarios públicos de cualquier página de Facebook, Instagram, Twitter que sea público. (Exportcomments, 2019)

**IBM Watson:** Plataforma de análisis de contenido y exploración cognitiva, podrá oír cómo hablan sus datos. Explore y analice contenido estructurado, no estructurado, interno, externo y público para descubrir tendencias y patrones que mejoren la toma de decisiones, el servicio al cliente y el retorno de la inversión. Los clientes de IBM han utilizado Watson Explorer para comprender el 100% de sus correos electrónicos y llamadas entrantes, mejorando así la calidad de la información y su capacidad para utilizar dicha información. Las funcionalidades cognitivas integradas — basadas en modelos de machine learning, procesamiento del lenguaje natural y APIs de próxima generación — pueden desbloquear valor de todos sus datos y ayudarle a

obtener una vista de 360 grados de sus clientes, en contexto, para crear mejores experiencias para sus clientes. (IBM, 2019).

**FreeLing:** Es una librería de código abierto para el procesamiento multilingüaje automático, que proporciona una amplia gama de servicios de análisis lingüístico para diversos idiomas. FreeLing ofrece a los desarrolladores de aplicaciones de Procesamiento del Lenguaje Natural funciones de análisis y anotación lingüística de textos, con la consiguiente reducción del coste de construcción de dichas aplicaciones. FreeLing es personalizable y ampliable, y está fuertemente orientado a aplicaciones del mundo real en términos de velocidad y robustez. Los desarrolladores pueden utilizar los recursos lingüísticos por defecto (diccionarios, lexicones, gramáticas, etc.), ampliarlos, adaptarlos a dominios particulares, o dado que la librería es de código abierto desarrollar otros nuevos para idiomas específicos o necesidades especiales de las aplicaciones. (Padró, 2019)

**KH Coder:** es un software gratuito para análisis cuantitativo de contenido o minería de texto. También se utiliza para la lingüística computacional. Puede analizar textos en catalán, chino (simplificado), holandés, inglés, francés, alemán, italiano, japonés, coreano, portugués, ruso, esloveno y español con KH Coder. (KH Coder, 2019)

**R:** Es un entorno de software libre para computación estadística y gráficos. Compila y se ejecuta en una amplia variedad de plataformas UNIX, Windows y MacOS. R es un conjunto integrado de

instalaciones de software para la manipulación de datos, el cálculo y la visualización gráfica. Incluye una instalación efectiva de manejo y almacenamiento de datos, un conjunto de operadores para cálculos en matrices, en particular matrices, una colección grande, coherente e integrada de herramientas intermedias para el análisis de datos, facilidades gráficas para el análisis y visualización de datos en pantalla o en papel, y un lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas por el usuario e instalaciones de entrada y salida. (The R Foundation, 2019)

### **7.2.1. Selección de la herramienta**

Una vez realizada la revisión de herramientas disponibles se propone el desarrollo de una herramienta informática construida en lenguaje R, entre sus ventajas se encuentran, es un lenguaje de distribución gratuita, bajo los términos de la GNU General Public Licence. Además, contiene los archivos necesarios para instalar R, ya sea desde las fuentes o binarios pre compilados, los cuales se distribuyen desde una gran comunidad almacenada en el sitio de internet Comprehensive R Archive Network (CRAN) junto con las instrucciones de instalación.

R posee muchas funciones para análisis estadísticos y gráficos; estos últimos pueden ser visualizados de manera inmediata en su propia ventana y ser guardados en varios formatos (jpg, png, bmp, ps, pdf, emf, pictex, xfig; los formatos disponibles dependen del sistema operativo). Los resultados de análisis estadísticos se muestran en la pantalla, y algunos resultados intermedios (como valores P-, coeficientes de regresión, residuales, . . .) se pueden guardar,

exportar a un archivo, o ser utilizados en análisis posteriores. El lenguaje R permite al usuario, por ejemplo, programar bucles para analizar conjuntos sucesivos de datos. También es posible combinar en un solo programa diferentes funciones estadísticas para realizar análisis más complejos. Usuarios de R tienen a su disponibilidad un gran número de programas escritos para S y disponibles en la red; la mayoría de estos pueden ser utilizados directamente con R. (Paradis, 2002).

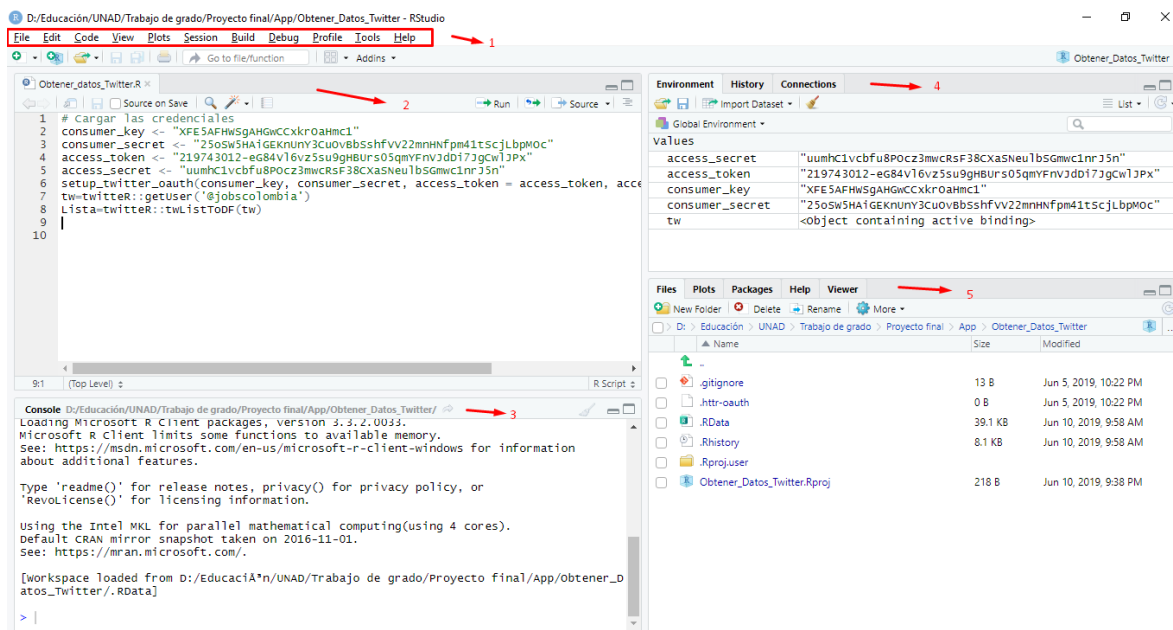
## 7.1 Captura de información

A partir de lo anterior se desarrolla una aplicación web interactiva, la cual se estructura bajo el lenguaje de programación R, utilizado como entorno de desarrollo integrado (IDE) a RStudio versión 1.2.1335 o superior (Figura 13), interfaz que permite acceder de manera sencilla a toda la potencia de R.

### Requerimientos

- Sistema operativo Windows 7 o superior
- Instalar R (versión 3.6.1 o superior)
- RsTools (Versión 3.5 o superior)

**Nota:** Para utilizar RStudio se requiere instalar R previamente.



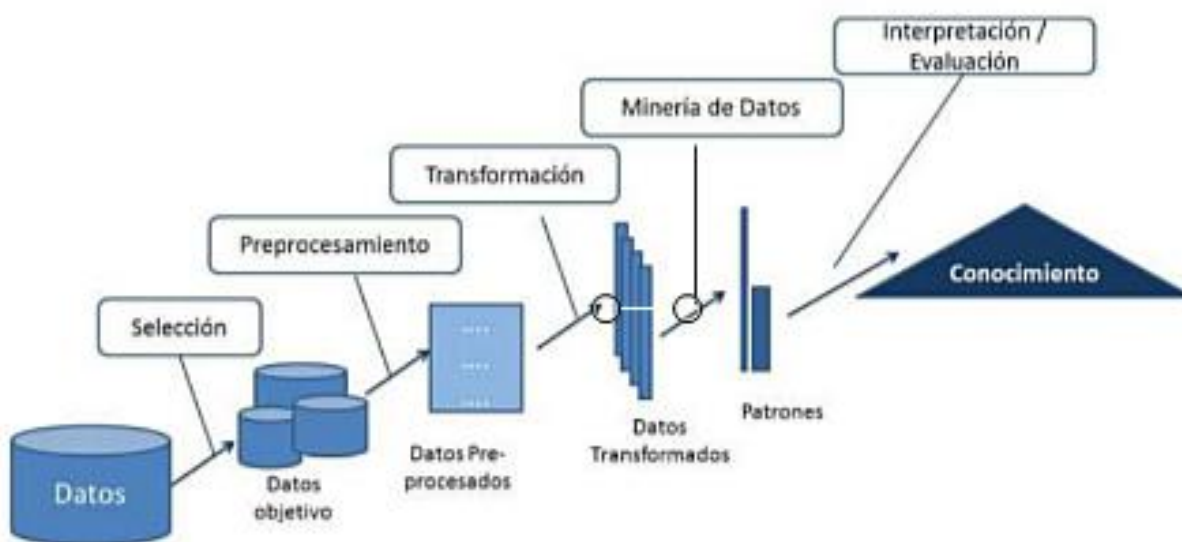
**Figura 13.** Entorno del IDE RStudio. Barra del menú principal (1), Editor (2), Consola (3), Historial de objetos almacenados en memoria (4), Directorio de trabajo (5).

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

La aplicación web se ejecuta en cuatro pasos fundamentales (Figura 14) como son:

- **Paso 1: Recolección** de datos: Se realiza través de API de Twitter y algoritmo en R.
- **Paso 2: Pre- procesamiento de datos:** Se realiza “limpieza” de datos y transformación al formato deseado.
- **Paso 3: Análisis de sentimiento:** Entrenar un clasificador (modelo) para clasificar tweets. Por ejemplo, positivo o negativo.

- **Paso 4: Predicción:** Utilizar estadísticas de las etiquetas de los tweets para medir el sentimiento positivo o negativo.

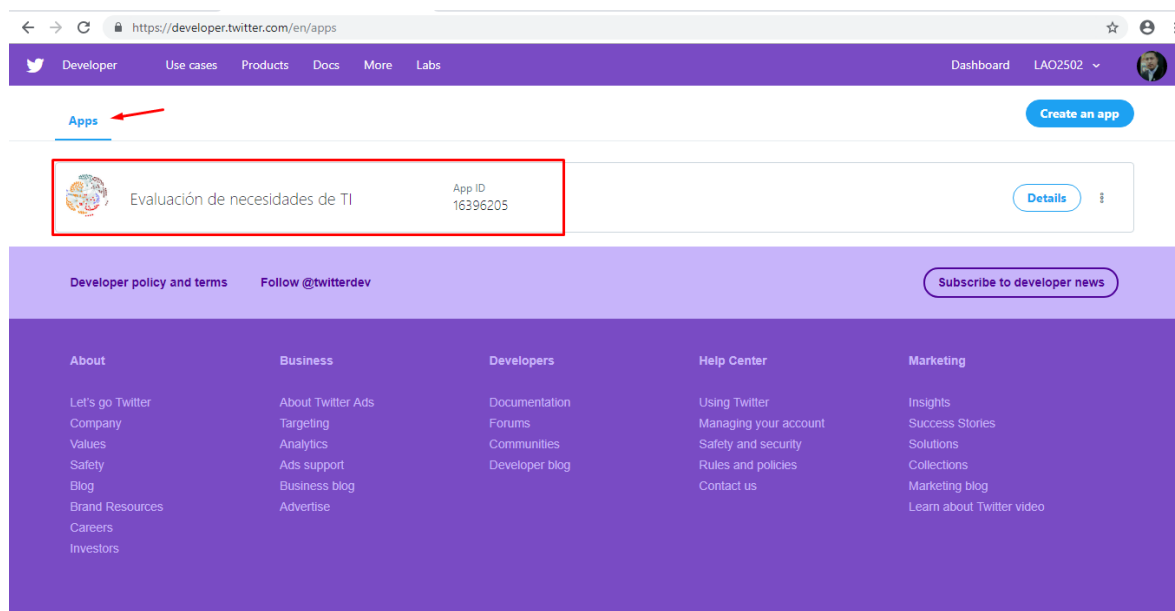


**Figura 14.** Descubrimiento de conocimiento en datos/textos

Fuente: (Lugo, 2019)

### Conexión

Se realiza el enlace a la base de datos de Twitter por medio de la API (Application Programming Interface) (Figura 15 y 16), para lo cual se utiliza programación con R, mediante el IDE RStudio versión libre (Figura 13).

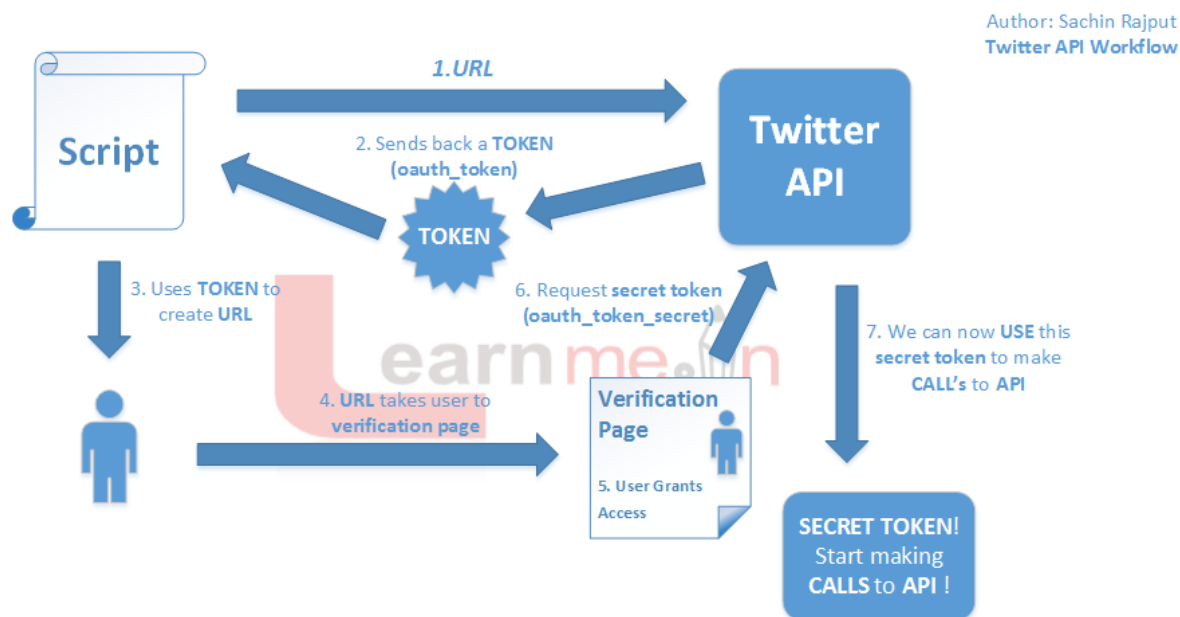


**Figura 15.** API de Twitter desde una cuenta de desarrollador.

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

### 7.3.1. Esquema de conexión

La captura de datos se realiza desde la plataforma Twitter. La siguiente imagen realiza una esquematización del modelo de enlace que se establece entre la aplicación de un desarrollador y la API de la plataforma Twitter.



**Figura 16.** Entorno del área de estudio.

Fuente: (Rajput, 2019)

### 7.3.2. Plataforma de red social utilizada

Como plataforma de recopilación de datos se escogió a Twitter, es una red social versátil y permite interactuar con otras aplicaciones (Aplican restricciones) debido a la disponibilidad de acceso por medio de una API. Twitter es actualmente una dinámica e ingente fuente de contenidos que, dada su popularidad e impacto, se ha convertido en la principal fuente de información para estudios de Social Media Analytics. Análisis de reputación de empresas, productos o personalidades, estudios de impacto relacionados con marketing, extracción de opiniones y predicción de tendencias son sólo algunos ejemplos de aplicaciones. (Amat, 2017).

Asimismo, de acuerdo con Web empresa, Twitter es una plataforma social, con un servicio de comunicación bidireccional con el que puede compartir información de diversos tipos de una forma rápida, sencilla y gratuita. (webempresa, 2018)

“Twitter es una lanzadera de pensamientos”.

En otras palabras, se trata de una de las redes de microblogging más populares que existen en la actualidad y su éxito reside en el envío de mensajes cortos llamados “tweets”.

Fue creada por Jack Dorsey y su equipo en 2006 y la idea se inspira en el envío de fragmentos cortos de texto (de 140 caracteres hasta 2018), donde puedes añadir un enlace, imágenes, vídeo, encuestas o incluso un gift.

Pero, ¿para qué sirve Twitter exactamente?

Es un formato muy completo, que permite obtener de forma inmediata información en formato de titular.

Rápido de leer y rápido de escribir. Así de fácil.

Esta plataforma tiene carácter de auto publicación basado en la inmediatez de sus mensajes.

Es una red perfectamente estructurada para compartir experiencias y vivencias en el momento en que suceden.

El tipo de contenido que se publica en Twitter es de diversa naturaleza: Se puede encontrar mensajes personales, fotografías, infografías, información corporativa, noticias, eventos, descuentos, publicidad, etc. (webempresa, 2018).

### **7.3.3. Planes y condiciones de uso**

**Precios:** Para el uso el desarrollo de la fase de campo se hace uso de la suscripción gratuita. El acceso API de Twitter se adapta a usted y a sus necesidades, para ello ofrece varios niveles de acceso a los datos para satisfacer las necesidades de las aplicaciones en todas las etapas de crecimiento. Puede comenzar con paquetes estándar y gratuitos, conseguir crecer con el acceso premium (Figura 17, Tabla 1) o considerar las suscripciones empresariales para obtener la mejor funcionalidad y acceso. (Twitter, Inc, 2019)

**Tabla 1.** *Niveles de suscripción de API Twitter.*

Fuente: (Twitter, 2019)

<b>Estándar</b>	<b>Premium</b>	<b>Empresariales</b>
<p>API estándar gratuitas son excelentes para comenzar, probar una integración o validar un concepto.</p> <p>Con:</p> <p>Fundamentos, API gratis, complejidad básica de consultas</p> <p>Acceso al foro</p>	<p>Ofrecen acceso escalable a los datos de Twitter para aquellos que buscan crecer, experimentar e innovar.</p> <p>Incluye:</p> <p>Acceso escalable a mayor cantidad de datos.</p> <p>Caja de arena gratis y contratos flexibles de mes a mes</p> <p>Acceso al foro</p>	<p>Nuestras API empresariales ofrecen el más alto nivel de acceso y confiabilidad a quienes dependen de los datos de Twitter.</p> <p>Incluye:</p> <p>API de nivel empresarial</p> <p>Paquetes a medida y contratos anuales</p> <p>Administradores de cuentas dedicados y soporte técnico</p>

Tweets	Estándar (gratis)	Prima	Empresa
Publicar y participar	✓		
Buscar tweets: 7 días	✓		
Buscar tweets: 30 días		✓	✓
Buscar tweets: archivo completo		✓	✓
Filtrar Tweets	✓		✓
Muestra de tweets	✓		✓
Lotes de Tweets			✓
<b>Mensajes directos</b>	✓		
<b>Cuenta y usuarios</b>	✓	✓	✓
<b>Métrica</b>			✓
<b>API de anuncios</b>	✓		
<b>Herramientas de editor y SDK</b>	✓		

**Figura 17.** Permisos de Acceso API según nivel de suscripción.

Fuente: (Twitter, 2019)

### 7.3.4 Método para captura de información

Es una herramienta informática web, la cual contiene un módulo para realizar recopilación de datos desde Twitter, su funcionamiento se establece mediante un clic desde un menú principal a la pestaña recopilar datos (Figura 18). La interfaz gráfica (Figura 19) está desarrollada por medio de una aplicación web, la cual se establece a través de una aplicación Shiny, que es una herramienta para crear aplicaciones web interactivas programada en lenguaje R.



**Figura 18.** Visualización y acceso del módulo de Recopilar datos

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

La siguiente ventana permite establecer conexión con Twitter a través de la API (Interfaz de programación de aplicaciones), además de realizar búsquedas de palabras clave.



**Figura 19.** Entorno principal del módulo de Recopilación

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Para lograr el funcionamiento de la plataforma anterior, lo primero que debe hacer es conectar con la API de Twitter, cargar todas las librerías, sacar la información de las palabras clave a analizar, conseguir sus tweets, convertirlo todo en un dataframe y descargarlo a un formato de texto CSV.

### **7.3.5 Codificación y parametrización**

A través de un script en R se establece la conexión y búsqueda en la base de datos de la plataforma de la red social, para ello se debe seguir esta secuencia:

- a) Instalar librerías TwitteR, ROAuth y httr

Para instalar desde la línea de comandos de RStudio escriba

```
install.packages("TwitteR")
```

```
install.packages("ROAuth")
```

```
install.packages("httr")
```

b) Ejecutar las librerías instaladas anteriormente.

Para ejecutar desde la línea de comandos de RStudio escriba

```
library(twitteR)
```

```
library(ROAuth)
```

```
library(httr)
```

Una vez realizado el paso anterior, se realiza la conexión por medio de la librería TwitteR con la base de datos de Twitter. Las líneas de código que permiten dicha conexión son las siguientes:

```
# Cargar las credenciales
```

```
consumer_key <- "xxx"
```

```
consumer_secret <- "xxx"
```

```
access_token <- "xxx"
```

```
access_secret <- "xxx"
```

```
# Valida las credenciales y realiza la conexión
```

```
setup_twitter_oauth(consumer_key, consumer_secret, access_token=access_token,  
access_secret=access_secret)
```

Ahora que se ha realizado la interfaz para conexión, solamente falta implementar las funciones para realizar la búsqueda. Estas son:

```
#Realiza una búsqueda con base a las palabras ingresadas en "..."
```

```
consulta <- searchTwitter("palabra a consultar", n=cantidad de datos a consultar)
```

```
#Creando un dataset de los datos obtenidos
```

```
Lista=twitterR::twListToDF(consulta)
```

```
#Establezco el directorio de trabajo
```

```
setwd("ruta de la carpeta donde se quiere exportar el documento")
```

*#Exportando datos – esta función realiza la escritura de un documento csv*

```
write.csv(Lista, file="busqueda.csv",row.names = F)
```

### 7.3.6 Validación del script

Para la interfaz gráfica se usa la librería Shiny, para este ejemplo se usa como palabra clave “Empleo Colombia”.

Se inicia con la ejecución de la aplicación y clic en la opción datos -> Recopilar datos, se muestra un formulario de conexión (Figura 20)

#### Establecer conexión

**Todos los campos son obligatorios**  
*\*Ingrese datos de enlace con la app-twitter y clic en conectar\**

**Ingrese clave (consumer key)**

**Ingrese clave secreta (consumer secret)**

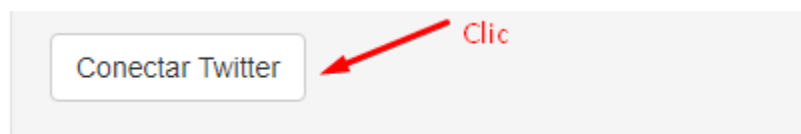
**Ingrese token de acceso (token access)**

**Ingrese token secreto de acceso (secret access)**

**Figura 20** Formulario de conexión

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Es obligatorio ingresar todos los datos solicitados en este formulario y posteriormente dar clic en Conectar Twitter (Figura 21). Si algún dato es incorrecto o faltante la conexión será rechazada.

**Figura 21.** Botón para establecer conexión con API Twitter

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Tras finalizar la conexión exitosamente se muestra un mensaje de confirmación (Figura 22).

## Establecer conexión

**Todos los campos son obligatorios**  
*\*Ingrese datos de enlace con la app-twitter y clic en conectar\**


**Ingrese clave (consumer key)**  
r51oloQdGaKsN30Hx8TvmU5Z8

**Ingrese clave secreta (consumer secret)**  
eVB9Vw2qTqjB7PQ7RVGNCGljz6UNam3nqbkiF

**Ingrese token de acceso (token access)**  
219743012-efZ5bpOckPPxlqr3TL1TBuEj0CYLvf9y

**Ingrese token secreto de acceso (secret access)**  
GVB3RzuGzK9tUiGHttiWpCJK1eKTn0L9eNuMwF

Conectar Twitter

Conexión realizada 

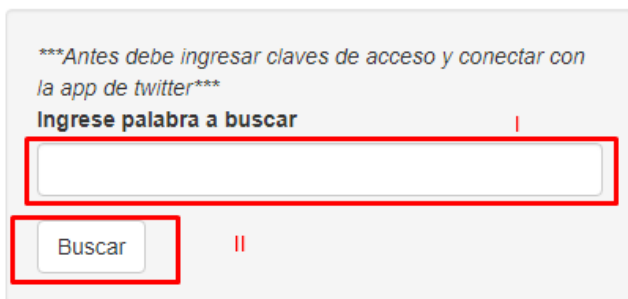
**Figura 22.** Notificación estado de conexión

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

- **Búsqueda**

Para este ejemplo se utiliza como palabra clave “Empleo Colombia”. Ingresar estos datos en el cuadro de búsqueda (Figura 23).

## Recopilación de datos



\*\*\*Antes debe ingresar claves de acceso y conectar con la app de twitter\*\*\*  
Ingrese palabra a buscar

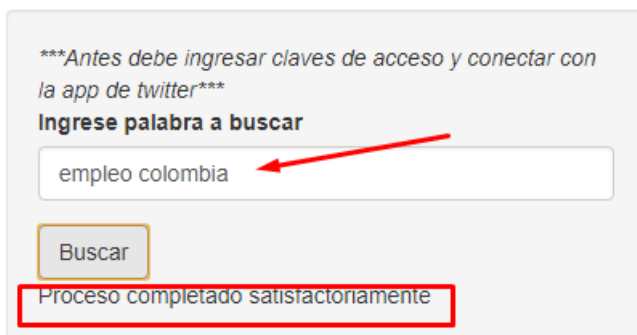
Buscar

**Figura 23.** Formulario de búsqueda

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Una vez finaliza el proceso de búsqueda se notifica con mensaje en pantalla “Proceso completado satisfactoriamente” (Figura 24)

## Recopilación de datos



\*\*\*Antes debe ingresar claves de acceso y conectar con la app de twitter\*\*\*  
Ingrese palabra a buscar

Buscar

Proceso completado satisfactoriamente

**Figura 24.** Notificación de finalización de consulta a través de API de Twitter

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Finalizada la consulta el sistema imprime una tabla en pantalla (Figura 25), y un archivo CSV con los tweets recopilados desde Twitter, este último se guarda en carpeta local (Ruta elegida por el implementador) (Figura 26), cuyo nombre es “busqueda.csv”. Para acceder al documento generado vaya a la carpeta indicada por el administrador. Ahora puede acceder al archivo así mismo puede copiarlo y guardarlo en su lugar preferido o entorno de trabajo.

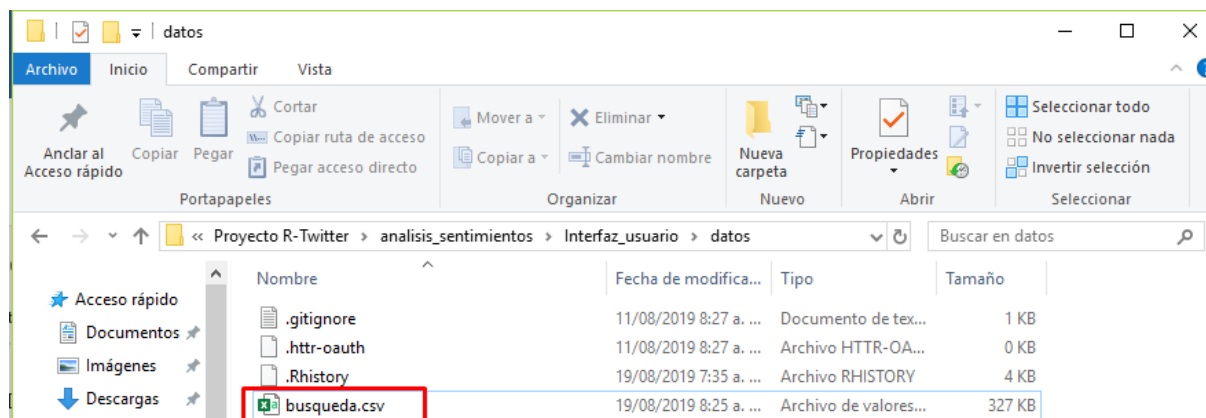
Este es el reporte en pantalla de la consulta realizada de forma exitosa

text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID	statusSource
RT @Gradiant: #WeAreHiring! Buscamos Ingeniero/a Senior de Sistemas de Comunicaciones ¿serás tú? ¡Queremos conocerte! Consulta todos los de...	FALSE						66699775381504	NA	<a href="http/rel="nofollow":
<div style="border: 1px solid #ccc; padding: 10px; background-color: #f9f9f9;"> <h3 style="margin: 0;">Recopilación de datos</h3> <p style="margin: 5px 0 0 20px;">***Antes debe ingresar claves de acceso y conectar con la app de twitter***</p> <p style="margin: 5px 0 0 20px;"><b>Ingrese palabra a buscar</b></p> <div style="border: 1px solid #ccc; padding: 5px; margin: 5px 0 0 20px; width: 80%;"> <input type="text" value="empleo ingeniero sistemas"/> </div> <div style="border: 1px solid #ccc; padding: 5px; margin: 5px 0 0 20px; display: inline-block;"> <p style="margin: 0;">Buscar</p> </div> <p style="margin: 5px 0 0 20px;">Proceso completado satisfactoriamente</p> </div>									
#WeAreHiring! Buscamos Ingeniero/a Senior de Sistemas de Comunicaciones ¿serás tú? ¡Queremos conocerte! Consulta to... https://t.co/l4z0wJ4r1X	FALSE						00277321768960	NA	<a href="https/rel="nofollow":
Ingeniero en Sistemas, Radio Caracas T, #Caracas, Distrito Capital, #Venezuela https://t.co/6QNKwJQ8L9 #Empleo	FALSE						16308697542657	NA	<a href="http/rel="nofollow":
RT @ing_empleo: Oferta #Empleo #ingeniero/a #graduado/a rama industrial – Procesos y Sistemas y de Puesta en Marcha #Bizkaia #jobs	FALSE	0.00	NA	1568801612.00	FALSE	NA	1174265187286884352	NA	<a href="https/rel="nofollow":

**Figura 25.** Tabla generada en pantalla de Tweets consultados

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Asimismo este es el directorio de destino con el archivo generado por el sistema y guardado localmente.



**Figura 26.** Documento exportado a carpeta local

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

#### 7.4. Preparación de información

Este proceso corresponde a la etapa previa a la importación de datos para realización de análisis de sentimientos, la cual requiere el desarrollo de los siguientes pasos:

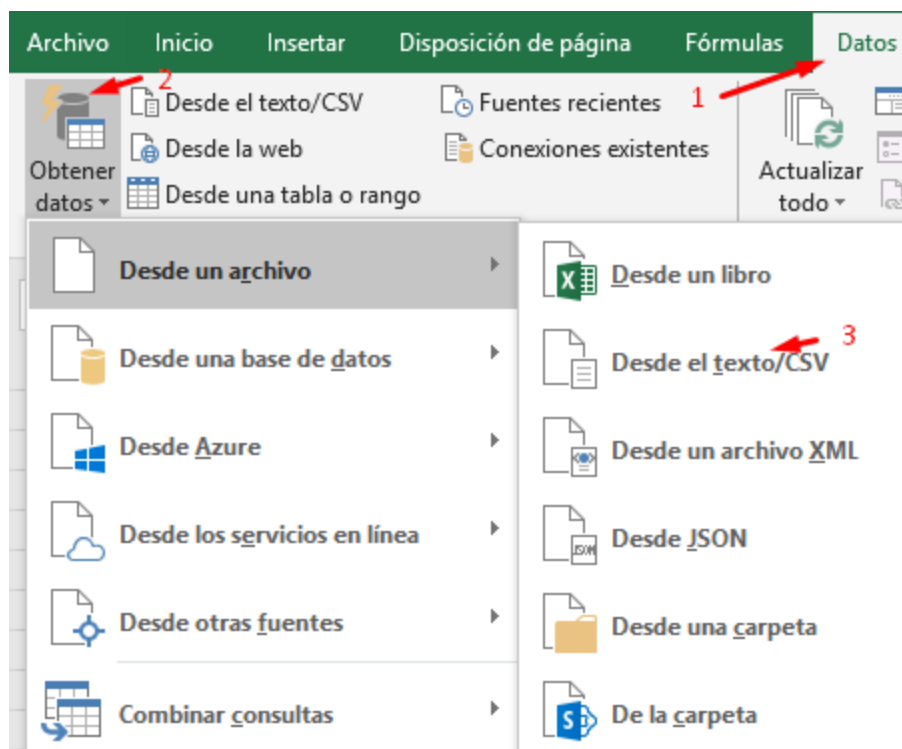
- Clasificación de los tweets
- Limpieza de datos

#### 7.4.1. Clasificación de tweets

Esta etapa requiere la clasificación manual de cada uno de los tweets de la base de datos de entrenamiento. Para lo cual es necesario añadir una columna al documento con nombre **Sentiment**, aquí se debe colocar a cada tweet su sentimiento (positivo = 1; negativo= -1)

Para ello abra Microsoft Excel e importe el documento CSV a trabajar. Para ello realice los siguientes pasos:

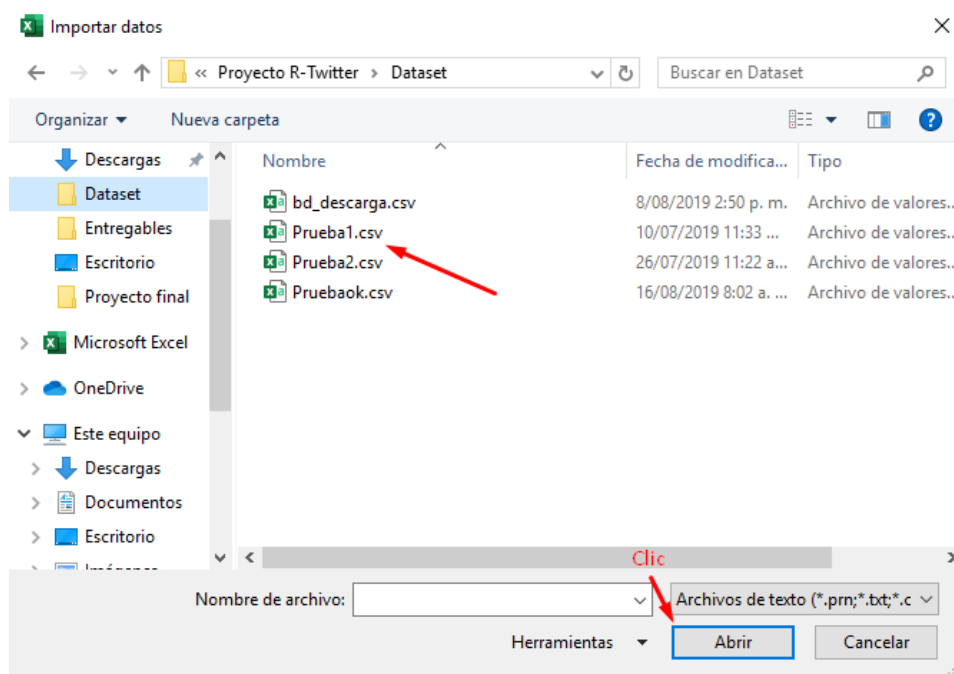
Desde la opción Datos -> Obtener datos -> Desde el texto/CSV. (Figura 27)



**Figura 27.** Ruta para importar datos CSV a Microsoft Excel

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Se abre la ventana para buscar el documento, seleccionarlo y clic en Abrir (Figura 28)



**Figura 28.** Pasos en ventana para importar documento

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

En la siguiente ventana (Figura 29) realice lo siguiente:

Origen de archivo: 1252: Europeo occidental (Windows) | Delimitador: Coma | Escoja el delimitador de los datos | Detección del tipo de datos | Basado en las primeras 200 filas

text	favorited	favoriteCount	replyToSN	created	truncated	replyTr
RT @ClauNaj: Esta Mendoza con dos bonos fiscales cre...	FALSE	0	NA	8/08/2019 7:49:29 p. m.	FALSE	NA
Yo una vez conocí una mesera de un restaurante de los...	FALSE	0	NA	8/08/2019 7:49:27 p. m.	TRUE	NA
RT @silviapaez: #Empleo Para quienes están buscando...	FALSE	0	NA	8/08/2019 7:49:24 p. m.	FALSE	NA
RT @majomalinis: WTF? Hasta cuando van a seguir con...	FALSE	0	NA	8/08/2019 7:49:14 p. m.	FALSE	NA
RT @IngeneoSAS: Estamos en búsqueda de un Analista...	FALSE	0	NA	8/08/2019 7:49:14 p. m.	FALSE	NA
@GlanceRockera Dooooo necesito el de Photoshop má...	FALSE	0	GlanceRockera	8/08/2019 7:49:13 p. m.	FALSE	115955117
Esta Mendoza con dos bonos fiscales creó más de 3500...	FALSE	1	ClauNaj	8/08/2019 7:49:10 p. m.	TRUE	115955203
RT @PartidoMIRA: Por el fortalecimiento del empleo, e...	FALSE	0	NA	8/08/2019 7:49:00 p. m.	FALSE	NA
RT @Kicillofok: Gobernar es cuidar la estructura produc...	FALSE	0	NA	8/08/2019 7:48:56 p. m.	FALSE	NA
@CNNChile Por lo que se infiere de la propuesta, quere...	FALSE	0	CNNChile	8/08/2019 7:48:50 p. m.	TRUE	115948260
Ey @mauriciomacri solo se banca un poco el grito si de...	FALSE	0	NA	8/08/2019 7:48:40 p. m.	TRUE	NA
¡Sin los mismos de siempre, avanzaremos como nunca!...	FALSE	0	NA	8/08/2019 7:48:26 p. m.	TRUE	NA
RT @Kicillofok: Gobernar es cuidar la estructura produc...	FALSE	0	NA	8/08/2019 7:48:19 p. m.	FALSE	NA
Quiero invitar a todas y todos los buscadores de emple...	FALSE	0	NA	8/08/2019 7:48:17 p. m.	TRUE	NA
Para solicitar presiona el siguiente enlace: https://t.co/...	FALSE	0	NA	8/08/2019 7:48:14 p. m.	FALSE	NA
RT @PodemosMostoles: En las últimas negociaciones,...	FALSE	0	NA	8/08/2019 7:48:12 p. m.	FALSE	NA
<U+0001F449> Con el propósito de mejorar las condici...	FALSE	0	NA	8/08/2019 7:48:12 p. m.	TRUE	NA
@JuanOrlandoH @radiohrn @HCHTelevDigital @diario...	FALSE	0	JuanOrlandoH	8/08/2019 7:48:11 p. m.	TRUE	115954468
RT @elcaporal_ver: Ni más ni menos que 107 mexicano...	FALSE	0	NA	8/08/2019 7:48:05 p. m.	FALSE	NA
RT @CCGranada: El presidente de Cámara de Comercio...	FALSE	0	NA	8/08/2019 7:48:02 p. m.	FALSE	NA

Clic → Cargar Transformar datos Cancelar

**Figura 29.** Ventana de carga y transformación de datos

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Finalice con clic en cargar, ahora tiene los datos en una tabla excel desde donde puede agregar, quitar y editar columnas y filas.

Aquí debe eliminar datos que no requiera o no sean necesarios , además de agregar la columna sentiment (Figura 30).

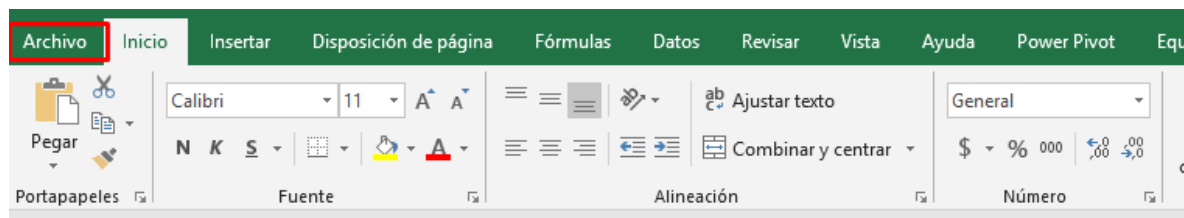
	A	B	C	D
	text	screenName	retweetCount	sentiment
2	Programadores(as) de Sistemas, L.A. Sistemas S, #Edo. Miranda (chacao), #Venezuela https://t.co/PKP2uthIQd #EmpleoEmpleateVeSist		0	1
3	programadores de sistemas, L.A. Sistemas S, #Edo. Miranda (chacao), #Venezuela https://t.co/Vup3qNpM7K #EmpleoEmpleateVeAdmon		0	1
4	Asistente de Sistemas. – Ecuador Laboral https://t.co/4RIXskDdSL #Guayaquil #Ecuador #Trabajo #EcuadorLaboral... http EcuadorLaboral		0	1
5	Administrador/a de Sistemas BBDD#empleo #trabajo #Alicante #trabajoalicanteInscribete aqui <U+0001F449> https://t. TrabajosAlacant		0	1
6	TIRADOR DE CABLE PARA INSTALACION DE SISTEMAS DE SEGURIDAD/ALARMAS#empleo #trabajo #Alicante #trabajoalica TrabajosAlacant		0	1
7	Especialista de Sistemas, Cadena de Farma, #Caracas, Distrito Capital, #Venezuela https://t.co/OKPMHC42u7 #EmpleoEmpleateVeSist		0	1
8	Nueva oportunidad de empleo Estudiante de ingeniería industrial / sistemas o administración de empresas:... https://t Accusys_Arg		0	1
9	Ingeniero de Redes y Sistemas @michaelpage_es #Barcelona #SystemCenter #empleo #tic #it https://t.co/j5qJxTHwSe tecnoempleo		0	1
10	Buscamos 2 técnicos de sistemas #mdm con experiencia en #microsoft #intune. Es imprescindible experiencia en otras Ibermatica_RRHH		0	1
11	Técnico/a de sistemas Linux#empleo #trabajo #Barcelona #empleobarcelonaInscribete aqui <U+0001F449> https://t.co/ Jobs_in_BCN		0	1
12	Oferta de #trabajo! En @CAI_Sistemas buscamos #programador/a #desarrollador/a #Delphi en Vigo. <U+27A1><U+FE0F: CAI_Sistemas		0	1
13	PANEL Sistemas Informáticos selecciona para León Analista Programador Java, C++, ADA (3 vacantes)... https://t.co/O6fi leonredcom		1	1
14	Oferta de empleo Responsable de Sistemas IT https://t.co/gRfXlhSDdO lwantic		0	1
15	Técnico/a sistemas de gestión integrado #MedioAmbiente y calidad – @LBconsultores1 #Barcelona... https://t.co/Ks6fk ISMedioambiente		0	1
16	La primera de las normas está relacionada con el empleo de los sistemas inalámbricos de alta velocidad en las banda... CruzYomil		0	1
17	Buscamos un Técnico de Sistemas Junior Unix para importante empresa. ¡Mándanos tu CV a seleccion@mensoft.es! #rr MensoftC		0	1
18	La primera de las normas está relacionada con el empleo de los sistemas inalámbricos de alta velocidad en las banda... radiorebeldecu		10	1
19	ingeniero de sistemas, Alimentos Dagan, #Caracas, Distrito Capital, #Venezuela https://t.co/nW0EWUIYnn #EmpleoEmpleateVeSist		0	1
174	Nueva oportunidad de empleo ADMINISTRADOR DE SISTEMAS WINDOWS: https://t.co/IH6BJXIMJq #Empleo #Trabajo h KaunisIaky		0	1
175	Para algunos ingenieros de sistemas la programación es un karma, para otros es una obligación, para otros es una ne... ¡FernandoNaviaB1		0	-1
176	<U+0001F6AB> El 72% del empleo doméstico es informal (no tiene aportes, vacaciones ni aguinaldo)Necesitamos:<U+00 mayrasmendoza		37	-1
177	"Es necesario racionalizar el uso de la Justicia oor los ciudadanos. debiendo incrementarse el empleo de sistemas a... h iuecesAPM		1	-1

**Figura 30.** Documento Excel con columna sentiment

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Finalmente, para completar este paso es necesario guardar el documento Excel en formato CSV.

Para ello debe ir a la opción Archivo -> Guardar como (Figura 31)

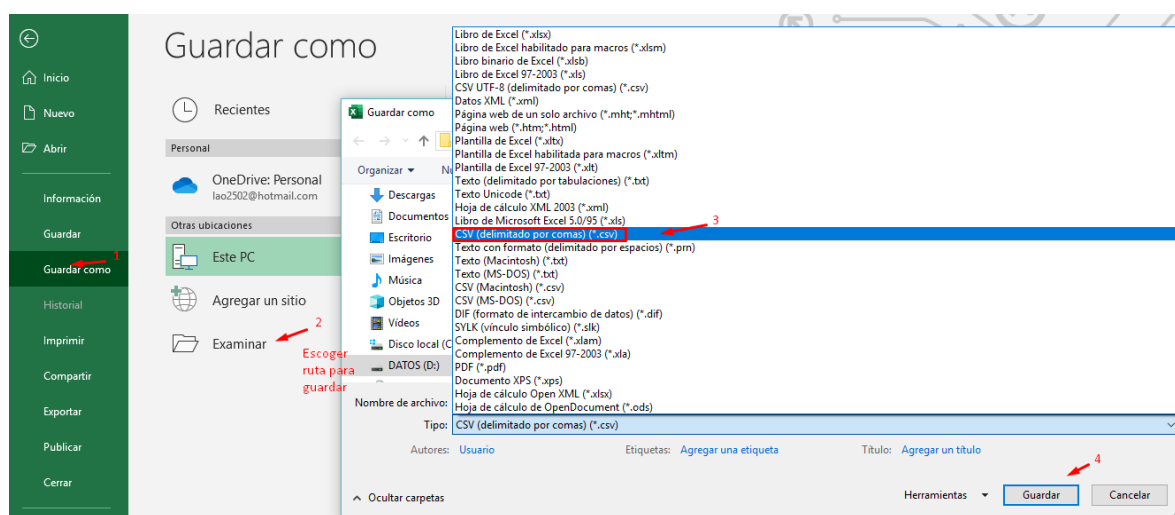


**Figura 31.** Cinta de opciones Microsoft Excel

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Y proceder como se muestra en la (Figura 32).

Esta figura muestra el paso a paso para guardar un documento.



**Figura 32.** Ventana importar datos en Microsoft Excel

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

Ahora tiene el documento listo para el siguiente paso.

#### 7.4.2. Minería de textos

Se realiza a través del segundo módulo de análisis desde el menú principal de la herramienta informática web, su acceso se establece mediante un clic desde un menú principal a la pestaña analizar (Figura 33). La interfaz gráfica está desarrollada por medio de una aplicación web, la cual se establece a través de una aplicación Shiny, que es una herramienta para crear aplicaciones web interactivas programada en lenguaje R.



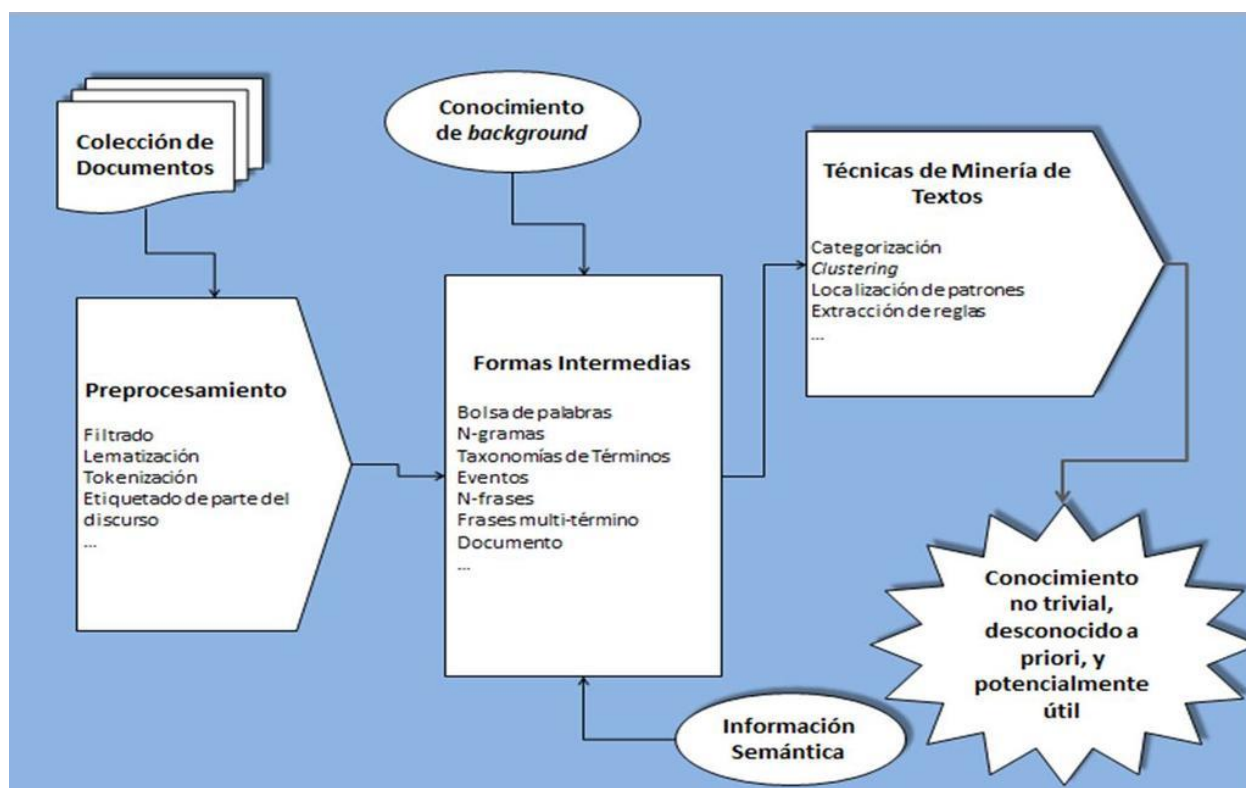
**Figura 33.** Visualización y acceso del módulo de Análisis

Fuente: (Ortiz Palma, Figuras de apoyo visual ARS Twitter, 2019)

En la anterior gráfica se despliega una ventana, la cual corresponde al entorno de incorporación de los datos a analizar en la plataforma.

Para realizar dicho análisis a nivel interno ocurren una serie de procesos. De allí que como en todo proyecto de ciencia de datos, la limpieza y el análisis exploratorio de estos resultan de vital

importancia para las siguientes etapas del mismo. De hecho, solo el pre procesado de los datos representa entre un 70 y un 80% de todo el trabajo. Se refiere a tareas como la corrección de codificaciones y tipos de datos erróneos, transformaciones de los datos o la detección y manejo adecuado de datos anómalos y faltantes. (Arcón, 2017)



**Figura 34.** Técnicas de tratamiento de datos del Corpus

Fuente: (Lugo, 2019)

Después de realizar la recolección de datos al Corpus (Colección de documentos) es necesario darles un tratamiento a los datos, algunos de los cuales se citan en la (Figura 34).

### 7.4.3. Script de limpieza

Este proceso se realiza desde R, utilizando la librería *tm* para minería de datos, para lo cual se ejecuta un Script para limpieza de datos, como el siguiente:

```
# Esta función convierte la tabla a texto para poder visualizarlo dentro del corpus

Corpus <- tm_map(Corpus, PlainTextDocument)
```

Teniendo en cuenta que los datos de los tweets son Datos NOSQL (No estructurados). Después de transformar el texto en un corpus (bolsa de palabras), se puede ejecutar varias funciones para caracterizarlo. El tipo más común de limpieza corresponde a eliminar todo aquello irrelevante o de poco aporte a los tweets.

```
# Esta función eliminar http

removeURL <- content_transformer(function(x) gsub("(f|ht)tp(s?):?/\\S+", "", x, perl=T))

# Esta función elimina las palabras que empiezan por "http." seguidas de cualquier carácter que no sea un espacio
```

```
Corpus <- tm_map(Corpus, removeURL)
```

```
#Esta función elimina la puntuación
```

```
Corpus <- tm_map(Corpus, removePunctuation)
```

```
#Esta función quita números
```

```
Corpus <- tm_map(Corpus, removeNumbers)
```

```
#Esta función transforma todas las palabras a minúsculas
```

```
Corpus <- tm_map(Corpus, content_transformer(tolower))
```

```
#Esta función remueve las stopwords (palabras muletillas) y otras palabras que no se deseen
```

```
Corpus <- tm_map(Corpus, removeWords, c(stopwords("spanish"), "las", "empleo"))
```

```
#Esta función acorta las palabras a su raíz "devolvieron" "devolver"
```

```
Corpus <- tm_map(Corpus, stemDocument, language="spanish")
```

```
#Esta función quita los espacios vacíos excesivos
```

```
Corpus <- tm_map(Corpus, stripWhitespace)
```

#### **7.4.4. verificación de datos**

Aquí se presentan los resultados de tendencias laborales y/o de formación obtenidos con la herramienta web ARS SIAVA, la cual fue desarrollada bajo el lenguaje R, utilizando como interfaz gráfica la librería Shiny.

Continuando con el ejemplo se observa que la muestra es de 218 tweets (Función 1) y que a nivel de sentimientos se distribuye con 6 para tweets negativos y 212 positivos (Tabla 2)

#### **Función 1. Código para conteo de tweets en R**

Fuente: (Rstudio, 2019)

```
> length(Corpus) #cuenta las palabras tiene el corpus
```

```
[1] 218
```

A continuación, se muestra el proceso de limpieza utilizando las diferentes funciones descritas anteriormente. Para validar el funcionamiento se utiliza como tweet impreso el #27.

Ahora bien, con la función `table(importdata$sentiment)` se muestra la relación que hay con los tweets positivos y negativos.

Básicamente, es el proceso permite determinar el tono emocional que hay detrás de una serie de palabras, y se utiliza para intentar entender las actitudes, opiniones y emociones expresadas en una mención de las redes sociales.

**Tabla 2.** *Relación de tweets vs sentimientos*

Fuente: (Rstudio, 2019)

-1	1
6	212

Convenciones: sentimiento positivo= 1; Negativo= -1

Lo anterior demuestra que los tweets sobre la consulta “trabajo ingeniero sistemas” realizada a través de Twitter es con tendencia a percepción positiva. Esto será de gran contribución más adelante, porque permitirán decidir sobre el tipo de datos a utilizar para el entrenamiento del modelo y posterior predicción de tendencias.

Para entrar en detalle de cada uno de los procesos de limpieza de datos se procede a revisar un tweet en particular.

```
> content(Corpus[[27]]) #imprime la posición 27
```

```
[1] "Proyecto de Ciencias Natutales con TIC tema Los Reinos de los seres Vivos. Se trabajo con Realidad aumentada y Codi... https://t.co/YTQxg310AT"
```

Ahora se observa lo que ocurre a lo largo del preprocesamiento o limpieza de los datos expuesta anteriormente.

El primer paso consiste en eliminar todo los que está pegado a un http. La función es:

```
> removeURL <- content_transformer(function(x) gsub("(f|ht)tp(s?):/\|S+", "", x, perl=T))
```

```
> Corpus <- tm_map(Corpus, removeURL)
```

Esta función elimina las palabras que empiezan por "http." seguidas de cualquier cosa que no sea un espacio), ahora se debe validar si el proceso se realiza. Nuevamente se llama a la función *content*.

```
> content(Corpus[[27]]) #imprime la posicion 27
```

```
[1] "Proyecto de Ciencias Natutales con TIC tema Los Reinos de los seres Vivos. Se trabajo con Realidad aumentada y Códi"
```

Se observa que efectivamente se realiza la eliminación al final del texto de la dirección web. En las siguientes líneas de código se realiza el mismo proceso de verificación. Por favor observe cuidadosamente el tweet inicial y el posterior a la ejecución de la función requerida.

Otra función interesante es la responsable de cambiar todas las palabras a minúsculas

```
> Corpus <- tm_map(Corpus, content_transformer(tolower))
```

```
> content(Corpus[[27]]) #imprime la posición 27
```

```
[1] "proyecto de ciencias natutales con tic tema los reinos de los seres vivos. se trabajo con realidad aumentada y códi"
```

De igual forma es ejecutada correctamente la función, cumpliendo con el objetivo de la misma.

Ahora se ejecuta una función para quitar la puntuación

```
> Corpus <- tm_map(Corpus, removePunctuation)

> content(Corpus[[27]]) #imprime la posición 27

[1] "proyecto de ciencias natutales con tic tema los reinos de los seres vivos se trabajo con
realidad aumentada y códi "
```

Otros datos de poca relevancia para este ejemplo son los números en el tweet, se pueden eliminar usando las siguientes líneas.

```
> Corpus <- tm_map(Corpus, removeNumbers) #quita los números

> content(Corpus[[27]]) #imprime la posición 27

[1] "proyecto de ciencias natutales con tic tema los reinos de los seres vivos se trabajo con
realidad aumentada y códi "
```

	Palabras											
frecuencia	abp	acompañ	adelant	adsum	aestahor	agency	agend	agradec	alamoconsulting	amp	ana.	
1	0	0	0	0	0	0	0	0		0	0	0
2	1	1	0	1	0	0	1	1		1	0	0
3	0	0	0	0	0	0	0	0		0	0	0
4	0	0	0	0	0	0	0	0		0	0	1

Palabras

**Figura 35.** Frecuencia de palaras por tweet

Fuente: (Rstudio, 2019)

En la tabla anterior (Figura 35) se puede observar cómo algunos términos más frecuentes en los tweets corresponden con artículos, preposiciones, pronombres en general, palabras que no aportan información relevante sobre el texto. Estas palabras se les conoce como stopwords. Para cada idioma existen distintos listados de stopwords, además, dependiendo del contexto, puede ser necesario adaptar el listado. En la tabla anterior aparece el término amp que procede de la etiqueta html &amp;. Con frecuencia, a medida que se realiza un análisis se encuentran palabras que deben incluirse en el listado de stopwords. (Amat, 2017)

La función encargada de realizar este proceso es:

```
> Corpus <- tm_map(Corpus, removeWords, stopwords("spanish")) #remueve las stopwords y
otras palabras que se deseen

> content(Corpus[[27]]) #imprime la posicion 27

[1] "proyecto ciencias natutales tic tema reinos seres vivos trabajo realidad aumentada
códi "
```

Después de realizar la verificación de los resultados se nota como existen algunas palabras irrelevantes o carentes de significado para la construcción de tendencias, es por ello que se fue necesario ajustar por medio de la inclusión de nuevas palabras a la función de stopwords. Por ejemplo: `tm_map(Corpus, removeWords, c(stopwords("spanish"), "amp", "leonesp", "barcelon", "leon", "hoy", "realiz" ))`

Del mismo modo se debe acortar las palabras a su raíz, esto con el fin de generar cuantificaciones erradas, por ejemplo, contar "devolvieron" y "devolver" como dos palabras distintas. La función es la siguiente:

```
> Corpus <- tm_map(Corpus, stemDocument, language="spanish")  
  
> content(Corpus[[27]]) #imprime la posicion 27  
  
[1] "proyect cienci natutal tic tem rein ser viv trabaj realid aument codi"
```

Finalmente, es necesario retirar los espacios vacíos excesivos, así:

```
> Corpus <- tm_map(Corpus, stripWhitespace)  
  
> content(Corpus[[27]]) #imprime la posicion 27  
  
[1] "proyect cienci natutal tic tem rein ser viv trabaj realid aument codi"
```

Tras concluir este proceso los datos se obtiene una base de datos más organizada y ajustada a las palabras que suministran información relevante para el análisis de sentimiento de los datos.

## **7.5. Análisis de resultados**

Una vez se han llevado a cabo los procedimientos de limpieza de datos y estos son consistentes, resulta necesario realizar un análisis de los mismos por medio de distintas pruebas estadísticas y gráficos que permitan describir las variables con las que se trabaja y determinar las posibles relaciones entre ellas. La calidad del conjunto de datos finalmente obtenido determinará la robustez y fiabilidad de los resultados del proyecto. (Arcón, 2017)

### **7.5.1. Análisis de sentimientos**

Para realizar análisis de sentimientos antes se deber haber realizado una clasificación manual del sentimiento antes de entrenar el modelo debido a el algoritmo debe aprender las reglas para clasificar los tweets, debe comprender por medio de la enseñanza de una persona cuales tweets están asociados a sentimientos positivos o negativos, cuales combinaciones se deben relacionar como tweets positivos. Es por ello que una manera eficaz de enseñarle a la máquina es por medio de ejemplos (aprendizaje supervisado) y posteriormente por medio de Machine Learning pueda clasificar tweets nuevos. No existe una regla sobre cuantos tweets se deben clasificar manualmente, solamente decir que, entre más, mejor será el aprendizaje del algoritmo y mayor la eficiencia de predicción.

Antes se debe entender como ingresan los textos de los tweets a la base de datos del programa o algoritmo propuesto, para ello veamos el siguiente ejemplo.

**Tweet:** Proyecto de Ciencias Natutales con TIC tema Los Reinos de los seres Vivos Se trabajo con Realidad aumentada.

**Sentimiento:** Clasificación manual = 1

**Tabla 3.** *Ingreso de tweets a la base de datos*

Adaptado de: (Martínez Gordillo , 2016)

Sentimi ento	proye cto	S ql	Cien cias	Natut ales	ti c	te ma	ofe rta	rein os	ser es	viv os	trab ajo	reali dad	aumen tada
1	1	0	1	1	1	1	0	1	1	1	1	1	1

Se puede observar cómo ingresan los tweets a la base de datos (Tabla 3), puede notar que las palabras o caracteres eliminados en el proceso de limpieza no se cargan. La tabla se continuará llenando según se presente una palabra en los demás tweets, si no se encuentra la agrega en una nueva columna. Respeto a la casilla de sentimientos este corresponde a la clasificación de cada tweet dada en la clasificación manual. Esta forma es como el software empezará a aprender.

En R se puede realizar un zoom a la matriz de datos del proyecto con la función:

*inspect(frequencies[15:20, 5:10])* , en las posiciones [15:20, 5:10]

Se observa siguiente:

<<*DocumentTermMatrix (documents: 6, terms: 6)*>>

*Non-/sparse entries: 22/14*

*Sparsity* : 39%

*Maximal term length: 10*

*Weighting* : *term frequency (tf)*

*Sample* :

A continuación, se observa cómo se llenó la matriz

Terms

Docs dat elbierz emple empleoleon learning

15 1 1 1 1 1 1

16 1 1 1 1 1 1

17	1	1	1	1	1	1
18	1	0	1	0	1	1
19	0	0	0	0	0	0
20	0	0	0	0	0	0

La organización de la base de datos anterior es indispensable para poder continuar con el análisis estadístico, Lo anterior porque este tipo de minería de datos solamente puede realizar cálculos a partir de datos numéricos, no desde textos tipo string.

### 7.5.2. Clasificación

A partir de esta etapa se empieza a trabajar con la variable sentiment. Es así como este proceso inicia con la creación de una nueva matriz con las frecuencias, se cuantifica las palabras y su frecuencia de ocurrencia en la base de datos. Para tal fin se hace uso de la siguiente línea de código: `frequencies <- DocumentTermMatrix(Corpus)`

Donde: `frequencies` es el objeto utilizado para imprimir los atributos de las frecuencias de la matriz;

`Documents`= Corresponde a las Filas de los tweets,

`terms`= Son las columnas con las palabras

La frecuencia es de gran utilidad a nivel estadístico, muestra la cantidad de veces que se repite un dato; además para esta etapa de clasificación permite realizar la selección de las palabras de acuerdo a un porcentaje de frecuencia específico.

El reporte para la matriz creada (Figura 36) da información detallada, para este ejemplo muestra los datos como una matriz con 218 tweets (documentos), 531 palabras (terms) y la longitud máxima de la palabra más larga es de 25 caracteres.

```
<<DocumentTermMatrix (documents: 218, terms: 531)>>  
Non-/sparse entries: 2130/113628  
Sparsity           : 98%  
Maximal term length: 25  
Weighting          : term frequency (tf)
```

**Figura 36.** Reporte de frecuencias de la matriz

Fuente: (Rstudio, 2019)

Ahora bien, la densidad de un grupo es importante en el momento de caracterizar un grupo de palabras, porque refleja si la temática que evidencia el mismo, está desarrollada o no. La densidad es una medida de la fuerza de las asociaciones internas de un grupo o clúster, se define como el promedio de los coeficientes de asociación entre las palabras clave dentro del grupo. Si  $S$  es un grupo creado, entonces su densidad  $DS$  es:

$$DS = \frac{1}{m'} \sum_{i \in S} \sum_{\substack{j \in S \\ j > 1}} E_{ij}$$

donde  $m'$  es el número de coeficientes de asociación internos no nulos. De esta forma, si las palabras dentro de un grupo aparecen con alta frecuencia de forma simultánea en diferentes documentos, significa que el grupo está representando a una temática elaborada y tendría una densidad alta. Por otro lado, si las palabras dentro del grupo están presentes de forma simultánea sólo en algunos documentos, pero además se encuentran en otros documentos asociadas con otras palabras, se dice que el grupo representa a una temática poco elaborada y, por lo tanto, su densidad es baja. (Rodríguez, 2007).

Es importante observar los términos de mayor frecuencia en la muestra, para ello se utiliza la función `findFreqTerms`, la cual permite hacer una inspección de la matriz.

```
> findFreqTerms(frecuencias, lowfreq = 50) # Ver Palabras con una mayor frecuencia igual o mayor a 50 veces
```

```
[1] "busc" "elbierz" "emple" "leon" "senior" "tic" "trabaj"
```

Por medio de `lowfreq` se puede seleccionar cuales datos se muestran. En el caso anterior la frecuencia mínima de palabras es aquellas con repeticiones mayores a 50 veces. Esta se puede ajustar en cualquier momento.

A nivel estadístico es importante manejar datos con significancia estadística y representen o muestren patrones del comportamiento de una muestra. Por lo anterior se filtran aquellas palabras con baja frecuencia, para este caso solamente se van a conservar de acuerdo a un porcentaje mayor o igual al 99.5%. Con esto se garantiza que las palabras de muy poca frecuencia (sparse) determinen el comportamiento en el análisis de sentimientos.

La función que realiza este trabajo es *SparseTerms(frecuencias, 0.995)* y se detalla en la (Figura 37).

```
> sparse #imprime sparse -- sparce=palabras que aparecen poco
<<DocumentTermMatrix (documents: 218, terms: 236)>>
Non-/sparse entries: 1835/49613
Sparsity           : 96%
Maximal term length: 20
weighting          : term frequency (tf)
```

**Figura 37.** Reporte de frecuencias de la matriz eliminado palabras de baja frecuencia

Fuente: (Rstudio, 2019)

Ahora al comparar las matrices (Figura 36 y 37) se observa como los tweets (documents) son los mismos pero el número de palabras cambian. Se encuentran menor número después del proceso de selección, pero menor porcentaje de ocurrencia (Figura 38).

<pre>&lt;&lt;DocumentTermMatrix (documents: 218, terms: 531)&gt;&gt; Non-/sparse entries: 2130/113628 Sparsity           : 98% Maximal term length: 25 weighting          : term frequency (tf)</pre>	<pre>&gt; sparse #imprime sparse -- sparse=palabras que aparecen poco &lt;&lt;DocumentTermMatrix (documents: 218, terms: 236)&gt;&gt; Non-/sparse entries: 1835/49613 Sparsity           : 96% Maximal term length: 20 weighting          : term frequency (tf)</pre>
Matriz inicial	Matriz optimizada

**Figura 38.** Comparación de matrices antes y después de filtrar por frecuencias

Adaptado de: (Rstudio, 2019)

Con el proceso anterior se garantiza que estas palabras son las más usadas y dan una mayor aproximación al modelo. Para continuar con el análisis, es necesario convertir este matriz de palabras a un dataframe (Hoja de cálculo para R), la cual permite realizar los cálculos y correr el modelo de análisis de sentimientos. La función es la siguiente:

```
tweetsSparse <- as.data.frame(as.matrix(sparse))
```

	analyst	busc	b.	dat	elbierz	emple	empleoleon	learning	leon	machin	scienc
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	1	0	0	0	1	0	0
8	0	0	0	0	1	1	0	0	1	0	0
9	0	0	0	0	0	1	0	0	1	0	0
10	0	0	0	0	1	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	1	0	0	0	1	0	0
13	0	0	0	0	1	0	0	0	1	0	0
14	0	0	0	0	0	1	0	0	1	0	0

Showing 1 to 15 of 218 entries, 237 total columns

**Figura 39.** Base de datos en formato R

Adaptado de: (Rstudio, 2019)

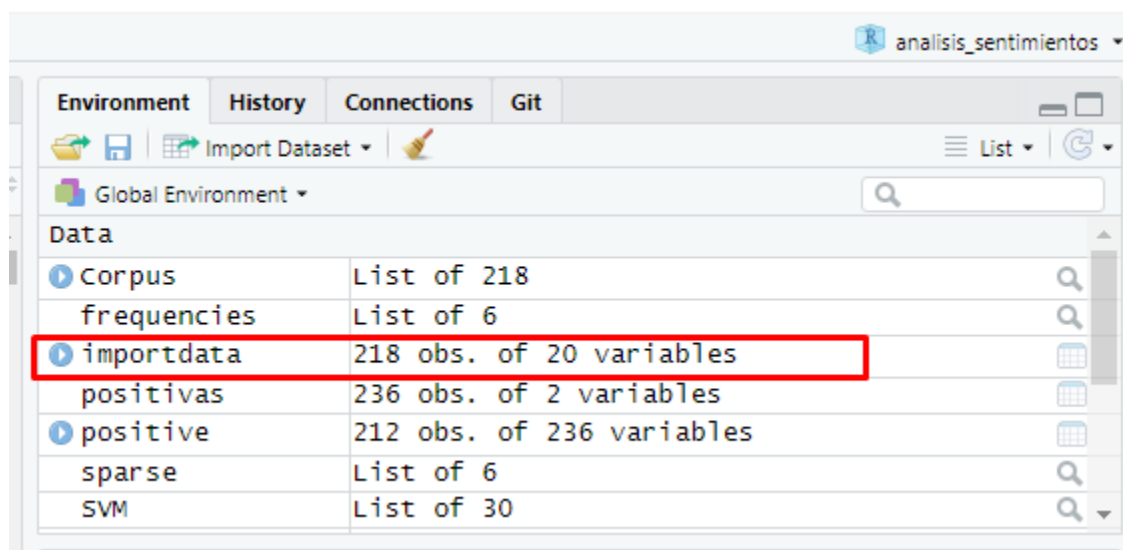
Es así como se retorna a la variable sparse una base de datos en formato R (Figura 39), además se debe asignarle los nombres de cada palabra al dataframe por medio de *colnames* como sigue:

```
> colnames(tweetsSparse) = make.names(colnames(tweetsSparse))
```

Como la base de datos creada ahora no contiene una columna de sentimientos (sentiment), se debe cargarle para poder realizar el análisis, para esto se puede aprovechar la primera base de datos creada al momento de importar los datos. Para este ejemplo es **impordata**, petición que se realiza así:

```
tweetsSparse$sentiment <- importdata$sentiment
```

La instrucción anterior examina la base de datos actual y le actualiza los sentimientos de la base de datos cargada inicialmente (Figura 40).



**Figura 40.** Base de datos inicial con tweets importados

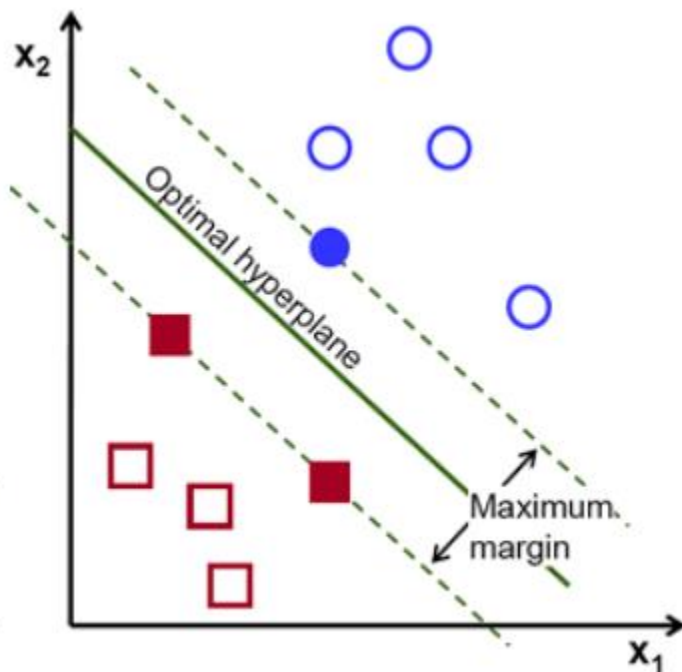
Adaptado de: (Rstudio, 2019)

A este paso se encuentra la base de datos ordenada y lista para correr el modelo de clasificación.

### 7.5.3. Modelo de clasificación

El modelo de clasificación utilizado es Support Vector Machine (SVM), ya que produce una precisión significativa con menos potencia de cálculo, además es de gran uso tanto para tareas de regresión como de clasificación. Es por ello una herramienta de gran utilidad para este objetivo de clasificación propuesto. (Gandhi, 2018)

El objetivo del algoritmo de máquina de vectores de soporte es encontrar un hiperplano en un espacio N-dimensional (N - el número de características) que clasifica claramente los puntos de datos. (Figura 41)

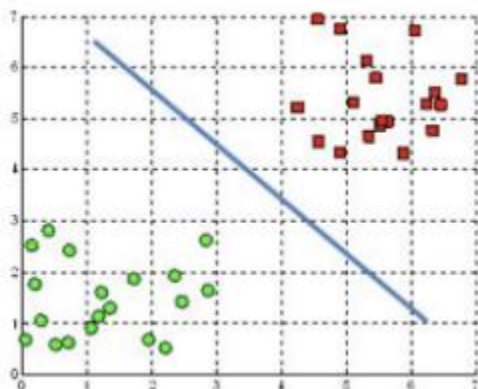


**Figura 41.** Visualización de la representación de un Hiperplano

Fuente: (Gandhi, 2018)

El fin de este ejercicio consiste en encontrar un plano que tenga el margen máximo (, es decir, la distancia máxima entre los puntos de datos de ambas clases (Figura 42). Maximizar la distancia de margen proporciona cierto refuerzo para que los puntos de datos futuros se puedan clasificar con más confianza. (Gandhi, 2018)

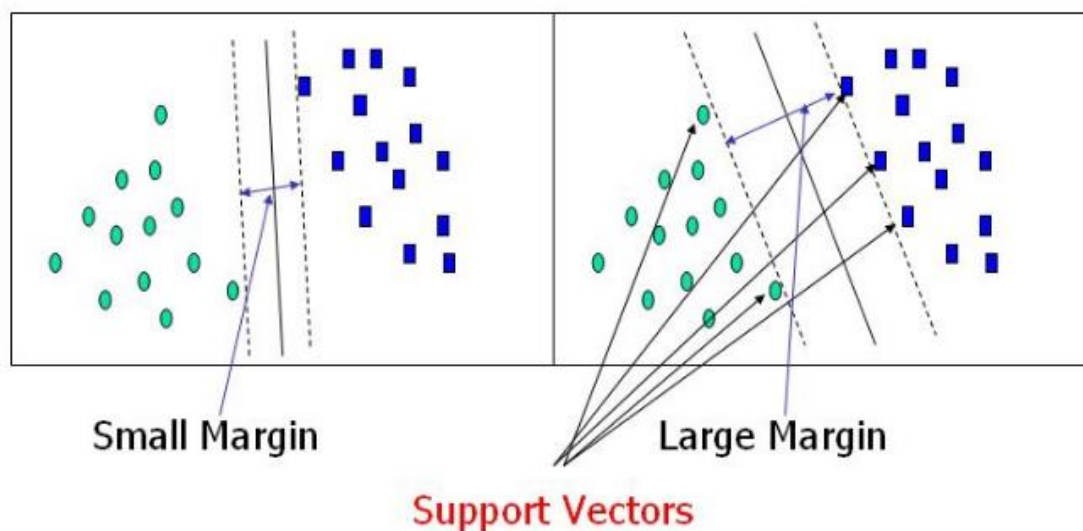
A hyperplane in  $\mathbb{R}^2$  is a line



**Figura 42.** Hiperplano en espacio de características 2D

Fuente: (Gandhi, 2018)

Los hiperplanos son límites de decisión que ayudan a clasificar los puntos de datos. Los puntos de datos que caen a ambos lados del hiperplano se pueden atribuir a diferentes clases.



**Figura 43.** Vectores de soporte

Fuente: (Gandhi, 2018)

Los vectores de soporte son puntos de datos que están más cerca del hiperplano e influyen en la posición y orientación del hiperplano (Figura 43). Usando estos vectores de soporte, maximizamos el margen del clasificador. Eliminar los vectores de soporte cambiará la posición del hiperplano. Estos son los puntos que ayudan a construir el SVM. (Gandhi, 2018)

Retornando al ejercicio de este proyecto y teniendo en cuenta el modelo anterior, se busca encontrar un hiperplano  $h$  de dimensión  $(n-1)$  que separe los tweets etiquetados con  $-1$  de los etiquetados con  $+1$  con un "margen máximo" ( $P$ ). Para ello se buscan los puntos más cercanos

entre varias clases, estos puntos se llaman "los vectores de soporte". A continuación, el SVM declara que la mejor línea de separación va a ser la línea que divide las dos clases y además es quién la maximiza la distancia del hiperplano a los vectores de soporte ( $r$ ). Para esto se utiliza la siguiente ecuación:

Ecuación del hiperplano:  $W^t X + b = 0$

Para el problema de optimización:

Encuentre  $\mathbf{w}$  y  $\mathbf{b}$ , tal que  $P = \frac{2}{[W]}$

### **Implementación de SVM en R**

Para la ejecución del modelo se debe crear un conjunto de datos para entrenamiento y otro para evaluación. El proceso de fragmentación requiere la librería *catools*. Para esto se va a dividir la base de datos de la siguiente manera:

- Entrenamiento: 80% de los datos. Estos datos son los encargados de calibrar y enseñarle al modelo SVM como debe realizar la clasificación.

- Evaluación: 20% de los datos. Permite evaluar cómo trabaja el modelo y cuál es su poder de predicción.

Además, es importante definir el punto de partida del Split, el punto de partida en R, para generar la fragmentación aleatoria de la base de datos en este caso será ajustado a 12.

Función: `set.seed(12)`

Ahora bien, por medio de esta función, se decide que observaciones se van para un lado u otro

```
split <- sample.split(tweetsSparse$sentiment, SplitRatio = 0.8)
```

El objeto Split va a contener cada uno de los datos, los mismos que contienen una asignación True o False de acuerdo al sentimiento de cada una de las observaciones de la muestra (Figura 44).

Object	Value
SVM	List of 30
TestSparse	43 obs. of 237 variables
trainSparse	175 obs. of 237 variables
tweetsSparse	218 obs. of 237 variables
Values	
predictSVM	Factor w/ 2 levels "-1","1": 2 2 2 2 2 2 2 2...
removeURL	function (x, ...)
split	logi [1:218] TRUE FALSE FALSE TRUE TRUE TRUE...

**Figura 44.** Creación y visualización del Split en el ambiente de trabajo de R

Adaptado de: (Rstudio, 2019)

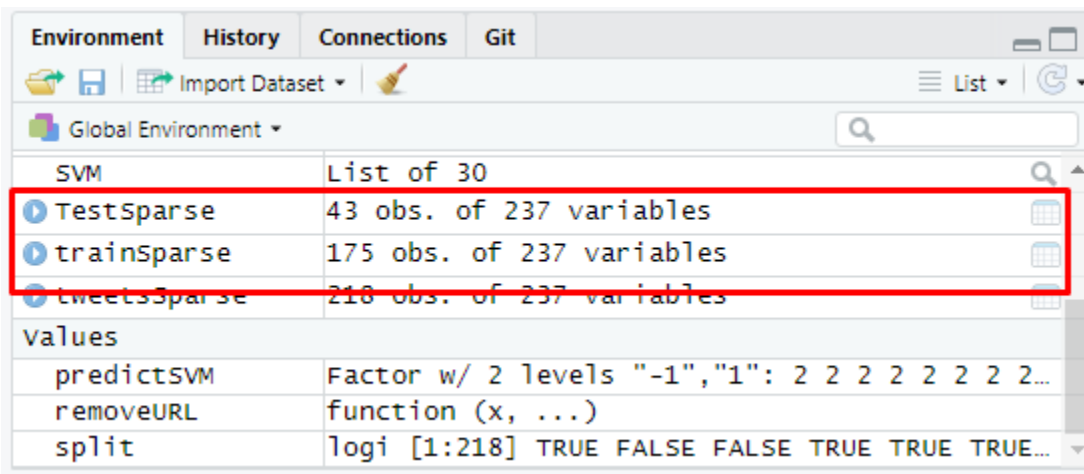
Como se puede observar los datos ahora están clasificados, en este momento es posible crear las dos bases de datos (entrenamiento y evaluación). Estas se generan de la siguiente manera.

- Creación de la base de datos de entrenamiento, función:

```
trainSparse = subset(tweetsSparse, split==TRUE)
```

- Creación de la base de datos de evaluación, función:

```
TestSparse = subset(tweetsSparse, split==FALSE)
```



**Figura 45.** Creación y visualización de las bases de datos entrenamiento y evaluación en el ambiente de trabajo de R

Adaptado de: (Rstudio, 2019)

La ejecución de estas líneas de código genera dos bases de datos (Figura 45), donde el conjunto de entrenamiento tiene como objetivo crear un modelo que clasifique los datos nuevos de forma correcta. Cuando un modelo de datos se clasifique de forma incorrecta el modelo de predicción no hará un trabajo adecuado algunas predicciones serán incorrectas. Es por ello de gran importancia el tamaño de la muestra y la clasificación manual correcta, Sin embargo, este modelo funciona de la misma manera tanto en los datos de evaluación como en los de entrenamiento. Es decir, este modelo no sobre ajusta los datos de entrenamiento.

A continuación, se muestran datos clasificados y representados en tablas (data set) de R.

The screenshot shows the RStudio interface with a window titled 'trainSparse' containing a data table. The table has 17 rows and 12 columns. The columns are: analyst, busc, b., dat, elbierz, emple, empleoleon, learning, leon, machin, and scienc. The data is binary, with values 0 or 1. The first row (index 1) has all 1s. The second row (index 4) has all 0s. The third row (index 5) has 0s for analyst, busc, b., and dat, and 1s for elbierz, emple, empleoleon, learning, leon, machin, and scienc. The fourth row (index 6) has 0s for analyst, busc, b., and dat, and 1s for leon, machin, and scienc. The fifth row (index 7) has 0s for analyst, busc, b., and dat, and 1s for elbierz, emple, empleoleon, learning, leon, machin, and scienc. The sixth row (index 8) has 0s for analyst, busc, b., and dat, and 1s for elbierz, emple, empleoleon, learning, leon, machin, and scienc. The seventh row (index 9) has 0s for analyst, busc, b., and dat, and 1s for leon, machin, and scienc. The eighth row (index 10) has 0s for analyst, busc, b., and dat, and 1s for elbierz, emple, empleoleon, learning, leon, machin, and scienc. The ninth row (index 11) has 0s for analyst, busc, b., and dat, and 1s for leon, machin, and scienc. The tenth row (index 13) has 0s for analyst, busc, b., and dat, and 1s for elbierz, emple, empleoleon, learning, leon, machin, and scienc. The eleventh row (index 14) has 0s for analyst, busc, b., and dat, and 1s for leon, machin, and scienc. The twelfth row (index 15) has 1s for all columns. The thirteenth row (index 16) has 1s for all columns. The fourteenth row (index 17) has 1s for all columns. The status bar at the bottom indicates 'Showing 1 to 15 of 175 entries, 237 total columns'.

	analyst	busc	b.	dat	elbierz	emple	empleoleon	learning	leon	machin	scienc
1	1	1	1	1	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	1	0	0	0	1	0	0
8	0	0	0	0	1	1	0	0	1	0	0
9	0	0	0	0	0	1	0	0	1	0	0
10	0	0	0	0	1	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	1	0	0
13	0	0	0	0	1	0	0	0	1	0	0
14	0	0	0	0	0	1	0	0	1	0	0
15	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1	1

**Figura 46.** Tabla de datos de entrenamiento

Adaptado de: (Rstudio, 2019)

La (Figura 46) representa los datos base para enseñarle al modelo de clasificación sobre el cómo interpretar una palabra desde el sentimiento que esta representa en un tweet.

	analyst	busc	b.	dat	elbierz	emple	empleoleon	learning	leon	machin	scienc
2	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	1	0	0	0	1	0	0
32	0	0	0	0	0	1	0	0	1	0	0
33	0	0	0	0	1	1	0	0	1	0	0
34	0	0	0	0	1	0	0	0	1	0	0
36	0	0	0	0	1	0	0	0	1	0	0
38	0	0	0	0	0	1	0	0	1	0	0
43	0	1	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	0	0	0
76	0	0	0	0	1	0	0	0	1	0	0

Showing 1 to 15 of 43 entries, 237 total columns

**Figura 47.** Tabla de datos de evaluación

Adaptado de: (Rstudio, 2019)

La (Figura 47) contiene la materia prima donde se realizan las predicciones de acuerdo al modelo SVM, determinando así el posible sentimiento de las palabras de un tweet.

A este punto se debe observar cómo está el sentimiento en la base de datos de evaluación (TestSparse), esto permite saber porque tipo de sentimiento realizar la predicción del modelo.

```
table(TestSparse$sentiment)
```

**Tabla 4.** *Distribución del sentimiento en base de datos de evaluación*

Fuente: (Rstudio, 2019)

-1	1
1	42

Para este proyecto, se observa predominancia del sentimiento positivo. Si calcula la precisión (sentimiento positivo/palabras TestSparse),  $42/43=0.98$ . Razón por la cual se procederá a trabajar únicamente con este sentimiento porque la hacerlo la precisión sería al menos de 98%. Por lo tanto, se descarta el sentimiento negativo.

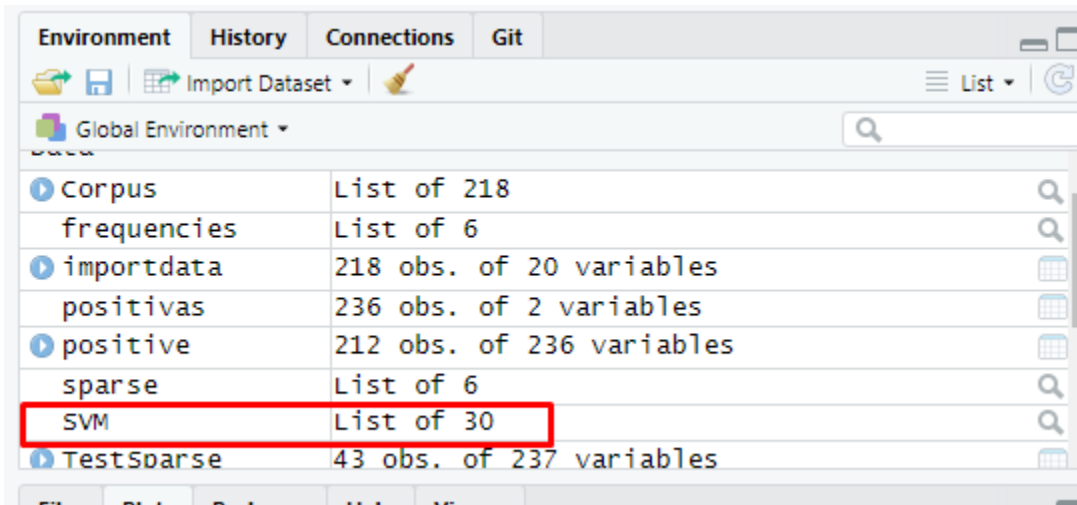
#### a) Entrenamiento

Para correr el algoritmo SVM de clasificación es necesario cargar en R estas librerías:

```
library(caret); library(e1071), library(plyr). (Martínez Gordillo , 2016)
```

```
SVM <- svm(as.factor(sentiment)~ ., data=trainSparse)
```

Con la función anterior se entrena al modelo de soporte vector machine para clasificar tweets nuevos. Como resultado se obtiene en pantalla una variable con los parámetros del modelo SVM (Figura 48).



**Figura 48.** Modelo de soporte vector machine

Adaptado de: (Rstudio, 2019)

Para ver su contenido se puede ejecutar la siguiente función:

```
summary(SVM)
```

Se observa un resumen con la descripción del modelo, como sigue:

*Call:*

```
svm(formula = as.factor(sentiment) ~ ., data = trainSparse)
```

Aquí se puede observar que es un el modelo va a predecir la variable *sentiment* utilizando todas las demás variables. Cuenta con los siguientes parámetros:

Parameters:

SVM-Type: C-classification

SVM-Kernel: radial

cost: 1

Number of Support Vectors: 29

(24 5)

Number of Classes: 2

Levels:

-1 1

Según esto, contiene un Kernel de tipo radial, posee 29 vectores de soporte y tiene dos clases a clasificar.

## b) Predicción

El modelo esta entrenado, ahora se debe evaluar su comportamiento realizando predicciones con los datos de evaluación. Para lo cual se genera las siguientes líneas de código:

```
predictSVM <- predict(SVM, newdata = TestSparse)
```

Esta función permite realizar la predicción de la nueva base de datos (TestSparse) a partir del modelo entrenado (SVM), la predicción se llevó a cabo, prueba de esto se crea un objeto con los tweets clasificados (Figura 49).

Environment	History	Connections	Git
Global Environment			
SVM	List of 30		
TestSparse	43 obs. of 237 variables		
trainSparse	175 obs. of 237 variables		
tweetsSparse	218 obs. of 237 variables		
values			
predictSVM	Factor w/ 2 levels "-1","1": 2 2 2 2 2 2 2 2...		
removeURL	function (x, ...)		
split	logi [1:218] TRUE FALSE FALSE TRUE TRUE TRUE...		

**Figura 49. Objeto con tweets y su predicción de sentimiento**

Adaptado de: (Rstudio, 2019)

Para evaluar la calidad de predicción y ver el desempeño de los tweets se hace uso de la función:

```
confusionMatrix(predictSVM,as.factor(TestSparse$sentiment))
```

Se obtienen los siguientes resultados:

```

Confusion Matrix and Statistics

      Reference
Prediction -1  1
-1         0  0
 1         1 42
Tweets clasificados correctamente

      Accuracy : 0.9767
      95% CI   : (0.8771, 0.9994)
No Information Rate : 0.9767
P-Value [Acc > NIR] : 0.7358

```

---

**Figura 50.** Reporte del modelo de predicción de tweets.

Adaptado de: (Rstudio, 2019)

Teniendo en cuenta los datos de la (Figura 50) se puede analizar lo siguiente:

**Primero:** Los tweets clasificados correctamente por el modelo SVM son los que se encuentran subrayados sobre la diagonal.

**Segundo:** El nivel de precisión es de 98% utilizando este modelo. Aquí si se tiene en cuenta el análisis de sentimientos inicial (Tabla 4) se observa que se mantiene, de allí que si se desea mejorar aún más el nivel de precisión se debe aumentar el tamaño de la muestra, a mayor enseñanza por medio de tweets clasificados al algoritmo de clasificación SVM, mayor será el aprendizaje.

### 7.5.4. Reportes y Graficas

Los resultados obtenidos se presentan de la siguiente manera:

- I. **Nube de palabras:** Ahora se va a crear una nube con las palabras más repetidas con la base de datos de tweets positivos. Para ello es necesario tener las palabras agrupadas por frecuencias.

Función:

```
positive <- subset(tweetsSparse, tweetsSparse$sentiment==1)
```

The screenshot shows the RStudio interface. On the left, a data table displays the frequency of words in positive tweets. The columns are labeled with words: analyst, busc, b., dat, elbierz, emple, empleoleon, learning, leon, machin, scienc. The rows are numbered 1 to 10. On the right, the Environment pane shows the workspace with several objects: Corpus (List of 218), frecuencias (List of 6), importdata (218 obs. of 20 variables), positivas (236 obs. of 2 variables), positive (212 obs. of 236 variables), sparse (List of 6), and SVM (List of 30). The 'positive' object is highlighted with a red box.

	analyst	busc	b.	dat	elbierz	emple	empleoleon	learning	leon	machin	scienc
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	1	0	0	0	1	0	0
8	0	0	0	0	1	1	0	0	1	0	0
9	0	0	0	0	0	1	0	0	1	0	0
10	0	0	0	0	1	0	0	0	1	0	0

**Figura 51.** Tabla de frecuencias con sentimiento positivo

Adaptado de: (Rstudio, 2019)

Se muestra una representación de todos los tweets con sentimiento positivo (Figura 51). Como la idea es realizar una tabla con frecuencias de las palabras y todas son de sentimiento positivo sobra la columna sentiment. Es por ello que ahora se debe proceder a eliminarla esta variable.

```
positive$sentiment <- NULL
```

Para visualizar estos datos se hace uso de la función `table(positivas)`

Palabras	
frecuencia	abp acompañ adelant adsum aestahor agency agend agradec alamoconsulting ana. anal
1	0 0 0 0 0 0 0 0 0 0 0 0 0
2	1 1 0 1 0 0 1 1 0 0 0 0
3	0 0 0 0 0 0 0 0 0 0 0 0 0
4	0 0 0 0 0 0 0 0 0 0 1 0 0

Palabras	
frecuencia	analyst anunc aplic aprend aquileomedinacr are arelanc articul aul aument ayud b.
1	0 0 0 0 0 0 0 0 0 0 0 0 1 0
2	0 0 0 1 1 0 0 0 0 1 1 0 0 0
3	0 0 1 0 0 0 1 0 0 0 0 0 0 0
4	0 0 0 0 0 0 0 1 0 0 0 0 0 0

**Figura 52.** Tabla de palabras y frecuencia

Fuente: (Rstudio, 2019)

A continuación, se generan las sumatorias de las frecuencias de cada una de las palabras contenidas en el objeto **positivas** (Figura 52), para lo cual se suma el total de para cada una de las palabras. Esto se logra con la función:

```
positivas <- as.data.frame(colSums(positive))
```

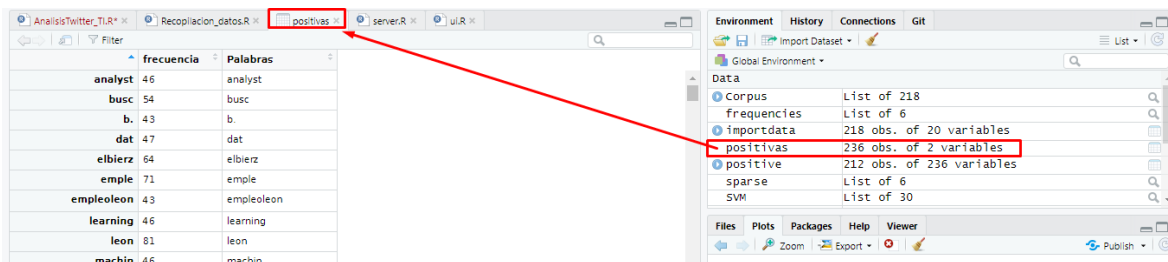
Luego se organizan las palabras como una nueva columna y no como un índice, para ello se crea una variable que se llama words, así:

```
positivas$words <- row.names(positivas)
```

Finalmente se nombran las dos columnas, como sigue:

```
colnames(positivas) <- c("frecuencia", "Palabras")
```

Se obtiene entonces una organización en tabla como en la (Figura 53), donde se observan las palabras agrupadas por frecuencias.



**Figura 53.** Palabras agrupadas por frecuencias

Adaptado de: (Rstudio, 2019)

En esta tabla (Figura 53) se presenta un resumen del número de palabras clave según su frecuencia. En esta tabla, la palabra que más se repite tiene frecuencia 81 y corresponde a la palabra clave empleo. Comparando estos datos con respecto a la tabla inicial del corpus se evidencia una gran dispersión de palabras clave, muchas palabras distintas y pocas que se repiten, esto fue básicamente lo también que debió filtrarse para evitar generar una mala predicción o análisis errados.

Como requisito para generar una nube de palabras se debe cargar las siguientes librerías:

```
library(wordcloud)
```

```
library("RColorBrewer")
```

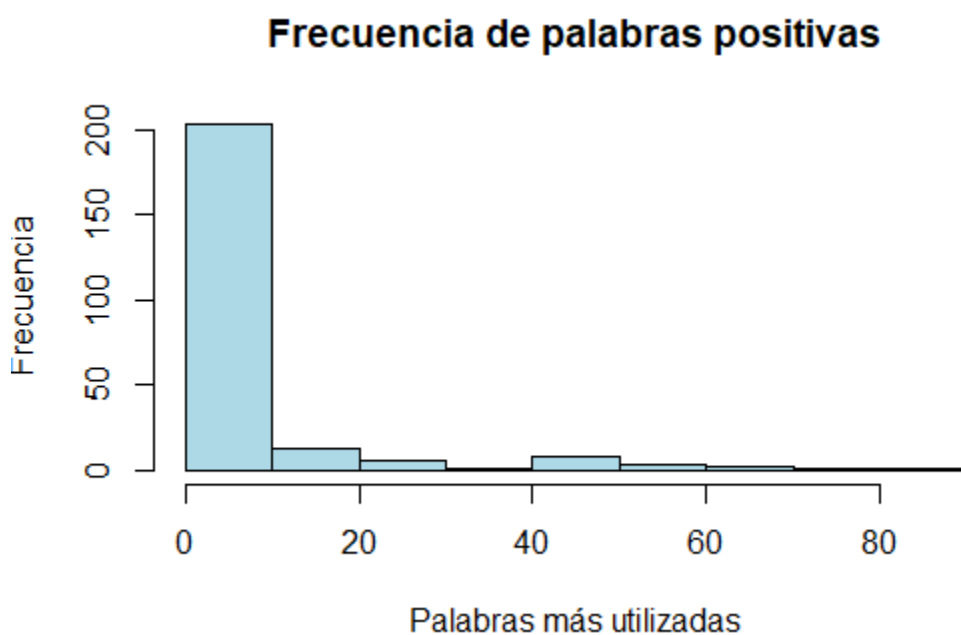
Y utilizar la función:

```
wordcloud(positivas$Palabras, positivas$frecuencia, random.order = FALSE, colors =  
brewer.pal(8, "Dark2"), max.words = 150)
```



Para crear histograma de frecuencia de palabras se utiliza la función:

```
hist(positivas$frecuencia, main = "Frecuencia de palabras positivas", xlab = "Palabras más utilizadas", ylab = "Frecuencia", col = "light blue", breaks = "Sturges")
```



**Figura 55.** Histograma de frecuencias tweets positivos

Fuente: (Rstudio, 2019)

Este histograma muestra un conjunto de datos, representados por medio de barras, las cuales representan las frecuencias con que aparecen las mediciones agrupadas en ciertos rangos o intervalos. La (figura 55) muestra la agrupación de los datos en clases, palabras y se cuenta

cuántas observaciones (frecuencia absoluta) hay en cada una de ellas, las mayores frecuencias corresponden a palabras como empleo (71), TIC (67), desarrollo de software senior y junior (52), ciencia de datos (46), machine learning (46), análisis de datos ( 46), base de datos SQL (15), redes (10 ) y conexión a internet (5) (Figura 56).

<b>frecuencia</b>	<b>Palabras</b>
46.00	analyst
54.00	busc
43.00	b.
47.00	dat
64.00	elbierz
71.00	emple
43.00	empleoleon
46.00	learning
81.00	leon
46.00	machin
46.00	scienc
52.00	senior
46.00	telic
3.00	cre
12.00	dec
7.00	arelanc
17.00	leonenredcom
3.00	plsql
28.00	program

**Figura 56.** Tabla de frecuencias análisis de sentimientos

Fuente: (RStudio, 2019)

A partir de las dos gráficas anteriores se puede observar, entre otras cosas, como las palabras “empleo” y “TIC” son las palabras más usadas, representando más del 69% del total de usos de las palabras en el Corpus. Lo anterior permite conocer las tendencias frente a temáticas, en este caso tendencias laborales en el sector TI con 29 vectores de soporte y dos clases a clasificar con un porcentaje de precisión del 98%.

**7.6. Diseño de guías****7.6.1. Guía de usuario**

Ver **anexo A**

**7.6.2. Manual técnico**

Ver **anexo B**

## 8. Conclusiones

Mediante el uso y aplicación de técnicas de análisis de información en la red social Twitter, se puede visualizar comportamiento y/o sentimientos en cada uno de los tweets de los usuarios, los mismos que permiten conocer tendencias y necesidades de la comunidad social, estos datos se pueden filtrar por zonas y así conocer por ejemplo necesidades del mercado, los cuales pueden ser de gran ayuda para la toma de decisiones en campos de formación como el sector de T.I. Para el ejemplo práctico en la etapa práctica de recopilación de tweets, se utiliza como palabra clave “trabajo para ingenieros”, utilizando la librería TwitterR y la API de Twitter donde se obtienen 800 tweets para el sector TI. Mientras que la fase de análisis de sentimientos, se obtienen 29 vectores de soporte y dos clases a clasificar con un porcentaje de precisión del 98%, desatándose tendencias laborales de mayor a menor frecuencias palabras como empleo (71), TIC (67), desarrollo de software senior y junior (52), ciencia de datos (46), machine learning (46), análisis de datos (46), base de datos SQL (15), redes (10) y conexión a internet (5), datos contenidos en 29 vectores de soporte y dos clases a clasificar con un porcentaje de precisión del 98%.

La identificación de la herramienta para la recopilación de información de la red social es clave para análisis exitoso de estudio en el entorno laboral, toda vez que ahorra recursos y permite obtener de primera mano datos estadísticos. El uso de R como interfaz de recopilación de datos brinda la posibilidad de hacer cálculos estadísticos y la generación de gráficos, ofreciendo una gran variedad de técnicas estadísticas y gráficas. Este es un entorno de análisis y programación completo con el que se añaden nuevas técnicas mediante la definición de funciones. Además, se

acopla muy bien con Twitter, debido que es una red social versátil y dispone de una API para acceso de terceros.

Existen una serie de pasos necesarios para la extracción y visualización de información de la red social Twitter entre los que se encuentran la Recolección de datos, fundamental para la obtención de la muestra; el Pre- procesamiento de datos donde se realiza “limpieza” de datos y transformación al formato deseado; el Análisis de sentimiento donde se entrena a un algoritmo para clasificar tweets y finalmente la etapa de predicción del modelo hace la clasificación de acuerdo al entrenamiento recibido. Estos pasos son fundamentales y el buen desarrollo de cada uno permite una mayor o menor precisión en el análisis de sentimientos de los datos consignados en la red social.

Es de gran utilidad el llevar a cabo una buena documentación de las técnicas para el análisis de la red social Twitter, debido a que el análisis ocurre en varias etapas y el usuario puede omitir algunos pasos, las cuales pueden repercutir en el resultado final.

## 9. Recomendaciones

Continuar con los estudios de análisis de sentimientos en redes sociales principalmente para Twitter, es un escenario donde se obtiene mucha información valiosa para el análisis de tendencias y necesidades. Además, a nivel cuantitativo se puede realizar una medición del número de posibles relaciones existentes entre los involucrados, como también la dirección y profundidad de la red existente.

Fomentar el uso de R para el análisis de redes sociales, es un lenguaje orientado a objetos para el cálculo estadístico y la generación de gráficos open source, que ofrece una gran variedad de técnicas estadísticas y gráficas. Es un lenguaje de programación completo con el que se añaden nuevas técnicas mediante la definición de funciones y la incorporación de cantidad de librerías.

Fortalecer los espacios de formación, desarrollo e intercambio de conocimiento a nivel intra e inter institucional, en áreas como Big data, analítica de datos, ciencia de datos, Machine Learning, inteligencia artificial, lenguajes de código abierto, programación estadística y demás tecnologías de inteligencia computacional, los mismos que sirvan para generar nuevos proyectos que faciliten la transferencia de conocimiento a nivel académico, social y empresarial.

Potenciar los semilleros de investigación ya que son un escenario para la enseñanza activa y constructiva, son espacios que permiten a estudiantes y docentes una participación real, controlada, guiada y procesual del binomio enseñanza-aprendizaje. Es además un espacio donde se prioriza la libertad, la creatividad y la innovación para el desarrollo de nuevos proyectos y métodos de aprendizaje.

## 10. Referencias

- Ahmed, W. (15 de 10 de 2015). *Análisis de redes sociales: Herramientas de recolección de datos de Twitter*. Obtenido de <http://ars-uns.blogspot.com/2015/10/el-uso-de-twitter-como-fuente-de-datos.html>
- Alcantara, H. X. (2015). *Análisis Egocéntrico De Redes Sociales: El Caso de los Representantes Sectoriales En El Consejo Territorial de Planeación Del Municipio De Arauca (2012 - 2015)*. Obtenido de <http://bdigital.unal.edu.co/10733/>
- Amat, J. (12 de 2017). *Text mining con R: ejemplo práctico Twitter*. Obtenido de RPubS: [https://rpubs.com/Joaquin\\_AR/334526](https://rpubs.com/Joaquin_AR/334526)
- Arcón, A. (23 de 06 de 2017). *5 librerías de R para la limpieza y exploración de datos*. Obtenido de <https://conocemachinelearning.wordpress.com/2017/06/23/5-librerias-de-r-para-la-limpieza-y-exploracion-de-datos/>
- Calvo, M. (2017). *Text Analytics para Procesado Semántico*. Obtenido de [http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_1475.pdf](http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1475.pdf)
- Cárdenas, J. (2016). *El análisis de redes: qué es, crecimiento y futuro. Pensando Sociología 12 (19)*. Obtenido de <http://networksprovidehappiness.com/analisis-de-redes-sociales-es/>
- Chiquiza, J. (5 de 05 de 2018). *Social Media en Colombia tiene un alcance de 83,8%, según ComScore*. Obtenido de El impacto promedio de estos medios en la región es de 84,5%: <https://www.larepublica.co/internet-economy/social-media-en-colombia-tiene-un-alcance-de-838-segun-comscore-2722542>
- CNA. (2013). *LINEAMIENTOS PARA LA ACREDITACIÓN DE PROGRAMAS DE PREGRADO*. Obtenido de [https://www.cna.gov.co/1741/articles-186359\\_pregrado\\_2013.pdf](https://www.cna.gov.co/1741/articles-186359_pregrado_2013.pdf)
- Criado, I., & Rojas, F. (2013). *Las Redes Sociales Digitales en la Gestión y las Políticas Públicas. Avances y Desafíos para un Gobierno Abierto*. Obtenido de [https://www.academia.edu/27641025/Las\\_Redets\\_Sociales\\_Digitales\\_en\\_la\\_Gesti%C3%B3n\\_y\\_las\\_Pol%C3%ADticas\\_P%C3%ABlicas.\\_Avances\\_y\\_Desaf%C3%ADos\\_para\\_un\\_Gobierno\\_Abierto](https://www.academia.edu/27641025/Las_Redets_Sociales_Digitales_en_la_Gesti%C3%B3n_y_las_Pol%C3%ADticas_P%C3%ABlicas._Avances_y_Desaf%C3%ADos_para_un_Gobierno_Abierto)
- Emprender Fácil. (2016). *Social Mention – Escucha lo que hablan de ti*. Obtenido de <https://www.emprender-facil.com/es/social-mention-escucha-hablan-de-ti/>

- Exportcomments. (2019). *Exportcomments.com*. Obtenido de <https://exportcomments.com/features>
- Fernández, J. M. (2016). *Análisis de contenidos en Social Media: Clasificación de mensajes e identificación de influyentes en el Banco Central Europeo*. Obtenido de [https://eprints.ucm.es/38544/1/TFM\\_javmateo\\_DEF\\_v3.pdf](https://eprints.ucm.es/38544/1/TFM_javmateo_DEF_v3.pdf)
- Gandhi, R. (7 de 06 de 2018). *Support Vector Machine — Introduction to Machine Learning Algorithms*. Obtenido de <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- García, L. M. (08 de 2014). *Análisis de sentimientos y predicción de eventos en Twitter*. Obtenido de [http://repositorio.uchile.cl/bitstream/handle/2250/130479/cf-montesinos\\_lg.pdf](http://repositorio.uchile.cl/bitstream/handle/2250/130479/cf-montesinos_lg.pdf)
- Hernández, R., Fernández, C., & Baptista, P. (2003). *Metodología de la investigación (3ª ed.)*. México: Editorial Mc Graw-Hill. Obtenido de <https://www.redalyc.org/pdf/1941/194118804003.pdf>
- Hootsuite. (2019). *Demuestra tu impacto en las redes sociales*. Obtenido de <https://hootsuite.com/es/plataforma/analytics#>
- HowSociable. (2019). *The World's Social Media Information Resource*. Obtenido de <https://howsociable.com/>
- IBM. (2019). *IBM Watson Explorer*. Obtenido de <https://www.ibm.com/co-es/products/watson-explorer>
- Jáuregui Fernández, C. (19 de 11 de 2018). *Análisis de sentimiento en R*. Obtenido de <https://www.linkedin.com/pulse/an%C3%A1lisis-de-sentimiento-en-r-carlos-j%C3%A1uregui-fern%C3%A1ndez/>
- KH Coder. (2019). *KH Coder*. Obtenido de <https://kncoder.net/en/>
- Lara, A. M., & Hidalgo, M. (2016). *Entorno Virtual de autoaprendizaje de la estadística. Departamento de Estadística e IO*. Obtenido de Universidad de Granada: <http://wpd.ugr.es/~bioestad/guia-r-studio/practica-1-r-studio/>
- Lugo, L. (23 de 08 de 2019). *Conferencia 2 Inteligencia social*. Obtenido de [conferencia2.unad.edu.co/investigacionzcbay](http://conferencia2.unad.edu.co/investigacionzcbay)
- Martínez Gordillo, J. D. (15 de 09 de 2016). *Primer Taller de Análisis de Sentimiento en Twitter con R*. Obtenido de <https://www.youtube.com/watch?v=nOIZnYLIPBo>

- Martínez, F. J. (14 de 09 de 2017). *Análisis de Sentimiento en Twitter de las*. Obtenido de [https://www.fundacionmapfre.org/documentacion/publico/i18n/catalogo\\_imagenes/image\\_n\\_id.cmd?idImagen=1108529](https://www.fundacionmapfre.org/documentacion/publico/i18n/catalogo_imagenes/image_n_id.cmd?idImagen=1108529)
- Meltwater. (2019). *Nosotros localizamos las tendencias antes de que estén en tendencia* . Obtenido de <https://www.meltwater.com/es/>
- Ministerio de Educación Nacional de Colombia. (2010). *Decreto No. 1295 de abril 20 de 2010*. Obtenido de <https://www.mineducacion.gov.co/1621/article-229430.html>
- Molina, J. L., Muñoz, J. M., & Domenech, M. (2002). *Redes de publicaciones científicas: un análisis de la estructura de coautorías*. Obtenido de <https://revistes.uab.cat/redes/article/view/v1-n1-molina-munoz-domenech>
- Montoya, J. B., & Carvalho, T. C. (2016). *Flujo documental entre áreas administrativas de una entidad bancaria: una aproximación desde el análisis de redes sociales (ARS)*. Obtenido de <https://www.redalyc.org/jatsRepo/161/16144742004/html/index.html>
- Navarro, L., & Salazar, J. (2007). *Análisis de redes sociales aplicado a redes de investigación en ciencia y tecnología*. Obtenido de <http://revistas.uach.cl/pdf/sintec/v3n2/art03.pdf>
- Ortiz Palma, L. A. (20 de 08 de 2019). Figuras de apoyo visual ARS Twitter. Popayán, Cauca, Colombia.
- Ortiz Palma, L. A. (21 de 08 de 2019). Figuras de apoyo visual ARS Twitter. Popayán, Cauca, Colombia.
- Ortiz Palma, L. A. (20 de 08 de 2019). Figuras de apoyo visual ARS Twitter. Popayán, Cauca, Colombia.
- Oviedo, M. (24 de 11 de 2015). *Ejecución de aplicaciones R*. Obtenido de <http://eio.usc.es/pub/moviedo/descargas/Shiny.html#12>
- Padró, L. (2019). *Analizadores Multilingues en FreeLing*. Obtenido de <http://nlp.lsi.upc.edu/publications/papers/padro11.pdf>
- Paradis, E. (2002). *R para Principiantes*. Obtenido de [https://cran.r-project.org/doc/contrib/rdebuts\\_es.pdf](https://cran.r-project.org/doc/contrib/rdebuts_es.pdf)
- Ponce, I. (2019). *Definición de redes sociales*. Obtenido de <http://recursostic.educacion.es/observatorio/web/eu/internet/web-20/1043-redes-sociales?start=1>
- QSR International. (2019). *Ncapture*. Obtenido de <http://help-ncapture.qsrinternational.com/>

- QSR International. (2019). *NVIVO Software para el análisis cuantitativo de datos*. Obtenido de <http://www.qsrinternational.com/nvivo-spanish>
- Rajput, S. (2019). *Twitter API Workflow*. Obtenido de Recuperado de <https://www.safedk.com/marketplace/sdks/oracle-twitter-api-me>
- Rodríguez Ruibal, A., & Santamaría Cristino, P. (2012). *Análisis del uso de las redes sociales en Internet: Facebook y Twitter en las Universidades españolas*. Obtenido de [https://www.researchgate.net/publication/277264445\\_Analisis\\_del\\_uso\\_de\\_las\\_redes\\_sociales\\_en\\_Internet\\_Facebook\\_y\\_Twitter\\_en\\_las\\_Universidades\\_espanolas](https://www.researchgate.net/publication/277264445_Analisis_del_uso_de_las_redes_sociales_en_Internet_Facebook_y_Twitter_en_las_Universidades_espanolas)
- Rodríguez, D. H. (12 de 2007). *Programación en R del método de las palabras asociadas*. Obtenido de [http://www.docentes.unal.edu.co/cepardot/docs/TrabajosGrado/TrabajoGrado\\_Daniel.pdf](http://www.docentes.unal.edu.co/cepardot/docs/TrabajosGrado/TrabajoGrado_Daniel.pdf)
- Romero, F. (2013). *Introducción al lenguaje de programación R. Dpto. de Ciencias de la Computación e Inteligencia Artificial*. . Obtenido de Universidad de Sevilla: [https://www.cs.us.es/~fran/curso\\_unia/introduccion\\_R.html](https://www.cs.us.es/~fran/curso_unia/introduccion_R.html)
- RStudio. (2019). *Características y funciones de RStudio IDE*. Obtenido de <https://www.rstudio.com/products/rstudio/features/>
- Rstudio. (22 de 08 de 2019). reportes R de IDE RStudio.
- Rstudio. (2019). *Shiny Server Professional v1.5.11 Administrator's Guide*. Obtenido de <https://docs.rstudio.com/shiny-server/#system-requirements>
- Santana, J. S. (27 de 11 de 2014). *El arte de programar en R: un lenguaje para la estadística*. Obtenido de [https://cran.r-project.org/doc/contrib/Santana\\_El\\_arte\\_de\\_programar\\_en\\_R.pdf](https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf)
- The R Foundation. (2019). *The R Project for Statistical Computing*. Obtenido de <https://www.r-project.org/>
- Tournery, M. (2015). *¿Por qué leer la guía de análisis y seguimiento de redes sociales?* Obtenido de [https://www.amic.media/media/files/file\\_352\\_815.pdf](https://www.amic.media/media/files/file_352_815.pdf)
- Tweetreach. (2018). *Twitter snapshot reports*. Obtenido de <https://tweetreach.com/twitter-analytics-report/>
- Twitter, I. (21 de 08 de 2019). *Twitter*. Obtenido de <https://developer.twitter.com/en/pricing>
- Twitter, Inc. (2019). *Pricing*. Obtenido de API access that scales with you and your solution.: <https://developer.twitter.com/en/pricing>

- UNAD. (2017). *PLAN DE GESTIÓN DEL SERVICIO IT*. Obtenido de <https://gidt.unad.edu.co/images/Documentos/Plan-de-Gestion-del-Servicio.pdf>
- UNAD. (2019). *Gerencia de Innovación y Desarrollo Tecnológico - GIDT*. Obtenido de <https://gidt.unad.edu.co/>
- UNAD. (2019). *Intranet*. Obtenido de <https://www.unad.edu.co/>
- UNAD. (2019). *SIG*. Obtenido de <https://sig.unad.edu.co/>
- UP2. (2018). *Collecting Facebook Data with Netvizz*. Obtenido de <http://www.up2.fr/index.php?n=Main.Netvizz>
- Uphill GmbH. (2019). *La solución completa para una gestión excepcional de las redes sociales*. Obtenido de <https://www.fanpagekarma.com/>
- webempresa. (01 de 03 de 2018). *¿Qué es twitter? ¿Cómo funciona? ¿Cómo puedo usarlo para mi organización?* Obtenido de <https://www.webempresa.com/blog/que-es-twitter-como-funciona.html>

## 11. ANEXOS

### **Anexo A – Guía de usuario**

En el siguiente link encuentra la guía de usuario:

[https://1drv.ms/w/s!A19G7kDqr\\_ECj3MB6Xcre3j7qApr?e=7y8EAh](https://1drv.ms/w/s!A19G7kDqr_ECj3MB6Xcre3j7qApr?e=7y8EAh)

### **Anexo B – Manual técnico**

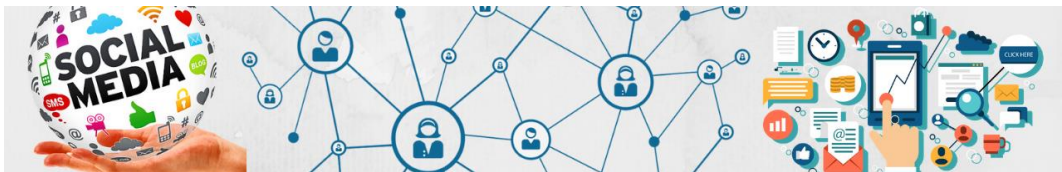
En el siguiente link encuentra el manual técnico:

[https://1drv.ms/w/s!A19G7kDqr\\_ECj3IwpONLKVL3D8Z8?e=uk6bC1](https://1drv.ms/w/s!A19G7kDqr_ECj3IwpONLKVL3D8Z8?e=uk6bC1)

### **Anexo C – Herramienta aplicación web ARS SIAVA**

En el siguiente link encuentra la aplicación web ARS SIAVA:


[https://1drv.ms/u/s!A19G7kDqr\\_ECj3ZNL18vt8OqHbWX?e=Yk6SMI](https://1drv.ms/u/s!A19G7kDqr_ECj3ZNL18vt8OqHbWX?e=Yk6SMI)



## **Software de Análisis de tendencias en red social Twitter**

### **Guía de Usuario**

Software de uso académico para la Universidad Nacional Abierta y a Distancia. Proyecto aplicado  
prerrequisito trabajo de grado Ingeniería de Sistemas

	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

### HOJA DE CONTROL


<b>Organismo</b>	Universidad Nacional Abierta y a Distancia – UNAD		
<b>Proyecto</b>	Aplicación de técnicas, herramientas de abstracción y visualización de datos, en TWITTER para el análisis de tendencias y necesidades de formación académica.		
<b>Entregable</b>	Guía de Usuario		
<b>Autor</b>	Luis Alberto Ortiz Palma		
<b>Versión/Edición</b>	01	<b>Fecha Versión</b>	30/08/2019
<b>Aprobado por</b>	Hermes Mosquera Angulo	<b>Fecha Aprobación</b>	02/10/2019
		<b>N° Total de Páginas</b>	47

### REGISTRO DE CAMBIOS

<b>Versión</b>	<b>Causa del Cambio</b>	<b>Responsable del Cambio</b>	<b>Fecha del Cambio</b>
01	Versión inicial	Luis Alberto Ortiz Palma	30/08/2019


### CONTROL DE DISTRIBUCIÓN

<b>Nombre y Apellidos</b>
Luis Alberto Ortiz Palma

	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

## INDICE

1	DESCRIPCIÓN DEL SISTEMA.....	4
1.1	Objeto .....	4
1.2	Alcance .....	4
1.3	Funcionalidad .....	4
2	MAPA DEL SISTEMA .....	6
2.1	Modelo Lógico .....	6
2.2	Navegación.....	6
3	DESCRIPCIÓN DEL SISTEMA.....	8
3.1	Módulo recopilación.....	8
3.1.1	Ventana principal .....	9
3.1.2	Ayudas contextuales .....	14
3.2	Módulo de Análisis.....	15
3.2.1	Ventana principal .....	15
3.2.2	Mensaje de notificación.....	19
3.2.3	Resultados.....	20
3.2.4	Ayudas contextuales .....	24
4	PREGUNTAS FRECUENTES .....	25
4.1	¿Como obtener y solicitar datos de acceso a la API de Twitter?.....	25
4.2	¿Como crear columna sentiment? .....	36
4.3	¿Como guardar mi documento Excel en formato CSV? .....	39
5	ANEXOS .....	41
6	GLOSARIO .....	44
7	BIBLIOGRAFÍA Y REFERENCIAS .....	46

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## **1 DESCRIPCIÓN DEL SISTEMA**

### ***1.1 Objeto***

Aplicar técnicas de análisis de sentimientos en redes sociales que permitan conocer tendencias, las cuales servirán de apoyo para atender y proyectar necesidades de formación académica.

### ***1.2 Alcance***

ARS SIAVA, es una aplicación web interactiva para análisis, recopilación y visualización estadística de redes sociales, el cual permite conocer tendencias, a partir del análisis de sentimientos de la información recopilada en redes sociales como Twitter. Este es un proyecto desarrollado dentro del semillero de investigación SIAVA CEAD Popayán, promovido por la Universidad Nacional Abierta y a Distancia - UNAD.

### ***1.3 Funcionalidad***


ARS SIAVA, permite realizar dos tareas fundamentales para análisis de redes sociales. El primer módulo permite obtener datos de los tweets publicados en la red social Twitter. El segundo módulo realiza análisis de sentimientos de los datos cargados y presenta reportes estadísticos.

#### **1.3.1 Módulo de recopilación**

Este módulo recopila datos desde Twitter por medio la API, para ello es necesario contar con claves y tokens de acceso las cuales puede solicitar en la página oficial de Twitter, la conexión API es un requisito obligatorio para uso de este módulo, así, como se debe tener presente que la extracción de información está supeditada por las condiciones, restricciones y/o políticas implementadas y en vigencia de la red social. El usuario debe realizar el ingreso de datos, por medio de palabras claves cortas y mediante un botón realiza la búsqueda de tweets relacionados. El proceso de búsqueda incluye validaciones interactivas, las cuales notifican al usuario por medio de mensajes en pantalla sobre errores y/o estados de la consulta. Como resultado del proceso se genera un documento formato CSV (del inglés comma Separated values), con la información recopilada.


#### **1.3.2 Módulo de Análisis**

Este módulo realiza análisis de sentimientos a una data. Estos datos deben encontrarse previamente organizados, clasificados con una columna de sentimientos y cargados en formato .CSV (desde Microsoft Excel), Como resultado de este análisis se presentan reportes

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

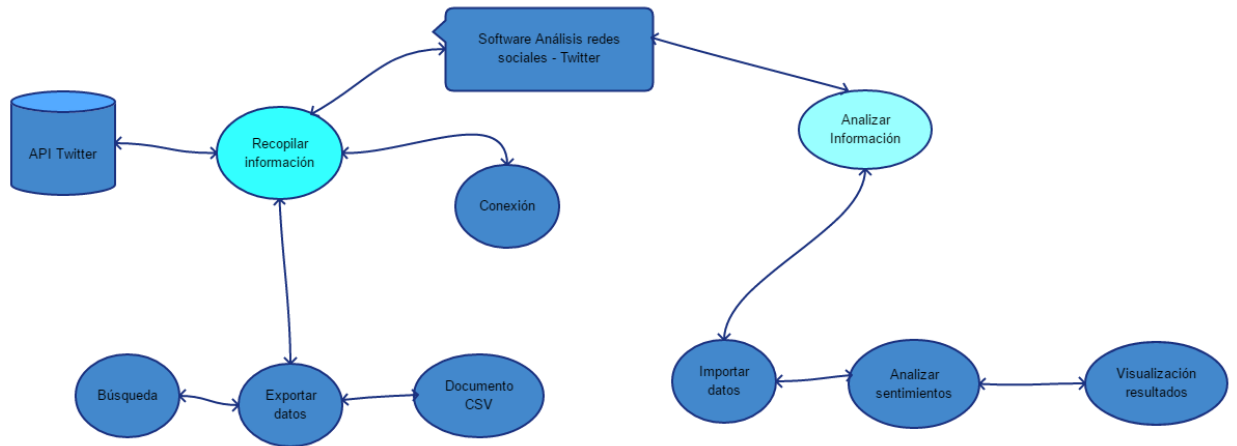
estadísticos entre ellos se encuentran frecuencias, media, moda, mediana y gráficas con histogramas y nubes de palabras. El usuario puede visualizar en forma selectiva cada uno de los resultados.

Para ejecutar la generación de reportes en forma exitosa es necesario que el proceso de validación de datos haya terminado sin errores.

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## 2 MAPA DEL SISTEMA

### 2.1 Modelo Lógico



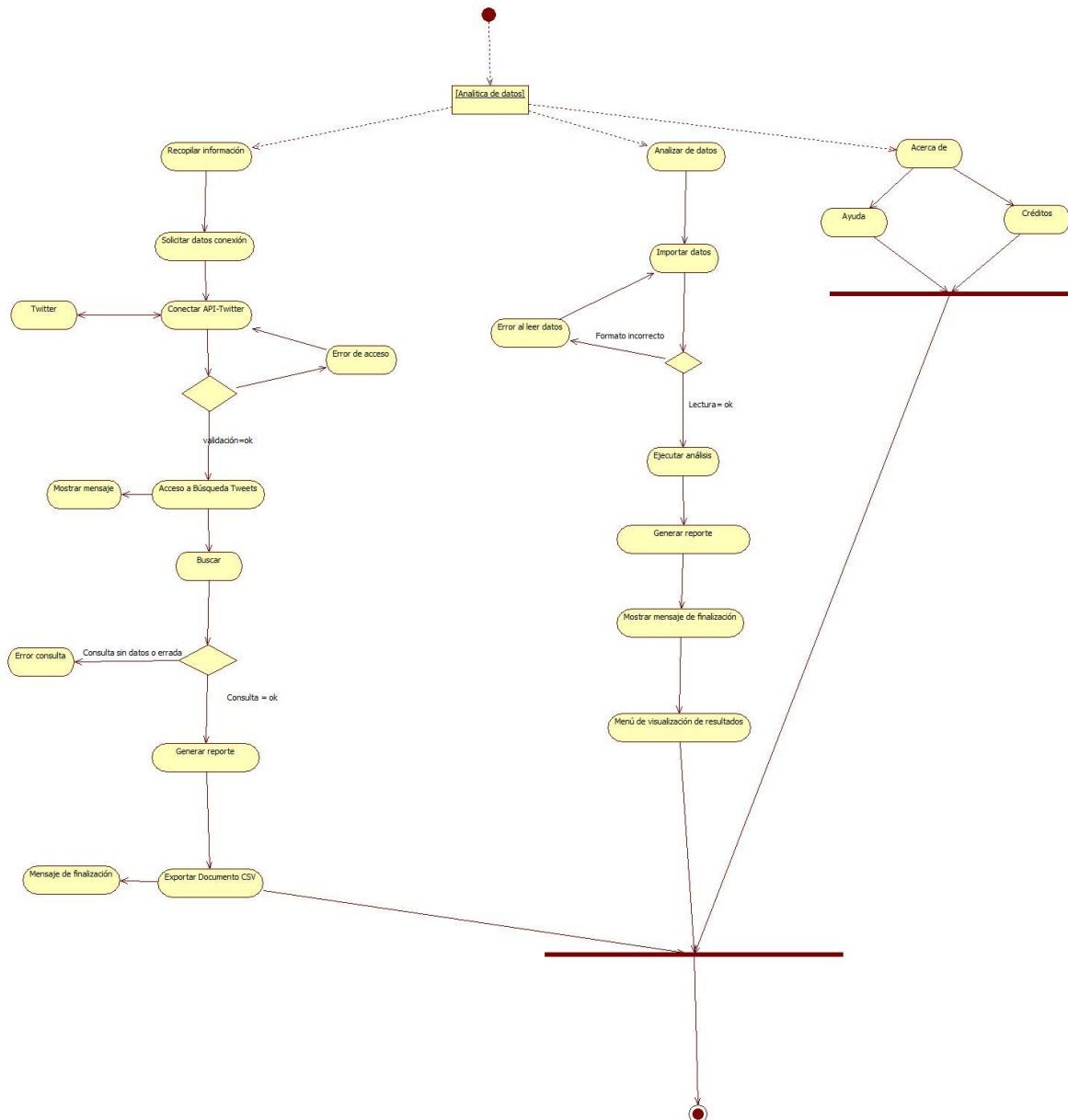
**Fig. 1 Diagrama de formato libre Tipo Top – Down**

Fuente: (Ortiz Palma, 2019)


El diagrama Tipo Top- Down (Fig. 1) muestra el levantamiento de la herramienta a desarrollar, se visualiza la interacción a través de las diferentes opciones de la plataforma (Se representan los caminos más significativos). En este diagrama se representarán las interacciones del sistema en círculos y mediante flechas las navegaciones entre las mismas.

### 2.2 Diagrama de actividades

En este diagrama (Fig.2) se observa el desarrollo de los diferentes procesos que ocurren en las diferentes etapas de la aplicación web ARS SIAVA.



**Fig. 2 Diagrama de actividades para la aplicación ARS SIAVA**  
Fuente: (Ortiz Palma, 2019)

	<p align="center"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p align="center"><b>UNAD Semillero de investigación S</b></p>
---	---	--

### 3 DESCRIPCIÓN DEL SISTEMA

En esta sección se describen la interfaz con las principales características de la aplicación. Se muestran las diferentes pantallas y acciones siguiendo los caminos lógicos que el usuario realizaría. Se explican las distintas pantallas y las acciones dependientes. Para ello se recogen los distintos subsistemas, y se expone la funcionalidad de forma estructurada. Asimismo, se explican los mensajes de error que pueden aparecer y las ayudas contextuales.

#### 3.1 Pantalla de inicio

El entorno inicial contiene una entrada por defecto (index), la cual muestra el menú principal con los módulos disponibles. (Fig.3)



**Fig. 3** Pagina Índice o principal de la aplicación web  
Fuente: (Ortiz Palma, 2019)


#### 3.2 MÓDULOS

##### 3.2.1 Módulo recopilación

Para ingresar al módulo se debe ubicar en el menú principal y dar clic en la pestaña recopilar datos (Fig. 4)



**Fig. 4** Visualización y acceso del módulo de Recopilar datos  
Fuente: (Ortiz Palma, 2019)

	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

Se despliega la siguiente ventana (Fig. 5)

### 3.2.1.1 Ventana principal



**Fig. 5 Entorno principal del módulo de Recopilación**


Fuente: (Ortiz Palma, 2019)

La anterior ventana se compone de dos formularios como sigue:

- **El formulario 1: Conexión con Twitter**

Es una ventana para establecer conexión con Twitter a través de la API (Interfaz de programación de aplicaciones). Para conocer como solicitar datos de acceso consulte sección de preguntas frecuentes.

Por medio de esta caja de texto (Fig. 6) el usuario puede establecer la conexión a la plataforma de Twitter.

	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

## Establecer conexión

**Todos los campos son obligatorios**  
*\*Ingrese datos de enlace con la app-twitter y clic en conectar\**

**Ingrese clave (consumer key)**

**Ingrese clave secreta (consumer secret)**

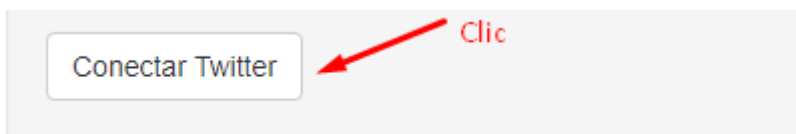
**Ingrese token de acceso (token access)**

**Ingrese token secreto de acceso (secret access)**

**Fig. 6 Formulario de conexión**

Fuente: (Ortiz Palma, 2019)


Es obligatorio ingresar todos los datos solicitados en este formulario y posteriormente dar clic en Conectar Twitter (Fig. 7). Si algún dato es incorrecto o faltante la conexión será rechazada.



**Fig. 7 Botón para establecer conexión con API Twitter**

Fuente: (Ortiz Palma, 2019)

Tras finalizar la conexión exitosamente se muestra un mensaje de confirmación (Fig. 8).

	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

## Establecer conexión


**Todos los campos son obligatorios**  
\*Ingrese datos de enlace con la app-twitter y clic en conectar\*

**Ingrese clave (consumer key)**

**Ingrese clave secreta (consumer secret)**

**Ingrese token de acceso (token access)**

**Ingrese token secreto de acceso (secret access)**

Conexión realizada 


**Fig. 8 Notificación estado de conexión**

Fuente: (Ortiz Palma, 2019)

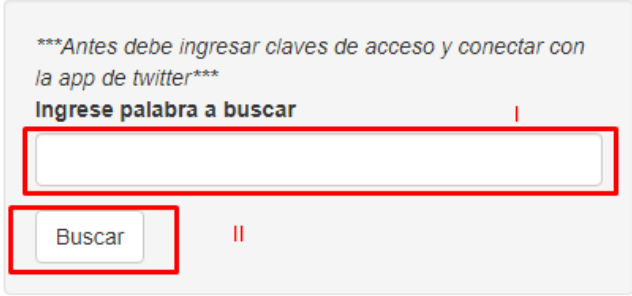
- **El formulario 2: Búsqueda**

Contiene un cuadro para consulta de palabras claves y búsqueda de tweets, los mismos que serán solicitados por medio de la acción de un botón de ejecución (Fig.9).

**Importante:** Antes de utilizar este módulo es necesario estar conectado con Twitter (Fig. 8), si se presenta algún error verificar conexión a la red y/o repetir los pasos del formulario I (Fig.6 y 7).

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## Recopilación de datos

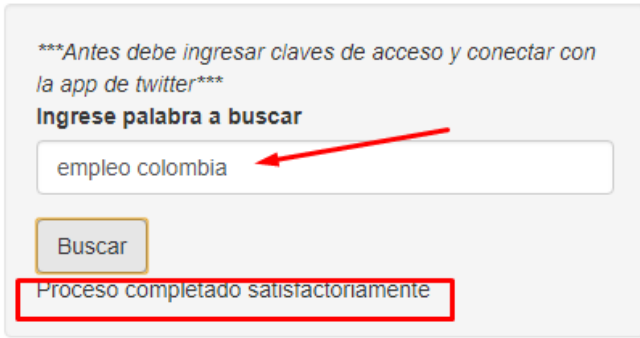


**Fig. 9 Formulario para búsqueda**

Fuente: (Ortiz Palma, 2019)

En el campo I (Fig. 9), se debe ingresar las palabras claves cortas (ejemplo: Empleo Colombia), posteriormente clic en botón buscar (II). Una vez finaliza el proceso de búsqueda se notifica con mensaje en pantalla “Proceso completado satisfactoriamente” (Fig. 10)

## Recopilación de datos

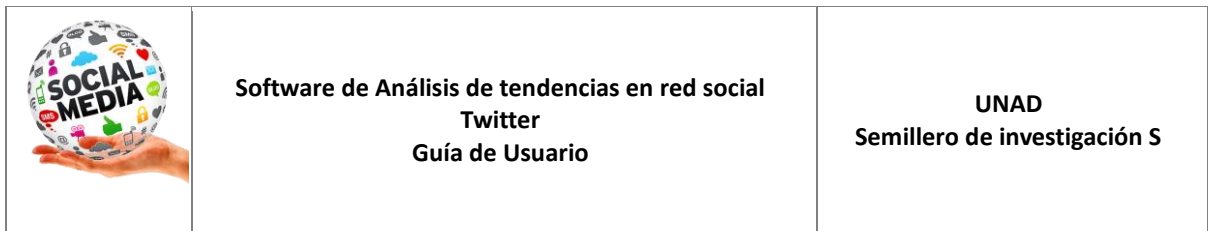


**Fig. 10 Notificación de finalización de consulta a través de API de Twitter**

Fuente: (Ortiz Palma, 2019)

Finalizada la consulta el sistema imprime una tabla en pantalla (Fig. 11), y un archivo CSV con los tweets recopilados desde Twitter, este último se guarda en carpeta local (Ruta elegida por el implementador) (Fig.12), cuyo nombre es “busqueda.csv”. Para acceder al documento generado vaya a la carpeta indicada por el administrador. Ahora puede acceder al archivo así mismo puede copiarlo y guardarlo en su lugar preferido o entorno de trabajo.

**Importante:** no se recomienda trabajar en la carpeta por defecto del programa, ya que este puede sobrescribir el archivo en cada consulta.

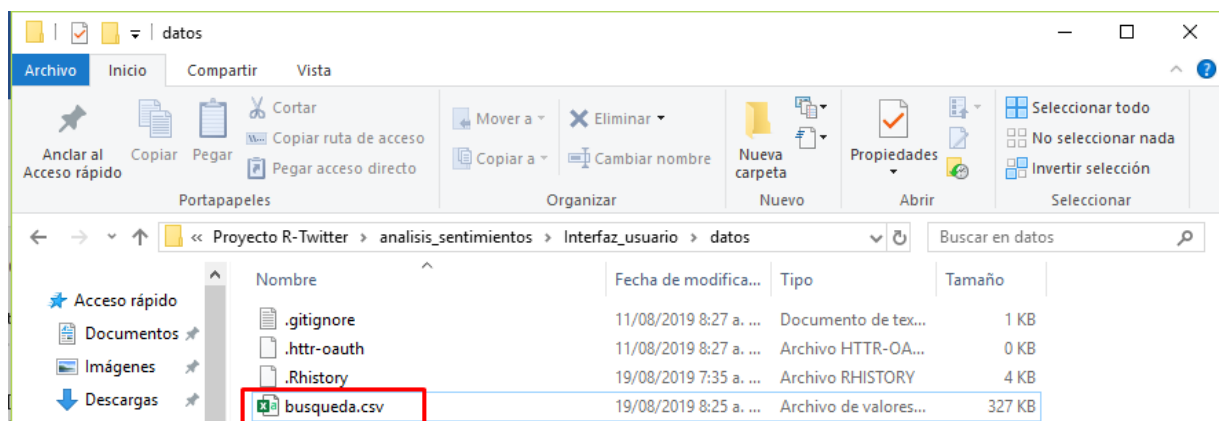


A continuación se muestra un reporte en pantalla de la consulta realizada de forma exitosa

text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID
RT @ViceColombia: @IvankaTrump estará en Colombia. La asesora del presidente @realDonaldTrump expresó a la Vicepresidente interés en apoy...	FALSE	0.00	NA	1566220884.00	FALSE	NA	1163440830767607814	NA
RT @ViceColombia: @IvankaTrump estará en Colombia. La asesora del presidente @realDonaldTrump expresó a la Vicepresidente interés en apoy...	FALSE	0.00	NA	1566220827.00	FALSE	NA	1163440590341603329	NA
Es dizque "mamerta" y buscando empleo le recomiendo que se acerque a la Colombia humana allí quizás le den... <a href="https://t.co/KdrQvzbvzI4">https://t.co/KdrQvzbvzI4</a>	FALSE	3.00	NA	1566220185.00	TRUE	NA	1163437898529738753	NA
Tengo una compañera de trabajo q me provoca es caerla a cofazo... Tiene su empleo, un hijo en Colombia, uno en Pana... <a href="https://t.co/frS3P7N0I4">https://t.co/frS3P7N0I4</a>	FALSE	1.00	NA	1566220178.00	TRUE	NA	1163437870457196545	NA
RT @alemustakis: Lo hicieron en Colombia!! No nos debemos quedar atrás. Más que nunca debemos apoyar a los emprendedores y pymes a que pue...	FALSE	0.00	NA	1566220088.00	FALSE	NA	1163437491426406400	NA
RT @ANI_Colombia: Ten cuidado, #NoTeDejesEngañar, están circulando por redes sociales y	FALSE	0.00	NA	1566219956.00	FALSE	NA	1163436940689063937	NA


**Fig. 11** Tabla generada en pantalla de Tweets consultados  
Fuente: (Ortiz Palma, 2019)

Aquí se muestra un directorio de destino con el archivo generado por el sistema y guardado localmente.



**Fig. 12** Documento exportado a carpeta local  
Fuente: (Ortiz Palma, 2019)

**Finalización de recopilación:** Si llego hasta este punto ha finalizado el proceso de recopilación de datos.

	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

### 3.2.1.2 Ayudas contextuales

A lo largo del trayecto del uso de este sistema, encontrará algunas alertas y mensajes de apoyo para el debido uso de esta herramienta.

La Fig. 13 muestra una ayuda apoyada en texto para recordarle al usuario a realizar los pasos de forma correcta. En este caso se recuerda la necesidad de establecer conexión antes de continuar con el proceso.

## Recopilación de datos

*\*\*\*Antes debe ingresar claves de acceso y conectar con la app de twitter\*\*\**

**Ingrese palabra a buscar**

**Fig. 13** Textos de apoyo de los requisitos previos para realizar la búsqueda.

Fuente: (Ortiz Palma, 2019)

Asimismo por medio en la Fig. 14, se resalta por medio de un texto la importancia de completar todos los campos en el formulario.

## Establecer conexión


**Todos los campos son obligatorios**  
*\*Ingrese datos de enlace con la app-twitter y clic en conectar\**

**Ingrese clave (consumer key)**

**Ingrese clave secreta (consumer secret)**

**Ingrese token de acceso (token access)**

**Ingrese token secreto de acceso (secret access)**

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social</b> <b>Twitter</b> <b>Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD</b> <b>Semillero de investigación S</b></p>
---	--	--

**Fig. 14** Textos de apoyo de los requisitos previos para realizar la conexión con la API  
Fuente: (Ortiz Palma, 2019)

Adicionalmente el sistema tiene habilitado una opción (Fig.15) dentro del menú principal donde el usuario puede conseguir ayuda como es el acceso a documentación relacionada, preguntas frecuentes y un link de contacto.



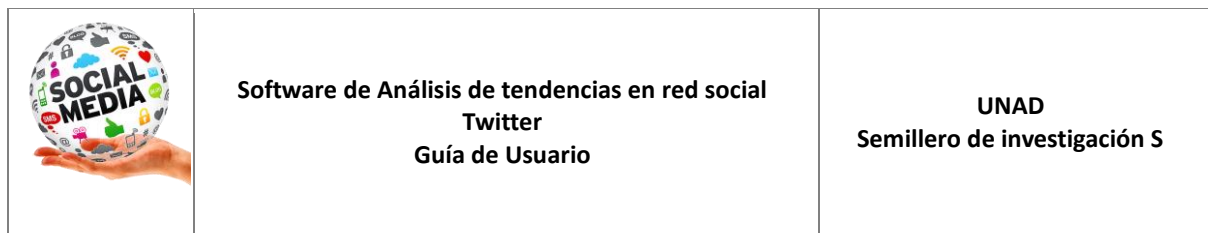
**Fig. 15** Menú de Ayuda sobre uso y soporte técnico  
Fuente: (Ortiz Palma, 2019)

### 3.2.2 Módulo de Análisis

Aquí puede realizar un estudio a sus datos y por medio del análisis de sentimientos de esta herramienta informática conocer y proyectar tendencias

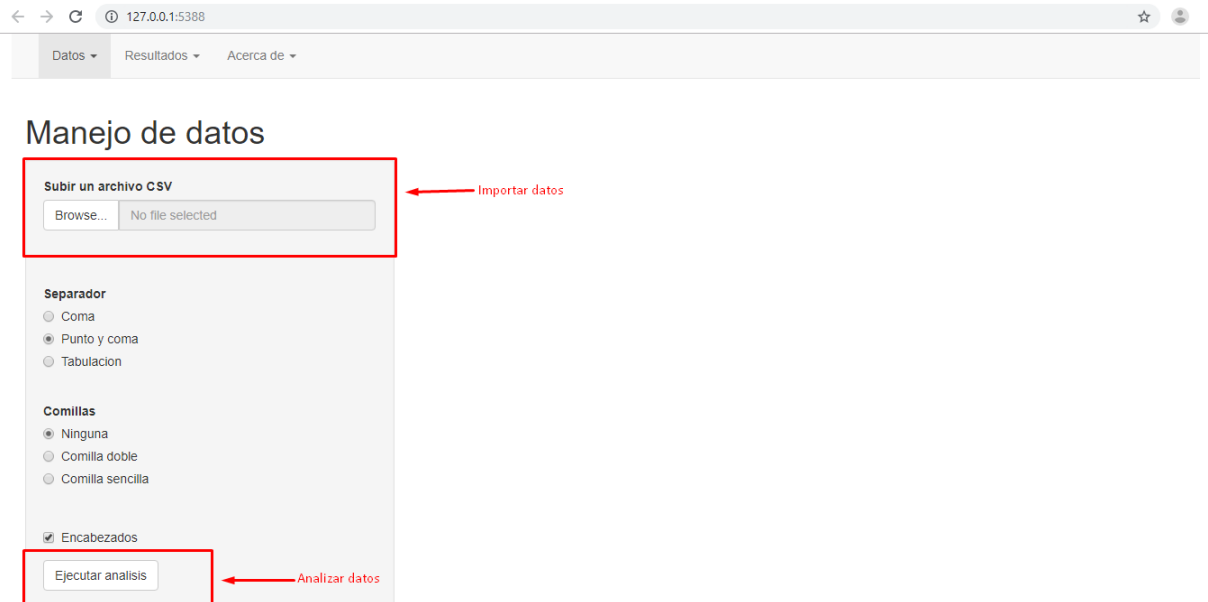
#### 3.2.2.1 Ventana principal

Para ingresar al módulo se debe ubicar en el menú principal y dar clic en la pestaña analizar datos (Fig. 16)



**Fig. 16 Visualización y acceso del módulo de Análisis**  
Fuente: (Ortiz Palma, 2019)

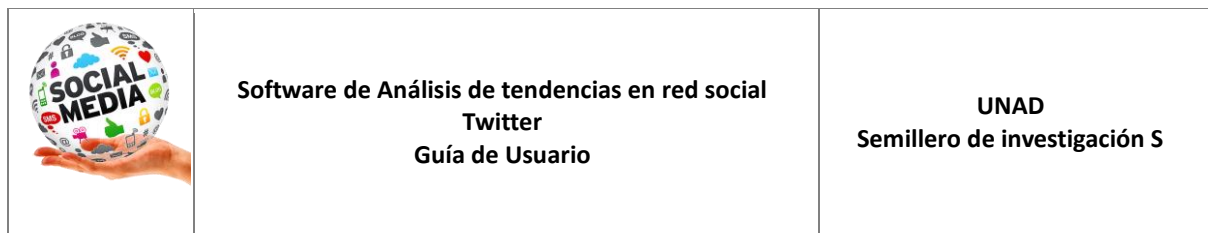
Se despliega una ventana (Fig. 17), la cual corresponde al entorno de incorporación de nuevos datos a la plataforma.



**Fig. 17 Página principal del módulo de análisis**  
Fuente: (Ortiz Palma, 2019)

**Importante:** Aquí debe cargar los datos con algunos requerimientos especiales como son:

- Documento en formato CSV
- El archivo debe contener una **columna** de nombre **sentiment**, la cual se debe clasificar de forma manual para cada uno de los tweets previamente (positivo = 1; negativo= -1).

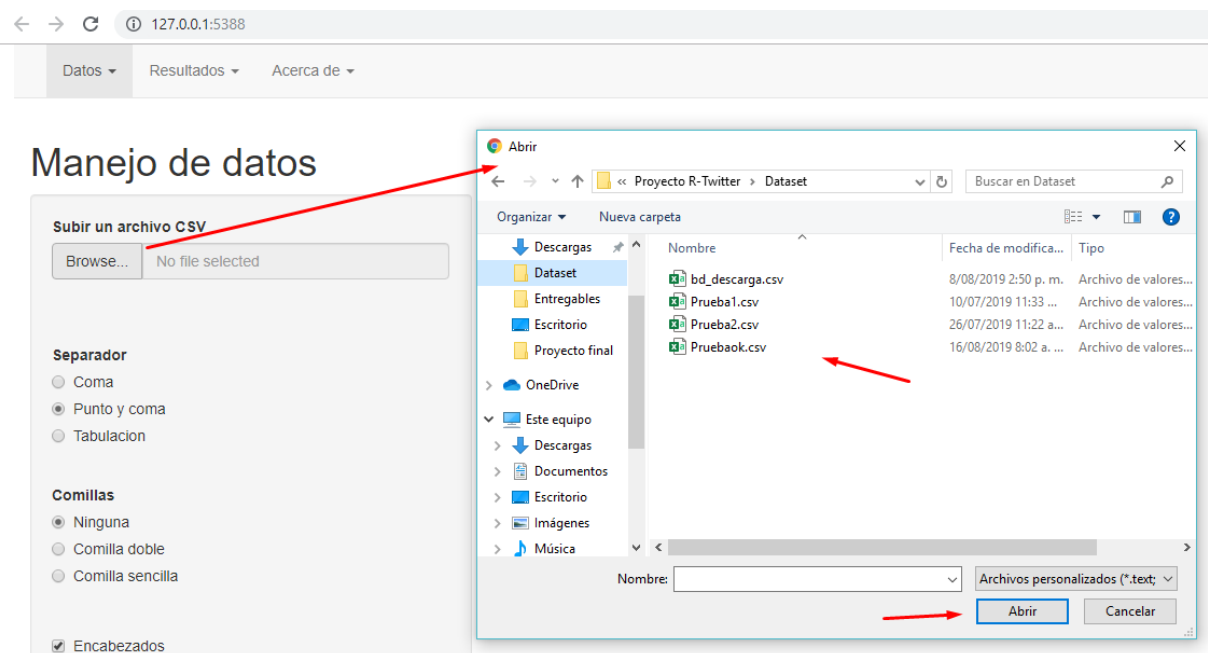


Ver procedimiento en sección [preguntas frecuentes](#).

Una vez se cumpla con los requerimientos antes mencionados continúe con la carga de los datos. Empiece dando clic en la opción **Subir archivo** (Fig. 17) para ello haga clic en **Browse**

Paso siguiente se despliega la ventana para importar un documento (Fig. 18). Siga estos pasos:

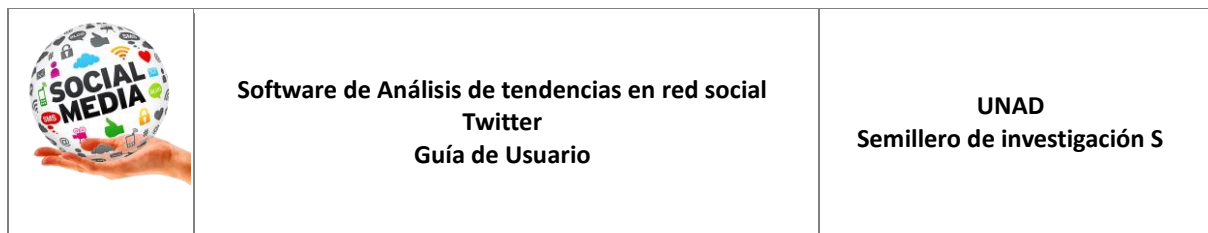
- Buscar el archivo en su equipo o medio desde donde desee cargarlo
- Seleccionar archivo
- Clic en Abrir



**Fig. 18** Paso a paso para importar documentos

Fuente: (Ortiz Palma, 2019)

Esperar hasta que el proceso de importación cargue completamente, una vez finalizado, se muestra mensaje de carga completa y una tabla con los datos (Fig. 19).



## Manejo de datos

**Subir un archivo CSV**

Browse... Prueba1.csv

Upload complete

**Separador**

Coma

Punto y coma

Tabulacion

**Comillas**

Ninguna

Comilla doble

Comilla sencilla

Encabezados

text	favorited	favoriteCount	replyToSN	created
RT @casaleantonio: ¿Hay algo mejor que el D1? jajajaja sólo que gane Millonarios. Hoy abrieron la tienda número MIL en Colombia <U+0001F44F><U+0001F3FB>. Son los...	FALSE	0	NA	9/07/2019 3:53
RT @casaleantonio: ¿Hay algo mejor que el D1? jajajaja sólo que gane Millonarios. Hoy abrieron la tienda número MIL en Colombia <U+0001F44F><U+0001F3FB>. Son los...	FALSE	0	NA	9/07/2019 3:49
@DanielPBayona @camillacasp89 @petrogustavo @lvanDuque @angelamrobledo @lvanCepedaCast @AidaAvellaE @GNavasTalero... https://t.co/ekNuiTX01j	FALSE	0	DanielPBayona	9/07/2019 3:49
RT @casaleantonio: ¿Hay algo mejor que el D1? jajajaja sólo que gane	FALSE	0	NA	9/07/2019 3:38

**Fig. 19 Proceso de cargue de documento**


Fuente: (Ortiz Palma, 2019)

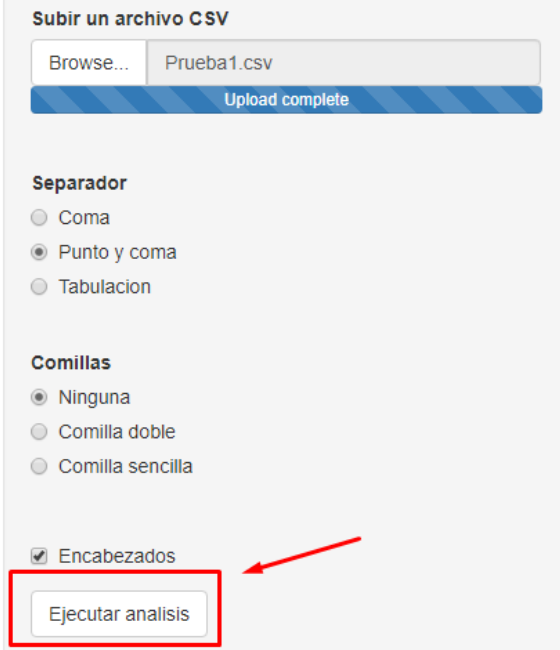
Lo anterior muestra la matriz de datos cargada al sistema.

**Importante:** verificar que la tabla contenga la columna **sentiment**

**Ejecutar análisis:** Ahora que todos los datos están cargados puede continuar con el proceso. Este paso es el encargado de realizar la clasificación de información, aplicar técnicas de análisis de sentimientos y generar tendencias con base a los datos suministrados.

Para proceder clic en el botón ejecutar análisis. (Fig. 20)

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---



**Fig. 20 Botón de acción para iniciar análisis de datos**  
Fuente: (Ortiz Palma, 2019)


Por favor espere...

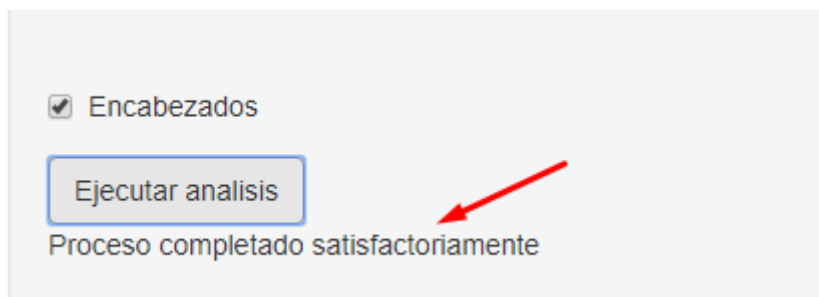
En esta etapa el sistema puede tardar un tiempo, lo cual depende de la cantidad de datos cargados.

Esperar hasta que reciba un mensaje de notificación y/o finalización del proceso antes de revisar los resultados.

### 3.2.2.2 Mensaje de notificación

Finalizado el proceso el software le notifica con el siguiente mensaje “ Proceso completado satisfactoriamente. (Fig. 21)

	<p align="center"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p align="center"><b>UNAD Semillero de investigación S</b></p>
---	---	--



**Fig. 21 Mensaje de notificación**

Fuente: (Ortiz Palma, 2019)

### 3.2.2.3 Resultados

Este entorno contiene un menú de opciones, donde el usuario puede consultar diferentes reportes. Como los que se describen a continuación.

**Importante:** Esto solamente es posible revisar si el proceso anterior terminó correctamente.

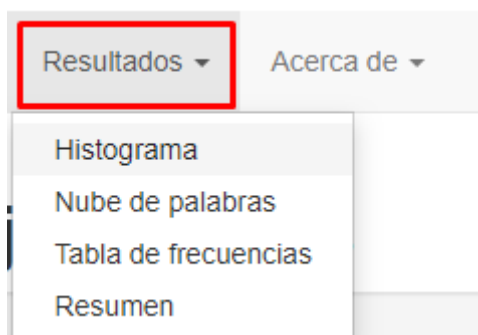
La ruta para acceder desde el menú principal es mediante clic en la opción de resultados (Fig. 22).

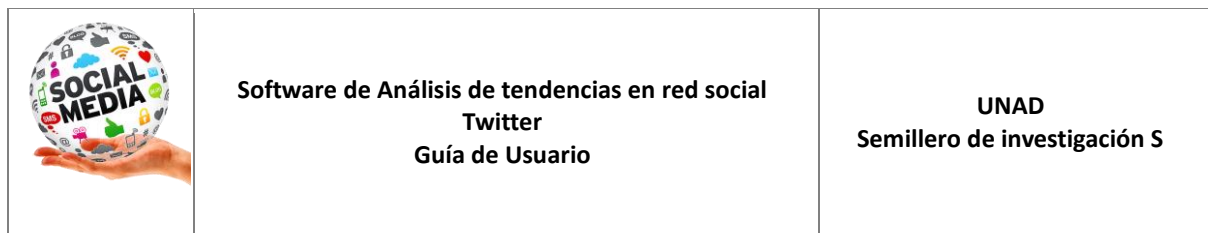


**Fig. 22 Visualización y accesos a pestaña de resultados**

Fuente: (Ortiz Palma, 2019)

En cual contiene reportes como: Histograma, Nube de palabras, tabla de frecuencias y un resumen (Fig. 23).





**Fig. 23 Menú de opciones de pestaña de Resultados**

Fuente: (Ortiz Palma, 2019)

- **Histograma:**

El histograma de un conjunto de datos es un gráfico de barras que representan las frecuencias con que aparecen las mediciones agrupadas en ciertos rangos o intervalos. (Correa & González, 2002)

Puede acceder en la siguiente ruta: Resultados -> Histograma (Fig. 24)



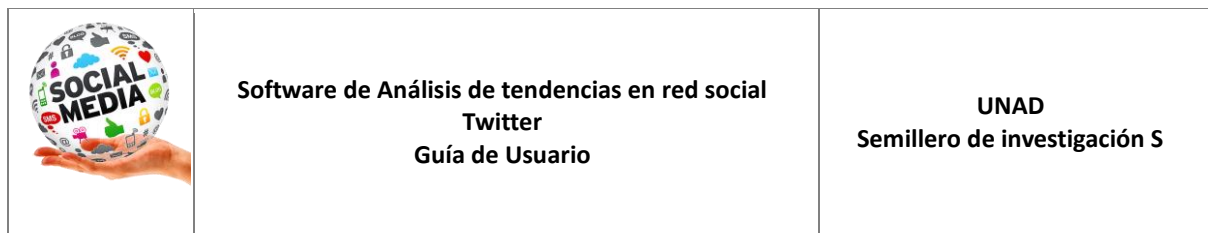
**Fig. 24 Visualización y acceso a Grafica – Histograma**

Fuente: (Ortiz Palma, 2019)

- **Nube de palabras**

Una nube de palabras o nube de etiquetas es una representación visual de las palabras que conforman un texto, en donde el tamaño es mayor para las palabras que aparecen con más frecuencia. Uno de sus usos principales es la visualización de las etiquetas de un sitio web, de modo que los temas más frecuentes en el sitio se muestren con mayor prominencia. Las etiquetas son palabras clave que suelen estar ordenadas alfabéticamente o, en ocasiones, agrupadas semánticamente. La importancia de una etiqueta se muestra con el tamaño de la fuente y/o color. (Martos, 2014)





**Fig. 26 Visualización y acceso a tabla de frecuencias**

Fuente: (Ortiz Palma, 2019)

- **Resumen**


Muestra un resumen del análisis de los datos entre los que se encuentra la frecuencia mínima y máxima de los datos, media, mediana y moda, primer y tercer cuartil, además de la longitud máxima de palabras encontrada.

Puede acceder en la siguiente ruta: Resultados -> Resumen (Fig. 27)



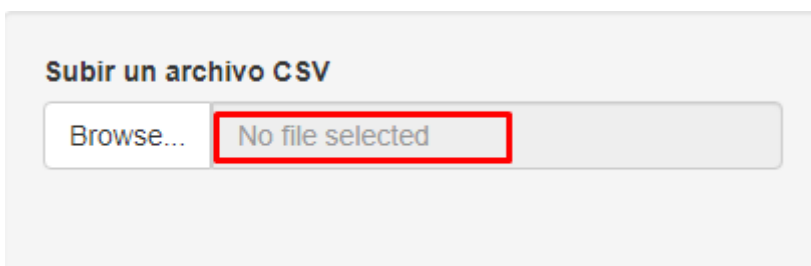
**Fig. 27 Visualización y acceso a Resumen**

Fuente: (Ortiz Palma, 2019)

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

### 3.2.2.4 Ayudas contextuales

Las ayudas contextuales se muestran cuando el usuario se encuentra en un punto y le permiten acceder a ayuda o indicaciones de la acción que se encuentre realizando en el momento. Por ejemplo, en la Fig. 28 se notifica que aún no se ha cargado un documento.



**Fig. 28 Notificación de documento no cargado**


Fuente: (Ortiz Palma, 2019)

Asimismo, se ha añadido una pestaña **Ayuda** en las opciones de Acerca de (Fig. 29). Aquí el usuario puede acceder a soporte y contextualización de los problemas más recurrentes.



**Fig. 29 Menú de Ayuda sobre uso y soporte técnico**

Fuente: (Ortiz Palma, 2019)

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## 4 PREGUNTAS FRECUENTES

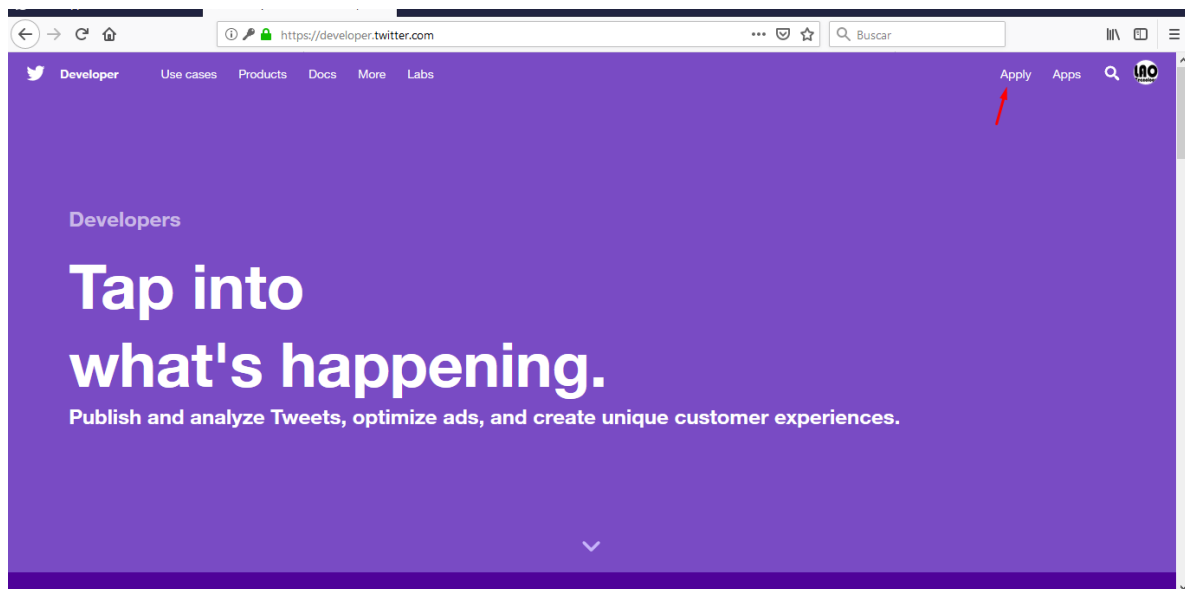
A continuación, se incluye una lista de las preguntas o dudas más frecuentes (Frequently Asked Questions) que pueden surgirle a un usuario del sistema junto a una explicación para cada una de ellas.

### 4.1 ¿Como obtener y solicitar datos de acceso a la API de Twitter?

#### Pasos para solicitud de modo desarrollador en twitter

##### 1. SOLICITUD DE API TWITTER

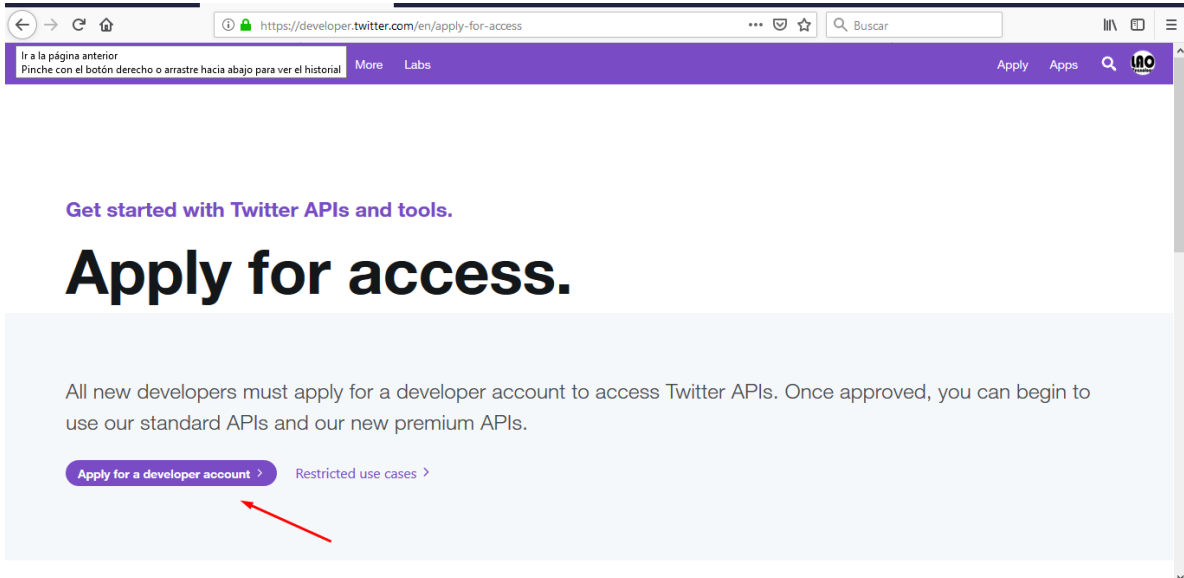
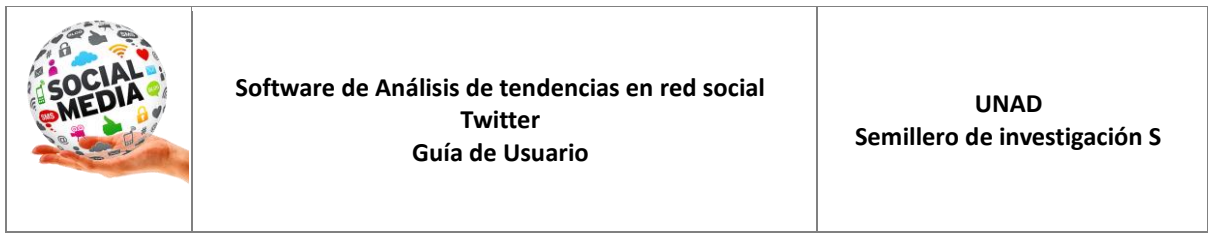
- Es necesario acceder a una cuenta de Twitter en caso de no disponer de ninguna acceder a <http://www.twitter.com> para crearla. (Fig. 30)
- Inicio de sesión <https://developer.twitter.com/en/apps>
- Darse de alta como desarrollador
- Llenar datos de uso y acceso



**Fig. 30. Ventana Twitter para aplicar como desarrollador**

Fuente: (Twitter, 2019)

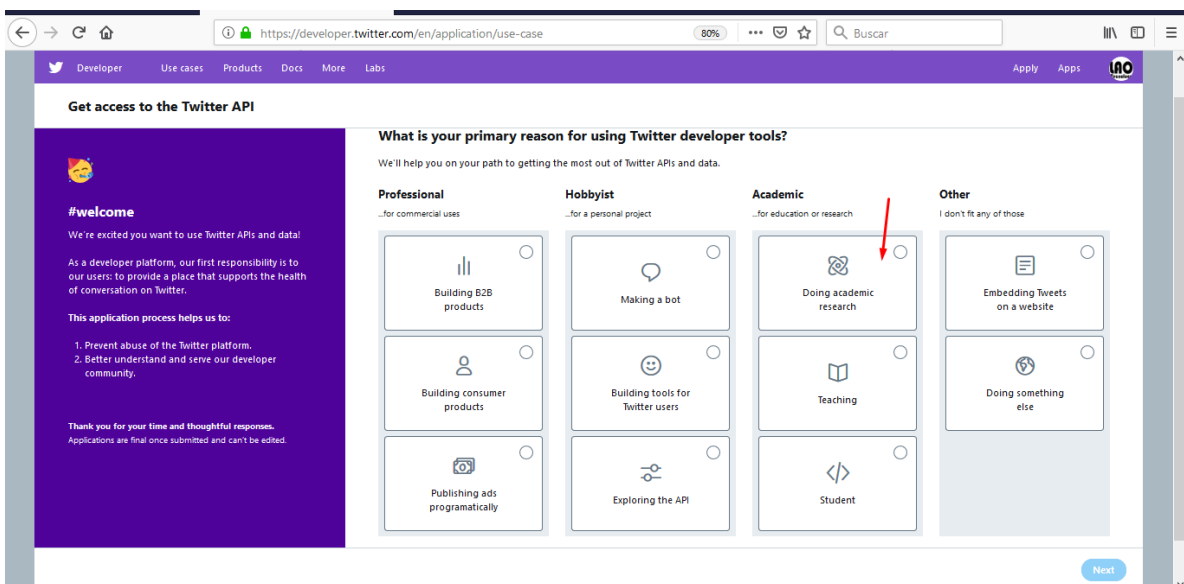
- Aplicar para una cuenta de desarrollador



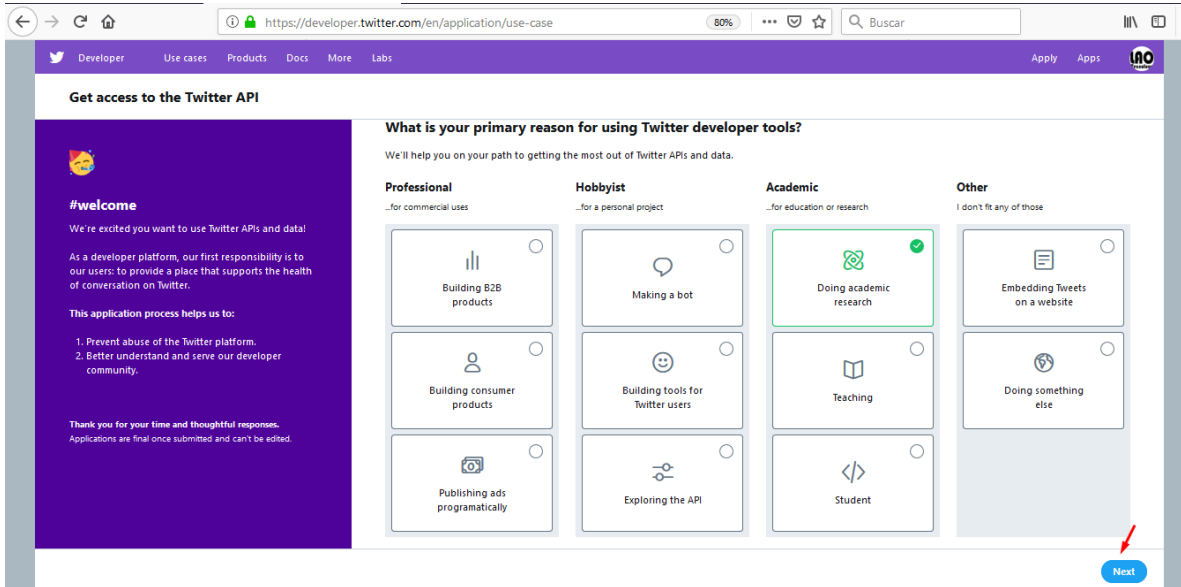
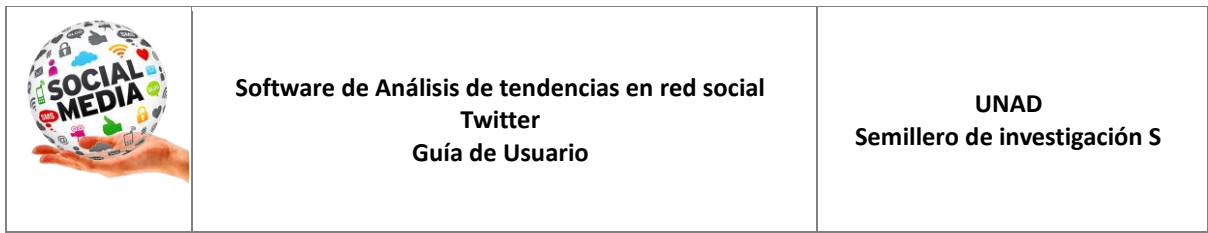
**Fig. 31. Ventana Twitter para aplicar a una cuenta como desarrollador**  
Fuente: (Twitter, 2019)

Esta solicitud es necesaria para poder obtener una API, en la (Fig. 31) se observa el paso necesario para darle de alta a una cuenta Twitter con opciones de desarrollador.

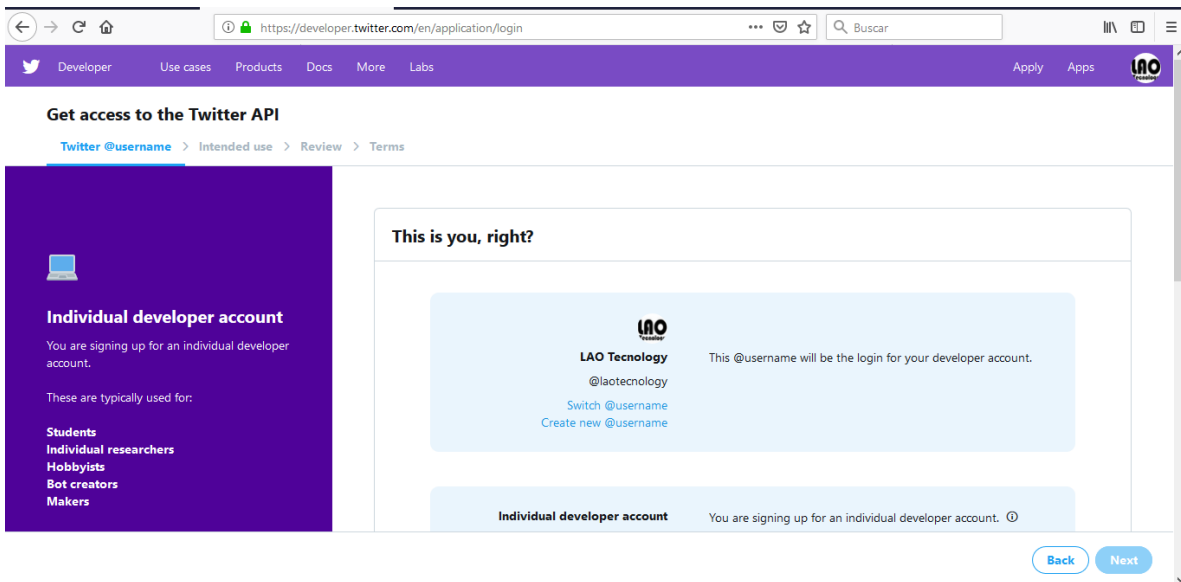
- Escoger el uso o fin de la cuenta



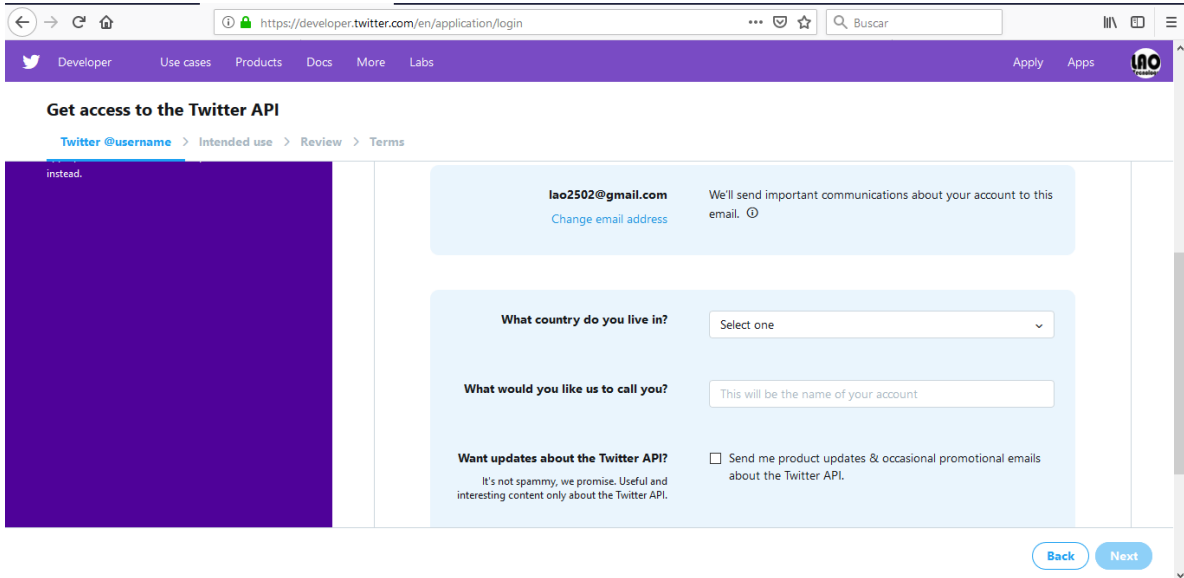
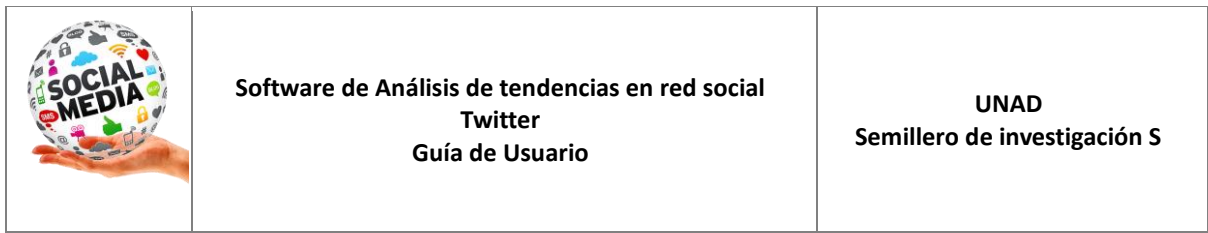
**Fig. 32. Opciones de uso de la API en la cuenta en Twitter**  
Fuente: (Twitter, 2019)



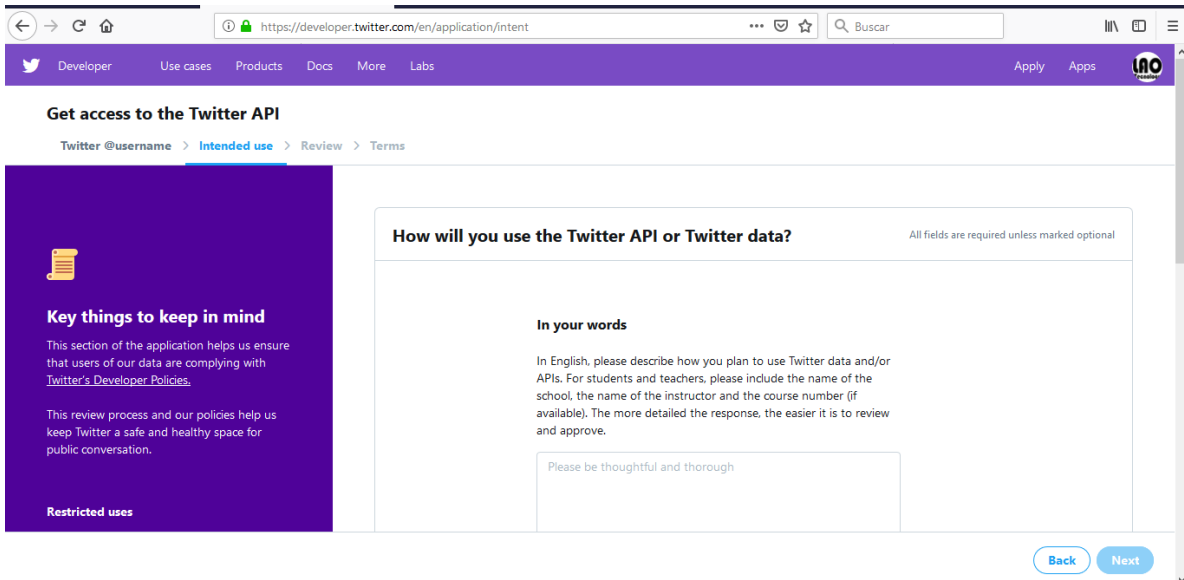
**Fig. 33. Botón de avance en la creación de API Twitter**  
Fuente: (Twitter, 2019)



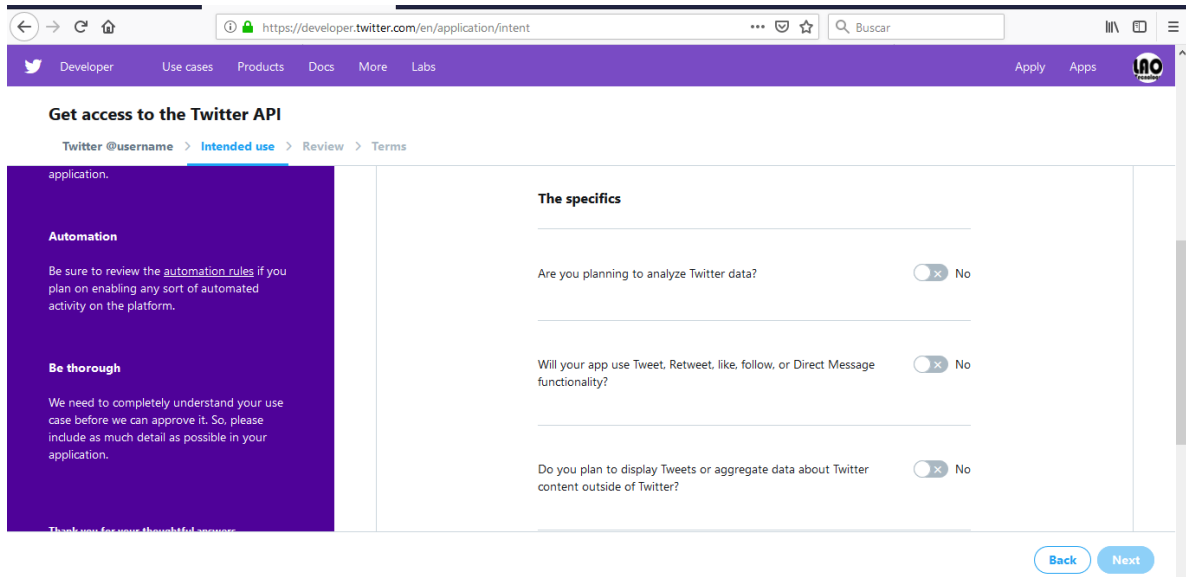
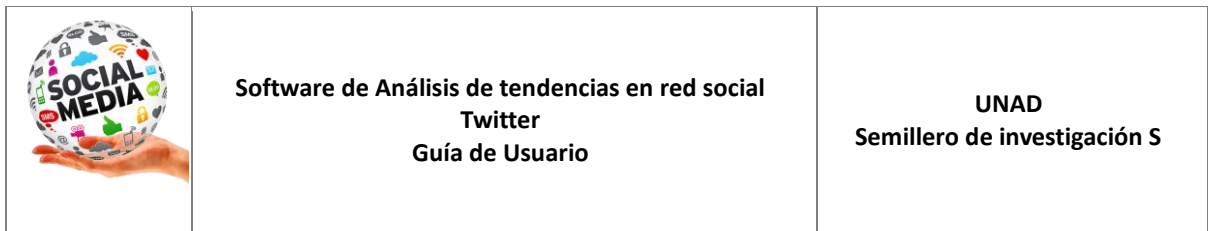
**Fig. 34. Ventana de verificación de la Cuenta utilizada para solicitud de API en Twitter**  
Fuente: (Twitter, 2019)



**Fig. 35. Datos de validación y configuración de la API Twitter**  
Fuente: (Twitter, 2019)

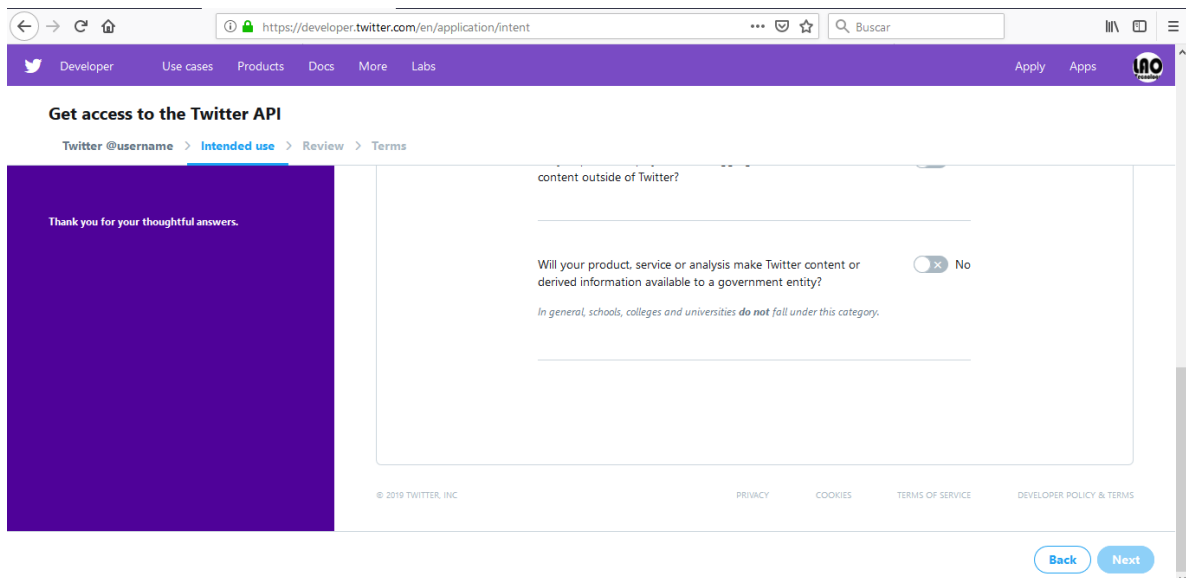


**Fig. 36. Verificación de seguridad y políticas de uso de la API Twitter**  
Fuente: (Twitter, 2019)



**Fig. 37. Verificación uso de la API en Twitter**

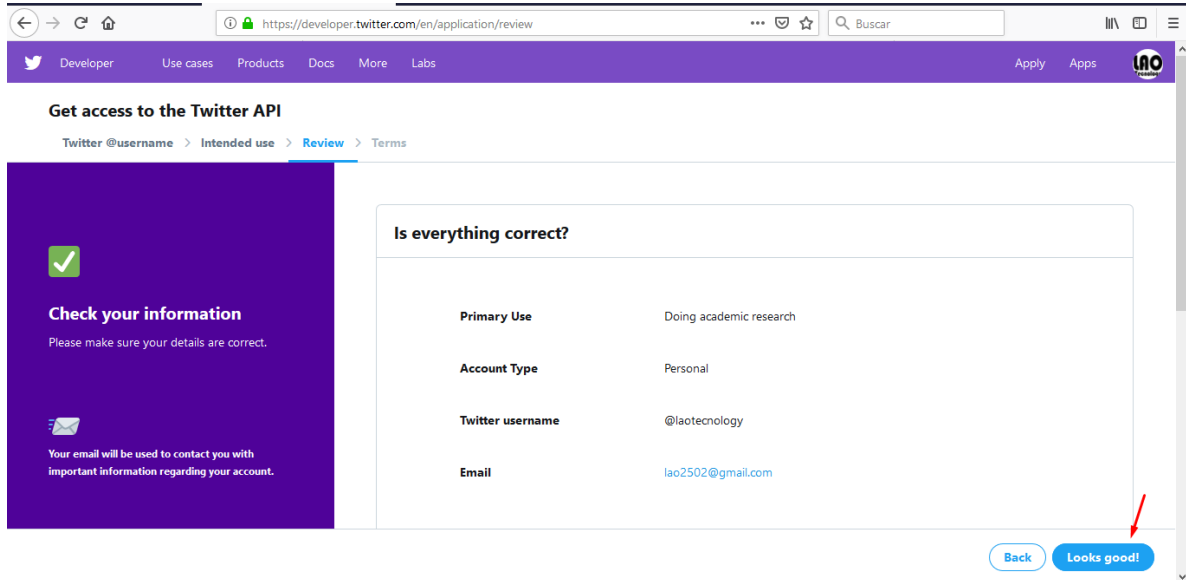
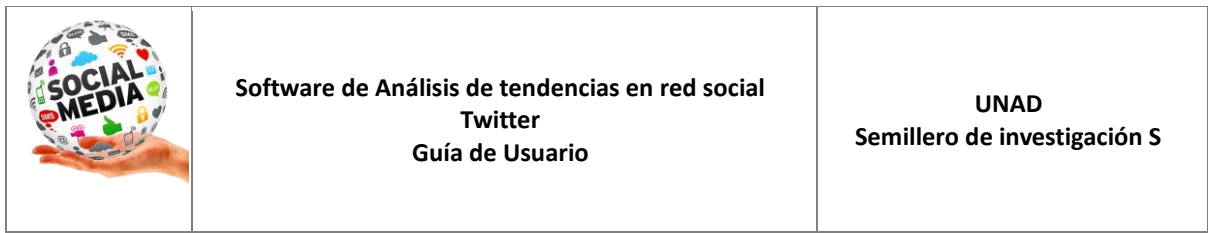
Fuente: (Twitter, 2019)



**Fig. 38. Verificación uso de la API en Twitter**

Fuente: (Twitter, 2019)

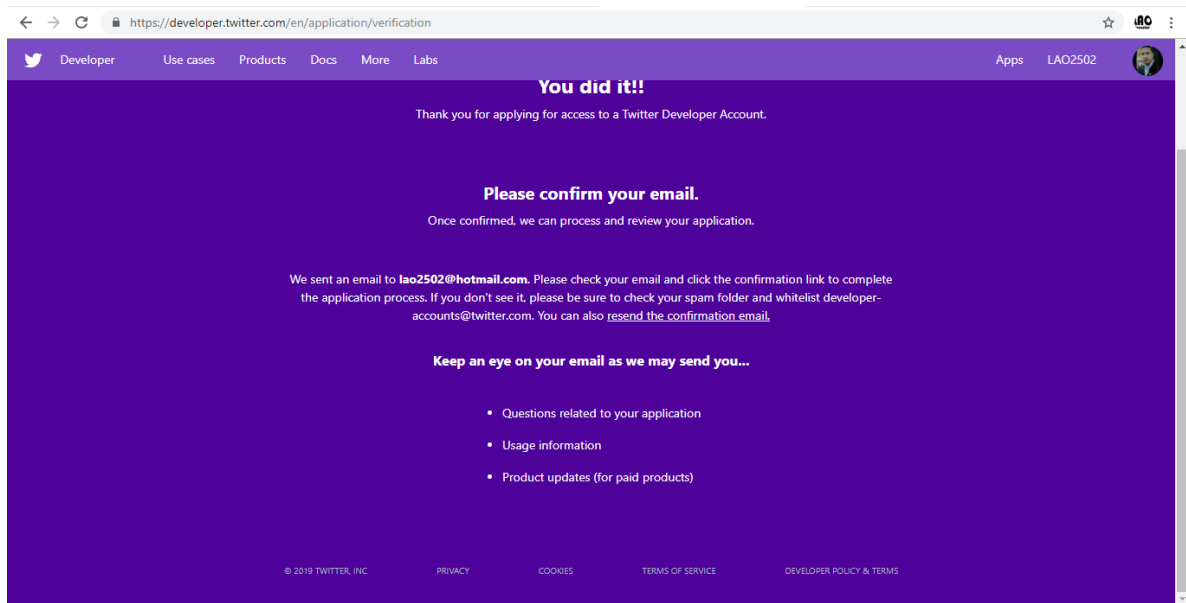
- Resumen de solicitud



**Fig. 39. Resumen de la solicitud de API Twitter**


Fuente: (Twitter, 2019)

- Aceptar términos de uso
- Envío de correo de confirmación

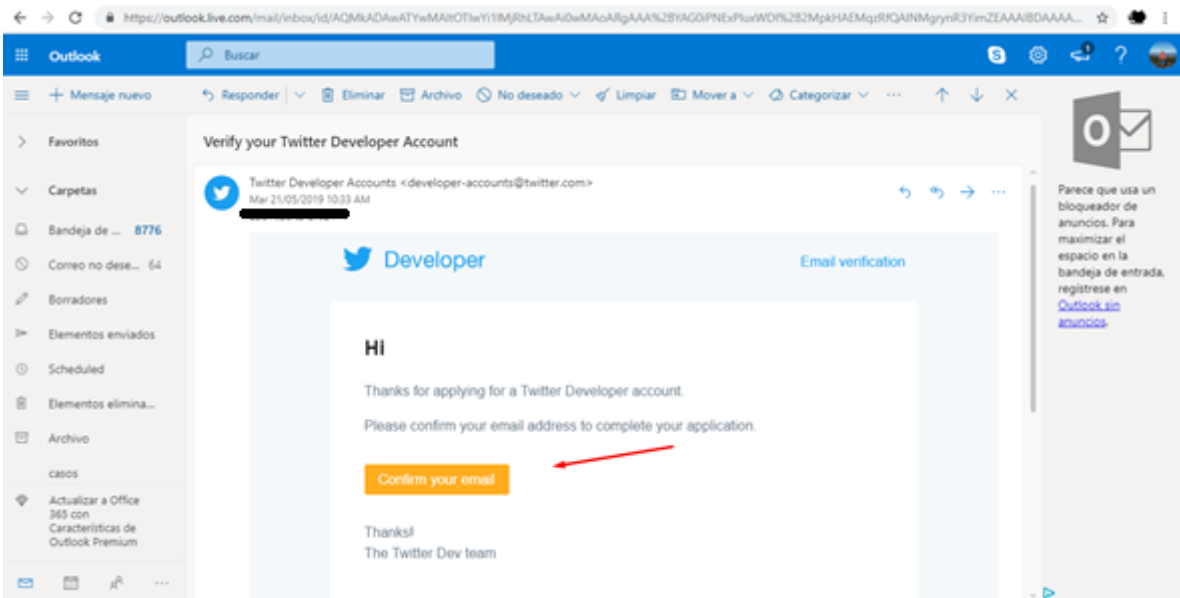


**Fig. 40. Envío de email de confirmación**

Fuente: (Twitter, 2019)

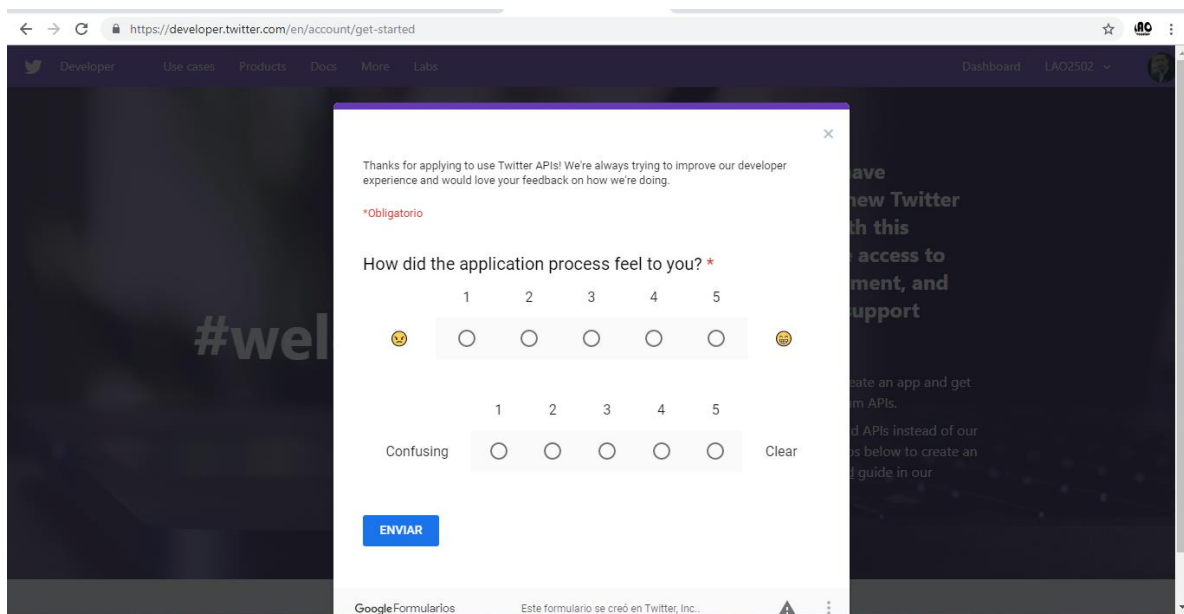
	<p><b>Software de Análisis de tendencias en red social</b>  <b>Twitter</b>  <b>Guía de Usuario</b></p>	<p><b>UNAD</b>  <b>Semillero de investigación S</b></p>
---	--	---

## 1. Email de verificación



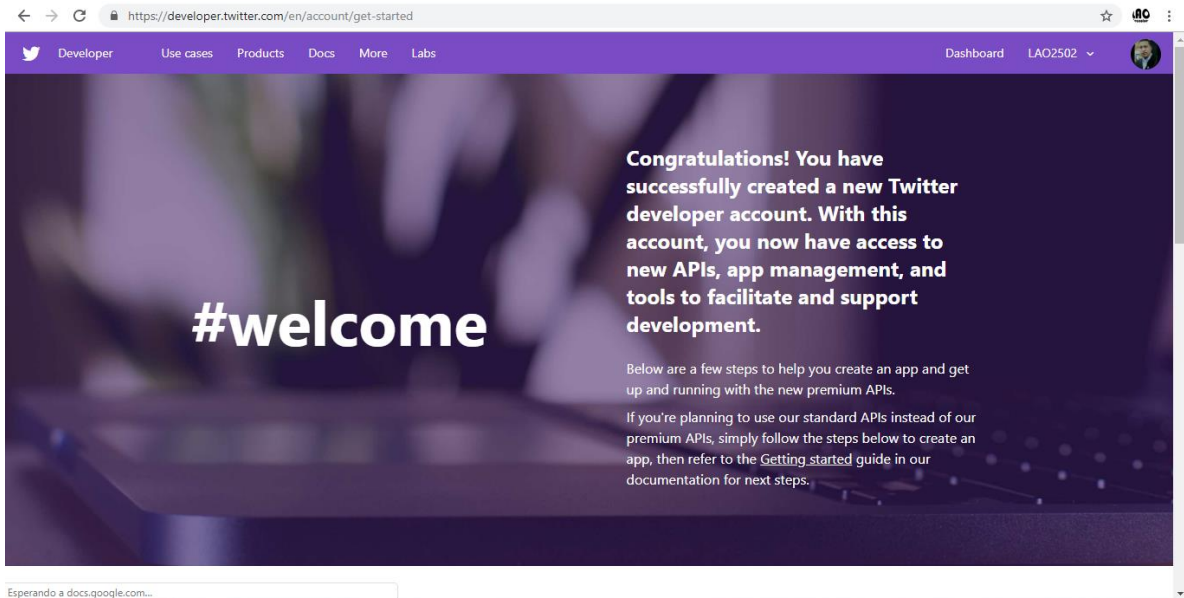
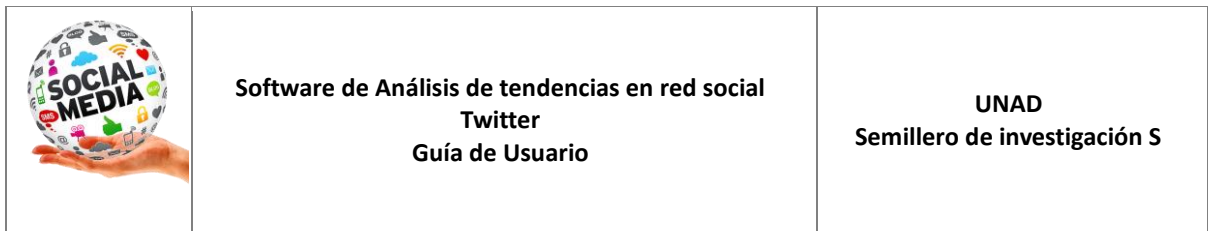
**Fig. 41. Mensaje de confirmación en correo electrónico**

Fuente: (Twitter, 2019)



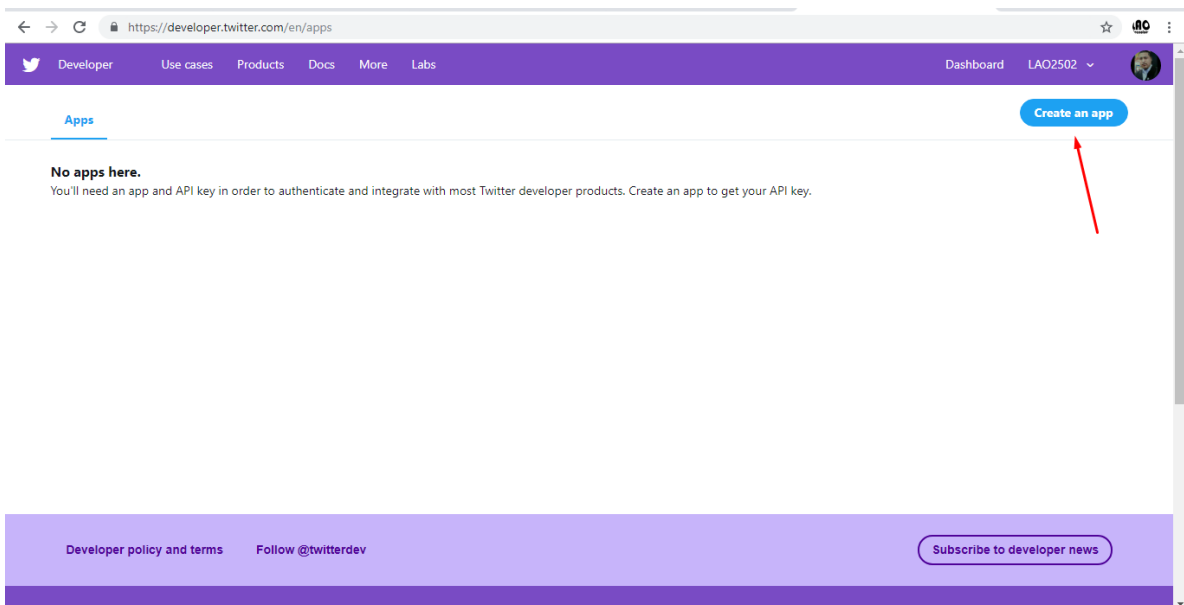
**Fig. 42. Respuesta de acción de confirmación de email**

Fuente: (Twitter, 2019)



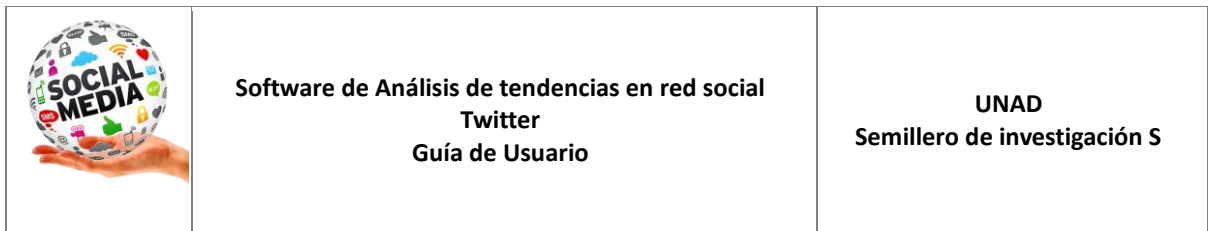
**Fig. 43. Mensaje de bienvenida y notificación de solicitud de API**  
Fuente: (Twitter, 2019)

## 2. Creación de aplicación



**Fig. 44. Botón de creación de API Twitter**  
Fuente: (Twitter, 2019)

Detalles de la aplicación

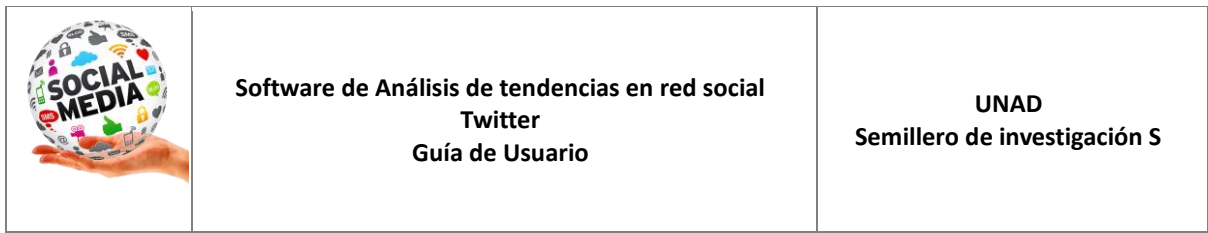


**Fig. 45. Formulario de creación de API Twitter**

Fuente: (Twitter, 2019)

**Fig. 46. Formulario de creación de API Twitter**

Fuente: (Twitter, 2019)



Nombre de la Organización

URL del sitio web de la organización

Diganos cómo se usará esta aplicación (requerido)

Este campo solo es visible para los empleados de Twitter. Ayúdanos a entender cómo se usará tu aplicación. ¿Qué le permitirá a usted y sus clientes hacer?

Please be detailed.

Caracteres mínimos: 100

Cancelar Crear

**Fig. 47. Formulario de creación de API Twitter**

Fuente: (Twitter, 2019)

### Resumen de la aplicación

Detalles de la aplicación

Editar

Icono de la aplicación

El icono de la aplicación es predeterminado, haga clic en editar para cargar.

Nombre de la aplicación

Evaluación de necesidades de TI

Descripción

APLICACIÓN DE TÉCNICAS DE ANÁLISIS DE INFORMACIÓN DE LA RED SOCIAL TWITTER, PARA LA VISUALIZACIÓN DE TENDENCIAS Y NECESIDADES LABORALES Y DE FORMACIÓN EN EL SECTOR DE T.I.

URL del sitio web

https://twitter.com/lao2502

Registrarse con Twitter

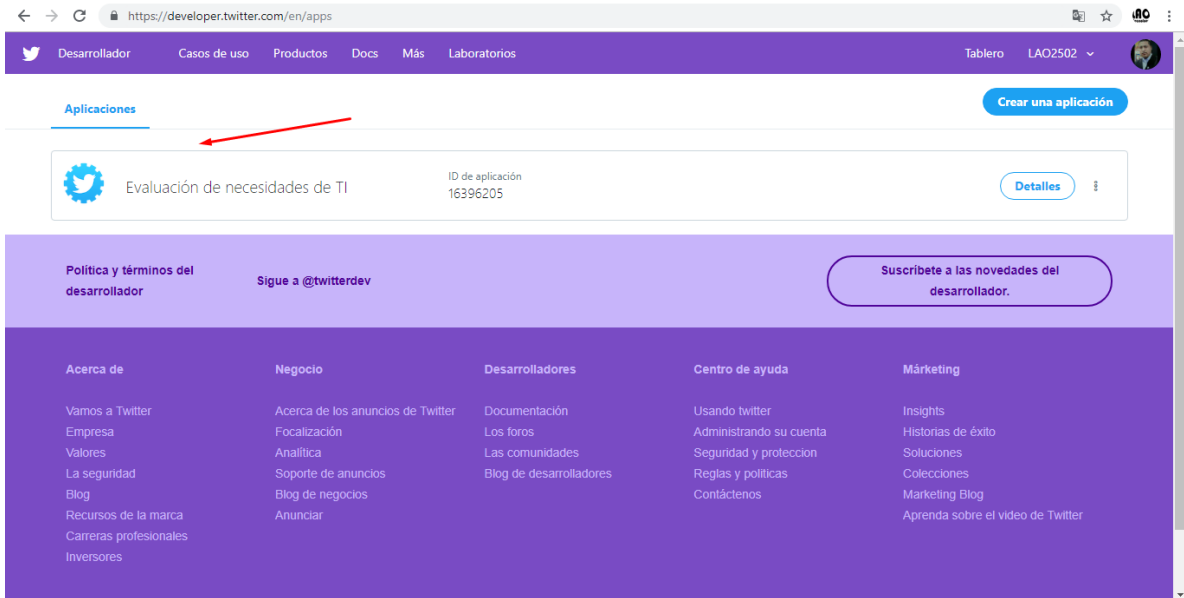
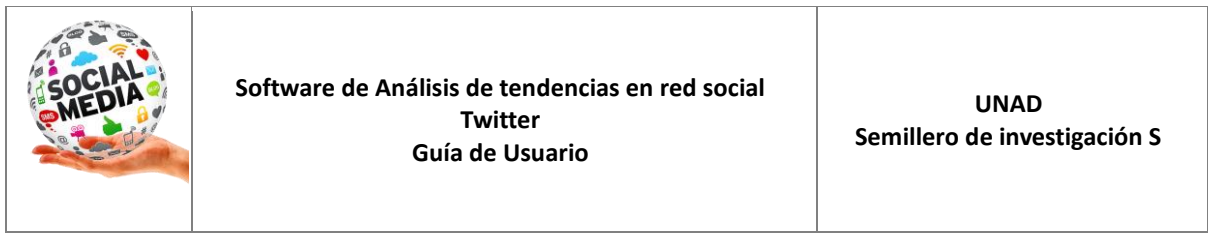
Discapacitado

URL de devolución de llamada

None

**Fig. 48. Resumen de la API creados en Twitter**

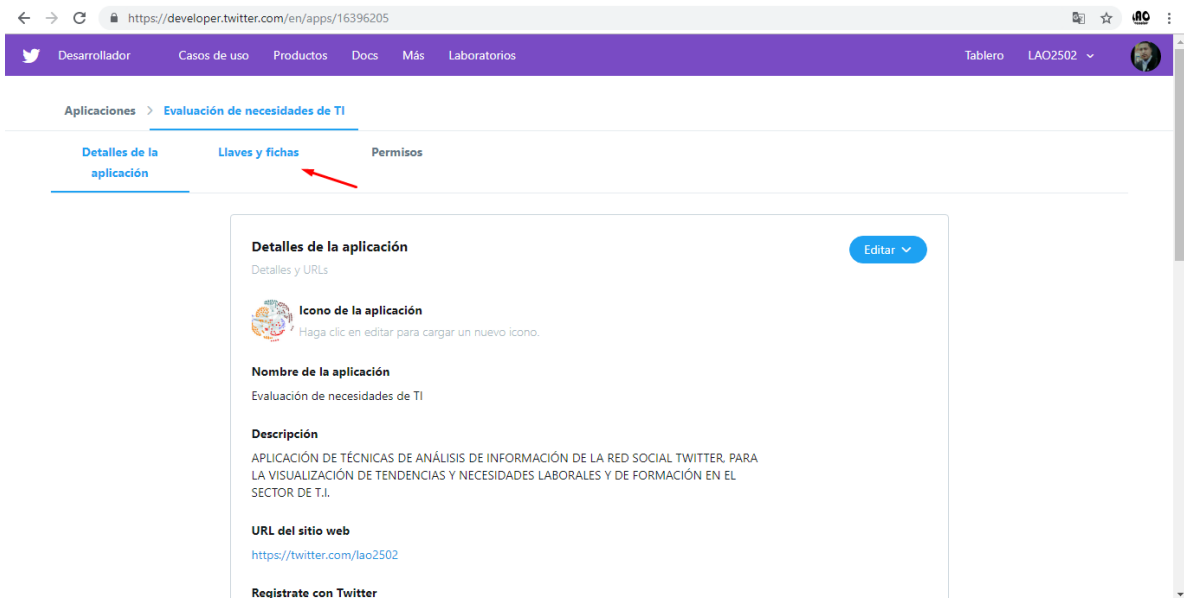
Fuente: (Twitter, 2019)



**Fig. 49. Visualización de API creada en Twitter**

Fuente: (Twitter, 2019)

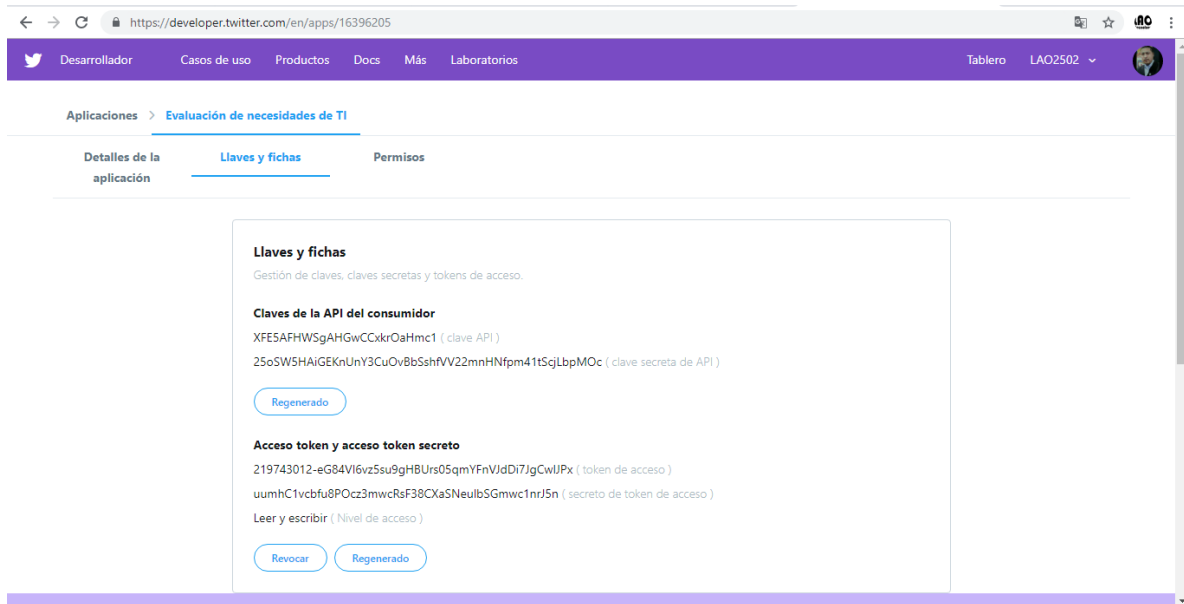
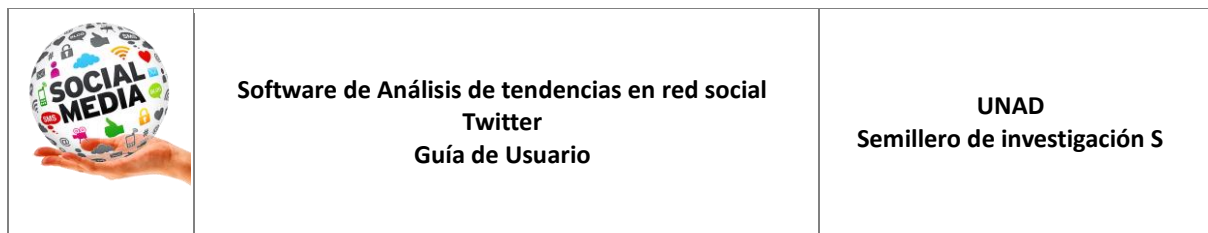
### Llaves y tokens



**Fig. 50. Acceso a Tokens y claves de la API Twitter**

Fuente: (Twitter, 2019)

Se debe Habilitar tokens




**Fig. 51. Tokens y claves de acceso a API Twitter**  
Fuente: (Twitter, 2019)

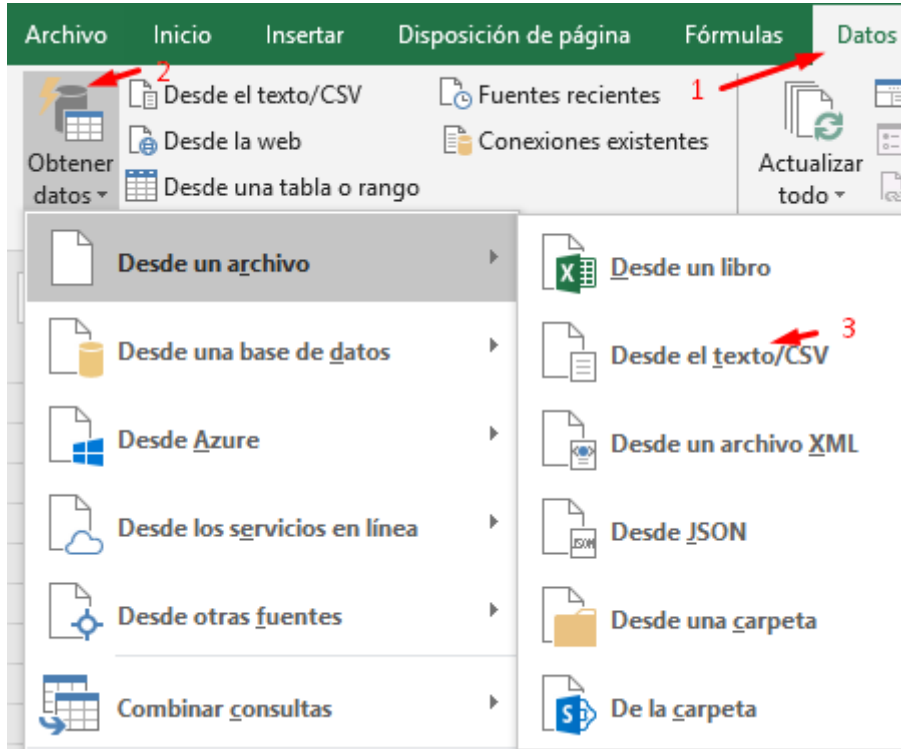
Terminado este proceso, usted tendrá los datos de acceso en la Opción Llaves y fichas (Fig. 51). Estos son los datos que debe usar para conectarse en la aplicación web ARS SIAVA.

#### **4.2 ¿Como crear columna sentiment?**

Para ello abra Microsoft Excel e importe el documento CSV a trabajar. Para ello realice los siguientes pasos:

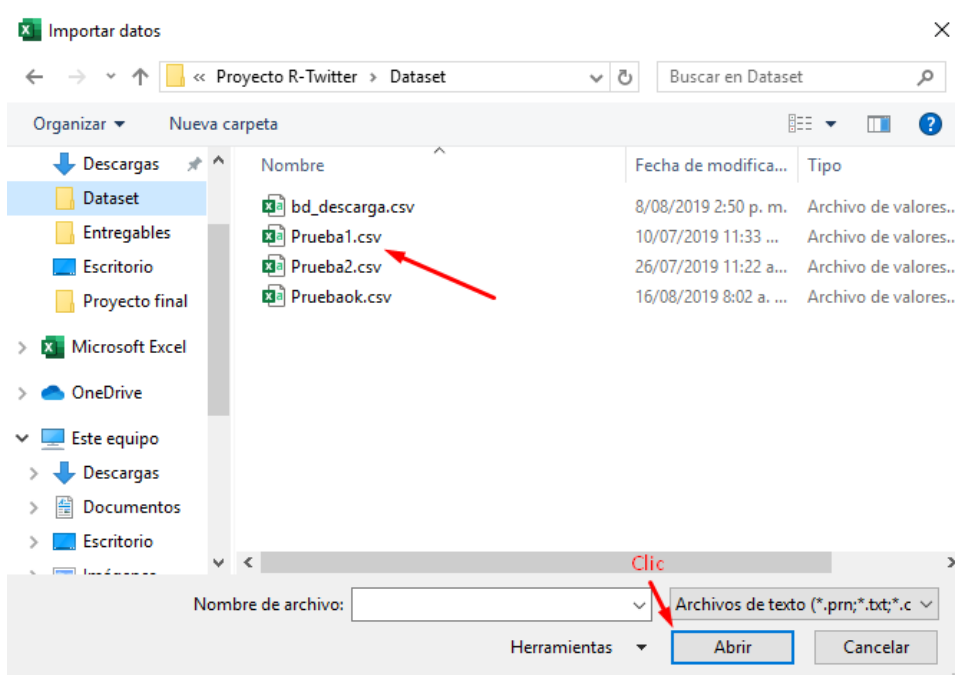
Desde la opción Datos -> Obtener datos -> Desde el texto/CSV. (Fig. 30)

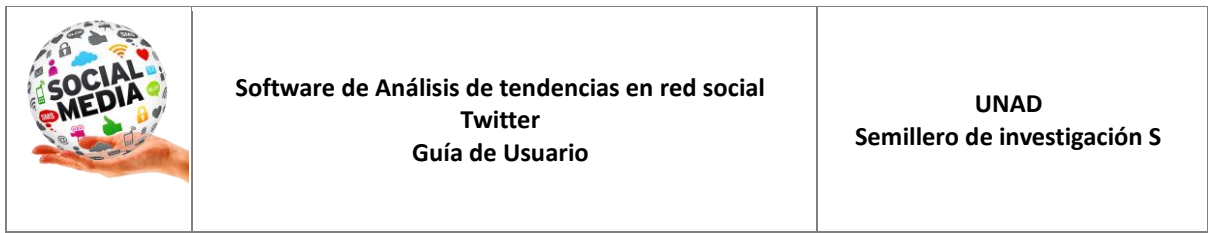
	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--



**Fig. 52. Ruta para importar datos CSV a Microsoft Excel**  
Adaptado de: (Microsoft, 2019)

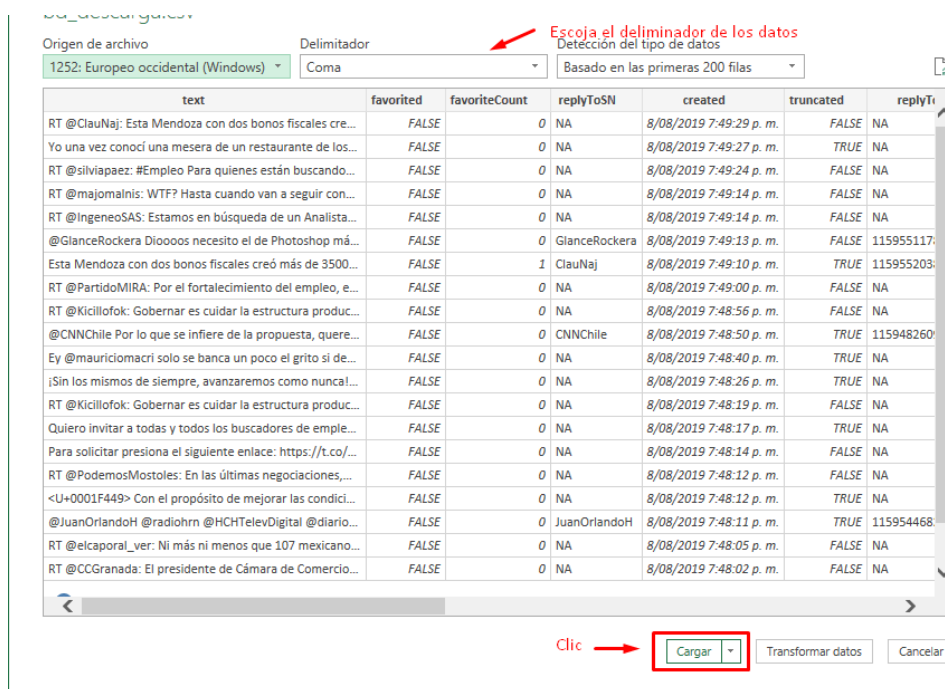
Se abre la ventana para buscar el documento, seleccionarlo y clic en Abrir





**Fig. 53. Pasos en ventana para importar documento**  
Adaptado de: (Microsft, 2019)

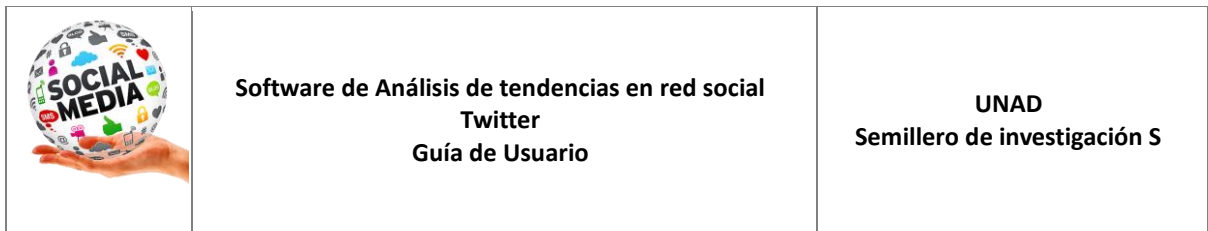
En la siguiente ventana realice lo siguiente:



**Fig. 54. Ventana de carga y transformación de datos**  
Adaptado de: (Microsft, 2019)

Finalice con clic en cargar, ahora tiene los datos en una tabla excel desde donde puede agregar, quitar y editar columnas y filas.

Aquí debe eliminar datos que no requiera o no sean necesarios , además de agregar la columna sentiment (Fig. 55).



1	text	screenName	retweetCount	sentiment	
2	Programadores(as) de Sistemas, L.A. Sistemas S, #Edo. Miranda (chacao), #Venezuela https://t.co/PKP2uthIQd #Empleo	EmpleateVeSist	0	1	
3	programadores de sistemas, L.A. Sistemas S, #Edo. Miranda (chacao), #Venezuela https://t.co/Vup3qNpM7K #Empleo	EmpleateVeAdmon	0	1	
4	Asistente de Sistemas. - Ecuador Laboral https://t.co/4RiXskDdSL #Guayaquil #Ecuador #Trabajo #EcuadorLaboral... http	EcuadorLaboral	0	1	
5	Administrador/a de Sistemas BBDD#empleo #trabajo #Alicante #trabajoalicanteInscribete aquí <U+0001F449> https://t.	TrabajosAlacant	0	1	
6	TIRADOR DE CABLE PARA INSTALACION DE SISTEMAS DE SEGURIDAD/ALARMAS#empleo #trabajo #Alicante #trabajoalica	TrabajosAlacant	0	1	
7	Especialista de Sistemas, Cadena de Farma, #Caracas, Distrito Capital, #Venezuela https://t.co/OKPMHC42u7 #Empleo	EmpleateVeSist	0	1	
8	Nueva oportunidad de empleo Estudiante de ingeniería industrial / sistemas o administración de empresas... https://t	Accusys_Arg	0	1	
9	Ingeniero de Redes y Sistemas @michaelpage_es #Barcelona #SystemCenter #empleo #tic #it https://t.co/j5qJxTHwSe	tecnoempleo	0	1	
10	Buscamos 2 técnicos de sistemas #mdm con experiencia en #microsoft #intune. Es imprescindible experiencia en otras	Ibermatica_RRHH	0	1	
11	Técnico/a de sistemas Linux#empleo #trabajo #Barcelona #empleobarcelonaInscribete aquí <U+0001F449> https://t.co/	Jobs_in_BCN	0	1	
12	Oferta de #trabajo! En @CAI_Sistemas buscamos #programador/a #desarrollador/a #Delphi en Vigo. <U+27A1><U+FE0F> CAI_Sistemas		0	1	
13	PANEL Sistemas Informáticos selecciona para León Analista Programador Java, C++, ADA (3 vacantes)... https://t.co/O6fi	leonenredcom	1	1	
14	Oferta de empleo Responsable de Sistemas IT https://t.co/gRfXlhSDdO	Iwantic	0	1	
15	Técnico/a sistemas de gestión integrado #MedioAmbiente y calidad - @LBconsultores1 #Barcelona... https://t.co/Ks6fk	ISMedioambiente	0	1	
16	La primera de las normas está relacionada con el empleo de los sistemas inalámbricos de alta velocidad en las banda...	CruzYomil	0	1	
17	Buscamos un Técnico de Sistemas Junior Unix para importante empresa. ¡Mándanos tu CV a seleccion@mensoft.es! #rr	MensoftC	0	1	
18	La primera de las normas está relacionada con el empleo de los sistemas inalámbricos de alta velocidad en las banda...	radiorebeldecu	10	1	
19	ingeniero de sistemas, Alimentos Dagan, #Caracas, Distrito Capital, #Venezuela https://t.co/nWOEWUIYnn #Empleo	EmpleateVeSist	0	1	
174	Nueva oportunidad de empleo ADMINISTRADOR DE SISTEMAS WINDOWS: https://t.co/IH6BJXIMJq #Empleo #Trabajo h	KaunisJaky	0	1	
175	Para algunos ingenieros de sistemas la programación es un karma, para otros es una obligación, para otros es una ne...	FernandoNaviaB1	0	-1	
176	<U+0001F6AB> El 72% del empleo doméstico es informal (no tiene aportes, vacaciones ni aguinaldo)Necesitamos:<U+00	mayrasmendoza	37	-1	
177	"Es necesario racionalizar el uso de la Justicia por los ciudadanos. debiendo incrementarse el empleo de sistemas a...	hijuecesAPM	1	-1	

Fig. 55. Documento Excel con columna sentiment

Adaptado de: (Microsoft, 2019)

### 4.3 ¿Como guardar mi documento Excel en formato CSV?

Para ello debe ir a la opción Archivo -> Guardar como (Fig. 56 y 57)

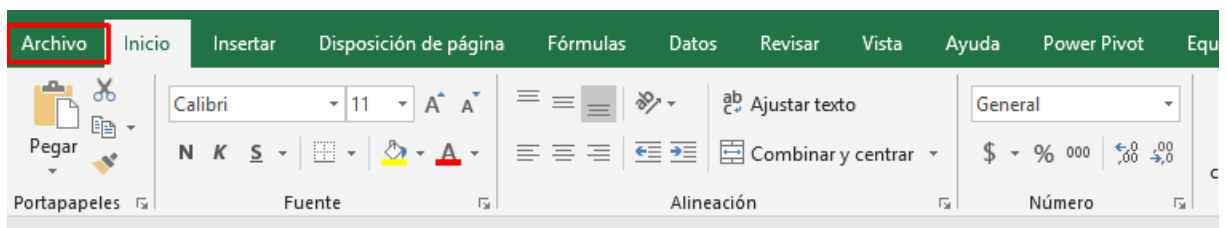
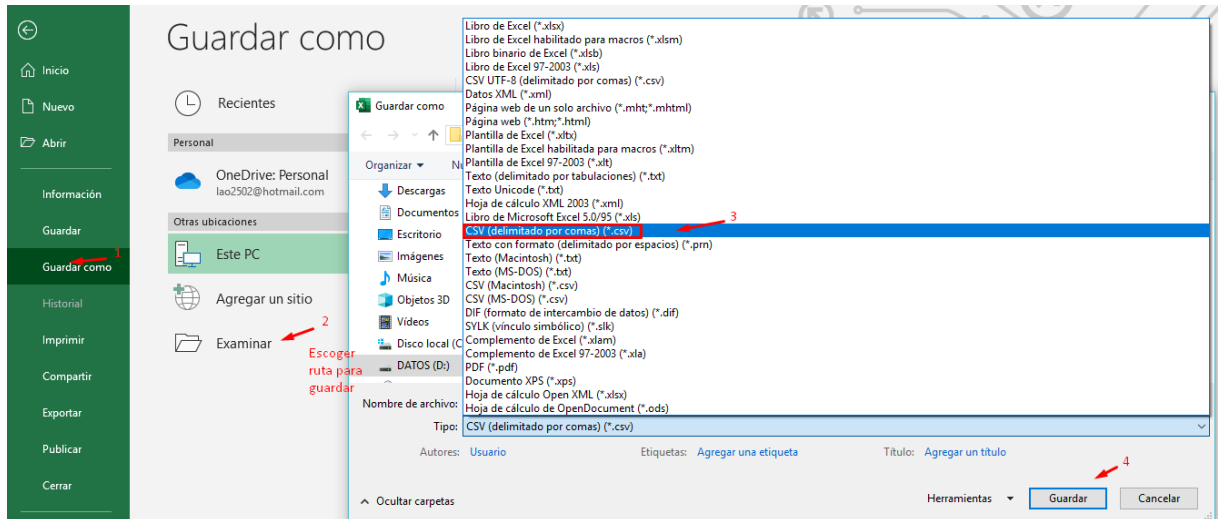
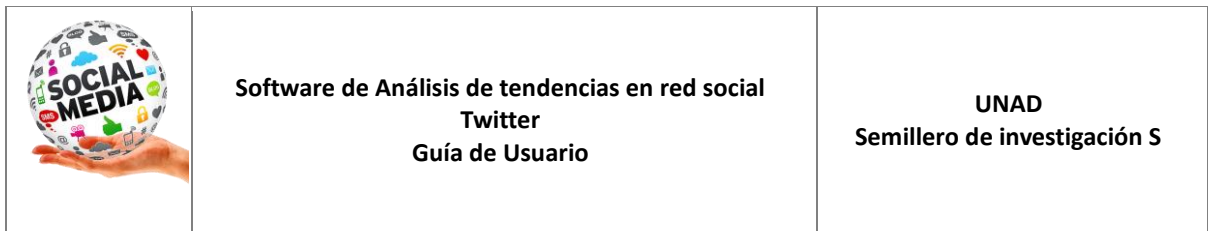


Fig. 56. Cinta de opciones Microsoft Excel


Adaptado de: (Microsoft, 2019)

Y proceder como sigue



**Fig. 57. Ventana importar datos en Microsoft Excel**  
Adaptado de: (Microsoft, 2019)

Ahora tiene un documento en formato CSV listo en la ruta que eligio para guardar.

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## 5 ANEXOS

Se anexan otras referencias de interés para la comprensión del sistema.

### ¿Qué es una API?

En un nivel superior, las API son la forma en que los programas informáticos "hablan" entre sí para solicitarse y enviarse información. Para esto, se le permite a la aplicación del software que llame a lo que se denomina **punto de conexión**: una dirección que corresponde a un tipo específico de información que proporcionamos (generalmente, los puntos de conexión son únicos, como los números telefónicos). Twitter permite acceder a partes de nuestro servicio mediante las API para permitirles a las personas crear software que se integre con Twitter, como una solución que ayuda a una empresa a responder a la opinión del cliente en Twitter.


Los datos de Twitter son únicos y se extraen a partir de datos de la mayoría de las otras plataformas sociales porque reflejan información que los usuarios deciden compartir de forma pública. Nuestra plataforma de API ofrece acceso amplio a los datos de Twitter que los usuarios han decidido compartir con el mundo. Twitter también es compatible con API que permiten a los usuarios administrar su propia información de Twitter que no es pública (como los Mensajes directos) y brindarla a los desarrolladores que ellos han autorizado a administrarla.

### Términos y condiciones de Twitter

#### Acceso a datos de Twitter

Cuando alguien quiere acceder a nuestras API, tiene que registrar una aplicación. De forma predeterminada, las aplicaciones solo pueden acceder a información pública en Twitter. Ciertos puntos de conexión, como aquellos responsables del envío o recepción de Mensajes directos, requieren que otorgues permisos adicionales antes de poder acceder a tu información. Estos permisos no se otorgan de forma predeterminada; tú eliges según cada aplicación si deseas otorgar este acceso y puedes controlar todas las aplicaciones autorizadas en tu cuenta.

Las API de Twitter incluye una gran variedad de puntos de conexión, que se dividen en cinco grupos principales:

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social</b> <b>Twitter</b> <b>Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD</b> <b>Semillero de investigación S</b></p>
---	--	--

## **Cuentas y usuarios**

Les permitimos a los desarrolladores administrar de forma programática el perfil y la configuración de una cuenta, silenciar o bloquear usuarios, administrar usuarios y seguidores, solicitar información sobre actividad de una cuenta autorizada y mucho más. Estos puntos de conexión pueden ayudar a servicios ciudadanos, como el Commonwealth of Virginia's Department of Emergency Management (Departamento de administración de emergencias del Commonwealth de Virginia) que ofrece información a los residentes sobre respuestas a emergencias y alertas de emergencias.


## **Tweets y respuestas**

Los Tweets y las respuestas públicas están a disposición de los desarrolladores, y les permitimos publicar Tweets a través de nuestra API. Los desarrolladores pueden acceder a los Tweets al buscar palabras clave específicas o solicitar una muestra de Tweets de cuentas específicas.

Las ONG, como la ONU, usan estos puntos de conexión para identificar, comprender y contrarrestar información errónea sobre iniciativas de salud pública. Por ejemplo, en Indonesia, existían rumores permanentes de que las vacunas contenían productos porcinos o provocaban infertilidad. Al comprender cómo se iniciaron los rumores y cómo se divulgaron, la ONU organizó un equipo en terreno para ayudar a disipar estos mitos, que causaban una gran inquietud en esta nación principalmente musulmana. De la misma forma, ayudamos a los investigadores a notar señales tempranas de síntomas de brotes de enfermedades y supervisar su contagio. Recientemente, un equipo en la Northeastern University desarrolló una técnica nueva para hacer seguimiento de la gripe mediante los datos de Twitter que prevé brotes hasta con seis semanas de anticipación, mucho antes que otros modelos, sin perder precisión.

## **Mensajes directos**

Nuestros puntos de conexión de Mensajes directos brindan acceso a las conversaciones por DM de usuarios que han otorgado permiso de forma explícita a una aplicación específica. No vendemos Mensajes directos. Nuestras API de DM proporcionan acceso limitado a desarrolladores para crear experiencias personalizadas en Twitter, como la campaña de Wendy's: March Madness bracket builder. En el caso de las cuentas que son de las empresas o cuentas que administran, pueden crear estas experiencias de diálogos en función de personas o chatbots para comunicarse directamente con clientes para ofrecerles atención, marketing o experiencias de interacción con la marca.

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## **Anuncios**


Ofrecemos un conjunto de API para permitirles a los desarrolladores, como Sprinklr, ayudar a las empresas a crear y administra automáticamente campañas de anuncios en Twitter. Los desarrolladores pueden usar Tweets públicos para identificar temas e intereses, y así brindar a las empresas herramientas para ejecutar campañas publicitarias a fin de alcanzar las diversas audiencias en Twitter.

## **Herramientas y SDK del editor**

Proporcionamos herramientas para desarrolladores y editores de software que permiten integrar las cronologías de Twitter, el botón para compartir y otros contenidos de Twitter en las páginas web. Estas herramientas les permiten a las marcas incluir conversaciones públicas en vivo de Twitter en su experiencia web, de forma de que los clientes puedan compartir fácilmente información y artículos de sus sitios.

Puedes obtener más información sobre nuestras API y los detalles de cada punto de conexión en nuestra documentación para desarrolladores.


En todas nuestras API y productos de datos, nos hacemos responsables de manera seria de proteger los datos de los usuarios. Mantenemos directivas y procesos estrictos para evaluar la forma en que los desarrolladores usan los datos de Twitter y restringimos el uso inadecuado de estos datos. Cuando detectamos que un desarrollador infringe nuestras directivas, tomamos la medida adecuada, que puede incluir la suspensión y terminación del acceso a las API de Twitter y productos de datos. (Twitter, 2019)

	<p align="center"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p align="center"><b>UNAD Semillero de investigación S</b></p>
---	---	--


## 6 GLOSARIO

Aquí encuentra la definición de todos los términos utilizados, y de interés para la comprensión del sistema

<b>Término</b>	<b>Descripción</b>
ARS	El análisis de redes sociales (ARS) es una metodología de investigación que ha venido cobrando gran importancia entre la comunidad académica, científica y de la sociedad civil, entre otros, por la posibilidad que ofrece de entender estructuras sociales que componen la sociedad. A partir de esta metodología, es posible comprender múltiples procesos sociales, económicos, empresariales, culturales, políticos, ambientales, comunitarios, etc. (Gómez, 2006).
Red social	Espacio donde los individuos interactúan, se comunican, coinciden, colaboran etc., a través de diversos procesos o acuerdos, que pueden ser bilaterales o multilaterales; de este modo la estructura que emerge de la interrelación se traduce en la existencia de una red social. Las redes sociales son, por tanto, conjuntos de relaciones sociales o interpersonales que ligan individuos u organizaciones en “grupos. (Sanz, 2003)
Clique	Es un subgrupo de una red en que sus actores están más cercanos y unidos que el resto de los actores de la red. Formalmente, un clique es un subgrupo de actores en el cual están presentes todos los vínculos posibles entre éstos.
N-clique	Es un subgrupo en el que todos los actores están separados a lo más por una distancia de N de los demás actores.
Puntos de corte	Son nodos que, si se remueven, dividen la red en dos o más componentes.
Twittero	Es cada usuario registrado. Se representa con @NombreDelUsuario.
Tweet (tuit)	Es cada uno de los mensajes que se publica. Recordar que cada uno de ellos contiene hasta 280 caracteres (se amplió de 140 caracteres en 2018) sin contar el material multimedia que se incluya en los contenidos.
Time Line (línea de tiempo)	Es la parte de una cuenta en la que se puede ver, por orden cronológico, los mensajes de los usuarios que determinada persona sigue.
Retweet (RT)	Es la republicación de un tweet lanzado por otro usuario.
Follower (seguidor)	Es el usuario que sigue tu cuenta. Este puede ver todos los tweets que usted publica.


	<b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b>	<b>UNAD Semillero de investigación S</b>
---	---	--

Following (siguendo a)	Es la acción de seguir a otro usuario dentro de esa red.
Lista	Es un listado que se puede configurar con tus cuentas favoritas. Puede crear la cantidad de listas como desee y otorgarles un nombre.
Me gusta	Está representado por un icono de corazón. Este se cliquea si a un usuario le gusta un tweet
Hashtag	El rey de Twitter. Se representa con un icono de almohadilla (#) y permite añadir tras él los términos que el usuario quiera. Se utiliza para facilitar búsquedas. Por ejemplo, usando #WordPress en el buscador, se puede encontrar un listado de los usuarios que han utilizado ese término en sus tweets
Trending Topic	Son los temas más comentados del momento, es decir, las palabras con más menciones de la red social en un determinado periodo de tiempo. (Webempresa, 2018)

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## 7 REFERENCIAS

Referencia	Título
Camargo, I. 2017. Distribución de frecuencias, histograma, polígono de frecuencias y ojiva en R. Recuperado de <a href="https://rstudio-pubs-static.s3.amazonaws.com/255905_642e1a8e4c3a4879a4856573a9b928de.html">https://rstudio-pubs-static.s3.amazonaws.com/255905_642e1a8e4c3a4879a4856573a9b928de.html</a>	Distribución de frecuencias, histograma, polígono de frecuencias y ojiva en R
Martos, G. 2014. Nubes de Palabras en R. Recuperado de <a href="https://rpubs.com/gabrielmartos/wordcloudnwsppr">https://rpubs.com/gabrielmartos/wordcloudnwsppr</a>	Nubes de Palabras en R
Gómez, H. 2006. ARS como metodología de investigación social. Recuperado de <a href="https://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/view/1374">https://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/view/1374</a>	ARS como metodología de investigación social
Sanz, L. 2003. Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes. Recuperado de <a href="http://ipp.csic.es/sites/default/files/content/workpaper/2003/dt-0307.pdf">http://ipp.csic.es/sites/default/files/content/workpaper/2003/dt-0307.pdf</a>	Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes
Correa, J. 2002. Gráficos estadísticos con R. Universidad Nacional – Sede Medellín. Recuperado de <a href="https://cran.r-project.org/doc/contrib/grafi3.pdf">https://cran.r-project.org/doc/contrib/grafi3.pdf</a>	Gráficos estadísticos con R
Trending Topic. Tomado de <a href="https://www.webempresa.com/blog/que-es-twitter-como-funciona.html">https://www.webempresa.com/blog/que-es-twitter-como-funciona.html</a>	Trending Topic

	<p style="text-align: center;"><b>Software de Análisis de tendencias en red social Twitter Guía de Usuario</b></p>	<p style="text-align: center;"><b>UNAD Semillero de investigación S</b></p>
---	--	---

## 8 BIBLIOGRAFÍA

- Brand, E., & Gómez, H. (2006). *ARS COMO METODOLOGÍA DE INVESTIGACIÓN SOCIAL*. Obtenido de <https://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/view/1374>
- Correa, J. C., & González, N. (2002). *Gráficos Estadísticos con R*. Obtenido de <https://cran.r-project.org/doc/contrib/grafi3.pdf>
- Martos, G. (10 de 2014). *Nubes de Palabras en R: otra forma del leer el periódico*. Obtenido de <https://rpubs.com/gabrielmartos/wordcloudnwsppr>
- Microsoft. (2019). Imágenes de apoyo.
- Ortiz Palma, L. A. (20 de 08 de 2019). Figuras de apoyo visual ARS Twitter. Popayán, Cauca, Colombia.
- Sanz, L. (2003). *Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes*. Obtenido de <http://ipp.csic.es/sites/default/files/content/workpaper/2003/dt-0307.pdf>
- Twitter. (2019). *Información sobre las API de Twitter*. Obtenido de <https://help.twitter.com/es/rules-and-policies/twitter-api>
- Twitter. (2019). *Twitter Developers*. Obtenido de <https://developer.twitter.com/>
- Webempresa. (2018). *¿Qué es twitter? ¿Cómo funciona? ¿Cómo puedo usarlo para mi organización?* Obtenido de <https://www.webempresa.com/blog/que-es-twitter-como-funciona.html>
- Yureicy, I. C. (4 de 03 de 2017). *Distribución de frecuencias, histograma, polígono de frecuencias y ojiva en R*. Obtenido de [https://rstudio-pubs-static.s3.amazonaws.com/255905\\_642e1a8e4c3a4879a4856573a9b928de.html](https://rstudio-pubs-static.s3.amazonaws.com/255905_642e1a8e4c3a4879a4856573a9b928de.html)



## **PROGRAMA DE INGENIERIA DE SISTEMAS SEMILLERO SIAVA**

### **Manual Técnico Sistema de Análisis de datos ARS - SIAVA**

#### **INTRODUCCIÓN**

Este documento busca brindar las instrucciones necesarias para la implementación de la herramienta tecnológica ARS SIAVA, la cual es una aplicación web interactiva para análisis, recopilación y visualización estadística de redes sociales, además permite conocer tendencias, a partir del análisis de sentimientos de la información recopilada en redes sociales como Twitter. Este es un proyecto desarrollado dentro del semillero de investigación SIAVA CEAD Popayán, promovido por la Universidad Nacional Abierta y a Distancia - UNAD.

**TABLA DE CONTENIDO**

Introducción.....	1
Tabla de contenido .....	2
1. Prerrequisitos de instalación del sistema .....	3
2. Pre-requisitos de instalación del sistema en el cliente.....	3
3. Frameworks y estándares.....	3
4. Diagrama de actividades.....	5
5. Scripts de instalación .....	6
5.1 Correr desde una computadora personal.....	6
5.1.1 Script interfaz de usuario .....	6
5.1.2 Script del servidor .....	12
5.1.3 Script Análisis de sentimientos .....	16
5.1.4 Ejecución.....	20
5.2 Correr desde un servidor.....	20
6. Referencias .....	21

## 1. PRERREQUISITOS DE INSTALACIÓN DEL SISTEMA

- Sistema operativo Windows 7 o superior/ GNU/Linux/ Macintosh/ Unix.
- Privilegios de administrador para instalar y ejecutar utilidades
- Conexión de red para recopilar datos a través de la red.
- Instalar R
- Instalar RStudio
- Instalar RsTools

*<Sistema operativo de los servidores de aplicaciones y base de datos, marca y versión de la base de datos, marca y versión de los servidores de aplicaciones, navegador, configuraciones de seguridad, etc., para que el sistema se pueda ejecutar correctamente>*

## 2. PRE-REQUISITOS DE INSTALACIÓN DEL SISTEMA EN EL CLIENTE

No requiere instalación solamente disponer de un navegador web, se recomienda las últimas versiones de Chrome, Firefox, Internet Explorer y Safari. Si utilizas un navegador web diferente, algunas características podrían no funcionar.

Aquí puede obtener la última versión de tu navegador.

Descarga Google Chrome <https://www.google.com/chrome>

Descarga Firefox <https://www.mozilla.org/firefox/>

Descarga Safari <https://www.apple.com/safari>

Descarga Internet Explorer <http://windows.microsoft.com/en-US/internet-explorer/download-ie>

## 3. FRAMEWORKS Y ESTÁNDARES

- Instalar Versión de R 3.6.1. Para ello se debe ir al siguiente link <https://cran.r-project.org/bin/windows/base/R-3.6.1-win.exe>
- Instalar Versión de RStudio 1.2. Para ello se debe ir al siguiente link <https://www.rstudio.com/products/rstudio/download/#download>
- Instalar Versión de RsTools 3.5. Para ello se debe ir al siguiente link <https://cran.r-project.org/bin/windows/Rtools/Rtools35.exe>

- La aplicación corre sobre entorno web a través de la librería Shiny para R. Shiny es un paquete R que facilita la creación de aplicaciones web interactivas directamente desde R.
- Instalar las siguientes librerías:

`library(shiny)` # Paquete R que facilita la creación de aplicaciones web interactivas

`library(twitteR)` # Paquete e R que proporciona acceso a la API de Twitter

`library(ROAuth)` # Proporciona una interfaz para la especificación OAuth 1.0 que permite a los usuarios autenticarse a través de OAuth en el servidor de su elección.

`library(httr)` # Herramientas para trabajar con URL y HTTP

`library(tm)` # Minería de texto

`library(SnowballC)` # Reduce las palabras a su raíz

`library(caTools)` # Herramientas estadísticas de ventanas móviles, GIF, Base64, ROC AUC, etc.

`library(caret)` # Ordenar datos

`library(e1071)` # Paquete para machine learning (Aprendizaje de máquina)

`library(plyr)` # Herramientas para dividir, aplicar y combinar datos

`library(magrittr)` # need to run every time you start R and want to use %>%

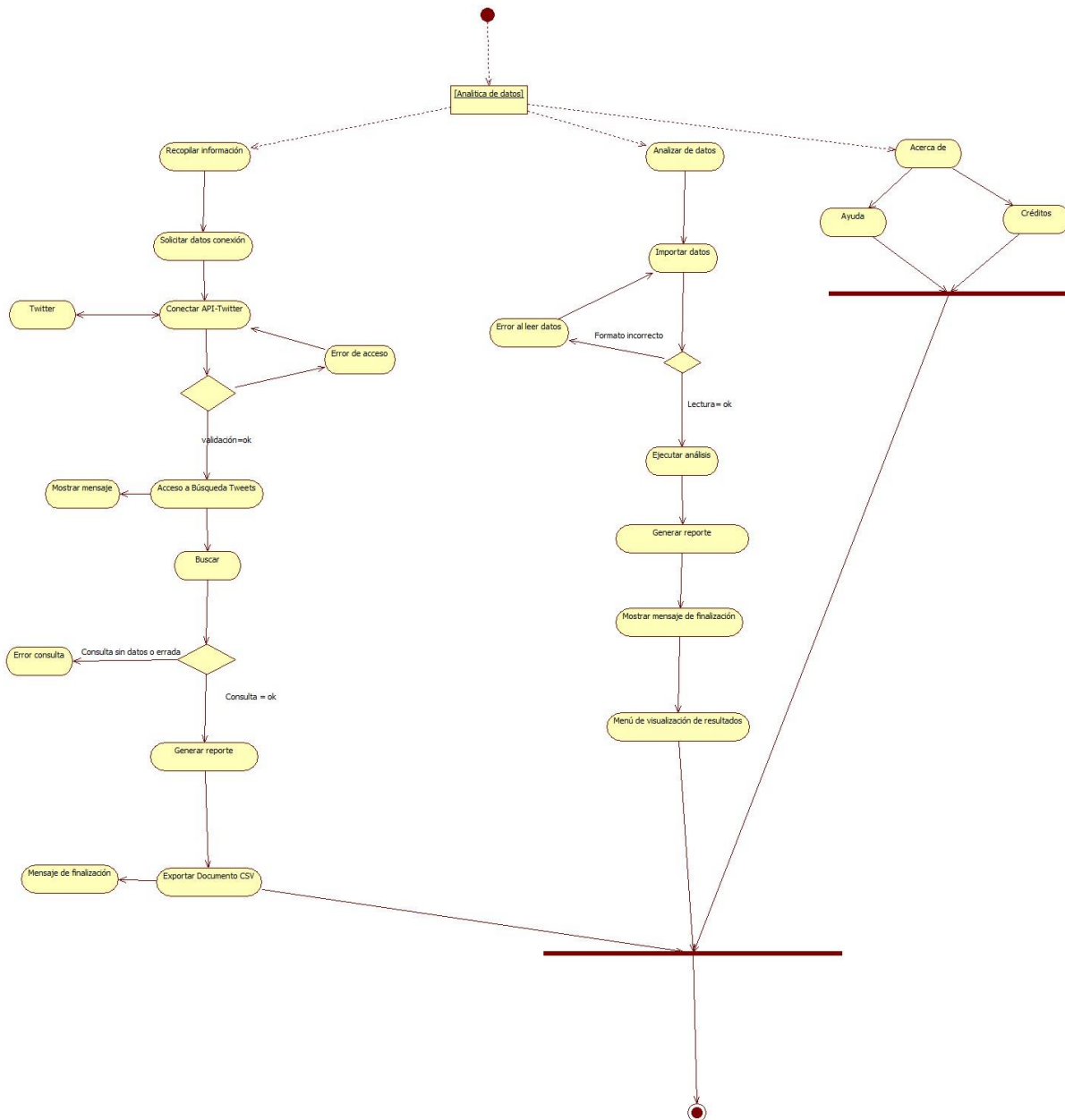
`library(dplyr)` # alternative, this also loads %>%

`library(wordcloud)` # Crear nubes de palabras, visualizar diferencias y similitudes entre documentos.

`library(RColorBrewer)` # Crea paletas de colores para aplicar a mapas temáticos y textos.

`library(ggplot2)` # Paquete de visualización de datos. Crea gráficas

#### 4. DIAGRAMA DE ACTIVIDADES



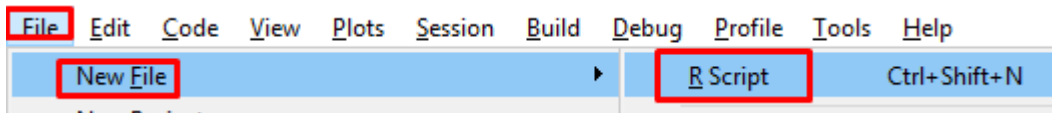
**Fig. 1 Diagrama de actividades para la aplicación ARS SIAVA**
  
 Fuente: (Ortiz Palma, 2019)

## 5. SCRIPTS DE INSTALACIÓN

### 5.1 Correr desde una computadora personal

Para correr de forma local es su computadora haga lo siguiente:

Crear una carpeta y en esta generar los scripts (Ver Figura 1) que se detallan a continuación, copiar el contenido en cada script como sigue.



**Figura 1. Ruta para crear nuevo script en R**

Fuente: (Rstudio, 2019)

#### 5.1.1 Script interfaz de usuario

Cree un nuevo script y copie el texto de abajo y guárdelo con el nombre: **“UI. R** “sin las comillas

```
library(shiny)

shinyUI(fluidPage(

  tags$img(src="Banner_ARS.png",width="100%", height="150px"),
  titlePanel("Software de Análisis de tendencias en redes sociales"),

  # crea una barra de navegación de nivel superior

  navbarPage(

    tags$head(

      tags$style(type="text/css",
        ".test_type {color: blue;
```

```
font-size: 20px;  
font-style: italic;}"  
)  
  
,  
  
# Crea un menu con varias pestañas - Datos  
#crea encabezados de selección de menú  
  
navbarMenu("Datos",  
  
# Crea una pestaña para ir a Recopilar datos  
  
tabPanel("Recopilar datos",  
  
tags$h2("Establecer conexión"),  
sidebarPanel(  
tags$strong("Todos los campos son obligatorios"),  
br(),  
tags$em(" *Ingrese datos de enlace con la app-witter y clic en conectar*"),  
br(),  
br(),  
textInput("clave", "Ingrese clave (consumer key)"),  
textInput("secreta", "Ingrese clave secreta (consumer secret)"),  
textInput("token", "Ingrese token de acceso (token access)"),  
textInput("stoken", "Ingrese token secreto de acceso (secret access)"),  
  
actionButton("conectar", "Conectar Twitter"),  
  
# Crear posición para el texto del lado del servidor  
  
textOutput("estado")  
),  
  
tags$h2 ("Recopilación de datos"),  
  
# formulario para búsqueda de datos  
  
sidebarPanel(  
  
tags$em("***Antes debe ingresar claves de acceso y conectar  
con la app de Twitter***"),
```

```

        textInput("búsqueda", "Ingrese palabra a buscar"),
        actionButton("Buscar", "Buscar"),
        textOutput("Consulta")
    ),
    mainPanel(
        tableOutput("contenido")
    )
),
# # Crea una pestaña para ir a Importar y analizar datos

tabPanel("Analizar",

    tags$h1 ("Manejo de datos"),

    # Widget para carga de datos externos formato csv
    sidebarLayout(
        sidebarPanel(
            fileInput("file1", "Subir un archivo CSV",
                accept = c(
                    "text/csv",
                    "text/comma-separated-values,text/plain",
                    ".csv")
            ),
            tags$br(),

            radioButtons(
                "sep", "Separador",
                choices = c(
                    "Coma" = ",",
                    "Punto y coma" = ";",
                    "Tabulación" = "\t"
                ),

                selected = ";"
            ),
            tags$br(),
            radioButtons(
                "quote", "Comillas",
                choices = c(
                    "Ninguna" = "",
                    "Comilla doble" = ""

```

```

        "Comilla sencilla" = ""
    ),
    selected = "
),
tags$br(),

checkboxInput("header", "Encabezados", TRUE),

#Crea boton para ejecutar análisis de datos

actionButton("analizar", label= "Ejecutar análisis"),

#salida de texto estado de análisis de datos
textOutput("analizar")

),

# Crear posición para el texto del lado del servidor

mainPanel(
    tableOutput("fileload")
)

)

),

# Crea un menú con varias pestañas - Resultados

navbarMenu("Resultados",

    # Crea una pestaña para ir a Histograma

    tabPanel("Histograma",

        sidebarLayout(
            sidebarPanel(
                sliderInput("rango", "Seleccione número de contenedores", min = 5, max =
50,value = 25)

```

```
    ),  
    mainPanel(  
      plotOutput("plot")  
    )  
  )  
,  
  
# Crea una pestaña para ir a Nube de palabras  
  
tabPanel("Nube de palabras",  
  sidebarLayout(  
    sidebarPanel(  
      hr(),  
      sliderInput("frec",  
        "Frecuencia mínima:",  
        min = 1, max = 50, value = 15),  
      sliderInput("max",  
        "Numero max. palabras:",  
        min = 1, max = 300, value = 100)  
    ),  
    # muestra nube de palabras  
    mainPanel(  
      plotOutput("plot2")  
    )  
  )  
,  
  
# Crea una pestaña para ir a Tabla de frecuencias  
  
tabPanel("Tabla de frecuencias", tableOutput("datafrec")),  
  
# Crea una pestaña para ir a Resumen  
  
tabPanel("Resumen", verbatimTextOutput("summary"))  
  
,  
  
# Crea un menú con varias pestañas - Acerca de  
  
navbarMenu("Acerca de",  
  
  # Crea una pestaña para ir a ayuda
```

```

tabPanel("Ayuda",
  h1("Solución de problemas y ayuda"),

  em('A continuación usted encontrara los enlaces de apoyo:

    De clic sobre la opción que desee. '),

  h4(HTML(paste("Documentación y problemas comunes",
a(href="https://www.unad.edu.co/", "Ver")))),
  h4(HTML(paste("Contacto:", a(href="mailto:lao2502@hotmail.com",
"Comenzar")))),
),

# Crea una pestaña para ir créditos

tabPanel("Créditos",
  mainPanel("",

    img(src="logo_unad.png"),

    h2('Herramienta tecnológica ARS SIAVA'),
    br(),

    p("ARS SIAVA, es una aplicación web interactiva para análisis,
recopilación y visualización estadística de redes sociales, el cual permite ver tendencias, a
partir del análisis de sentimientos de la información recopilada. Este es un proyecto
desarrollado dentro del semillero de investigación SIAVA CEAD Popayán, promovido por la
Universidad Nacional Abierta y a Distancia - UNAD."),

    br(),
    strong ("Director proyecto: Hermes Mosquera - Msc. Ingeniero de
sistemas"),br(),
    strong ("Desarrollado por: Luis Alberto Ortiz Palma - Ingeniería de
sistemas"),

    br(),
    br(),

    img(src="Semilleros_unad.png", height=150, weight=360)

  )
)

```

```
)  
)  
)  
)  
)
```

### 5.1.2 Script del servidor

Cree un nuevo script y copie el texto de abajo y guárdelo con el nombre: “**server. R**” sin las comillas

```
library(shiny)  
library(twitteR)  
library(ROAuth)  
library(httr)  
  
shinyServer(function(input, output) {  
  
  # Crear acción cuando se hace clic en actionButton de conexión  
  
  observeEvent(input$conectar,  
  
    {  
  
      consumer_key <- input$clave  
      consumer_secret <- input$secret  
      access_token <- input$token  
      access_secret <- input$stoken  
  
      output$estado <- renderText({  
        if (is.null(consumer_key, consumer_secret, access_token, access_secret))  
          return(NULL)  
      })  
    }  
  )  
})
```

```
})

setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

# imprime aviso tras completar el proceso anterior
output$estado = renderText("Conexión realizada")
}

)

# Crea accion buscar para pestaña de Recopilación de datos
observeEvent( input$Buscar,

  {
    #Establezco el directorio de trabajo

    setwd("D:/Estudios/UNAD/Trabajo de grado/Proyecto final/Proyecto R-
    Twitter/analisis_sentimientos/Interfaz_usuario/datos")

    palabra <-input$busqueda

    #Seleccionamos tweets o la búsqueda a realizar
    consulta <- searchTwitter(palabra, n=1000)

    #Creando un dataset de los datos obtenidos

    Lista=twitteR::twListToDF(consulta)

    #Exportando datos

    write.csv(Lista, file="busqueda.csv",row.names = F)
```

```
output$contenido <- renderTable(  
  
  Lista  
  
  )  
output$Consulta = renderText("Proceso completado satisfactoriamente")  
  
  })  
  
# Ejecutar análisis de datos mediante evento clic del botón analizar (ejecutar análisis)  
observeEvent( input$analizar,  
  
  {  
    setwd("D:/Estudios/UNAD/Trabajo de grado/Proyecto final/Proyecto R-  
Twitter/analisis_sentimientos/Interfaz_usuario")  
  
    source("AnalisisTwitter_TI.R")  
  
    # imprime aviso tras completar el proceso anterior  
    output$analizar = renderText("Proceso completado satisfactoriamente")  
  
  }  
  
  )  
  
#Establezco el directorio de trabajo  
  
  setwd("D:/Estudios/UNAD/Trabajo de grado/Proyecto final/Proyecto R-  
Twitter/analisis_sentimientos/Interfaz_usuario/datos")  
  
# Importación y análisis de datos  
  
  output$fileload <- renderTable({  
    # input $ file1 será NULL inicialmente. Después de que el usuario seleccione  
    # y cargue un archivo, este será un dataframe con columnas "nombre",
```

```
# "tamaño", "tipo" y "ruta de datos". La columna 'ruta de datos'  
# contendrá los nombres de archivos locales donde pueden  
# encontrarse los datos  
  
indata <- input$file1  
  
if (is.null(indata))  
  return(NULL)  
  
datafile<- input$file1$datapath  
  
dataup<- read.csv(datafile,  
  header = input$header,  
  sep = input$sep,  
  quote = input$quote  
  
)  
#write.csv2(dataup, file = "importado.csv", sep = input$sep, quote = FALSE, row.names =  
FALSE)  
  
# #Exportando datos  
write.csv2(dataup, file="importado.csv", quote = FALSE, row.names = FALSE)  
  
output$fileload <- renderTable(  
  dataup  
)  
  
})  
  
#{  
#  
# # Puede acceder al valor del widget con input$action, por ejemplo  
# output$value <- renderPrint({ input$analizar })  
#  
# }  
  
# para mostrar los datos de frecuencias positivas en la página "frecuencias"  
output$datafrec <- renderTable({
```

```
positivas
})

# para mostrar el histograma en la página "Histograma"

output$plot <- renderPlot({
  hist(positivas$frecuencia, main = "Frecuencia de palabras positivas", xlab = "Palabras
más utilizadas", ylab = "Frecuencia", col = "light blue", breaks=input$rancho )
})

# para mostrar nube de palabras en la página "Nube de palabras"

output$plot2 <- renderPlot({
  wordcloud(positivas$Palabras, positivas$frecuencia, random.order = FALSE, colors =
brewer.pal(8,"Dark2"), max.words = 300)
})

# para la visualización de estadísticas de resumen del conjunto de datos en "página Resumen"
output$summary <- renderPrint({
  summary(positivas)
})
})
```

### 5.1.3 Script Análisis de sentimientos

Cree un nuevo script y copie el texto de abajo y guárdelo con el nombre: “**AnalisisTwitter. R**”  
“sin las comillas

```
library(tm) # para minería de texto
library(SnowballC) # para reducir una palabra a su raíz
library(caTools)
library(caret) # ordenar datos
library(e1071) # machine learning
library(plyr)
library(magrittr) # Necesita correrla cada vez que usted inicie R and quiera usar %>%
```

```
library(dplyr) # alternative, this also loads %>%
library(wordcloud)
library(RColorBrewer)
library(ggplot2)

# posicionamiento del directorio de trabajo

setwd("D:/Estudios/UNAD/Trabajo de grado/Proyecto final/Proyecto R-
Twitter/analisis_sentimientos/Interfaz_usuario/datos")

#retorna el directorio sobre el cual se está trabajando

getwd()

importdata <- read.csv("importado.csv", sep =";") #cargar los datos al objeto tweets ("nombre
del archivo.csv", separador)

table(importdata$sentiment) #Cuantos tweets positivos y negativos hay
Corpus= Corpus(VectorSource(importdata$text)) #objeto corpus se le asigna el objeto de tweets
// lee los tweets
length(Corpus) #cuenta las palabras tiene el corpus
content(Corpus[[27]]) #imprime la posición 27

#Preprocesamiento

removeURL <- content_transformer(function(x) gsub("(f|ht)tp(s?)://\\S+", "", x, perl=T)) #
funci?n para eliminar http
Corpus <- tm_map(Corpus, removeURL) # Elimina las palabras que empiezan por "http."
seguidas de cualquier cosa que no sea un espacio)
content(Corpus[[27]]) #imprime la posición 27
Corpus <- tm_map(Corpus, removePunctuation) #quita la puntuación
content(Corpus[[27]]) #imprime la posición 27
Corpus <- tm_map(Corpus, removeNumbers) #quita los números
content(Corpus[[27]]) #imprime la posición 27
Corpus <- tm_map(Corpus, content_transformer(tolower)) #procesamiento cambia las palabras
a minúsculas
content(Corpus[[27]]) #imprime la posición 27
stopwords("spanish")[1:50] #preposiciones etc... [1:50] imprime las 50 primeras /
stopwords("spanish")
Corpus <- tm_map(Corpus, removeWords, c(stopwords("spanish"), "amp","leonesp",
"barcelon","leon","hoy","reaz" )) #remueve las stopwords y otras palabras que se deseen
content(Corpus[[27]]) #imprime la posición 27
```

```
Corpus <- tm_map(Corpus, stemDocument, language="spanish") #acorta las palabras a su raiz  
"devolvieron" "devolver"  
content(Corpus[[27]]) #imprime la posición 27  
Corpus <- tm_map(Corpus, stripWhitespace) #quita los espacios vacíos excesivos  
content(Corpus[[27]]) #imprime la posición 27
```

*#clasificación*

```
frequencies <- DocumentTermMatrix(Corpus) #crear matriz de palabras  
frequencies #imprime atributos de frequencies de la matriz; Documents= Filas -tweets, terms=  
palabras - columnas  
inspect(frequencies[15:20, 5:10]) #zoom de la matriz / posiciones [15:20, 5:10]  
findFreqTerms(frequencies, lowfreq = 50) # Ver Palabras con mayor frecuencias -- iguales o  
mayores a 50  
sparse <- removeSparseTerms(frequencies, 0.995) #quitar palabras que son mencionadas muy  
poco / deja las palabras más usadas  
sparse #imprime sparse -- sparse=palabras que aparecen poco  
tweetsSparse <- as.data.frame(as.matrix(sparse)) # retornar la variable sparse como una base  
de datos en formato R  
colnames(tweetsSparse) = make.names(colnames(tweetsSparse)) #asignar los nombres de cada  
palabra al dataframe  
tweetsSparse$sentiment <- importdata$sentiment #examina con la base de datos de sentimientos  
precargada
```

*#modelo de clasificación*

*#Objetivo: Encontrar un hiperplano  $h$  de dimensión  $(n-1)$  que separe los ejemplos etiquetados con  $-1$  de los etiquetados con  $+1$*

*#con un "margen máximo"  $(P)$ .*

*#los SVM buscan los puntos más cercanos entre varias clases. Estos puntos se llaman "los vectores de soporte"*

*#SVM luego declara que la mejor línea de separación va a ser la línea que divide las dos clases y que al mismo tiempo*

*#maximiza la distancia del hiperplano a los vectores de soporte  $(r)$*

*#Ecuación del hiperplano:  $WX+b=0$*

*#Problema de optimización: Encuentre  $W$  y  $b$  tal que  $P=2/[[W]]$*

*#partir la base de datos en entrenamiento y evaluación*

*# (entrenamiento: enseñarle a la BD como funcionar, evaluación: evaluar como lo hace el modelo y su poder de predicción)*

```
set.seed(12)
```

```
split <- sample.split(tweetsSparse$sentiment, SplitRatio = 0.8) #decide que observaciones se  
van para un lado u otro
```

```
#plitRatio= 0.8 / el 805 se va a entrenamiento
trainSparse = subset(tweetsSparse, split==TRUE) #partición modelo de entrenamiento
TestSparse = subset(tweetsSparse, split==FALSE) #partición modelo de evaluación
table(TestSparse$sentiment)
165/328
#algoritmo de clasificación

# Se debe cargar: library(caret); library(e1071), library(plyr)

SVM <- svm(as.factor(sentiment)~ ., data=trainSparse) # Entrena al modelo soporte vector
machine
summary(SVM) #Describe el modelo
predictSVM <- predict(SVM, newdata = TestSparse)
confusionMatrix(predictSVM,as.factor(TestSparse$sentiment)) #evaluar el desempeño de los
tweets

# ahora se va a crear una nube con las palabras más repetidas

positive <- subset(tweetsSparse, tweetsSparse$sentiment==1) #crea base de datos de tweets
positivos
positive$sentiment <- NULL # Eliminamos la variable sentiment porque no se necesita en la
nube de palabras

positivas <- as.data.frame(colSums(positive)) #generar frecuencias de las palabras positivas --
suma de columnas (cada palabra)
positivas$words <- row.names(positivas) #crea variable que se llame words
colnames(positivas) <- c("frecuencia", "Palabras")
table(positivas)

# Crear una nube de palabras
# Se debe cargar:
# library(wordcloud)
# library("RColorBrewer")

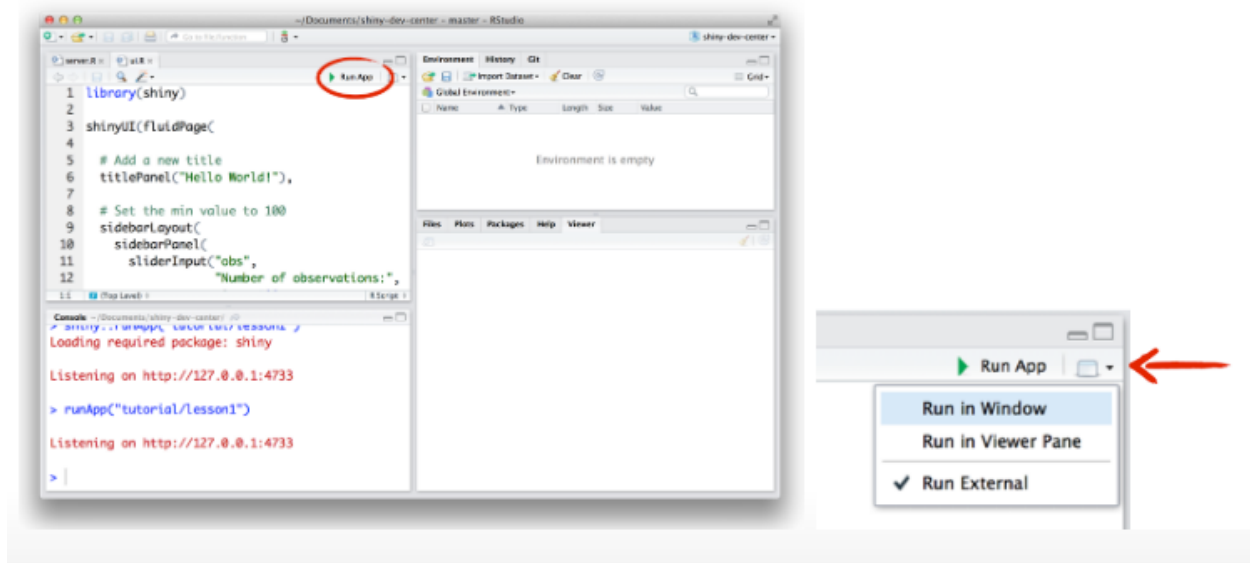
wordcloud(positivas$Palabras, positivas$frecuencia, random.order = FALSE, colors =
brewer.pal(8,"Dark2"), max.words = 150)

# Crear histograma de frecuencia de palabras

hist(positivas$frecuencia, main = "Frecuencia de palabras positivas", xlab = "Palabras más
utilizadas" , ylab = "Frecuencia", col = "light blue", breaks = "Sturges")
```

### 5.1.4 Ejecución

Puede correr la aplicación desde cualquiera de los dos archivos en el editor de texto RStudio, mediante el botón que se marca debajo (RStudio reconoce automáticamente que se trata de un código de una aplicación)



**Figura 2. Ejecutar la aplicación en RStudio**

Fuente: (Oviedo, 2015)

### 5.2 Correr desde un servidor

Shiny Server permite a los usuarios alojar y administrar aplicaciones Shiny en Internet. Shiny es un paquete R que utiliza un modelo de programación reactiva para simplificar el desarrollo de aplicaciones web con tecnología R. Shiny Server puede administrar procesos R que ejecutan varias aplicaciones Shiny en diferentes URL y puertos. El uso de Shiny Server ofrece una variedad de beneficios sobre simplemente ejecutar Shiny en R directamente. (Rstudio, 2019)

Puede conocer el paso a paso en la siguiente dirección web: <https://docs.rstudio.com/shiny-server/#system-requirements>

## 6. REFERENCIAS

- Ortiz Palma, L. A. (20 de 08 de 2019). Figuras de apoyo visual ARS Twitter. Popayán, Cauca, Colombia.
- Oviedo, M. (2015). *Shiny: crear una aplicación web interactiva desde R*. Obtenido de Ejecutar la aplicación (en RStudio)
- Rstudio. (22 de 08 de 2019). reportes R de IDE RStudio.
- Rstudio. (2019). *Shiny Server Professional v1.5.11 Administrator's Guide*. Obtenido de <https://docs.rstudio.com/shiny-server/#system-requirements>