



Trabajo de Fin de Grado

Integración de sistemas para el análisis de
redes sociales utilizando procesamiento de
lenguaje natural

*System integration for the analysis of social networks using
natural language processing*

Constanza Polette León Baritussio

La Laguna, 29 de Junio de 2018

D. **Dagoberto Castellanos Nieves**, con N.I.F. 79234766-L Profesor Contratado Doctor adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor.

CERTIFICA (N)

Que la presente memoria titulada:

“Integración de sistemas para el análisis de redes sociales utilizando procesamiento de lenguaje natural”

ha sido realizada bajo su dirección por Dña. **Constanza Polette León Baritussio**, con N.I.F. X-6337312-F.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 29 de Junio de 2018.

Agradecimientos

A mis padres Sandra y Sandro, por el apoyo y por siempre estar presente en los momentos difíciles. Siempre habéis estado ahí para mí.

A mi director de proyecto Dagoberto, por darme la orientación y motivación, así como las herramientas necesarias para llevar a cabo este proyecto, además de su valiosa ayuda.

A todos los profesores, que embarcaron conmigo estos años universitarios, que gracias a los conocimientos brindados, he podido desarrollar tanto este trabajo como mi carrera profesional.

A mis amigos y compañeros, por haber compartido muchos momentos. No solo habéis estado ahí para apoyarnos mutuamente y superar juntos los obstáculos, sino también por todos esos momentos de alegrías.

Finalmente a Juan Carlos, por creer en mí y hacerme ver a las bases de datos desde otra perspectiva.

Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

Resumen

El objetivo de este trabajo ha sido, desarrollar una integración de sistemas, en concreto un sistema de alerta y un sistema de análisis de sentimiento, orientado al ámbito del turismo.

Actualmente, las redes sociales son un medio en donde se puede acceder a la información, compartir opiniones además de crear contenido. Asimismo, están compuestas por miles de usuarios que interaccionan unos con otros, por lo que se crean millones de datos. Estos datos generan una gran cantidad de información, en donde se tratan temas de actualidad, como pueden ser los destinos turísticos. Uno de los destinos más buscados y mencionados en las redes sociales por los usuarios son las Islas Canarias, que gracias a su localización geográfica, ofrecen un microclima muy agradable para los turistas. Por otro lado, cuenta con un agradable clima. Canarias posee un alto valor cultural y natural gracias sus Parques Nacionales y la sostenibilidad con el medio ambiente.

Para abarcar el procesamiento de todos los datos generados en las redes sociales por los usuarios, nace la necesidad de utilizar mecanismos como los sistemas de alerta y sistemas de análisis de sentimientos. Los sistemas de alerta, son sistemas que ante uno o varios eventos, son capaces de generar una notificación o alerta, para indicar a un usuario o grupo de usuarios sobre un acontecimiento. Los sistemas de análisis de sentimientos o también conocidos como “minería de opinión”, utilizan el procesamiento del lenguaje natural y la similitud semántica, para proporcionar información del estado de ánimo o grado de satisfacción de un determinado usuario. Estos sistemas, aplicados en el sector turístico, proporcionan un medio rápido y efectivo a la hora de comunicar a las personas, por ejemplo, acerca de la situación climatológica creando una alerta por oleaje en el caso de los sistemas de alerta. Y en el caso de los sistemas de análisis de sentimientos, opiniones positivas o negativas dependiendo del léxico empleado, basándose en las propias experiencias de los usuarios.

Palabras clave: *Redes sociales, sistema de alerta, sistema de análisis de sentimiento, integración, procesamiento, lenguaje natural.*

Resumo

O objetivo deste trabalho, foi desenvolver uma integração de sistemas, em particular um sistema de alerta e um sistema de análise de sentimentos, orientado para o campo do turismo.

Atualmente, as redes sociais são um meio onde se pode acessar a informação, compartilhar opiniões e criar conteúdo. Além disso, elas são compostas de milhares de usuários que estão em constante interação entre si, causando então milhões de dados gerados. Esses dados produzem uma grande quantidade de informação, que lida com questões atuais, como: política, deportes o destinos turísticos. Um dos destinos mais procurados e mencionados em redes sociais por usuarios, são as Ilhas Canárias, que graças à sua localização geográfica, oferecem um microclima muito agradável para os turistas. Por outro lado, além de ter um clima agradável, as Ilhas Canárias têm um alto valor cultural e natural graças aos seus Parques Nacionais e à sustentabilidade com o meio ambiente.

Para cobrir o processamento de todos os dados gerados nas redes sociais pelos usuários, nasce a necessidade de utilizar mecanismos como sistemas de alerta e sistemas de análise de sentimentos. Os sistemas de alerta são sistemas que, na presença de um ou mais eventos, são capazes de lançar uma notificação ou alerta para indicar a um usuário ou grupo de usuários sobre um evento. Os sistemas de análise de sentimentos, também conhecidos como "prospecção de opinião", usam processamento de linguagem natural e semelhança semântica, para fornecer informações sobre o humor ou grau de satisfação de um determinado usuário. Esses sistemas, aplicados no setor de turismo, provêm gerar um meio rápido e eficaz de comunicar às pessoas, por exemplo, sobre a situação climática que produz um alerta de ondas no caso de sistemas de alerta. E no caso dos sistemas de análise de sentimentos, opiniões positivas ou negativas, dependendo do léxico utilizado, com base nas próprias experiências dos usuários.

Palavras-chave: Redes sociais, sistema de alerta, sistema de análise de sentimentos, integração, processamento, linguagem natural.

Índice General

Capítulo 1	11
Introducción	11
1.1 Internet	12
1.2 Web 2.0.....	13
1.3 Redes Sociales	14
1.3.1 Clasificación de las redes sociales.....	15
1.4 Redes Colaborativas.....	20
1.4.1 Tipologías de redes organizacionales.....	21
1.4.2 Similitudes y diferencias con respecto a las redes sociales.....	22
1.5 Tecnologías	23
1.5.1 Procesamiento del lenguaje natural (PLN).....	23
1.5.2 RStudio.....	25
1.5.3 Lenguaje R.....	26
1.5.4 ShinyApps	28
1.5.5 GitHub.....	29
Capítulo 2	30
Sistemas de alerta y sistemas de análisis de sentimientos	30
2.1 Sistemas de alerta	30
2.1.1 Sistemas de alertas de vigilancia tecnológica.....	30
2.1.2 Sistemas de Alerta Temprana.....	32
2.2 Sistemas de análisis de sentimientos.....	32
2.3 Propuesta de proyecto	33
Capítulo 3	35
Análisis de requisitos y caso de uso	35
3.1 Requisitos.....	35
3.1.1 Servidor.....	35
3.1.2 Cliente	35
3.2 Diagramas de casos de uso	36
Capítulo 4	38
Integración de los sistemas de alerta y de análisis de sentimientos propuestos	38
4.1 Selección de datos.....	38

4.2 Extracción de datos	38
4.3 Tratamiento de los datos obtenidos.....	44
4.4 Análisis de los datos.....	46
4.5 Alertas enviadas	48
4.6 La aplicación con ShinyApp	48
Capítulo 5.....	54
Conclusiones y líneas futuras	54
5.1 Conclusiones	54
5.2 Líneas Futuras.....	55
Capítulo 6.....	57
6.1 Conclusões.....	57
6.2 Estudios futuros.....	58
Capítulo 7.....	60
Presupuesto.....	60
7.1 Presupuesto del trabajo realizado.....	60
Referencias.....	63

Índice de figuras

Figura 1. Comparativa Deportes.....	18
Figura 2. Comparativa Universidades y facultades.....	19
Figura 3. Comparativa gobierno de España.....	20
Figura 4: RStudio	26
Figura 5: Logo lenguaje R.....	27
Figura 6: ShinyApps.....	29
Figura 7: Github.....	29
Figura 8: Caso de uso del servidor.....	36
Figura 9: Caso de uso del cliente.....	37
Figura 10: Registro en red social Twitter	39
Figura 11: Formulario de registro Twitter	39

Figura 12: Cuenta de Twitter SAYAS creada	40
Figura 13 Crear nueva app en la API de Twitter.....	40
Figura 14: Formulario de registro de la API de Twitter	41
Figura 15: Aplicación creada en la API de twitter.....	41
Figura 16: Access Token para la autenticación	42
Figura 17: Código para autenticación con OAuth.....	43
Figura 18: PIN para proceso de autorización de la API	44
Figura 19: Dataframe de extracción de datos	44
Figura 20: Datos en fichero CSV.....	45
Figura 21: Reorganización de datos en fichero .csv y eliminación de duplicados	45
Figura 22: Limpieza de datos.....	46
Figura 23: Diccionario de palabras positivas y negativas	47
Figura 24: Cuenta palabras positivas o negativas	47
Figura 25: Evaluación de los tweets,hashtags	48
Figura 26: Código para envío de alertas en el servidor	48
Figura 27: Aplicación Sistema de Alerta y Análisis de Sentimientos (SayAS)	49
Figura 28: Estructura SayAS en el servidor	50
Figura 29: Formulario de SayAS con la búsqueda del hashtag #playa.....	51
Figura 30: Alerta recibida en el correo	51
Figura 31: Tabla con tweets y su usuario correspondiente con la búsqueda del hashtag #playa	52
Figura 32: Tabla en el servidor	52
Figura 33: Gráfica de análisis de sentimientos.....	53
Figura 34: Gráfico de análisis en el servidor	53

Índice de tablas

Tabla 1: Librerías utilizadas	28
Tabla 2. Tabla resumen de los Tipos.....	61
Tabla 3: Presupuesto hardware utilizado	61
Tabla 4: Presupuesto del software utilizado	62
Tabla 5: Tabla resumen del presupuesto	62

Capítulo 1

Introducción

Internet, es una herramienta imprescindible en la sociedad actual, ya que, crea puentes de comunicación entre los individuos, eliminando así las barreras de espacio y tiempo. Con la llegada de la Web 2.0 (término acuñado por Tim O'Reilly), que está basada en comunidades virtuales de usuarios, en donde estos pueden colaborar e interactuar entre sí, además de crear contenidos. Y una gama especial de servicios, como son las redes sociales los blogs y las wikis.

La creación de las redes sociales supone una nueva forma de comunicación. Las personas pueden expresar opiniones digitales en foros, debates y hacer que estas lleguen a un público muy amplio y diverso.

Dado que en las redes sociales hay un continuo intercambio de información, los métodos tradicionales empleados para el tratamiento de los datos, ya no resultan efectivos a la hora de procesar estos datos. No son capaces de afrontar los grandes volúmenes de datos. Existen diversas investigaciones [1] [2] [3] realizadas en el ámbito de las redes sociales para intentar solventar este problema. Estos trabajos, hacen uso de métodos más actuales para el tratamiento de los datos, por ejemplo, mediante el procesamiento del lenguaje natural (PLN), son capaces de realizar nuevos métodos para el análisis de los datos obtenidos de las redes sociales.

En nuestro proyecto, se pretende analizar la información de las redes sociales para posteriormente, procesarla con las tecnologías más actuales. Como consecuencia de esto, se persiguen una serie de objetivos que se detallan a continuación:

(OB1). Estudiar del estado del arte de las redes sociales.

(OB2). Analizar distintas metodologías para sistemas de alerta y sistemas de análisis de sentimientos.

(OB3). Diseñar e implementar un sistema de alerta y un sistema de análisis sentimientos con los datos propuestos.

(OB4). Testear, validar e integrar en una sola herramienta, un sistema de alerta y un sistema de análisis de sentimientos.

(OB5). Proponer un caso de uso para la herramienta que tendrá integrada un sistema de alerta y un sistema de análisis de sentimientos.

(OB6). Elaborar documentación técnica.

La memoria está compuesta por los siguientes siete capítulos: En el capítulo 1, se recoge en una introducción, conceptos clave como: Internet, Web 2.0, redes sociales, redes colaborativas y tecnologías necesarias para llevar a cabo el proyecto como: procesamiento del lenguaje natural, lenguaje de programación R y sus librerías, entornos de desarrollo como RStudio y ShinyApps. En el capítulo 2, se pretende definir qué es y para qué sirven un sistema de alerta y un sistema de análisis de sentimientos. Además expondremos nuestra propuesta de proyecto. En el capítulo 3, se realiza un análisis de los requisitos que debe cumplir nuestro sistema y una propuesta de caso de uso. En el capítulo 4, se explica cómo llevaremos a cabo la aplicación Web de nuestro proyecto. La intención de este capítulo es explicar cómo y con qué metodologías, se realizó la elaboración de una herramienta, que permite la integración de un sistema de alerta y un sistema de análisis de sentimientos. En el capítulo 5 y 6, se exponen las conclusiones a las que se llegaron, posteriormente a la realización de la aplicación Web, así como las líneas futuras que podría seguir el proyecto. En el capítulo 7, se coteja una estimación aproximada del coste del proyecto, en base a las horas empleadas, el software utilizado y el hardware empleado.

1.1 Internet

El nombre Internet, proviene de las palabras en inglés “Interconnected Networks”, que significa “redes interconectadas”. Sus orígenes se remontan a 1969, cuando se estableció la primera conexión entre ordenadores, conocida como ARPANET.

Internet, es una arquitectura de sistema que ha revolucionado las comunicaciones y los métodos de comercio. Permite que varias redes de computadoras estén interconectadas. Proporciona un mecanismo potente y accesible por cualquier individuo, ya que, ofrece una amplia información que puede ser utilizada para cualquier propósito en general. La difusión de textos digitalizados, imágenes y grabaciones de audio y vídeo, ha dado lugar a una explosión de información [4].

Internet, se ha convertido en uno de los medios más importantes en la sociedad desde finales del siglo XX y principios del XXI. Trajo consigo una nueva forma de comunicación entre las personas y una mejora en la transmisión y acceso de la información eliminando así las barreras de espacio/tiempo. Surge así la World WideWeb [5] o también conocida como Web 1.0. Mediante el concepto denominado la “webvolución” [6], en otras palabras, la evolución de la Web, el enfoque del concepto de la Web ha ido transformándose de “acceso y hallazgos” (Web 1.0), a “compartir, participar y colaborar” (Web 2.0), hasta llegar a la “colaboración inmersiva y co-creación” (Web 3.0) [7] [8].

1.2 Web 2.0

Después de la llegada de la World Wide Web en los 90 y la Web 2.0 [9], para referirse a una segunda generación en la historia de la Web, la Web 2.0 se puede entender cómo –“todas aquellas utilidades y servicios de Internet que se sustentan en una base de datos, la cual puede ser modificada por los usuarios del servicio, ya sea en su contenido (añadiendo, modificando o borrando información o asociando datos a la información existente), pues bien en la forma de presentarlos, o en contenido y forma simultáneamente”- [10]. Se puede afirmar, que el término “Web 2.0” representa la evolución de las aplicaciones tradicionales hacia aplicaciones Web enfocadas en el usuario final. No se trata de una nueva tecnología, sino de una actitud de colaboración y participación de las personas para proporcionar mejores datos, nuevos servicios y aplicaciones on-line [11]. En donde los usuarios pueden colaborar e interactuar entre sí [12], creando contenidos dentro de una comunidad virtual y conectándose con otros individuos.

Aunque, todos los sitios de la Web están interconectados entre sí, con el fin de organizar los recursos digitales [12], proponen una taxonomía compuesta por cuatro líneas fundamentales que ayudarán al usuario en la selección de aquellas aplicaciones que cubran de una mejor manera sus necesidades e intereses. La estructura propuesta se denomina “los cuatro pilares de la Web 2.0”, que son [13]:

- **Redes sociales.** Describe todas aquellas herramientas diseñadas para la creación de espacios, que promuevan o faciliten la formación de comunidades e instancias de intercambio social.

- **Contenidos.** Hace referencia a aquellas herramientas que favorecen la escritura en línea, así como su distribución e intercambio.
- **Organización social e inteligente de la información.** Herramientas y recursos para etiquetar, syndicar e indexar, que faciliten el orden y almacenamiento de la información, así como otros recursos disponibles en la Red.
- **Aplicaciones y servicios (mashups).** Dentro de esta clasificación, se incluye numerosas herramientas software, plataformas en línea y un híbrido de recursos creados para ofrecer servicios de valor al usuario final.

1.3 Redes Sociales

Como se describía en la sección anterior, la Web 2.0 está compuesta por comunidades de usuarios y servicios, como: las redes sociales, los blogs y las wikis.

Para definir a las redes sociales, existen diferentes perspectivas: las que se centran en la comunicación y colaboración o las que se basan en estudio de las características estructurales de sus herramientas [14]. Algunos autores como [15], definen a las redes sociales como: herramientas que facilitan el construir un perfil público dentro de un sistema delimitado, pudiendo así, articular una lista de diferentes usuarios con los que comparte una conexión, y además poder ver las conexiones que suceden dentro de su sistema. Otros autores [16], las define como: una estructura social que se puede representar como nodos conectados por aristas, en donde los nodos representan a los individuos y las aristas las relaciones existentes entre ellos. Para el autor [17], se centra en un contexto educativo, señala que son un entorno que permite la discusión de diferentes problemáticas en espacios interactivos y flexibles de aprendizaje. Nosotros, definiremos a las redes sociales como una estructura compuesta por un grupo de individuos u organizaciones y que están relacionados entre sí, ya sea, mediante relaciones laborales, amistad, familiar. Además, de un punto de encuentro en donde los usuarios comparten, crean y acceden a la información.

En la sociedad actual, las redes sociales despiertan un inusitado interés social y profesional. En el ámbito académico, el interés se centra en conocer los efectos que estas provocan, en el éxito y desarrollo educativo de cualquier

estudiante, sobre todo a niveles universitarios [18]. En el resto de ámbitos como el laboral, social y relaciones afectivas, la creación de las redes sociales supone una nueva forma de relacionarse. Estas herramientas ofrecen un espacio virtual para que, miles de personas puedan compartir contenidos multimedia con intereses similares [19]. Se consideran puntos de encuentro, en los que es posible acceder a la información, compartir impresiones, consultar archivos y recursos disponibles a tiempo real [20].

Redes sociales como, Facebook, Twitter, Orkut, son estructuras compuestas por una infinidad de usuarios relacionados entre sí, independientemente que se conozcan o no. En estas redes sociales, al tener un gran número de usuarios, se manejan grandes volúmenes de datos, ya que, los individuos comparten, buscan y difunden información a través de texto, imágenes o audio. Gracias a las redes sociales, las personas pueden expresar opiniones digitales en foros, debates y hacer que estas lleguen a un público muy amplio y diverso.

En el año 2002, comenzaron a aparecer los primeros sitios Web, promocionando las redes de círculos de amigos en línea. La popularidad de estos sitios creció rápidamente y compañías como por ejemplo, Myspace y LinkedIn surgieron en el año 2003, Google lanzó Orkut en 2004 y en el mismo año se dio a conocer Facebook, y en el año 2006 se lanzó Twitter.

Uno de los objetivos de la creación de estas redes, es crear un sitio de interacción virtual en donde miles de personas de todo el mundo puedan compartir, comunicarse y cooperar entre sí.

1.3.1 Clasificación de las redes sociales

Las redes sociales según su tipología, pueden clasificarse entre redes sociales horizontales o generales, y verticales o especializadas.

1. Horizontales

Las redes sociales horizontales, no tienen una temática definida, están orientadas al público en general y se centran en las relaciones de contacto entre usuarios. Proveen herramientas para que las personas se relacionen entre sí y todas comparten características: crear un perfil, compartir contenidos, generar lista de contactos [21]. Algunas de ellas son: Facebook, Orkut, Google+.

2. Verticales

En cuanto a las redes sociales verticales, tienden a especializarse según los intereses de sus usuarios, está orientada a un público específico y diariamente ganan miles de seguidores. Cualquier red social que se centre en un tema en concreto, puede dar origen a una nueva categoría de clasificación y es por ello que destacaremos las más relevantes, como por ejemplo las redes sociales según su temática, actividad y por último según el contenido compartido [21].

3. Temática

- **Profesionales.** Son redes enfocadas en temas de negocios y actividades comerciales. Los usuarios definen un perfil en donde detallan su ocupación, estudios académicos, experiencia laboral. La más conocida es LinkedIn.
- **Viajes.** Estas redes sociales, conectan viajeros que comparten experiencias, sirviendo así de guía para otros viajeros. Un ejemplo de ellas sería Travellerspoint, tripadvisor.
- **Relaciones personales.** Son redes sociales propias para relaciones personales de pareja. Por ejemplo: meetic, Tinder.

4. Actividad

- **Microblogging.** Estas redes sociales, ofrecen un servicio de envío y publicación de mensajes breves de texto. También, permiten seguir a otros usuarios, aunque esto no establece necesariamente una relación recíproca, como los seguidores. La red social más destacada es Twitter.
- **Geolocalización.** Permiten, mostrar la localización de una persona o establecimiento. Mediante estas redes sociales los usuarios pueden localizar el contenido digital que comparten. Por ejemplo: Foursquare, Trippy.

5. Contenido

- **Fotos.** Estos servicios, ofrecen la posibilidad de almacenar, ordenar, buscar y compartir fotografías. Las más importantes en número de usuarios son: Instagram, Flickr, Fotolog.

- **Vídeos.** Los sitios Web de almacenamiento de vídeos, se han popularizado de tal manera, que en los últimos años incorporan la creación de perfiles y listas de amigos para la participación colectiva mediante los recursos de los usuarios, y los gustos sobre los mismos. Algunos son: Youtube, Dailymotion.
- **Documentos.** Por la red navegan documentos de todo tipo en formatos diversos, en estas redes sociales podemos encontrar, publicar y compartir los textos definidos por nuestras preferencias de una manera fácil y accesible. Su mayor exponente es Scribd, GoogleDrive.

Nosotros, haremos una clasificación por medio de gráficas para representar a las redes sociales más importantes en la red. Pondremos énfasis en lo que respecta a la compartición de información, mediante la opinión de los usuarios. Y a los temas tratados con especial relevancia, como por ejemplo temas de actualidad que son: la política, Internet y telecomunicaciones, destinos turísticos, noticias en general, deportes y así una infinidad de temáticas.

Según las estadísticas del último año, en todo el mundo a lo que deportes se refiere, Twitter es la más utilizada a la hora de divulgar opiniones acerca de los partidos, competiciones y campeonatos [22] (Véase figura 1).



Figura 1. Comparativa Deportes

En donde usuarios comparten sus opiniones, en partidos, competiciones y campeonatos realizados en el último año.

En el caso de las facultades y universidades que utilizan redes sociales, la más destacada es Twitter frente a redes sociales como Orkut o Instagram [23] (Véase figura 2).



Figura 2. Comparativa Universidades y facultades

Comparativa Universidades y facultades, del último año en donde la red social más utilizada es Twitter frente a Google+ o Instagram.

Comparativa en España de las redes de mensajería rápida más utilizadas, a la hora de compartir contenidos a cerca del gobierno español [24] (Véase figura 3).



Figura 3. Comparativa gobierno de España

Comparativa del gobierno de España, en donde el medio de comunicación de mensajería corta más utilizado, es Twitter en comparación con otras como Telegram o Whatsapp.

1.4 Redes Colaborativas

En la sección anterior, se ha definido el concepto de redes sociales y sus clasificaciones. En este apartado se definirá que es una red colaborativa y su similitud, diferencias con respecto a las redes sociales.

Las Redes Colaborativas (RC) tienen una especial importancia. Podemos definir a una Red Colaborativa como una red compuesta por varias entidades autónomas, dispersas geográficamente y diferentes en términos de: entorno, cultura, capital social y objetivos. En las RC, la colaboración entre los miembros de la red hace posible alcanzar los objetivos marcados y que de manera independiente, estos objetivos serían más costosos de alcanzar para las organizaciones [25]. Otros autores como [26], describe a las RC como, una red que se compone de un conjunto de vínculos, que pueden representar relaciones, o ausencia de ellas. Los nodos pueden ser actores, unidades organizativas y organizaciones. El objeto de intercambio define el tipo de red

a estudiar (bienes, servicios, influencias, afecto). Las relaciones se asumen como perdurables [27].

1.4.1 Tipologías de redes organizacionales

En las redes colaborativas, se identifican cuatro escenarios posibles en los que analizar los patrones de relaciones entre organizaciones, clasificados en [28]: relación binaria, relación estelar, red de trabajo y por último, entre redes. Nosotros haremos una segunda clasificación: análisis centrado en una organización y análisis orientado a la red.

1. Análisis centrado en una organización

- **Relación binaria.** El análisis de la relación se realiza respecto a dos entidades que establecen un acuerdo de colaboración. El contenido de la relación se centra en definir:
 - **Predisposición, o no.** Para actuar de manera conjunta, mediante el intercambio de bienes o servicios.
 - **Dependencia mutua.** Resultado de la acción conjunta de las dos organizaciones.
 - **Organización del vínculo.** Representa el grado de colaboración que puede medirse en términos de: formalización del acuerdo, intensidad, reciprocidad y estandarización.
- **Relación estelar.** Es un modelo centralizado en el que la organización que predomina, analiza de manera dual, su actividad de cooperación y colaboración con cada uno de las otras organizaciones o entidades.

2. Análisis Orientado a la red

- **Red de trabajo.** Se trata de un conjunto de organizaciones, que se movilizan de manera cooperativa con la finalidad de alcanzar un objetivo determinado. Aunque, para lograrlo no necesariamente deben adquirir una entidad. La duración de esta colaboración, depende del objetivo consensuado y predeterminado cuyo alcance de manera individual no sería posible. Este tipo de

colaboración, no se enfoca hacia las relaciones entre los participantes, sino en el comportamiento organizativo global. Para conseguirlo, los participantes deben determinar [29]:

- El número de actores implicados.
 - El ámbito de poder al que accede la organización escogida como líder.
 - La similitud de valores y actitudes entre los miembros.
 - El efecto que tiene el comportamiento de otros equipos frente a éste.
- **Entre Redes.** La colaboración entre redes, representa a un conjunto de redes, generalmente con un número elevado de integrantes, que mantienen algún tipo de relación entre ellas. Diferenciando con la red de trabajo, la colaboración entre redes, posee una naturaleza dinámica y su extensión temporal puede ser indefinida [30].

1.4.2 Similitudes y diferencias con respecto a las redes sociales

Anteriormente, hemos definido a las redes sociales como una estructura compuesta por un grupo de individuos u organizaciones y que están relacionados entre sí. Y en donde los usuarios comparten, crean y acceden a la información. Las redes colaborativa al igual que las redes sociales, son una estructura compuesta por personas u organizaciones que colaboran entre sí. Ya sea, en algún proyecto o alguna investigación para lograr objetivos comunes, independientemente del ámbito en donde se desarrollen, en donde interactúan mediante los ordenadores.

La diferencia entre las redes sociales con respecto a las redes colaborativas, es que, las redes colaborativas se crean con el propósito de unir a personas, u organizaciones, para que colaboren entre sí. Mediante el proceso colaborativo compuesto por: **colaboración, cooperación, coordinación y networking**. Con un fin en común el “alcanzar un objetivo final” que puede ser el resolver un problema o algún proyecto de investigación. En cambio las redes sociales, simplemente se crean para que las personas se creen un perfil en la red y de esta manera, conectar usuarios, compartir contenido si así lo quieren y difundirlo.

1.5 Tecnologías

En las secciones anteriores, se ha definido a las redes sociales, a las redes colaborativas, sus similitudes y diferencias. Además, en cómo están compuestas por miles de usuarios. Las redes sociales, al tener un gran número de usuarios, están constantemente compartiendo y difundiendo información a través de imágenes, texto o audio. Por lo que se generan grandes volúmenes de datos. Para poder analizar esa gran cantidad de datos, es necesario la utilización de tecnologías como: el procesamiento del lenguaje natural (PLN), lenguajes de programación como el lenguaje R y entornos de desarrollo como RStudio y servicios para el despliegue de aplicaciones como ShinyApps.

1.5.1 Procesamiento del lenguaje natural (PLN)

Para entender el procesamiento del lenguaje natural, primero definiremos en que consiste el lenguaje natural (LN). Un lenguaje natural, es aquel que ha evolucionado con el tiempo para fines de comunicación humana, como el español o alemán [31]. El lenguaje natural, es la vía de comunicación que utilizan las personas para comunicarse entre ellas habitualmente.

El Procesamiento de Lenguaje Natural (PLN), se entiende como la habilidad que tiene una máquina para procesar la información comunicada, no simplemente las letras o los sonidos del lenguaje. Un traductor automático sería un ejemplo de dispositivo o software de PLN [32].

El PLN tiene aplicaciones muy variadas, ya que, su alcance es muy extenso, algunas aplicaciones son [31].

- Traducción automática.
- Recuperación de la información.
- Extracción de Información y Resúmenes.
- Tutores inteligentes.
- Reconocimiento de Voz.

La arquitectura de un sistema de PLN, se fundamenta en una definición del LN por niveles, que son: fonológico, morfológico, sintáctico, semántico y pragmático [31].

- **Nivel Fonológico.** Trata de cómo las palabras se relacionan con los sonidos que representan.

- **Nivel Morfológico.** Trata de cómo las palabras se construyen a partir de unas unidades de significado más pequeñas llamadas morfemas.
- **Nivel sintáctico.** Trata de cómo las palabras pueden unirse para formar oraciones, fijando el papel estructural que cada palabra juega en la oración y que sintagmas son parte de otros sintagmas.
- **Nivel semántico:** trata del significado de las palabras y de cómo los significados se unen para dar significado a una oración, también se refiere al significado independiente del contexto, es decir, de las oraciones aisladas.
- **Nivel Pragmático.** Trata de cómo las oraciones se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones. Se reconoce un subnivel recursivo: discursivo, que trata de cómo el significado de una oración se ve afectado por las oraciones inmediatamente anteriores.

Existen cuatro componentes necesarios para usar y manejar de manera más productiva el conocimiento plasmado en textos. Estos son: la digitalización, la recuperación, la presentación de la información y su uso directo por el software [32].

1. **Digitalización de la información.** Existe muchísimos documentos, libros, manuales, en donde su contenido está plasmado en papel y es difícil de acceder a la información. Por ello, es importante que a la hora de usar el PLN el contenido este en formato digital. Por digitalización podemos entender como, la obtención de la información en formato digital en donde el texto es una secuencia de letras, por ejemplo los correos electrónicos, documentos en periódicos digitales, información en las redes, etc.
2. **Recuperación de la información.** También conocida como búsqueda de la información, tiene la problemática que un mismo concepto o idea puede expresarse de maneras distintas, utilizando diferentes palabras. Por ejemplo: el usuario expresa su interés con la frase “la derrota del Madrid” y el documento relevante para tal petición es “la victoria del Barcelona”. Los dos textos no tienen ninguna palabra en común, pero un humano, usando su experiencia lingüística (derrota – victoria) y su

conocimiento del mundo (Madrid – Barcelona) fácilmente detectaría la relevancia del documento para la petición. Se han alcanzado avances significativos para que los programas puedan usar este tipo de razonamiento para así, satisfacer de una mejor manera las necesidades de los usuarios.

3. **Presentación de la información.** Es importante que la información contenida en los textos, este representado de una manera eficaz. Un ejemplo sería, la construcción automática de resúmenes, es decir, un generador automático de resúmenes, a partir de un texto largo, detectar lo más trascendente y presentarlo en un texto corto que se pueda leer en un tiempo más razonable. Otra forma de resumir la información para hacerlos más manejables sería, agrupar y clasificar los documentos que sean semejantes entre sí. La minería de texto, es otra manera de resumir el contenido textual (encontrar opiniones expresadas en los textos, las tendencias de cambio de estas opiniones o las relaciones inesperadas entre los eventos descritos en los textos), la extracción de información (llenar bases de datos sobre un tema específico, leyendo los textos) y sistemas de soporte a la toma de decisiones (buscar, sintetizar y presentar de manera eficiente la información relevante para un directivo).
4. **Uso de la información.** El uso de la información contenida en los textos por el mismo software para resolver tareas más complejas, permite a la máquina encontrar el conocimiento necesario de los textos disponibles. Por ejemplo: los artículos científicos o los libros de texto.

1.5.2 RStudio

RStudio, es un entorno de desarrollo integrado (IDE) para R. Incluye una consola, un editor de resaltado de sintaxis, que soporta la ejecución de código directo, así como herramientas para: la depuración y gestión de espacio de trabajo. RStudio está disponible para: Windows, Mac y Linux se puede ejecutar en el escritorio o en un navegador conectado a RStudio Server o RStudio Server Pro (Debian / Ubuntu, RedHat / CentOS, y SUSE Linux) [33].

RStudio tiene como objetivo proporcionar un entorno de desarrollo informático estadístico R. Permite el análisis y desarrollo para que se lleven a cabo los análisis correspondientes con el lenguaje R (véase figura 4). Las principales características son:

- Acceso RStudio localmente.
- El resaltado de sintaxis, completado de código, y la sangría inteligente.
- Ejecutar código R directamente desde el editor de código fuente.
- Saltar rápidamente a la definición de funciones.
- Gestionar fácilmente múltiples directorios de trabajo utilizando proyectos.
- R ayuda integrada y documentación.
- Depurador interactivo para diagnosticar y corregir los errores rápidamente.
- Paquete de herramientas de desarrollado por usuarios.



Figura 4: RStudio

1.5.3 Lenguaje R

El lenguaje R, además de ser un lenguaje de programación, también es un entorno con un enfoque al análisis y gráficos. Es un proyecto GNU similar al lenguaje de programación S. Fue desarrollado en Bell Laboratories anteriormente AT & T, ahora Lucent Technologies), por John Chambers y colegas. R puede considerarse como una implementación diferente a S. Entre estos dos lenguajes de programación, existen algunas diferencias importantes, aunque, en gran medida el código escrito para S, se ejecuta bajo R.

R, proporciona una gran variedad de modelos estadísticos (modelos lineales y no lineales, pruebas estadísticas clásicas, clasificaciones, agrupaciones, etc.) y técnicas gráficas, además es muy extensible. El lenguaje R, proporciona facilidad a la hora de crear publicaciones bien diseñadas incluyendo símbolos matemáticos y fórmulas donde se es necesario. Asimismo, es uno de los lenguajes más utilizados por la comunidad de programadores, y es muy popular en campos como: la minería de datos, investigaciones biomédicas, la bioinformática, y finanzas. R está disponible como software libre bajo la licencia de GNU, compila y se ejecuta en una

amplia variedad de plataformas como: UNIX, Windows, MacOS. Ofrece la posibilidad de trabajar con distintas bibliotecas o librerías que ofrecen funciones para analizar los datos que se le provee [34] (Véase figura 5).



Figura 5: Logo lenguaje R

1. Características

R es un conjunto integrado de paquetes para la manipulación de datos, el cálculo y la visualización gráfica [35]. Incluye:

- Una instalación eficaz de tratamiento y almacenamiento de datos.
- Una serie de operadores para cálculos matemáticos.
- Una gran colección de herramientas integradas para el análisis de datos.
- Instalaciones gráficas para el análisis de datos y visualización.
- Proporciona gráficas para análisis de datos que funcionan directamente sobre pantalla o impresora.
- Un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas.
- Permite que los usuarios lo extiendan definiendo sus propias funciones.
- Orientado a objetos, característica que hereda de S.

2. Librerías

R, es miembro de un proyecto colaborativo y abierto, en donde sus usuarios pueden publicar paquetes y extender su configuración básica. R se puede extender fácilmente con los paquetes, existen unos ocho paquetes suministrados con la distribución de R y muchos de ellos están disponibles en CRAN (<https://www.r-project.org/>).

En nuestro proyecto para su la elaboración utilizaremos algunas librerías como (Véase tabla 1):

Conexión y extracción de datos	Tratamiento de datos	Sistema de Alerta	Sistema de Análisis de Sentimientos	Visualización y despliegue de la app
ROAuth	dplyr	mailR	leaflet	shiny
base64enc	tm	-	stringr	shinythemes
twitterR	stringr	-	-	base64enc
streamR	reader	-	-	ggthemes
RCurl	plyr	-	-	rsconnect
-	-	-	-	ggplot
-	-	-	-	RPushbullet

Tabla 1: Librerías utilizadas

1.5.4 ShinyApps

ShinyApps, es un servicio de hosting utilizado para el despliegue de aplicaciones Shiny. Aunque, existen otras posibilidades para llevar a cabo el despliegue como:

- En la máquina local mediante el IDE RStudio y bajo cualquier sistema operativo.
- En un servidor dedicado con el Server Shiny de RStudio.
- En algún servicio de hosting dedicado especializado para Shiny.

Para este proyecto se utilizara el servicio de hosting ofrecido por ShinyApps, ya que, es fácil de utilizar y no supone coste alguno. Solo se necesita una cuenta de usuario para hacer el despliegue. Además la integración del paquete shiny y RStudio con ShinyApps, permite realizar el despliegue de forma automática, sin tener que configurar y desplegar de forma manual [36](Véase figura 6).



Figura 6: ShinyApps

Nuestro proyecto se encuentra desplegado en el siguiente enlace:
<https://alu0100673647.shinyapps.io/SAYAS/>

1.5.5 GitHub

GitHub es una plataforma Web de desarrollo colaborativo y alojamiento de repositorios basada en Git (software de control de versiones). Ofrece toda la funcionalidad de Git e integra diversas herramientas de control de acceso, colaboración, trazabilidad, gestión de tareas y control de proyectos. Además, proporciona una interfaz Web que permite al usuario registrado crear repositorios vacíos o por clonación de otro repositorio hospedado en GitHub (fork en la terminología de GitHub), enviar solicitudes de cambio entre repositorios hospedados (pull request en la terminología de GitHub), y gestionar dichas solicitudes. Asimismo, los repositorios hospedados en GitHub pueden actuar como repositorios remotos de repositorios locales [37] (Véase figura 7).



Figura 7: Github

El código de nuestro proyecto así como esta memoria se encuentra alojado en el siguiente repositorio: <https://github.com/Alu0100673647/TFG>

Capítulo 2

Sistemas de alerta y sistemas de análisis de sentimientos

En el capítulo anterior, se daba una introducción acerca de qué es Internet, la llegada de la Web 2.0, las redes sociales, las redes colaborativas y las distintas tecnologías utilizadas en el desarrollo de este proyecto.

En este capítulo, definiremos en qué consiste un sistema de alerta y un sistema de análisis de sentimientos, mostrando ejemplos de cada uno de ellos. Y por último detallaremos el sistema que queremos proponer en nuestro proyecto.

2.1 Sistemas de alerta

Un sistema de alerta, se podría definir como: sistema que ante uno o varios eventos, es capaz de generar una notificación o alerta, para indicar a un usuario o grupo de usuarios sobre un suceso, acontecimiento que está ocurriendo en ese momento o simplemente retornar una información.

Los sistemas de alerta han sido implementados en diversos campos como: la vigilancia tecnológica, la climatología (SAT), educación, sanidad, entre otros.

2.1.1 Sistemas de alertas de vigilancia tecnológica

En un sistema de alertas de vigilancia tecnológica, como por ejemplo "ALERTAS", ofrece información de actualidad sobre: ciencia, tecnología e innovación. A través de estas alertas, los usuarios pueden tener acceso a: eventos actualizados, noticias diarias, artículos recientes, ferias y congresos. Este sistema para poder ofrecer el servicio de alerta, necesita que el usuario este registrado y que además en el formulario de registro, especifique sus preferencias. Posteriormente el sistema, recogerá esta información para llevar a cabo y de manera diaria el proceso de detección, clasificación y priorización de resultados para así, ofrecer al usuario un resumen de noticias, eventos y novedades.

En estos sistemas, se incorpora tecnologías del lenguaje natural para la recuperación, extracción y clasificación de información documental, desarrolladas bajo los estándares de la Web semántica. Con la suma de estrategias y técnicas es posible presentar al usuario solo aquella información que cubra sus necesidades. Estas son: la creación y mantenimiento de ontologías, la asignación semi-automática de metadatos y generación de vinculaciones y relaciones semánticas entre estos para la extracción de información textual.

Permiten, la clasificación de la información procedente de fuentes relevantes en el ámbito científico. Extrae solamente la información relevante dependiendo de las preferencias de los usuarios y la almacena en bases de datos para su posterior procesamiento. Está compuesto por varias partes [38].

- **Clasificación inteligente.** Consiste en hacer que el sistema aprenda a diferenciar entre la información relevante de la que lo no es. Y para ello utiliza técnicas de aprendizaje automático y categorización de textos, y a diferenciar entre unas temáticas y sectores de otros.
- **Rastreo automático.** El sistema tecnológico rastrea periódicamente una selección de dominios y fuentes de noticias especializadas. Cuando se detectan nuevos contenidos, el sistema los descarga, categoriza automáticamente y selecciona los resultados relevantes, descartando los no relevantes. Estos contenidos actualizados y relevantes son las denominadas Alertas de Vigilancia. Éstas son almacenadas en el escritorio de la plataforma virtual y accesible para todos los usuarios, previamente registrados. Cada usuario recibirá entonces un conjunto de alertas específicas y personalizables según sus preferencias seleccionadas.
- **Ranking y envío por correo.** Según el medio de consulta y la frecuencia escogida, el usuario recibirá un resumen personalizado con las diez alertas más relevantes del conjunto de todas las alertas que el sistema tecnológico ha localizado, clasificado y priorizado. Para ello, se elabora un ranking automático por usuario sector de actividad, países de interés y temáticas escogidas en su perfil.

2.1.2 Sistemas de Alerta Temprana

Actualmente, se está viviendo en una época en donde se están sufriendo cambios climáticos a nivel global, en donde los distintos países del mundo están intentando adaptarse a las nuevas condiciones climatológicas.

Un Sistema de Alerta Temprana (SAT), se puede definir como un sistema de colección de información variada que, mediante monitoreo constante, permite advertir sobre situaciones amenazantes a la seguridad alimentaria y a la seguridad civil [39].

El objetivo principal de un sistema de alerta temprana (SAT), es permitir alertar y dirigir en caso de desastres naturales o climatológicos, para que comunidades amenazadas puedan actuar a tiempo y de manera apropiada para reducir la posibilidad de pérdidas de vidas humanas, daños a la propiedad y al ambiente y pérdida del sustento [26].

Existen estudios realizados [27] [40], en donde se quiere minimizar mediante la prevención de riesgo los daños que sufre la sociedad al tener que enfrentarse a desastres naturales como son: los desprendimientos de tierra, terremotos, inundaciones, etcétera. Estos fenómenos se originan comúnmente en ríos de montaña en donde existen fuertes vertientes, zonas con actividad volcánica.

2.2 Sistemas de análisis de sentimientos

Los sistemas de análisis de sentimientos o también conocidos como “minería de opinión”, utilizan el procesamiento del lenguaje natural y la similitud semántica, para proporcionar información del estado de ánimo o grado de satisfacción de un determinado usuario analizando las frases, comentarios, hashtags, etcétera, que pública.

La minería de opiniones es una reciente sub-disciplina entre la recuperación de información y la lingüística computacional. Esta sub-disciplina no trata de detectar el tema abordado en un texto sino las opiniones y sentimientos expresados en el mismo [41]. Determina si los comentarios en foros online, blogs o redes sociales sobre un tema en particular (producto, libro, película, etc.) son positivos, negativos o neutros. Las opiniones son muy importantes cuando alguien desea conocer la opinión de los otros antes de tomar una decisión.

Tradicionalmente, se pueden distinguir dos metodologías principales en minería de opiniones. Por un lado, hay enfoques para minería de opiniones que tienen como objetivo clasificar documentos enteros con polaridad

positiva o negativa. Por otro, lado hay enfoques que se centran en distinguir las frases con contenido de sentimiento dentro de los textos de opinión de los usuarios. El inconveniente de ambas metodologías, es que, un mismo documento o una sola frase podrían contener diferentes opiniones sobre las distintas características de un producto o servicio [42].

Cuando se habla de minería de opiniones, también llamada análisis del sentimiento, se refiere a la aplicación de una serie de técnicas del procesamiento del lenguaje natural, lingüística computacional y minería de textos, que tienen como objetivo la extracción de información subjetiva a partir de contenidos generados por los usuarios, como comentarios en blogs u opiniones en revistas de productos [43] [44] [45]. Con este tipo de tecnologías, es posible extraer un valor tangible y directo a partir de un comentario en lenguaje natural. Se podrían monitorizar y gestionar las opiniones sobre personas o entidades en foros y redes sociales concretas. Así como abordar temas como las tendencias [46].

Existen investigaciones realizadas [47] [48] [49], que para solventar el problema de la gran cantidad de información generada en los medios sociales, hace uso de técnicas de procesamiento del lenguaje natural (PLN). Estos trabajos estudian y proponen métodos para determinar la polaridad sobre opiniones, hashtags, tweets, así como determinar el tópico del mismo. Además de las investigaciones mencionadas, también hay herramientas que son capaces de analizar un tweet y determinar si son de carácter positivo o negativo, tales como:

- Mr. Tuit (<http://www.mrtuit.com/>)
- KEYHOLE (<http://keyhole.co/>)

2.3 Propuesta de proyecto

En nuestro proyecto tras analizar las distintas redes sociales, nos hemos decantado por trabajar con la red social Twitter. Por sus miles de usuarios compartiendo información diariamente, a través de los tweets, hashtags. En esta red social, se manejan grandes volúmenes de datos con diversidad de ámbitos como: la política, el turismo, opiniones digitales, etcétera. Para abarcar el procesamiento de todos los datos generados en la red social Twitter, utilizaremos el lenguaje R, ya que, es un lenguaje muy potente a la hora de trabajar con la minería de datos. También aplicaremos el

procesamiento del lenguaje natural (PLN), para poder analizar los datos extraídos de la red social.

Para abarcar el procesamiento, de todos los datos generados en las redes sociales por los usuarios, nace la necesidad de utilizar mecanismos como los sistemas de alerta y sistemas de análisis de sentimientos. Los sistemas de alerta, son sistemas que ante uno o varios eventos, son capaces de generar una notificación o alerta, para indicar a un usuario o grupo de usuarios sobre un suceso, acontecimiento. Los sistemas de análisis de sentimientos o también conocidos como “minería de opinión”, utilizan el procesamiento del lenguaje natural y la similitud semántica, para proporcionar información del estado de ánimo o grado de satisfacción de un determinado usuario. Estos sistemas, aplicados en el sector turístico, proporcionarían un medio rápido y efectivo a la hora de comunicar a las personas, por ejemplo, acerca de la situación climatológica generando una alerta por lluvia en el caso de los sistemas de alerta. Y en el caso de los sistemas de análisis de sentimientos, opiniones positivas o negativas dependiendo del léxico empleado, basándose en las propias experiencias de los usuarios.

Nuestra propuesta consiste realizar una integración de un sistema de alerta y un sistema de análisis de sentimientos orientando en el ámbito del turismo. Para ello trabajaremos con un grupo de tecnologías descritas en el capítulo 1, sección 1.5. Para la obtención y la extracción de los datos utilizaremos como fuente de información a la red social Twitter.

En esta propuesta, se quiere elaborar un sistema de alerta y un sistema de análisis de sentimientos, aplicados en el sector turístico, para así proporcionar un medio rápido y efectivo a la hora de comunicar o alertar a las personas de algún suceso. Por ejemplo, acerca de la situación climatológica generando una alerta por lluvia en el caso de los sistemas de alerta. Y en el caso de los sistemas de análisis de sentimientos, opiniones positivas o negativas dependiendo del léxico empleado, basándose en las propias experiencias de los usuarios.

Capítulo 3

Análisis de requisitos y caso de uso

En el capítulo anterior, definíamos en qué consistían los sistemas de alerta y de análisis de sentimiento, así como la propuesta de sistema para nuestro proyecto.

En este capítulo, se muestran los requisitos funcionales y no funcionales establecidos para la integración de sistemas, desarrollado para nuestra aplicación Web. También se recoge un diagrama del caso de uso de la aplicación (Véase figuras 8 y 9).

3.1 Requisitos

3.1.1 Servidor

- Requisitos funcionales:
 - **Recepción de peticiones de cliente:** el servidor será capaz de alojar la aplicación.
 - **Publicación de los datos:** el servidor tendrá acceso a una lista en donde se verán reflejados la información buscada.
- Requisitos no funcionales:
 - **Acceso a fichero con datos:** el servidor ha de acceder a los datos (usuarios, tweets) realizados en la búsquedas para utilizarlo en el cumplimiento de los requisitos funcionales.

3.1.2 Cliente

- Requisitos funcionales:
 - **Importación de conjuntos de datos del servidor:** la aplicación cliente, deberá ser capaz de acceder a la lista de datos que proporciona el servidor bajo los ficheros.csv.

- **Gráfica de tweets:** la aplicación deberá mostrar una gráfica distinguiendo en tres colores, las distintas clasificaciones de los tweets, según su naturaleza.
- **Tabla de información:** la aplicación deberá proporcionar una tabla, en donde se reflejan los datos (usuario, tweets) sobre los que se realiza el análisis.
- **Formulario para generación de búsquedas:** deberá existir un formulario que el usuario deberá rellenar con los siguientes campos:
 - **Búsqueda:** con el término que se desea buscar.
 - **Número de tweets:** entero que establecerá el número de tweets sobre el que se realizara la búsqueda.
 - **Tipo de alerta:** en donde el usuario elegirá el tipo de notificación que desea recibir.
- Requisitos no funcionales
 - **Plataformas:** será una aplicación Web, compatible con los navegadores más populares.
 - **Interfaz:** la interfaz deberá ser accesible y usable.

3.2 Diagramas de casos de uso

- Servidor

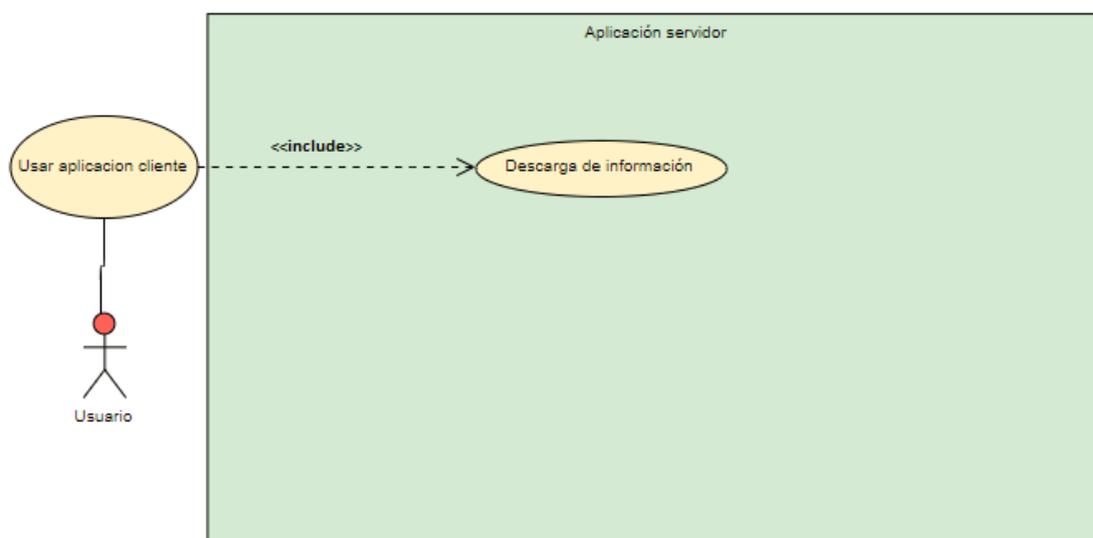


Figura 8: Caso de uso del servidor

- Cliente

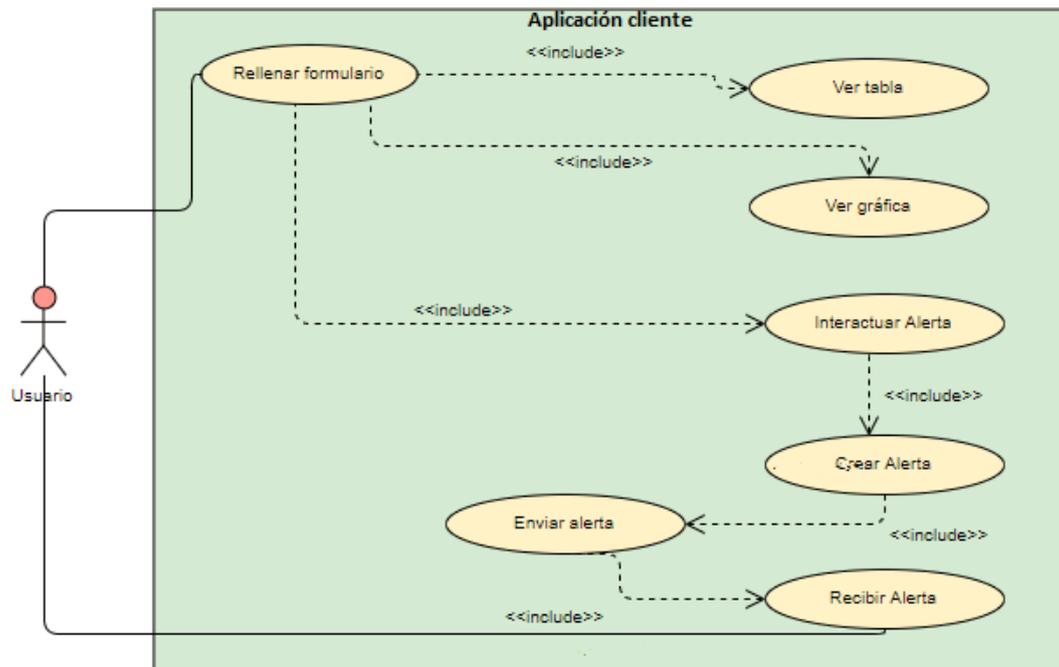


Figura 9: Caso de uso del cliente

Capítulo 4

Integración de los sistemas de alerta y de análisis de sentimientos propuestos

En el capítulo anterior, hicimos un análisis de los requisitos para la integración de sistemas, así como la aplicación de un caso de uso.

En este capítulo, detallaremos cómo hacer una integración de estos dos sistemas. Empezando en como seleccionamos y extraemos los datos, para su posterior procesamiento y formato de salida.

4.1 Selección de datos

La fuente de donde obtendremos los datos es la red social Twitter, ya que, abarca mucha información de miles de usuarios. Esta información está representada en distintos formatos, texto, imágenes, videos, etc. Nosotros, utilizaremos la información en formato texto, porque a la hora de procesar estos datos, al estar en formato plano y digitalizado, el procesamiento del lenguaje natural, será más favorable. Por este motivo, el contenido que recogeremos de la red social, serán: hashtags, tweets, retweets y menciones.

4.2 Extracción de datos

Para obtener los datos de Twitter, lo haremos mediante el lenguaje de programación R, bajo el entorno de trabajo RStudio. Durante la extracción de estos datos, se han seguido una serie pasos, que se detallarán a continuación.

1. Cuenta en Twitter

Es necesario disponer de una cuenta en Twitter, la que utilizaremos para el desarrollo de la aplicación es: Sistemas de Alerta y Análisis de Sentimientos “SAyAS” (<https://twitter.com/SAYAS91995364>). (Véase figuras 10,11,12).



Iniciar sesión

**Descubre lo que está
pasando en el mundo en
este momento**

Únete hoy a Twitter.

Registrarse

Iniciar sesión

Figura 10: Registro en red social Twitter



Siguiente

Crea tu cuenta

SAyAS

5/50

6666666666

[Usar correo](#)

Figura 11: Formulario de registro Twitter



Figura 12: Cuenta de Twitter SAYAS creada

2. Registro de una API en Twitter

Para crear la aplicación, debemos registrarla en la página: (<https://dev.twitter.com/apps>) y cumplimentar los datos de registro. El nombre de la aplicación que hemos registrado es: SAYAS-2017 (Véase figuras 13, 14, 15).

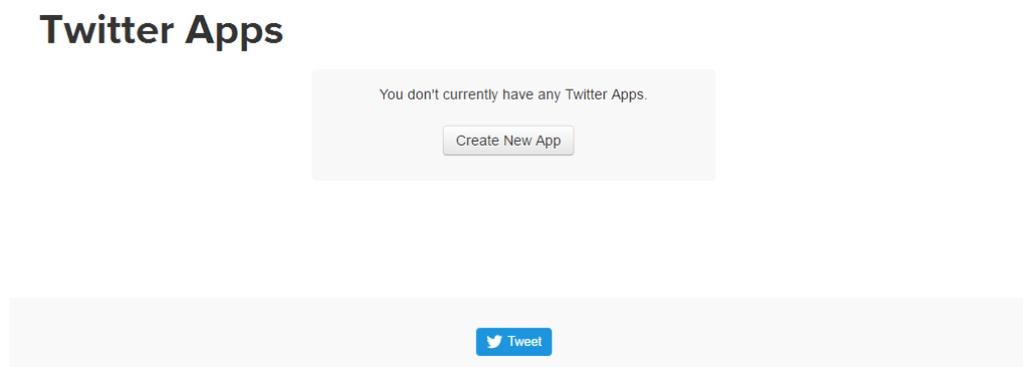


Figura 13 Crear nueva app en la API de Twitter

Create an application

Application Details

Name *
Servicio de alerta
Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *
APP que ofrece un servicio de alerta para las noticias mas relevantes
Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *
https://twitter.com/SAYAS91995364
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later)

Callback URL
Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Figura 14: Formulario de registro de la API de Twitter

Your application has been created. Please take a moment to review and adjust your application's settings.

Servicio de alerta

Test OAuth

Details Settings Keys and Access Tokens Permissions

 APP que ofrece un servicio de alerta para las noticias mas relevantes
<https://twitter.com/SAYAS91995364>

Organization

Information about the organization or company associated with your application. This information is optional.

Organization	None
Organization website	None

Figura 15: Aplicación creada en la API de twitter

3. Autenticación de Twitter con R

Para la autenticación en Twitter, accederemos a la parte de configuración en la API de Twitter y copiaremos los datos de la sección de “Your Access Token”. (Véase figura 16).

Servicio de alerta

Test OAuth

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	[REDACTED]
Consumer Secret (API Secret)	[REDACTED]
Access Level	Read and write (modify app permissions)
Owner	SAYAS91995364
Owner ID	1102083228

Application Actions

[Regenerate Consumer Key and Secret](#) [Change App Permissions](#)

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	[REDACTED]
Access Token Secret	[REDACTED]
Access Level	Read and write
Owner	SAYAS91995364
Owner ID	1102083228

Figura 16: Access Token para la autenticación

El siguiente código, utiliza los datos obtenidos de “Acces Token” para realizar la autenticación en twitter mediante la API. Haremos uso de la función `setup_twitter_oauth()`, en donde le pasaremos nuestros datos para autenticarnos (Veáse en la figura 17).

```
# Cargar parametros de configuracion

reqURL <- "https://api.twitter.com/oauth/request_token"
accessURL <- "https://api.twitter.com/oauth/access_token"
authURL <- "https://api.twitter.com/oauth/authorize"
options(httr_oauth_cache=T)

# Cargar las credenciales obtenidas del paso anterior

consumer_key <-
consumer_secret
access_token <-
access_secret <
options(httr_oauth_cache = TRUE)

# Creamos la conexión a twitter

setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
# streamR authentication

credentials_file <- "my_oauth.Rdata"
if (file.exists(credentials_file)){
  load(credentials_file)
} else {
  cred <- OAuthFactory$new(consumerKey = consumer_key, consumerSecret =
                           consumer_secret, requestURL = reqURL, accessURL = accessURL,
                           authURL = authURL)
  cred$handshake(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl"))
  save(cred, file = credentials_file)
}
```

Figura 17: Código para autenticación con OAuth

Una vez ejecutado el código, se abrirá una pestaña en el navegador, solicitando el acceso a la aplicación, pulsaremos “permitir acceso” y nos devolverá un PIN Este PIN, es el que introducimos por consola en RStudio, y con eso se nos creará el fichero “My_Oauth.data” para el login (Veáse en la figura 18).



Figura 18: PIN para proceso de autorización de la API

4. Construcción del dataframe

Para la construcción de nuestro dataframe, que es, una estructura parecida a una matriz pero que puede contener diferentes tipos de datos. Utilizamos un grupo de funciones: `searchTwitter()`, `twListToDF()`, de las librerías citadas en la sección 1.5.3. Con ello obtenemos un dataframe en donde tenemos los datos ordenados por tweet y usuario que creó dicho tweet (Véase en la figura 19).

```
tweets <- searchTwitter(input$busqueda, n = 100, lang="es",
                        geocode = "28.5,-16.5,10000km")
tweets.df <- twListToDF(tweets)
```

Figura 19: Dataframe de extracción de datos

4.3 Tratamiento de los datos obtenidos

Como los datos obtenidos en la extracción son de diferentes tipos, recibirán un procesamiento para que así, la información pueda manipularse de manera más sencilla. Para ello, realizaremos una serie de pasos: 1. Volcar datos obtenidos del dataframe en un fichero con formato CSV, 2. Limpiar los datos obtenidos en la extracción.

1. **Volcar datos obtenidos del dataframe en un fichero con formato CSV**

Una vez tenemos nuestro dataframe, introduciremos los datos en nuestro fichero para la posterior depuración de los datos. Para ello, realizaremos una comprobación en donde si el fichero no existe pues lo creamos (Véase en la figura 20).

```
if (file.exists(paste(input$busqueda, '_stack.csv')) == FALSE)
  write.csv(tweets.df, file = paste(input$busqueda, '_stack.csv'), row.names = F)
```

Figura 20: Datos en fichero CSV

2. Limpiar los datos obtenidos en la extracción

Los datos almacenados en nuestro fichero, contienen elementos que no nos interesan por lo que es conveniente eliminarlos. Como por ejemplo links, símbolos de puntuación, números y duplicados. Para eliminar los duplicados y obtener unos datos “limpios” haremos uso de un grupo de funciones: `rbind()`, `!duplicated()`, `subset()`, `gsub()`. (Véase en la figura 21 y 22).

```
stack <- read.csv(file = paste(input$busqueda, '_stack.csv'))
stack <- rbind(stack, tweets.df)
stack <- subset(stack, !duplicated(stack$text))
write.csv(stack, file = paste(input$busqueda, '_stack.csv'), row.names = F)
```

Figura 21: Reorganización de datos en fichero .csv y eliminación de duplicados

```

# remueve links
sentence <- gsub('http\\w+', '', sentence)
sentence <- gsub('[:punct:]', '', sentence)
sentence <- gsub('[:cntrl:]', '', sentence)
sentence <- gsub('\\d+', '', sentence)

# Transforma a minuscula
sentence <- tolower(sentence)

# divide en palabras

word.list <- str_split(sentence, '\\s+')

# sometimes a list() is one level of hierarchy too much
words <- unlist(word.list)
words <- enc2utf8(words)

```

Figura 22: Limpieza de datos

4.4 Análisis de los datos

Una vez obtenidos los datos de la red social Twitter y posteriormente de realizar un tratamiento a estos, en donde se realizan distintos procedimientos para así obtener un texto al que se le pueda aplicar el siguiente análisis de sentimientos.

Para realizar dicho análisis, sobre los tweets obtenidos en la extracción de datos, la estrategia empleada, consiste en contrastar cada una de las palabras de cada tweet con las contenidas en los ficheros positive-words.txt [50] y negative-words.txt [51]. Estos archivos, funcionan como una “Base de Datos” los cuales encierran palabras en el idioma “Español” con significado positivo y negativo, respectivamente.

El objetivo que se persigue con esto, es el de establecer una calificación a cada mensaje “tweet” que pueda ser representada. Así pues, las puntuaciones positivas equivalen a tweets cuyos mensajes tienen connotación positiva (mayor a cero), las negativas (menor que cero) tienen connotación negativa y las notas que sean “0” son tweets neutros, los cuales no entran dentro de los dos grupos anteriormente mencionados.

Primero, realizamos la carga de nuestros ficheros de palabras positivas y negativas, en donde pasaremos palabra por palabra a la función `scores()` para su posterior uso. También crearemos un dataset, con los tweets. (Veáse figura 23).

```
# Diccionarios con las palabras positivas y negativas
pos <- scan('positive-words.txt', what = 'character', comment.char = ';')
neg <- scan('negative-words.txt', what = 'character', comment.char = ';')
posWords <- c(pos)
negWords <- c(neg)
Dataset <- stack
Dataset$text <- as.factor(Dataset$text)
```

Figura 23: Diccionario de palabras positivas y negativas

En segundo lugar, contamos el número de las palabras positivas y negativas que contiene cada palabra de un tweet o hashtag. Y retornaremos el valor positivo o negativo dependiendo del resultado que nos de, al realizar la resta. Por ejemplo: si el valor de `sum(pos.matches) = 10` y el valor de `sum(neg.matches) = 3`, cuando hagamos la resta de $10 - 3 = 7$, por lo tanto el valor será positivo. Construimos un dataframe, que contendrá el valor numerico de la variable “score” y los tweets o hashtags. (Veáse figura 24).

```
# Analisis de sentimiento

scoreSentiment <- function(sentences, pos.words, neg.words, .progress='none') {
  scores <- laply(sentences, function(sentence, pos.words, neg.words) {
    pos.matches <- match(words, pos.words)
    neg.matches <- match(words, neg.words)
    pos.matches <- !is.na(pos.matches)
    neg.matches <- !is.na(neg.matches)
    score <- sum(pos.matches) - sum(neg.matches)
    return(score)
  }, pos.words, neg.words, .progress=.progress)

  scores.df <- data.frame(score = scores, text = sentences)
  return(scores.df)
}
```

Figura 24: Cuenta palabras positivas o negativas

Posteriormente, realizamos la evaluación de los tweets o hashtags, mediante la función `mutate()`, en donde analizamos la palabra. Si el valor de “score” es mayor a cero, entonces sera una palabra postiva, si es menor a cero es negativa, en otro caso sera una palabra neutra. Finalmente, agrupamos los tweets con las funciones: `group_by()` y `summarise()`, y escribimos el resultado en el fichero “_opin.csv”.(Veáse figura 25).

```

# Evaluacion
stat <- scores
stat$created <- stack$created
stat$created <- as.Date(stat$created)
stat <- mutate(stat, tweet=ifelse(stat$score > 0, 'positivo',
                                ifelse(stat$score < 0, 'negativo', 'neutro')))

by.tweet <- group_by(stat, tweet, created)
by.tweet <- summarise(by.tweet, number=n())
write.csv(by.tweet, file=paste(input$busqueda, '_opin.csv'), row.names=TRUE)
return(by.tweet)
})

```

Figura 25: Evaluación de los tweets,hashtags

4.5 Alertas enviadas

A la hora de realizar las alertas se ha optado por realizarlas vía correo electrónico. Mediante la utilización de la librería “mailR”, la cual permite el envío de correos electrónicos con lenguaje R (véase en la figura 27).

```

shinyServer(function(input, output) {

  dataInput <- reactive({

    if(input$gmail != "example@gmail.com") {

      recipient = "Recipient 1 ";|
      i = "<";
      d = ">";
      r = paste(i,input$gmail,d,sep = "",collapse = NULL);
      para = paste(recipient,r, sep = "", collapse = NULL);
      # #

      send.mail(from = "alu0100673647@ull.edu.es",
               to = c(para),
               subject = "alerta SAYAS3",
               body = "alerta sobre busqueda en la app https://alu0100673647.shinyapps.io/SAYAS/ ",
               smtp = list(host.name = "aspmx.l.google.com.", port = 25),
               authenticate = FALSE,
               send = TRUE)

    }
  })
}

```

Figura 26: Código para envío de alertas en el servidor

4.6 La aplicación con ShinyApp

En la implementación de la aplicación Web, se ha utilizado la librería “shiny”, que facilita la creación de aplicaciones Web interactivas directamente desde R. También, tiene la posibilidad de extender sus

aplicaciones con temas CSS, HTML widgets y acciones JavaScript. Para el despliegue, se ha utilizado el servicio de hosting que ofrece ShinyApps:

- <https://alu0100673647.shinyapps.io/SAYAS/>

A continuación se muestra la apariencia final de nuestra aplicación Web, denominada “Sistema de Alerta y Análisis de Sentimientos” (SAyAS). Podemos observar que consta de un formulario, una tabla con contenido y una gráfica (Véase en la figura 28 y 29).



Figura 27: Aplicación Sistema de Alerta y Análisis de Sentimientos (SayAS)

```

shinyUI(fluidPage(theme = shinytheme("cerulean"),
  # Application title
  tags$head(
    tags$style(HTML("
      @import url('//fonts.googleapis.com/css?family=Lobster|Cabin:400,700');

      h1 {
        font-family: 'Lobster', cursive;
        font-weight: 500;
        line-height: 1.1;
        color: #00aced;
      }
    )))
),
headerPanel("SAyAS (Sistema de Alerta y Análisis de Sentimientos)"),
sidebarLayout(
  sidebarPanel(
    textInput("busqueda", label = "Búsqueda:", value = "#playa"),
    numericInput("obs", "Número de tweets:", 15, min = 3, max = 20),
    submitButton("Buscar"),
    textInput("gmail", label = "Gmail:", value = "example@gmail.com"),
    submitButton("email")
  ),
  mainPanel(
    tabsetPanel( type = "tabs",
      tabPanel("Tweets analizados", tableOutput("table"))
    ),
    plotOutput('graphic2'),
    plotOutput("histogram")
  )
)
))

```

Figura 28: Estructura SayAS en el servidor

Contemplamos que en la aplicación SAYAS, posee un formulario del lado del cliente, en donde el usuario rellenará dicho formulario con el término que quiera buscar (#hashtag, @usuario, tweet). También tiene la opción de elegir la cantidad que se desea buscar, al igual que si desea recibir una alerta (Véase figura 30 y 31).

Búsqueda:

Número de tweets:

Buscar

Gmail:

email

Figura 29: Formulario de SayAS con la búsqueda del hashtag #playa

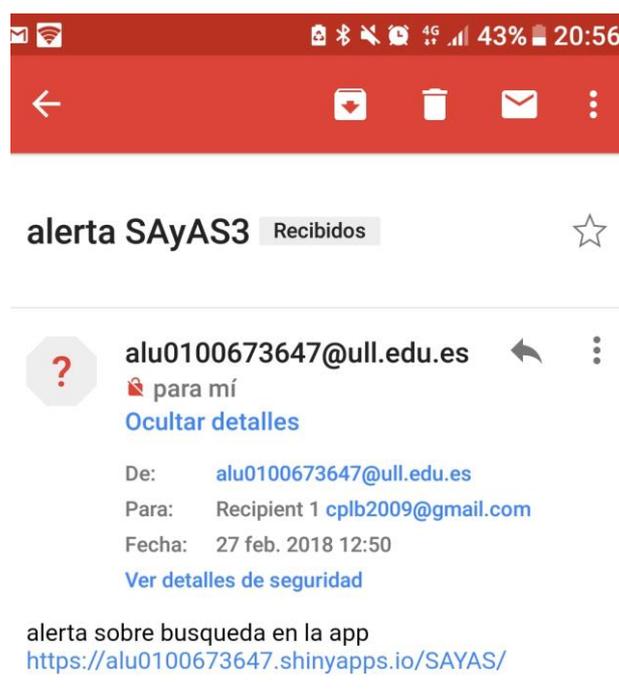


Figura 30: Alerta recibida en el correo

A continuación, se puede apreciar que la aplicación SayAS, cuenta con una tabla de los tweets analizados, en donde, los datos que muestran son los propios tweets de los usuarios así como, el usuario que los ha realizado (Véase en la figura 32 y 33).

Tweets analizados	
Tweet	User
RT @TeneriFeliz: El Verodal (La Frontera, #ElHierro), la #playa más occidental de la Unión Europea, entre las 11 playas españolas únicas po	ElHierroTurismo
Un verano más sin balneario en la playa de la Cicer https://t.co/3Om0L9sBFY #playa	LasCanteras
A muchos nos gusta ir a la playa a relajarnos, descansar, tomar el sol o estar en contacto con la naturaleza, pero https://t.co/AuMDAlOvvy	Sies55Bioecorec
RT @LasCanteras: Comienzan las obras para peatonalizar la calle Luis Morote ...y se forma un caos en el tráfico... evita la zona https	caterf
RT @LasCanteras: Se intentan abrir claros. Hace calor. Gran bajamar al atardecer #Felizmiercoles Todo lo que necesitas saber si hoy vie	camo_estelena
Se intentan abrir claros. Hace calor. Gran bajamar al atardecer #Felizmiercoles Todo lo que necesitas saber si https://t.co/itidsX6Bkd	LasCanteras
El Verodal (La Frontera, #ElHierro), la #playa más occidental de la Unión Europea, entre las 11 playas españolas ún https://t.co/IYLyGLqgA	TeneriFeliz
RT @LasCanteras: Comienzan las obras para peatonalizar la calle Luis Morote ...y se forma un caos en el tráfico... evita la zona https	GustavodeDios
RT @LasCanteras: Comienzan las obras para peatonalizar la calle Luis Morote ...y se forma un caos en el tráfico... evita la zona https	cleosagas1
Comienzan las obras para peatonalizar la calle Luis Morote ...y se forma un caos en el tráfico... evita la zona https://t.co/9hmJjBU9rO	LasCanteras

Figura 31: Tabla con tweets y su usuario correspondiente con la búsqueda del hashtag #playa

```
# Crea la tabla con los tweets

output$table <- renderTable({
  head(dataInput()[, c("Tweet", "User")], n = input$obs)
})
output$value <- renderPrint({ output$table })
```

Figura 32: Tabla en el servidor

Como última sección de SayAS, se distingue una gráfica, en donde podemos visualizar el análisis de sentimientos. Existe una relación entre el eje de las 'X' que son los días sobre los que hacemos el análisis, hemos fijado este valor en los últimos cinco días. Y en el eje de las 'Y' el número máximo de tweets que ha realizado un usuario. Se distinguen 3 colores en la leyenda, en

azul los tweets negativos, en gris los tweets neutros y en verde los positivos (Véase en la figura 34 y 35).

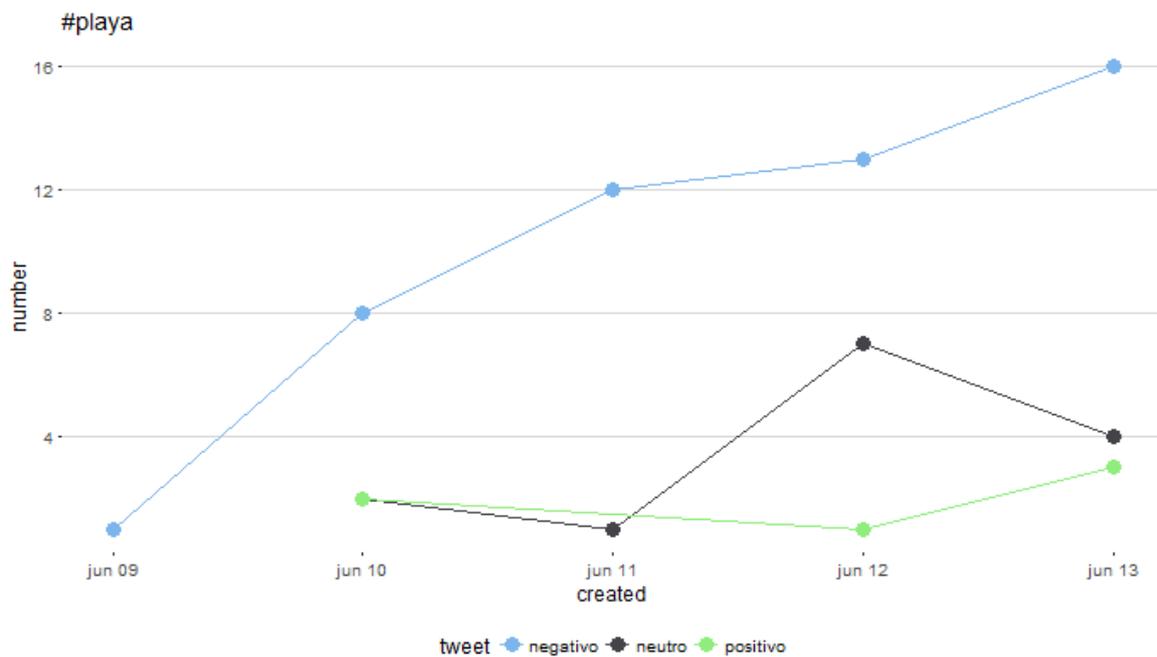


Figura 33: Gráfica de análisis de sentimientos

```
# Crear grafico de analisis

output$graphic2 <- renderPlot({
  data <- analysis()
  ggplot(data,aes(x = created, y = number, group = tweet, color = tweet)) +
    geom_line() +
    geom_point(size = 4) +
    ggtitle(input$busqueda) +
    theme_hc() +
    scale_colour_hc()
})
```

Figura 34: Gráfico de análisis en el servidor

Capítulo 5

Conclusiones y líneas futuras

En este capítulo se describen las conclusiones obtenidas posteriormente a la realización del Trabajo de Fin de Grado. Tiene como propósito, el revisar los objetivos que fueron establecidos en el capítulo 1 para la realización de este proyecto y para comprobar el grado de satisfacción en cada uno de ellos. A posteriori de este análisis, se cotejan algunas líneas futuras para otros posibles Trabajos de Fin de Grado, en los que se pudiera optimizar o ampliar las funcionalidades del sistema de alerta y de análisis de sentimientos como resultado del trabajo presente.

Este sistema se ha basado en una integración de sistemas. En concreto, en un sistema de alerta y un sistema de análisis de sentimientos. Este sistema, denominado “SAyAS (Sistema de Alerta y Análisis de Sentimientos), permite a los usuarios, a través de una aplicación Web, obtener un análisis a cerca de una palabra buscada, así como recibir una notificación de esa búsqueda.

5.1 Conclusiones

En el capítulo 1 de la presente memoria, se han expuesto una serie de objetivos a perseguir, cuyos resultados y conclusiones son los siguientes:

1. Se ha logrado una recopilación lo suficientemente completa y concisa de la evolución de lo que se conoce como Internet, como surgieron las redes sociales, redes colaborativas y su similitud con las redes sociales.
2. Se han analizado diferentes metodologías para los sistemas de alerta, así como para los sistemas de análisis de sentimientos.
3. Se ha diseñado e implementando un sistema de alerta y un sistema de análisis de sentimientos con los datos obtenidos de la red social Twitter.
4. Se ha llevado a cabo la integración en una sola herramienta de un sistema de alerta y un sistema de análisis de sentimientos
5. Se ha propuesto un caso de uso en donde se puede emplear la herramienta desarrollada.
6. Se ha creado la memoria final de este Trabajo de Fin de Grado.

Durante su realización, fue imprescindible estudiar y comprender como funcionan los sistemas de alerta y de análisis de sentimientos para después realizar una integración de ambos. Así como, el procesamiento del lenguaje natural (PLN). Por otro lado, fue importante aprender a usar las herramientas como RStudio, lenguaje R, y comprender muchas de sus librerías como Shiny, twitterR, entre otras.

En este Trabajo de Fin de Grado, se pudo apreciar la potencia que tienen los sistemas de alerta, para generar las notificaciones. En cuanto a los de sistemas de análisis de sentimientos, son sistemas muy complejos, ya que, obtienen un resultado favorable o no dependiendo de la información contrastada.

Estamos en la era de la información y de los grandes volúmenes de datos, todo gira en torno de las redes sociales y la opinión digitaliza. Este recurso es de vital importancia, no solo para obtener información acerca de los usuarios en las redes, sino para conocer a las personas y poder así, procesar dicha información y transformarla en un bien que pueda mover un sector como es el sector turístico en Canarias. Y paralelamente al sector económico.

Construir un sistema de tales características ha sido muy enriquecedor e interesante para mí, la experiencia y conocimientos adquiridos han sido clave, para iniciarme en este mundo del Big Data.

Por lo tanto, como se han cumplido todos los objetivos propuestos, podemos dar por concluido este Trabajo de Fin de Grado.

5.2 Líneas Futuras

Como líneas futuras, se detalla una lista de posibles líneas de trabajo para la integración de sistemas, resultante de este Trabajo de Fin de Grado. Estas consisten en nuevas características y funcionalidades que podrían estudiarse para mejorar y optimizar así el trabajo realizado en este proyecto:

1. **Otros formatos de alerta:** se podría añadir una funcionalidad en donde el usuario además de recibir una alerta vía Gmail, pudiera recibir una alerta vía Twitter o SMS.
 - a. Envío de alerta similar a la búsqueda cada equis tiempo. Junto a una Gráfica con el análisis.

2. **Sistema de geolocalización:** además de los sistemas ya integrados, se podía incluir otro sistema de geolocalización, en donde dependiendo de la ubicación del usuario, pues reciba una alerta del tiempo que pueda hacer en esa zona. Otra funcionalidad de un sistema de geolocalización es la posibilidad de añadir una dirección en donde se quiera ir y que este le indique la experiencia de otros usuarios y el clima en la zona.
3. **Añadir otros idiomas:** como es una aplicación orientada al ámbito de turismo, y como gran parte de los turistas en Canarias son extranjeros, sería una buena idea, añadir la opción de traducir la interfaz y los resultados de la aplicación a otros idiomas como el Inglés, Alemán o Portugués.

Capítulo 6

Considerações finais

Este capítulo narra, as conclusões alcançadas após do desenvolvimento do Trabalho de Final de Grau. Sua finalidade é revisar os objetivos que foram estabelecidos no Capítulo 1 para a realização deste projeto, e verificar o grau de satisfação em cada um deles. Subsequentemente a este análise, se cotejam alguns estudos futuros para outros possíveis Trabalhos de Final de Grau, nos quais as funcionalidades do sistema de alerta e de análise de sentimentos poderiam ser otimizadas ou estendidas como resultado do presente trabalho.

Este sistema foi fundamentado na integração de dois sistemas. Especificamente, em um sistema de alerta e um sistema para a análise de sentimentos. Este sistema, denominado "SAyAS" (Sistema de Alerta e Análise de Sentimentos), permite aos usuários, através de uma aplicação Web, obter uma análise sobre uma palavra pesquisada, assim como receber uma notificação dessa busca.

6.1 Conclusões

No capítulo 1 deste relatório, uma série de objetivos a serem perseguidos foram estabelecidos, cujos resultados e conclusões são os seguintes:

1. Se alcançou uma recompilação suficientemente completa e concisa da evolução do que é conhecido como a Internet, assim como o surgimento das redes sociais, as redes colaborativas e sua similaridade com redes sociais.
2. Se analisaram diferentes metodologias, para os sistemas de alerta, bem como para sistemas de análise de sentimentos.
3. Foram projetados e implementados um sistema de alerta e um sistema de análise de sentimentos com os dados obtidos da rede social Twitter.
4. Foi realizado uma Integração em uma única ferramenta, de um sistema de alerta e um sistema de análise de sentimentos.
5. Foi apresentado um caso de uso onde a ferramenta desenvolvida pode ser usada.
6. Foi documentado o relatório final deste Trabalho de Final de Grau.

Durante a sua realização, foi essencial estudar e entender os sistemas de alerta e de análise de sentimentos para depois integrá-los. Também o processamento de linguagem natural (PLN). Por outro lado, era importante aprender a usar ferramentas como o RStudio, a linguagem R e entender muitas de suas livrarias como Shiny, twitterR, entre outras.

Neste Trabalho de Final de Grau, pudemos ver a potência que os sistemas de alerta têm ao gerar notificações. Em Quanto aos sistemas de análise de sentimentos, são sistemas muito complexos, pois, conseguem obter um resultado favorável ou não, dependendo da informação contrastada.

Estamos na era da informação e grandes volumes de dados. Tudo gira em torno de redes sociais e opiniões digitalizadas. Este recurso é de vital importância, não só para obter informações sobre os usuários nas redes, mas também para conhecer as pessoas, e assim poder processar essas informações e transformá-las em um bem, que possa movimentar um setor como é o setor turístico nas Ilhas Canárias, em paralelo ao setor econômico.

Construir um sistema com essas características foi muito enriquecedor e interessante para mim, a experiência e o conhecimento adquiridos foram fundamentais para começar neste mundo de Big Data.

Portanto, uma vez que todos os objetivos propostos foram cumpridos, pode-se dar por concluído este Trabalho de Final de Grau.

6.2 Estudos futuros

Como estudos futuros, se detalha uma lista de possíveis recomendações para integração de sistemas, resultante deste Trabalho de Final de Grau. Estes consistem em novos recursos e funcionalidades que podem ser estudados para melhorar e otimizar o trabalho realizado neste projeto:

1. **Outros formatos de alerta:** se poderia adicionar uma funcionalidade onde o usuário, além de receber um alerta via Gmail, poderia receber um alerta via Twitter ou SMS.

a. Envio de alerta semelhante à pesquisa de quando em quando. Ao lado de um gráfico com a análise.

2. **Sistema de Geolocalização:** além dos sistemas já integrados, outro sistema de geolocalização pode ser incluído, dependendo da localização do usuário, pois ele recebe um alerta do tempo que pode fazer nessa área. Outra funcionalidade de um sistema de geolocalização é a possibilidade de adicionar um endereço onde você quer ir e que isso indica a experiência de outros usuários e o clima na área.

3. **Adicionar outras línguas:** como é uma aplicação orientada para o campo do turismo, e uma vez que a maioria dos turistas nas Ilhas Canárias são estrangeiros, seria uma boa ideia adicionar a opção de traduzir a interface e os resultados da aplicação, para outras línguas, como Inglês, Alemão ou Português.

Capítulo 7

Presupuesto

En este capítulo se realiza una estimación aproximada del proyecto, en función de las horas dedicadas y de los recursos utilizados para su elaboración. El presupuesto estimado para la realización de este proyecto, se basa en la duración de horas de este, junto a su implementación. Cada hora es estimada en un total de 10,5 euros.

7.1 Presupuesto del trabajo realizado

Tiempo empleado

Objetivos	Descripción	Horas totales	Coste
Estudio del estado del arte	Estudio del estado del arte relacionado con las redes sociales, así como los distintos sistemas de alerta y de análisis de sentimientos en la red social twitter.	120	1.260
Análisis de metodologías	Análisis de métodos para sistemas de alerta y sistemas de análisis de sentimientos utilizando a la red social Twitter como fuente de información.	80	840
Diseño e implementación	Diseñar e implementar un sistema de alerta y un sistema de análisis sentimientos con los datos propuestos	150	1575
Testeo, validación e	Testear, validar e integrar en una sola herramienta, un sistema de	70	735

integración	alerta y un sistema de análisis de sentimientos.		
Caso de uso	Proponer un caso de uso para la herramienta que tendrá integrada un sistema de alerta y un sistema de análisis de sentimientos	50	525
Documentación	Elaboración de documentación técnica, entregables y memoria del proyecto.	140	1.470
Total		610	6405

Tabla 2. Tabla resumen de los Tipos

Hardware utilizado

Dispositivo	Descripción	Coste(€)
ASUS A53S	Ordenador portátil	425

Tabla 3: Presupuesto hardware utilizado

El coste estimado del ordenador portátil, no se suma al presupuesto de tiempo invertido, ya que, ya se disponía de este antes de la realización del proyecto.

Software utilizado

Software	Coste(€)
GitHub	0
Windows 10	135
Google Chrome, Mozilla Firefox, Microsoft Edge	0
Servidor ShinyApps	0

Tabla 4: Presupuesto del software utilizado

El sistema operativo Windows 10, está incluido en el ordenador utilizado para la realización del Trabajo de Fin de Grado, por lo que su coste no afecta al presupuesto real de tiempo invertido.

Presupuesto total del proyecto

Recursos	Coste(€)
Tiempo empleado	6.046
Software	0
Hardware	0
Total	6406

Tabla 5: Tabla resumen del presupuesto

Referencias

- [1] J. M. Cotelo, C. Fermín, F. Ortega y J. A. Troyano, «Explorando Twitter mediante la integración de información estructurada y no estructurada,» *Procesamiento del Lenguaje Natural*, nº 55, 2015.
- [2] M. d. P. Salas Zárate, J. Medina Moreira, K. Lagos Ortiz, H. Luna Aveiga, M. Á. Rodríguez García y R. Valencia García, «Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach,» *Departamento de Informática y Sistemas, Universidad de Murcia, 30100. Universidad de Guayaquil, Cda. Universitaria Salvador Allende, Guayaquil, Ecuador 3 Computational Bioscience and Technology, 4700 KAUST, P.O.Box 2882, Thuwal 23955-6900, Saudi Arabia*, 2017.
- [3] J. Kim y H. Makarand, «Social network analysis,» *International Journal of Information Management: The Journal for Information Professionals* 38.1, pp. 86-96, 2018.
- [4] E. B. Britannica Academic, «Internet,» 2016.
- [5] T. Berners-Lee y R. Cailliau, «WorldWideWeb: Proposal for a HyperText project,» *Retrieved on February*, vol. 26, p. 2008, 1990.
- [6] K. Kapp y T. O'Driscoll, «Learning in 3D,» *Adding a new dimension to enterprise learning and collaboration*, Pfeiffer, San Francisco, CA, 2010.
- [7] G. Domínguez Fernández, «LA EDUCACIÓN SOCIAL Y LA WEB 2.0: NUEVOS ESPACIOS DE INNOVACIÓN E,» *Pixel-Bit. Revista de Medios y Educación*, pp. 105-114, 2009.
- [8] M. d. C. Llorente Cejudo, «Assessing personal learning environments,» *University of Alicante*, 2013.
- [9] T. O'REILLY, *What is web 2.0.*, 2005.
- [10] X. Ribes, «La Web 2.0. El valor de los metadatos y de la inteligencia colectiva,» *Telos*, vol. 73, pp. 36-43, 2007.

- [11] J. J. Flores Cueto, J. J. Morán Corzo y J. J. Rodríguez Vila, «Las redes sociales,» *Universidad de San Martín de Porres*, pp. 1-15, 2009.
- [12] C. Cobo y H. P. Kuklinski, *Planeta Web 2.0.: Inteligencia colectiva o medios fast food*, LMI, 2000.
- [13] L. B. Prato, *Aplicaciones Web 2.0: Redes Sociales*, Buenos Aires: Eduvim-Editorial Universitaria Villa María, 2010.
- [14] J. Cabero, J. Barroso, M. Llorente y V. Marín, *Las redes sociales como instrumentos para la formación. Percepciones de los alumnos universitarios hacia el trabajo en grupo*, Sevilla, España: Grupo de Investigación Didáctica. Universidad de Sevilla, 2013.
- [15] Charles, N. B. ELLISON y C. LAMPE, «Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*,» 2008.
- [16] J. J. De Haro Ollé, «Redes sociales para la educación,» *Anaya Multimedia. Madrid*, 2010.
- [17] J. Cabero, «Las redes sociales en el entramado educativo de la Web. 2.0,» *E. Navas, Web*, vol. 2, pp. 11-28, 2012.
- [18] L. Abramson, «Can social networking keep students in school,» *NPR: Morning edition*, 2011.
- [19] M. S. Granovetter, «The strength of weak ties,» *American journal of sociology*, vol. 78, nº 6, pp. 1360-1380, 1973.
- [20] I. N. De Ledo, «Las redes sociales,» *Revista Venezolana de Oncología*, vol. 23, nº 3, p. 133, 2011.
- [21] «Observatorio Tecnológico,» 2012. [En línea]. Available: <http://recursostic.educacion.es/observatorio/web/en/internet/web-20/1043-redes-sociales?start=3>. [Último acceso: 26 04 2017].
- [22] G. trends, «Google trend Deportes,» <https://trends.google.es/trends/explore?cat=20&q=%2Fm%2F0289n8t,orkut,%2Fm%2F0glpjll>, Todo el mundo.
- [23] G. F. Trends, «Google Trends Facultades y universidades,»

<https://trends.google.es/trends/explore?cat=372&q=%2Fm%2F0289n8t,orkut,%2Fm%2F0glpjl>, Todo el mundo.

- [24] G. G. E. Trends, «Google Trends Gobierno España,» <https://trends.google.es/trends/explore?cat=19&date=today%203-m&geo=ES&q=%2Fm%2F0289n8t,%2Fm%2F0zwk75g,%2Fm%2F0gwzvs1>, España.
- [25] L. M. Camarinha-Matos y H. Afsarmanesh, «Collaborative networks: a new scientific discipline,» *Journal of intelligent manufacturing*, vol. 16, n° 4-5, pp. 439-452, 2005.
- [26] Y. C. Betancourt y R. C. Menéndez, «Sistemas de Alerta Temprana Escolar en caso de inundaciones en la parroquia Panaquire, estado Miranda. Una propuesta educativa ambiental en gesti{\'}n de riesgo,» *REVISTAS DE INVESTIGACIÓN*, vol. 38, n° 82, 2014.
- [27] J. Rosales, F. Campos, V. Córdova, C. Matos, J. Corimanya, W. Palomino y J. Castillo, «UNA PROPUESTA DE MODELO DE DIFUSIÓN DE ALERTA TEMPRANA PARA DESLIZAMIENTOS DE TIERRA EN EL PERÚ USANDO LA TELEVISIÓN DIGITAL TERRESTRE,» *Revista Científica TECNIA*, vol. 25, n° 1, p. 53, 2017.
- [28] H. ALDRICH y D. A. WHETTEN, «Organization-sets, action-sets, and networks: Making the most of simplicity. Handbook of organizational design,» pp. 385-408, 1981.
- [29] S. P. Borgatti y P. C. Foster, «The Network Paradigm in Organizational Research: A Review and Typology,» *Journal of Management*, vol. 29, n° 6, pp. 991-1013, 2003.
- [30] R. D. Franco Pereyra, «Arquitectura, metodología y plataforma tecnológica para la ingeniería y operación de redes colaborativas. Una aproximación basada en servicios digitales,» pp. 41-69, 2012.
- [31] A. Cortez Vásquez, J. Pariona Quispe y A. Huayna, «Procesamiento de Lenguaje Natural,» *Revista de investigación de Sistemas e Informática*, vol. 6, n° 2, pp. 45-54, 2014.
- [32] A. Gelbukh, «Procesamiento de lenguaje natural y sus aplicaciones,» *Korpus*

Sapiens. Sociedad Mexicana de inteligencia artificial, vol. 1, 2010.

- [33] Rstudio, *Rstudio*.
- [34] «Rproject,» [En línea]. Available: <https://www.r-project.org/about.html>. [Último acceso: 2017].
- [35] A. G. y Silvia González, «R Development Core Team,» The address of the publisher, 2000.
- [36] J. Mulero, «"Aplicaciones interactivas diseñadas con Shiny.",» 2016.
- [37] F. J. Lopez-Pellicer, «GitHub como herramienta docente. Actas de las XXI Jornadas de la Enseñanza Universitaria de la Informática,» pp. 66-73, 2015.
- [38] OVTT, *OVTT*.
- [39] L. C. Bolaños Martínez y F. J. C. Caicedo, «Evaluación de la implementación de sistemas automatizados de alerta temprana en tiempo real dirigidos a disminuir los riesgos que representan las amenazas por inundaciones y avenidas torrenciales en la subcuenca Rio Molino, municipio de Popayán, departamento del Cauca,» 2016.
- [40] M. A. A. Coll, «Sistemas de Alerta Temprana (SAT) para la Reducción del Riesgo de Inundaciones Subterráneas y Fenómenos Atmosféricos en el Área Metropolitana de Barranquilla,» *Scientia et Technica*, vol. 18, n° 2, pp. 303-308, 2013.
- [41] A. Esuli y F. Sebastiani, «Determining the Semantic Orientation of Terms Through Gloss Classification,» de *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2005.
- [42] E. Cambria, B. Schuller, W. Liu, H. Wang y C. Havasi, «Knowledge-Based Approaches to Concept-Level Sentiment Analysis,» *IEEE Intelligent Systems*, vol. 28, n° 2, pp. 15-21, 2013.
- [43] F. L. Cruz, J. A. Troyano, F. Enríquez, F. J. Ortega y C. G. Vallejo, «Long autonomy or long delay? The importance of domain in opinion mining,» *Expert Systems with Applications*, vol. 40, n° 8, pp. 3174-3184, 2013.
- [44] A. Moreo, M. Romero, J. Castro y J. M. Zurita, «Lexicon-based comments-

- oriented news sentiment analyzer system,» *Expert Systems with Applications*, vol. 39, n° 10, pp. 9166-9180, 2012.
- [45] M. R. Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez y L. Ureña-López, «Experiments with SVM to classify opinions in different domains,» *Expert Systems with Applications*, vol. 38, n° 12, pp. 14799-14804, 2011.
- [46] I. Peñalver Martínez, «Minería de opiniones basada en características guiada por ontología,» *Proyecto de investigación*, 2015.
- [47] M. d. P. Salas-Zárate, «Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach. Computational and Mathematical Methods in Medicine,» 2017.
- [48] F. Pla y L.-F. Hurtado, «ELiRF-UPV en TASS-2013: Análisis de sentimientos en Twitter,» de *XXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2013)*. TASS, 2013.
- [49] J. M. ;. C. M. F. ;. O. F. J. ;. T. J. J. A. Cotelo, «Explorando Twitter mediante la Integración de Información Estructurada y No Estructurada. Procesamiento del lenguaje natural,» 2015.
- [50] rparatodosxs, «positive.words,» <https://gist.github.com/rparatodxs/29b2bd5a5e4fb265654fd67a83e4cdc4#file-positive-words-txt>, 2017.
- [51] rparatodosxs, «negative-words,» <https://gist.github.com/rparatodxs/7055224448e170761451d683511a8b2c#file-negative-words-txt>, 2017.
- [52] A. C. Vásquez, J. P. Quispe, A. M. Huayna y others, «Procesamiento de lenguaje natural,» *Revista de investigación de Sistemas e Informática*, vol. 6, n° 2, pp. 45-54, 2009.
- [53] J. Musser y T. O'reilly, «Web 2.0,» *Principles and Best Practices.[Excerpt]*. oO: O'Reilly Media, 2006.
- [54] J. M.-M. K. L.-O. H. L.-A. M. Á. R.-G. María del Pilar Salas-Zárate y R. Valencia-García, «Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach, Computational and Mathematical Methods in Medicine,»

2017.

- [55] A. Gelbukh, «Procesamiento de Lenguaje Natural y sus Aplicaciones,» *Komputer Sapiens*, 2010.
- [56] N. B. Ellison y others, «Social network sites: Definition, history, and scholarship,» *Journal of Computer-Mediated Communication*, vol. 13, nº 1, pp. 210-230, 2007.
- [57] D. J. Brass, J. Galaskiewicz, H. R. Greve y W. Tsai, «Taking stock of networks and organizations: A multilevel perspective,» *Academy of management journal*, vol. 47, nº 6, pp. 795-817, 2004.
- [58] J. M. d. N. S. M. L. Silva, «Gestão de informação de desempenho em redes colaborativas de organizações,» 2012.
- [59] J. C. Almenara, J. B. Osuna, M. d. C. L. Cejudo y C. Y. Cabrera, «Redes sociales y Tecnologías de la Información y la Comunicación en Educación: aprendizaje colaborativo, diferencias de género, edad y preferencias,» *Revista de Educación a Distancia*, nº 51, 2016.
- [60] M. d. C. Llorente Cejudo, «Assessing personal learning environments (PLEs). An expert evaluation,» 2013.
- [61] C. C. Romaní y H. P. Kuklinski, «PLANETA WEB 2.0,» *Inteligencia colectiva o medios fast food. Group de Recerca d'Interaccions Digitals, Universitat de Vic*, 2007.
- [62] <https://cran.r-project.org/>, «CRAN».