



Universidad Autónoma de
Querétaro
Facultad de Informática
Licenciatura en Informática

Título del tema de tesis registrado

**Análisis de contenido generado en Twitter y desarrollo de un módulo social analítico para la
detección de oportunidades de negocio**

Que como parte de los requisitos para obtener el Grado de

Licenciado en Informática

Presenta

Reiji Takane Paredes

Dirigido por:

M.S.I José Alejandro Vargas Díaz

M.S.I José Alejandro Vargas Díaz
Presidente

M.S.I.D Carlos Alberto Olmos Trejo
Secretario

L.I Anabel Palacios Martínez
Vocal

M.S.I Eduardo Aguirre Caracheo
Suplente

M.S.I Diego Octavio Ibarra Corona
Suplente

Centro Universitario, Querétaro, Qro. Mayo 2019, México

RESUMEN

La aparición del internet ha dejado una huella positiva en la sociedad, permitiendo el acceso a grandes fuentes de información cultural, científica, cursos educativos, entretenimiento y compra de artículos. A su vez, diversas redes sociales han emergido y la interacción con estas se ha vuelto parte esencial de nuestra vida cotidiana, a tal nivel que las acciones que se lleven a cabo dentro de estas tienen impactos en la sociedad a escala global. Muchas empresas han decidido conocer la opinión del consumidor final mediante las redes sociales, sin embargo, el volumen de comentarios es tan grande que les es imposible realizar una revisión, adicionalmente la comunidad es influenciada por diversos editores, analistas y tech partners especializados, dedicados a evaluar productos o servicios de un área en específico. Como resultado a estos factores surge el concepto llamado Social CRM. Social CRM es una estrategia que es apoyada por una variedad de herramientas y tecnologías con el fin de desarrollar negocios colaborativos. Diversos softwares CRM han empezado a implementar este concepto, sin embargo, estos no aprovechan la información que obtienen de los medios sociales. En la presente investigación se analizaron distintos CRM con el objetivo de desarrollar una herramienta enfocada al Social CRM.

(Palabras clave: Sistemas de Información, Marketing, Administración, Redes Sociales, CRM, Social CRM, Análisis de Texto)

ABSTRACT

Internet's invention has contributed in many ways to society, granting unlimited access to scientific, cultural, educational and entertainment content, also, a wide variety of social networking sites emerged. Interaction with social networking sites have increased in a way that every action made, generates effects in the real world. In order to improve product quality and service, many companies have tried to reach product consumers via social networking sites but with the continuous generation of user content, it is not feasible to review the large volumes of content, also, various product review specialists influence in product consumer choices. As a result of all this factors, Social CRM emerged. Social CRM is a strategy which makes use of a variety of tools in order to develop collaborations. Many CRM companies have began implementing social network compatibility, however, they don't fully exploit the information they get from social networking sites. The following research reviews different CRM systems in order to develop Social CRM oriented tool.

Keywords: Information Technology, Marketing, Social Networking, CRM, Social CRM, Text Analysis

DEDICATORIA

A mi mamá, que siempre me ha apoyado a realizar mis metas e invitarme a realizar actividades nuevas.

A Verónica, Michell, Nadia, Brenda y Lupita, por acompañarme durante toda la carrera y por su apoyo.

A Yasmin, Fernanda, Andrea, Jheus y Rodrigo por el tiempo que brindaron para escuchar mis ideas y jugar UNO.

A Whitey por brindarme momentos de risa y compañía durante la redacción de este documento.

Dirección General de Bibliotecas UAQ

AGRADECIMIENTOS

Al M.S.I. José Alejandro Vargas Díaz por el apoyo brindado desde el curso propedéutico hasta el día de hoy.

Al M.I.S.D. Carlos Alberto Olmos Trejo por el apoyo y paciencia brindada durante la revisión y reescritura del presente documento.

A la L.A. Anabel Palacios Martínez por el apoyo y guía brindada desde el inicio de la carrera, así como por forjar parte de mis conocimientos e impulsar mi desarrollo personal y profesional.

Al M.S.I. Eduardo Aguirre Caracheo y al M.S.I. Diego Octavio Ibarra Corona por brindarme consejos e ideas para el desarrollo de nuevos proyectos, así como su participación en la revisión de este documento.

Dirección General de Bibliotecas UAO

TABLA DE CONTENIDOS

1	INTRODUCCIÓN.....	10
2	OBJETIVOS.....	11
2.1	Temas a tratar.....	11
2.2	Generales.....	11
2.3	Específicos.....	11
2.4	Productos tangibles obtenidos.....	11
3	REDES SOCIALES.....	12
3.1	Breve historia de las Redes Sociales.....	12
3.2	Redes Sociales en México.....	13
3.3	Impacto del social media en México.....	15
3.4	Social Media Marketing.....	16
3.5	Social Influence Marketing.....	17
3.6	Crisis en los medios sociales.....	18
3.7	Conclusión del capítulo.....	21
4	CRM.....	22
4.1	Breve Historia del CRM.....	22
4.2	Análisis de CRM Tradicional y CRM Social.....	23
4.3	La integración social de Skyline Boston.....	28
4.4	Conclusión del Capítulo.....	28
5	INTEGRACIÓN DE REDES SOCIALES EN DIVERSOS CRM.....	28
5.1	Nimble.....	29
5.2	Zoho.....	35
5.3	vTiger.....	36
5.4	Conclusión del capítulo.....	37

6	EXTRACCIÓN Y MINERÍA DE TEXTO.....	37
6.1	Minería de texto	39
6.2	Minería de Opiniones.....	40
7	PROPUESTA DE MÓDULO	47
7.1	Requisitos	50
7.1.1	Requisitos Funcionales	50
7.1.2	Requisitos no funcionales, Interfaces externas.....	54
7.1.3	Requisitos no Funcionales Interfaz de Usuario	55
7.1.4	Requisitos de rendimiento	55
7.2	Diagrama de clases.....	56
7.3	Interfaz de usuario	59
8	CASO PRÁCTICO	61
9	CONCLUSIONES.....	79
10	BIBLIOGRAFÍA.....	80

ÍNDICE DE TABLAS

TABLA. 7-1 REQUISITO FUNCIONAL 1.....	50
TABLA. 7-2 REQUISITO FUNCIONAL 2.....	50
TABLA. 7-3 REQUISITO FUNCIONAL 3.....	50
TABLA. 7-4 REQUISITO FUNCIONAL 4.....	51
TABLA. 7-5 REQUISITO FUNCIONAL 5.....	51
TABLA. 7-6 REQUISITO FUNCIONAL 6.....	51
TABLA. 7-7 REQUISITO FUNCIONAL 7.....	51
TABLA. 7-8 REQUISITO FUNCIONAL 8.....	51
TABLA. 7-9 REQUISITO FUNCIONAL 9.....	52
TABLA. 7-10 REQUISITO FUNCIONAL 10.....	52
TABLA. 7-11 REQUISITO FUNCIONAL 11.....	52
TABLA. 7-12 REQUISITO FUNCIONAL 12.....	52
TABLA. 7-13 REQUISITO FUNCIONAL 13.....	52
TABLA. 7-14 REQUISITO FUNCIONAL 14.....	52
TABLA. 7-15 REQUISITO FUNCIONAL 15.....	53
TABLA. 7-16 REQUISITO FUNCIONAL 16.....	53
TABLA. 7-17 REQUISITO FUNCIONAL 17.....	53
TABLA. 7-18 REQUISITO FUNCIONAL 18.....	53
TABLA. 7-19 REQUISITO FUNCIONAL 19.....	53
TABLA. 7-20 REQUISITO FUNCIONAL 20.....	54
TABLA. 7-21 REQUISITO FUNCIONAL 21.....	54
TABLA. 7-22 REQUISITO NO FUNCIONAL 1.....	54
TABLA. 7-23 REQUISITO NO FUNCIONAL 2.....	54
TABLA. 7-24 REQUISITO NO FUNCIONAL 3.....	54
TABLA. 7-25 REQUISITO NO FUNCIONAL 3.....	55

TABLA. 7-26 REQUISITO NO FUNCIONAL 4.....	55
TABLA. 7-27 REQUISITO NO FUNCIONAL 5.....	55
TABLA. 7-28 REQUISITO NO FUNCIONAL DE INTERFAZ DE USUARIO 1.....	55
TABLA. 7-29 REQUISITO NO FUNCIONAL DE INTERFAZ DE USUARIO 2.....	55
TABLA. 7-30 REQUISITO DE RENDIMIENTO 1.....	56
TABLA. 7-31 REQUISITO DE RENDIMIENTO 2.....	56
TABLA. 7-32 REQUISITO DE RENDIMIENTO 3.....	56
TABLA. 7-33 REQUISITO DE RENDIMIENTO 4.....	56
TABLA. 8-1 INFORMACIÓN DE RESULTADOS GENERAL.....	68
TABLA. 8-2 LISTA DE TWEETS CON TFIDF ALTO.....	70
TABLA. 8-3 HASHTAGS CON TFIDF ALTO.....	72
TABLA. 8-4 USUARIOS CON TFIDF ALTO.....	74
TABLA. 8-5 TÉRMINOS CON TFIDF ALTO.....	76

ÍNDICE DE FIGURAS

FIGURA. 3-1 CLASIFICACIÓN DE LOS MEDIOS SOCIALES. FUENTE: KAPLAN.....	12
FIGURA. 3-2 HÁBITOS DEL USO DE INTERNET. FUENTE: ESTUDIO HÁBITOS DEL USUARIO 2016. AMIPCI.....	13
FIGURA. 3-3 PENETRACIÓN REDES SOCIALES. FUENTE: ESTUDIO HÁBITOS DEL USUARIO 2016. AMIPCI.....	15
FIGURA. 3-4 EL ECOSISTEMA DE LOS MEDIOS SOCIALES. (SCHULTZ, 2007).....	16
FIGURA. 3-5 RESPUESTA DE NESTLÉ A LOS COMENTARIOS EN REDES SOCIALES. RECUPERADO DE HTTP://WWW.CBSNEWS.COM/NEWS/NESTLES-FACEBOOK-PAGE-HOW-A-COMPANY-CAN-REALLY-SCREW-UP-SOCIAL-MEDIA/	19
FIGURA. 3-6 TWEET DE HAWKERS MÉXICO ANTE LAS ELECCIONES. (GALEANO, 2016). RECUPERADO DE HTTP://MARKETING4ECOMMERCE.NET/POLEMICA-DE-HAWKERS-EN-MEXICO/	20
FIGURA. 3-7 TWEET DE SERGIO CONTRA HAWKERS. RECUPERADO DE HTTPS://TWITTER.COM/SCHECOPEREZ/STATUS/796476888390701056?REF_SRC=TWSRC%5ETFW	20
FIGURA. 3-8 TWEET DE DIABLOS ROJOS CONTRA HAWKERS. RECUPERADO DE HTTPS://TWITTER.COM/DIABLOSROJOSMX/STATUS/796511534180696065?REF_SRC=TWSRC%5ETFW	21
FIGURA. 4-1 PROCESO DEL CRM SOCIAL. FUENTE. (MORGAN. J, 2010).....	26
FIGURA. 5-1 MATRIZ DE CRM. EXTRAÍDO DE HTTPS://WWW.G2CROWD.COM/CATEGORIES/CRM?UTF8=%E2%9C%93&ORDER=TOP_SHELF	29
FIGURA. 5-2 VISUALIZACIÓN DE FOLLOWERS EN EL CRM.....	30
FIGURA. 5-3 VISUALIZACIÓN DE INFORMACIÓN DE CONTACTO.....	30
FIGURA. 5-4 VISUALIZACIÓN DE LA LÍNEA DE TIEMPO.....	31
FIGURA. 5-5 LISTA DE CONTACTOS QUE SIGUE EL USUARIO.....	32
FIGURA. 5-6 LISTA DE SEGUIDORES DEL CONTACTO.....	33
FIGURA. 5-7 LISTA DE ACTIVIDAD ENLAZADA AL CRM.....	34
FIGURA. 5-8 PERMISOS DEL USUARIO.....	35
FIGURA. 5-9 TABLA DE CONTENIDOS.....	35
FIGURA. 5-10 TABLA DE MENSAJES.....	36
FIGURA. 5-11 BÚSQUEDA DE STREAMS.....	36
FIGURA. 5-12 TABLA DE STREAMS.....	36
FIGURA. 6-1 PROCESO BUSINESS INTELLIGENCE. EXTRAÍDO DE HTTPS://WWW.INTECHOPEN.COM/SOURCE/HTML/18527/MEDIA/IMAGE8.JPEG	37
FIGURA. 6-2 MODELO DE COMUNICACIÓN REST EXTRAÍDO DE HTTPS://DEV.TWITTER.COM/STREAMING/OVERVIEW	38
FIGURA. 6-3 MODELO DE COMUNICACIÓN STREAMING EXTRAÍDO DE HTTPS://DEV.TWITTER.COM/STREAMING/OVERVIEW	39
FIGURA. 6-4 PROCESO DE ANÁLISIS DE TEXTO.....	41
FIGURA. 6-5 ETIQUETARIO EMPLEADO EN GRAMPAL. CONSULTADO EL 21 DE MAYO 2017. EXTRAÍDO DE HTTP://CARTAGO.LLLF.UAM.ES/GRAMPAL/GRAMPAL.CGI?M=ETIQUETARIO	42

FIGURA. 6-6 CATEGORIZACIÓN LÉXICA DE LA ORACIÓN. EXTRAÍDO DE HTTP://LICEU.UAB.CAT/~JOAQUIM/LANGUAGE_TECHNOLOGY/NLP/PLN_ANALISIS.HTML#TOKENIZACI_N ..	43
FIGURA. 6-7 ESTRUCTURA DE UN ÁRBOL CON ANÁLISIS SUPERFICIAL. PADRÓ, L. (2016).....	43
FIGURA. 6-8 ESTRUCTURA DE UN ÁRBOL CON ANÁLISIS PROFUNDO. PADRÓ, L. (2016).	44
FIGURA. 6-9 ANÁLISIS DE DEPENDENCIAS.....	44
FIGURA. 6-10 ROLES SEMÁNTICOS.....	45
FIGURA. 7-1 <i>DESCRIPCIÓN GRÁFICA DEL ÍNDICE INVERTIDO. (STANFORD, 2008)</i>	48
FIGURA. 7-2 DIAGRAMA DE CLASES UTILIZADAS EN HADOOP.....	57
FIGURA. 7-3 DIAGRAMA DE CLASES MAP-REDUCE.	58
FIGURA. 7-4 DIAGRAMA DE CLASES CORRESPONDIENTE A COMPONENTES DEL DRIVER.	59
FIGURA. 7-5 INTERFAZ DE USUARIO.....	60
FIGURA. 8-1 ARCHIVOS GENERADOS POR EL EXTRACTOR.	61
FIGURA. 8-2 DIAGRAMA DE FLUJO PARA UNIR ARCHIVOS JSON.	62
FIGURA. 8-3 ESTRUCTURA DE UN TWEET.	63
FIGURA. 8-4 DIAGRAMA DE FLUJO PARA SEPARAR LOS VALORES EN DIFERENTES ARCHIVOS.....	64
FIGURA. 8-5 ENTRADA Y SALIDA DEL PRIMER PASO.	65
FIGURA. 8-6 ENTRADA Y SALIDA DEL SEGUNDO PASO.....	66
FIGURA. 8-7 ÍNDICE DE FRECUENCIA DE TÉRMINOS ANTES Y DESPUÉS DE LA FASE DE FILTRADO.	67
FIGURA. 8-8 GRAFICA DE TWEETS DESTACADOS POR TFIDF.....	71
FIGURA. 8-9 GRÁFICA DE HASHTAGS DESTACADOS.....	72
FIGURA. 8-10 NUBE DE HASHTAGS.	73
FIGURA. 8-11 GRÁFICA DE USUARIOS DESTACADOS.....	74
FIGURA. 8-12 NUBE DE USUARIOS MENCIONADOS.....	75
FIGURA. 8-13 GRÁFICA DE TÉRMINOS DESTACADOS.....	77
FIGURA. 8-14 NUBE DE TÉRMINOS.....	78

1 INTRODUCCIÓN

Con la aparición de las redes sociales la interacción entre los consumidores y las empresas ha cambiado, cada día, nuevos canales de comunicación como las redes sociales han surgido a tal grado que las empresas han asignado puestos de trabajo dedicados al manejo de estas redes.

De acuerdo a la plataforma internet live stats, se estima que cada día se generan 500 millones de tweets al día en todo el mundo. En estas redes sociales diversos usuarios comparten experiencias de productos o servicios que han utilizado, esto ha llevado a las empresas a tratar de obtener retroalimentación para mejorar sus productos, sin embargo, el volumen de comentarios es tan grande que les es imposible realizar una revisión.

Diversas empresas dedicadas al desarrollo de plataformas CRM han implementado módulos que permiten la conexión a las redes sociales para facilitar la interacción de las empresas con sus clientes, dando como resultado el CRM Social. Sin embargo, dado el volumen que se genera diariamente por estos medios, es imposible obtener o realizar un análisis con las herramientas que se tienen.

Con base a esta problemática, se busca analizar distintos CRM con el objetivo de desarrollar una herramienta enfocada al Social CRM.

2 OBJETIVOS

2.1 Temas a tratar

En la presente investigación, los temas principales que se contemplarán son los siguientes:

- Redes Sociales
- CRM
- Minería de Opinión

2.2 Generales.

Desarrollar un módulo social de CRM que ayude al mercado en la detección de necesidades de consumo, lugares y preferencias en el Estado de Querétaro, mediante el análisis de texto, con el fin de aportar datos que permitan detectar oportunidades de negocio.

2.3 Específicos.

- Realizar un análisis de las plataformas populares de CRM.
- Realizar un análisis de las plataformas de monitoreo de Twitter.
- Buscar metodologías de análisis de sentimientos para detectar el sentido de un mensaje expresado dentro de una red social.
- Desarrollo de un prototipo de un módulo social analítico que ayude a la detección de oportunidades de negocio.

2.4 Productos tangibles obtenidos.

- Prototipo de un módulo social de CRM.
- Reporte de actividad de una tendencia generada en Twitter.

3 Redes Sociales

Kaplan (2009) define los medios sociales (Social Media) como un grupo de aplicaciones web construidas con base a los fundamentos tecnológicos e ideológicos de la Web 2.0 y permiten la creación e intercambio del contenido generado por el usuario. Los medios sociales se dividen en 6 categorías considerando la presencia social (Social presence), flujo de información (Media richness), el grado de influencia a otros usuarios (Self-Presentation) y el grado de privacidad (Self-Disclosure):

		Social presence/ Media richness		
		Low	Medium	High
Self-presentation/ Self-disclosure	High	Blogs	Social networking sites (e.g., Facebook)	Virtual social worlds (e.g., Second Life)
	Low	Collaborative projects (e.g., Wikipedia)	Content communities (e.g., YouTube)	Virtual game worlds (e.g., World of Warcraft)

Figura. 3-1 Clasificación de los medios sociales. Fuente: Kaplan.

Según Boyd (2008) las redes sociales son servicios web que permiten a individuos:

Construir un perfil público o semi público en un sistema enlazado; Visualizar una lista de usuarios con los que comparte una relación; Visualizar las conexiones que ha realizado él y otros dentro del sistema. Cohen (2009) las define como un sistema para compartir información y servicios entre individuos o grupos que tienen un interés en común. La nomenclatura y naturaleza de estas conexiones varía de sitio a sitio (Boyd, 2008).

3.1 Breve historia de las Redes Sociales

SixDegrees.com fue la primera red reconocida bajo las definiciones mencionadas anteriormente, esta página permitía a los usuarios crear perfiles y visualizar una lista de sus contactos (Boyd, 2008). Diversos servicios de redes sociales empezaron soportando nichos demográficos antes de expandirse a un mayor público. MiGente, AsianAvenue, BlackPlanet surgieron entre los años 1997 – 2001, que permitían crear perfiles personales, profesionales y de citas. En mayo de 2003, LinkedIn es abierta al público con un crecimiento lento, el primer mes empezando con 2,500 usuarios y en el tercer bimestre con 37,000 usuarios. Dos años después tenía más de 1.7 millones de profesionales registrados (Ken, 2013). En ese mismo año MySpace empieza a competir con sitios como Friendster, Xanga y AsianAvenue. Debido a que algunas bandas de rock

que fueron expulsadas de Friendster, ellos, empezaron a utilizar MySpace para promocionar pases VIP, esto provocó que adolescentes empezaran una migración a MySpace. Mientras MySpace acaparaba la atención en los Estados Unidos, gran variedad de redes sociales proliferaba en el mundo. Orkut en Brasil, Hi5 en América Latina, diversos servicios de comunicación y blogging empezaron a implementar funciones de redes sociales. Facebook empezó a inicios del 2004 como una red exclusiva para la Universidad de Harvard, en septiembre 2005 Facebook abrió sus puertas para estudiantes de preparatoria, empresarios y eventualmente a todos. En Marzo de 2006 Twitter es abierto al público y en febrero de 2012 anuncia servicio de publicidad para pequeños negocios basado en tweets promocionales. De acuerdo con el sitio web Alexa y la AMIPCI (2015) las redes sociales más utilizadas en México son Facebook, LinkedIn, Instagram, Twitter, Pinterest y Youtube.

3.2 Redes Sociales en México

Al año 2015 la población internauta es del 59.8% de la población, es decir, 65 millones de personas, con un tiempo promedio diario de conexión de 7 horas y 14 minutos. En las actividades online realizadas el 79% de los internautas acceden a redes sociales. En promedio un internauta está inscrito a 5 redes sociales. (AMIPCI, 2015).

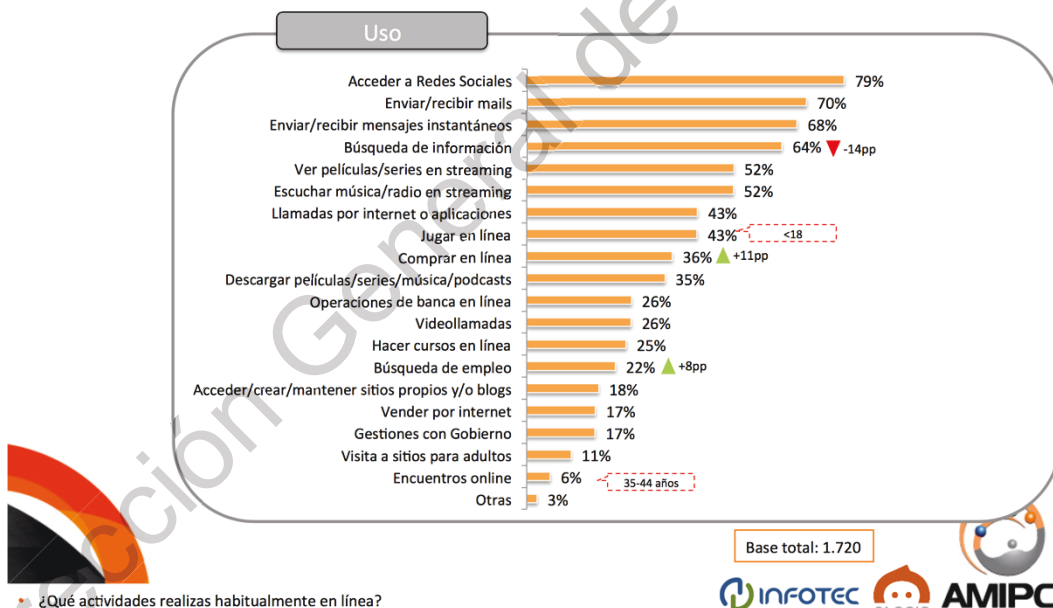


Figura. 3-2 Hábitos del uso de internet. Fuente: Estudio hábitos del usuario 2016. AMIPCI

A continuación, se describen algunas de las redes sociales más utilizadas en México (AMIPCI, 2016), cada una con un segmento de usuarios específico.

Facebook

Es una red que conecta a amigos, familiares y socios de negocios. Permite buscar amigos, compartir recursos e incluso crear encuestas. Todas estas actividades se registran en la línea de tiempo de la biografía del usuario y son visibles para más o menos gente. (Rubín, 2016).

Youtube

Es la comunidad más grande de videos que hay en internet, en ella los usuarios pueden distribuir contenido creado por el usuario y ofrece un foro para que los usuarios se conecten, informen e inspiren. (Youtube, sf.).

Twitter

Es una red social a base de mensajes cortos de no más de 280 caracteres, permitiendo agregar un video o imagen. Permite suscribirse a los Tweets de los usuarios y viceversa. (Rubín, 2016).

Instagram

Permite tomar fotos, transformarlas y compartir en diversas redes sociales o enviar por correo electrónico. También permite conocer y compartir las imágenes publicadas por amigos y conocidos. (Rubira, 2013).

LinkedIn

Permite conectar usuarios con profesionales de un sector específico y crear un currículum vitae digital teniendo en cuenta aspectos como estudios, experiencia, áreas de especialización y conocimientos (Florida, 2016).

Penetración Redes Sociales

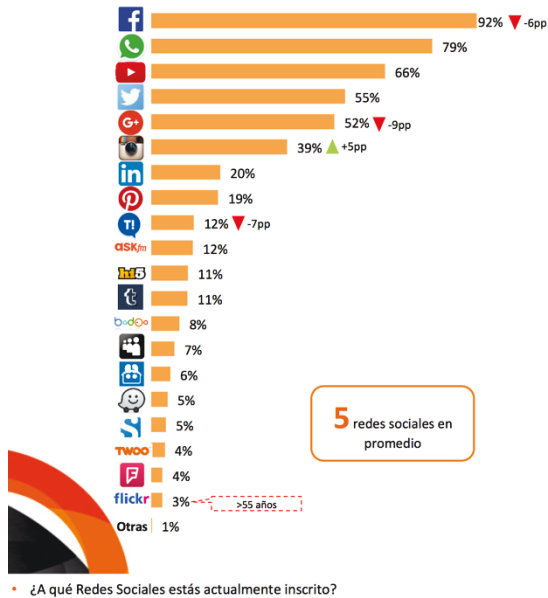


Figura. 3-3 Penetración Redes Sociales. Fuente: Estudio hábitos del usuario 2016. AMIPCI.

3.3 Impacto del social media en México

El impacto del social media en México es tan grande que ha sido utilizado para mover masas. El movimiento #yosoy132 surge a raíz de una protesta el 11 de mayo en el campus Ciudad de México de la Universidad Iberoamericana contra el candidato presidencial Enrique Peña Nieto. El político fue cuestionado por los estudiantes de la institución en referencia a los disturbios y represión que tuvo cuando era gobernador del Estado de México. El candidato tuvo que abandonar la Universidad rodeado por un cordón de seguridad, este acontecimiento fue grabado por varios estudiantes e inmediatamente publicados en las redes sociales alcanzando gran difusión. La cobertura de los hechos por parte de los medios tradicionales de comunicación y declaraciones de políticos indignaron más a los estudiantes. Estos fueron calificados de violentos y se les pretendió desvincularlos de la comunidad estudiantil negando su adscripción a la universidad. Ante estas acusaciones 131 estudiantes publicaron un video donde mostraban sus identificaciones que los acreditaban como estudiantes de dicha universidad, desmintiendo las acusaciones. A raíz del video la frase "131 Alumnos de la Ibero" se convirtió en el tema más comentado en México y en el mundo en la red social Twitter. Posteriormente surge el hashtag #yosoy132 que expresa solidaridad con los 131 estudiantes del video y da nombre al movimiento. El 18 de mayo los estudiantes convocaron una marcha contra el sesgo informativo que partió de la Universidad Iberoamericana hasta las oficinas de Televisa Santa Fe. Al día siguiente se convoca a través de las redes sociales una marcha contra el candidato Peña Nieto en varias ciudades del país reuniendo a unas 46,000 personas en el Distrito Federal (Candón. M, 2013).

3.4 Social Media Marketing

Existen diversos factores por los que está cambiando las comunicaciones de marketing. En esta era digital los consumidores están mejor informados, pueden utilizar el internet y otras tecnologías para encontrar información que no le pertenece al proveedor, pueden comunicarse con más facilidad con otros consumidores para compartir experiencias acerca de un producto. (Kotler. P, 2014).

El ecosistema se centra en la experiencia del consumidor, el concepto de un ecosistema de medios sociales permite al marketing pensar en una estrategia general con el fin de identificar objetivos, crear contenido adecuado y cómo propagarlo en el ecosistema (Hanna. R, 2011).

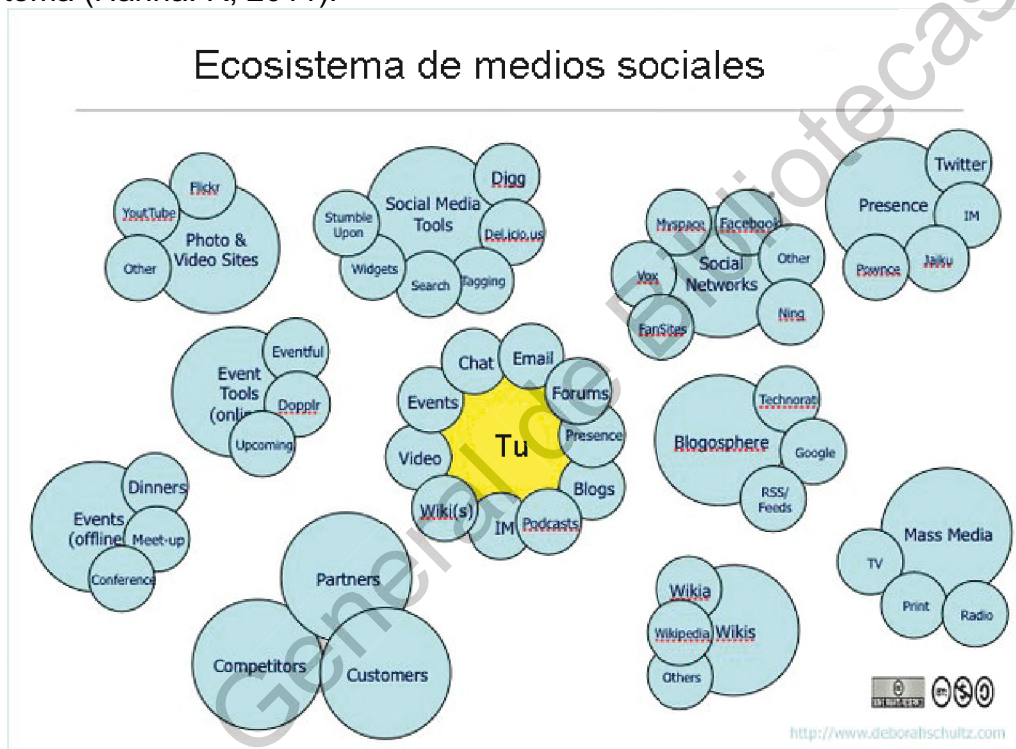


Figura. 3-4 El ecosistema de los medios sociales. (Schultz, 2007).

Corcoran (2009) divide el ecosistema en tres tipos de medios: medios propios *Owned Media*; medios de paga *Paid Media* y medios ganados *Earned Media*.

Los medios propios son los canales que se pueden controlar, existen los medios propios y los medios parcialmente propios. Los medios propios crean portabilidad en la marca. Medios ganados es el contenido creado de los clientes, es esencial aprender a cómo escuchar y responder a los medios positivos y negativos. Medios de paga se refiere a los anuncios que están en diversos canales de comunicación.

Li, Bernoff (2008) segmenta los participantes del ecosistema en cinco tipos de comportamientos diferentes: Creadores, Críticos, Recolectores, Conectores y

Espectadores. Los participantes en el ecosistema interactúan de forma distinta en cada plataforma y entre ellos (Hanna. R, 2011).

3.5 Social Influence Marketing

Social Influence Marketing es una técnica de marketing que emplea los medios sociales como los vlogs, redes sociales, podcasts y wikis con el fin de promocionar un servicio. Los medios sociales son el contenido creado por y para el consumo de los usuarios de una red social, están conformados por comentarios, fotografías y conversaciones que publican. (S.Singh, 2010).

Para comprender cómo funcionan los influenciadores, recordemos que las personas buscan opiniones y puntos de vista al momento de buscar un producto, algunas veces la compra de una persona inspira a que otra compre el producto, en especial si esa persona realiza comentarios positivos de acerca de su compra. Los influenciadores sociales son personas que persuaden a una persona a comprar un producto, de acuerdo al contexto, el influenciador social puede ser un familiar, amigo, compañero de trabajo e incluso alguien que el individuo no conoce. (S.Singh, 2010).

Tres tipos de influenciadores existen:

Influenciadores Referentes

Son usuarios que participan activamente en las plataformas sociales compartiendo videos o comentarios acerca de su experiencia con un producto. Estos son los usuarios que más influyen en la decisión de compra (S.Singh, 2010).

Influenciadores Expertos

También conocidos como conocidos como *key influencers*, son autoridades de un área de productos específicos. Estos usuarios de redes sociales poseen blogs y gran cantidad de seguidores en sus redes sociales (S.Singh, 2010). Ellos representan una fuente de información para usuarios desinformados que desean realizar una compra de un bien que puede representar un riesgo alto (T.L.Childers, 1992).

Influenciadores Posicionales

Los influenciadores posicionales son las personas más cercanas al consumidor y a la decisión de compra, ellos pueden persuadir directamente al consumidor en el momento de la compra. Podemos clasificarlos en dos subtipos de influenciadores, que

determinan la preferencia y decisión compra de acuerdo al comportamiento que este desarrolle (S.Singh, 2010):

Influenciadores Familiares

Miembros de la familia son usualmente de más importancia al individuo a causa de que se identifican con estos. La familia incluye un impacto en las normas, actitudes, y valores del individuo. La comunicación intrafamiliar puede influenciar las preferencias por marcas, búsqueda de información, sensibilidad de precios y su adherencia a los principios de precio-calidad, estos patrones varían de acuerdo al tipo de familia (T.L.Childers, 1992).

Influencia de Amistades

A medida que el individuo avanza hacia su adolescencia, la influencia parental tiende a ser menos significativa. Este desarrolla conexiones y relaciones con sus amistades y compañeros. Gran parte del tiempo libre es invertido en la compañía de sus compañeros, valorándolos como personas de confianza con los que comparten valores, actitudes e intereses (T.L.Childers, 1992).

3.6 Crisis en los medios sociales

Una crisis en los medios sociales se define como una situación negativa que se origina o se amplifica en un medio social y tiene como resultado pérdidas financieras, cambios en el proceso de negocio o contenido negativo en los medios. (Owyang. J, 2011). Owyang, categoriza las crisis en tres tipos:

Nivel 1: Resultan en contenido negativo de la empresa circulando por los medios.

Nivel 2: Resultan en contenido negativo de la empresa circulando por los medios y una respuesta de cambio en la empresa.

Nivel 3: Resultan en una pérdida financiera a corto plazo.

A continuación se presentan casos de crisis, donde no se tuvo un manejo adecuado de la situación.

Caso Nestlé – “Nestlé vs Greenpeace”

En el año 2010 Greenpeace lanzó una protesta en las redes sociales contra Nestlé, en la página oficial de Facebook de Nestlé se empezó a llenar de comentarios acusando a la empresa de sus prácticas no éticas.(Magee. K, 2010).

El problema no se concentraba en la protesta pero en la forma en que el community manager de Nestlé se expresaba hacia los usuarios.

Minuto a minuto oleadas de mensajes negativos aparecían, el community manager quiso controlar el flujo de los comentarios escribiendo: “sus comentarios son bienvenidos pero por favor no publiquen comentarios utilizando una versión alterada de nuestros logos como foto de perfil o serán eliminados”. (Brodia. R, 2010).



Figura. 3-5 Respuesta de Nestlé a los comentarios en redes sociales. Recuperado de <http://www.cbsnews.com/news/nestles-facebook-page-how-a-company-can-really-screw-up-social-media/>

Como resultado Nestlé anunció que dejarían de usar productos que ocasionan la destrucción de selvas tropicales (Greenpeace, 2010) y la contratación de una nueva agencia para estructurar su presencia en las redes sociales (Fox. E, 2010).

Caso Hawkers – “Hawkers México”

Hawkers es una empresa fundada en España dedicada a la venta de gafas a bajo costo de gran calidad y diversidad de forma online.

La crisis comienza el día de la victoria de Donald Trump a las elecciones presidenciales de Estados Unidos. Entre las declaraciones que este realizó, comentó la construcción de un muro en México para separarlo de Estados Unidos. Cuando se confirmaron los resultados de las elecciones, se publicó en la cuenta de Hawkers México el siguiente Tweet (Figura 3-6):



Figura. 3-6 Tweet de Hawkeros México ante las elecciones. (Galeano, 2016). Recuperado de <http://marketing4ecommerce.net/polemica-de-hawkers-en-mexico/>

Este Tweet generó la ruptura con el piloto de Fórmula 1 Sergio “Checo” Pérez (Figura 3-7), con el que se esperaba que saliese al mercado una edición especial con su nombre, a su vez Los Diablos Rojos del México de la Liga Mexicana de Beisbol también rompieron con Hawkeros, retirando de su tienda artículos de Hawkeros (Figura 3-8). Tras la polémica, Hawkeros lanzó un comunicado para la creación de la Fundación Querido Sergio x Hawkeros, aportando \$500,000 pesos y el total de ventas en lentes del pasado 17 de noviembre a esta fundación (Hawkeros).



Figura. 3-7 Tweet de Sergio contra Hawkeros. Recuperado de https://twitter.com/SChecoPerez/status/796476888390701056?ref_src=twsrc%5Etfw



Figura. 3-8 Tweet de Diablos Rojos contra Hawkercs. Recuperado de https://twitter.com/DiablosRojosMX/status/796511534180696065?ref_src=twsrc%5Etfw

3.7 Conclusión del capítulo

Los medios sociales representan una herramienta de doble filo, al utilizarlas de forma adecuada se puede obtener información que puede ser de utilidad para la mejora o creación de un nuevo producto o servicio. Se debe tomar a consideración el uso de herramientas de monitoreo de medios sociales y capacitación adecuada en el manejo de casos de crisis, gestión de contenido y uso de herramientas de monitoreo, con el fin de evitar generar una imagen negativa de la marca en los medios sociales o pérdidas económicas.

4 CRM

La definición de CRM Customer Relationship Management es muy amplia, es muy difícil realizar una sola definición sobre qué es ya que muchos vendedores, consultores y compañías tienen su propia definición de CRM (Payne. A, 2005). Estas definiciones están divididas en tres tipos de definiciones: Tecnológicas, estratégicas y enfocadas al flujo de clientes, (Kellen. V, 2007). A continuación, se mostrarán definiciones de diferentes autores.

De acuerdo al manual de la Asociación Americana de Control de Producción e Inventarios (APICS), CRM se define como una filosofía de marketing basada en priorizar al cliente, consiste en el análisis y diseño de la información con el fin de que los departamentos de ventas y marketing identifiquen las necesidades potenciales del cliente. Este análisis incluye la administración de cuentas, ordenes de catálogo, procesos de pago y crédito. Xu (2002) define que CRM es un término en la industria para las metodologías, software y algunas capacidades de internet para apoyar a la empresa a administrar las relaciones con los clientes. Yi (2010) define un CRM como una estrategia de negocio a largo plazo enfocada a establecer relaciones estables con los clientes con el fin de identificar las demandas de consumo de acuerdo a diferentes segmento de clientes. Altwegri (2015), CRM es un proceso que una empresa utiliza para organizar y registrar toda la información e interacciones con el cliente.

4.1 Breve Historia del CRM

Durante 1850, los negocios podían vender cualquier bien que ofrecían, produciendo un mercado enfocado al vendedor y a la producción. A inicios del siglo XX, nuevos competidores surgieron y los negocios notaron que tenían que ofrecer lo que los clientes querían, esto marcó los inicios de una orientación de mercado, enfocada a ofrecer bienes para segmentos de mercado en específico (Bose. R, 2002). A finales de 1980 la llegada masiva de PCs y arquitecturas cliente/servidor dio paso a un crecimiento en el desarrollo de software para recolección y organización de datos de los clientes llamado Database Marketing, no era tan intrincado como un CRM y era parecido a un rolodex. Los noventas representaron un gran avance en términos del CRM, Brock Control Systems contribuyó en el desarrollo lo que permitió a que el Contact Sales Management evolucionara a Sales Force Automation (SFA). Tom Siebel abandonó Oracle para crear Siebel Systems, posteriormente volviéndose en el líder del mercado SFA. En 1995 SFA y Contact Management evolucionaron para parecerse un poco al CRM, sin embargo, este producto no tenía el nombre como tal, a finales de este año se acuñó el término CRM en la industria. Múltiples vendedores de Enterprise Resource Planning ERP como Oracle y Baan entraron al mercado CRM, con la esperanza de dominar el mercado. 1999 marcó un gran paso para los CRM realizando su primer acercamiento al mercado móvil con la introducción de Siebel Handheld, con el uso frecuente del intranet, extranet e internet los proveedores de CRM ofrecieron un nivel de colaboración intraorganizacional, algo nunca antes disponible en esta industria. A pesar de sufrir gravemente de la crisis económica de la burbuja dot-com, CRM continuó incrementando su potencial. En 2005 Microsoft entra al mercado CRM con Dynamics

CRM y Oracle adquirió Siebel entre otros vendedores de aplicaciones. El CRM Social surgió en el mercado con la introducción de ComcastCares, una aplicación para crear relaciones más profundas con los clientes cambiando del enfoque de una transacción a una interacción. Salesforce hace un gran cambio en el 2007 presentando un CRM en la nube. Desde el final de la primera década hasta el día de hoy, los CRM en la nube continúan incrementando su popularidad, gracias a su costo inicial y su integración con dispositivos móviles.

4.2 Análisis de CRM Tradicional y CRM Social

CRM Tradicional

Los sistemas CRM están divididos en 3 aspectos (Figura. 4-1) que definen su funcionalidad y propósito, los cuales son: el aspecto analítico, el aspecto operacional y el aspecto colaborativo. Es de importancia señalar que un sistema CRM puede implementar estos tres aspectos de forma modular (Behboudi. M, 2011).



Figura. 4-1 Estructura del CRM Tradicional Fuente. (Behboudi, 2011).

El aspecto analítico tiene el fin de predecir que segmentos de clientes son propensos a utilizar un producto o servicio de acuerdo a sus preferencias e intereses (Wu, Ye, Yang, & Wang, 2009), este extrae información de Data Mart, Data Warehouse, Bases de datos de clientes y otros sistemas de información que posteriormente procesa con diversos algoritmos y técnicas de minería de datos produciendo información para crear una estrategia (Behboudi. M, 2011).

El aspecto operacional se refiere a servicios que ofrecen soporte a procesos de negocio que están relacionados con los clientes, están divididos en Marketing, Ventas y Servicio al cliente (Ngai. E, 2005).

El propósito principal de la colaboración es incrementar la calidad del servicio al cliente dando como resultado la satisfacción y lealtad a la empresa (Ngai. E, 2005). Se compone de los canales de comunicación tradicionales de la compañía como e-mail, fax, sitio web y call centers. La integración de la información entre diversas áreas de la empresa como ventas, marketing y soporte permite compartir entre ellos la información que obtienen mediante la interacción con los clientes. Por ejemplo, la información adquirida por el área de soporte técnico puede contener información para el área de marketing acerca de productos y servicios que podrían ser de interés para el consumidor. (Behboudi. M, 2011).

CRM Social

Tradicionalmente la relación entre organizaciones y consumidores ha sido optimizada hacia la organización, no el consumidor. Sin embargo, la rápida adopción de redes sociales ha cambiado el balance del poder hacia el consumidor (Owyang. J, 2010). Algunas de ellas intentan integrarse a la interacción para ganar de nuevo el control, pero esto es imposible a causa de que el consumidor busca recomendaciones de personas que utilizan servicios o productos similares a él, personas que influyen al consumidor de utilizar un producto (Greenberg. P, 2010).

Behboudi (2011) define el CRM Social como una filosofía y estrategia soportada por una plataforma tecnológica, reglas, procesos y características sociales diseñadas para comprometer al cliente en una colaboración con el fin de obtener valor mutuo. La combinación de un networking social y el CRM representa una gran oportunidad para enriquecer las interacciones con el cliente, a su vez brinda al negocio una forma de administrar y medir la comunicación con sus clientes. Al combinar el networking social con el CRM los tres aspectos de este que se mencionaron anteriormente adquieren un nuevo paradigma que se muestra en la Figura. 4-2.

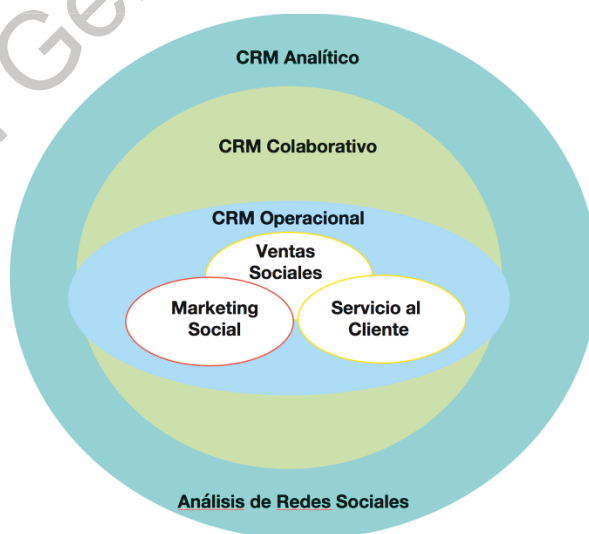


Figura. 4-2 Estructura del CRM Social. Fuente. (Behboudi, 2011).

CRM Operacional Social – Al aplicar el paradigma social, Behboudi (2011) define estos servicios como componentes del aspecto operacional:

- Marketing Social (Social Marketing)
- Ventas Sociales (Social Sales)
- Servicio al Cliente (Customer Service)

El marketing social es la combinación del marketing con herramientas tecnológicas, diversas herramientas para el marketing social se encuentran en desarrollo temprano. Hay herramientas que son útiles para el social marketing, diseñadas para interactuar con redes sociales, enfocadas a interpretar los datos del usuario que permiten enviarle catálogos de ofertas y productos. Además, el módulo de ventas sociales trasciende las funciones de automatización de control de inventario, procesamiento de ventas, seguimiento de cliente y análisis de desempeño, señalando el contenido ideal para el cliente mediante el análisis de información de ventas y marketing.

Cualquier nueva forma de comunicación que permita al cliente estar en contacto con la empresa tiene importancia para obtener datos e información acerca del cliente y sus interacciones, por ejemplo el uso de blogs, mensajes de texto, foros de discusión, comentarios de redes sociales y videos son parte de del grupo de herramientas que usa el cliente para compartir su pensamiento, utilizando estos canales de comunicación se puede obtener información acerca de los clientes.

Social Analytical CRM

Comprende las funciones tradicionales de la parte analítica del CRM sumándosele el análisis de redes sociales que consiste en identificar a los diferentes tipos de participantes que se encuentran en el ecosistema, quién realiza decisiones, quién es influyente y como interactúan dentro de un grupo social, que pueden llevar al éxito de un producto. La visualización de las interacciones en los grupos sociales es conocido como un grafo social. El análisis de las relaciones entre los grupos sociales y encontrar clientes potenciales en estos grupos es de gran valor para negocios B2B, es decir, para empresas que realizan transacciones entre ellas.

Wu Propone una serie de métodos para el análisis de grupos sociales:

Una de las propiedades más importantes de las redes sociales con los grupos, cuya estructura es un conjunto denso de usuarios conectados por enlaces de interés, familia o amistad. El objetivo de la detección de grupos es dividir la red social en sub grupos o clusters ya que el análisis sólo se enfocará a un grupo específico con el fin de identificar como los miembros de este sub grupo se comunican con sus vecinos y de qué forma están enlazados a ellos. Comprender la evolución de una red social es útil para inferir las modas y patrones de un contexto en particular, la evolución de grupos es un factor crítico para la precepción del mercado y el diseño de estrategias para la detección de grupos potenciales. Adicionalmente, Behboudi recomienda en un CRM Social usar

ciclos de vida para ilustrar el proceso evolucionario de diferentes comunidades. Después de seguir los grupos evolucionarios, se deben segmentar en clústers para explorar sus patrones a detalle, con el fin de identificar tipos de usuarios.

Morgan desarrolló un diagrama de flujo de datos del CRM Social, describe de forma básica las etapas, herramientas y acciones que se involucran dentro del CRM Social para llevar a cabo la conversación. El flujo de datos comienza desde la generación de información por parte de la comunidad compuesta en el aspecto tradicional por registros de llamadas, emails, encuestas y quejas que son registradas en el CRM. La información generada en los medios sociales es recolectada y monitoreada por medio de API que ofrecen varias redes sociales o herramientas de monitoreo de social media disponibles en el mercado, esta información es almacenada y analizada por el módulo analítico del CRM que identifica perfiles, interacciones, preferencias, genera información para ser utilizada por un equipo social que decidirá si se realiza una decisión macro o micro con base a las reglas del negocio que afectará el comportamiento de la comunidad.

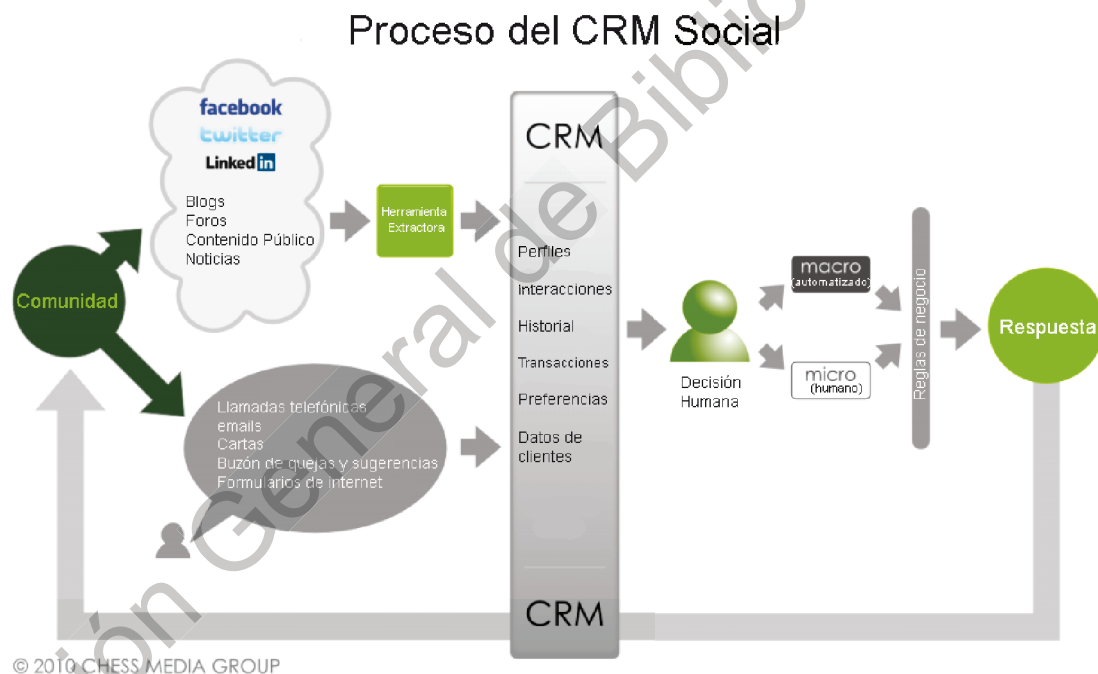


Figura. 4-3 Proceso del CRM Social. Fuente. (Morgan. J, 2010).

En la figura 4-1 se identifican los 5 factores clave que describe Owyang como las 5Ms del CRM Social que son Monitoreo, Mapeo, Middleware, Management (administración) y Medición que son necesarios para desarrollar el CRM Social en cualquier organización sin importar el departamento al que pertenezca.

Monitoreo: Se refiere a un software que permita monitorear los medios sociales, con el fin de saber qué están diciendo los usuarios de una marca o servicio. El software recolecta la información que se encuentra dispersa en una red social tomando como referencia palabras clave, para ser analizada por el CRM analítico. Owyang sugiere que

para un monitoreo efectivo se deben elegir palabras clave justas para que el monitoreo no sea corto o demasiado amplio.

Mapeo: Tiene la finalidad de identificar las relaciones entre los clientes. Muestra a las marcas qué grupos son sus clientes, que gustos tienen y como se relacionan. Behboudi (2011) mencionaba la importancia que tenía seguir la evolución de un grupo social ya que se podía identificar hacia que tendencias de consumo se dirigía un grupo en específico. Para lograr esto es necesario acceder a los perfiles públicos de cada usuario para la recolección de datos.

Management (administración): enlaza los medios sociales y canales de comunicación con los procesos de negocio, sin un propósito, los datos sociales pierden sentido. Los procesos y reglas de negocio son necesarios para triangular la información adecuada con los departamentos adecuados.

Middleware: Se refiere a tecnologías cuyo fin es enlazar el mundo social con la empresa, permite el flujo de la información por los sistemas y la integración de los datos.

Medición: Consiste en herramientas analíticas que proveen de datos de medición basados en KPI Indicadores clave de rendimiento que la empresa toma como medida de desempeño. También permite la visualización de diferentes KPI en forma de dashboards, permitiendo a la organización tomar decisiones.

Como ejemplo el caso de crisis de Nestlé, la ONG Greenpeace realizó una campaña con el fin de proteger las selvas tropicales de Indonesia de la tala y explotación para algunos productos, en Facebook se realizaron oleadas de mensajes donde se le pedía a Nestlé que dejara de comprar aceite de palma no sostenible, desafortunadamente la persona encargada de la comunicación por este medio, amenazó con borrar los mensajes, que como consecuencia provocó que los usuarios de las redes arremetieran más contra la compañía generando una imagen negativa de esta.

Greenpeace realizó la campaña para invitar a la gente a protestar por diferentes medios de comunicación tradicionales (email, llamadas telefónicas, quejas y sugerencias, página de internet) y sociales (Facebook, Twitter), también en forma de protestas en las oficinas de Nestlé. El módulo social del CRM generaría un reporte para que un equipo social generara una estrategia adecuada para responder a la campaña de Greenpeace generando uno de los siguientes escenarios:

A.- Nestlé anunciaría por sus medios sociales que se tomarían cartas en el asunto, evitando así una crisis por sus medios de comunicación sociales y manteniendo una imagen de marca ética y civilizada.

B.- Nestlé anunciaría una junta con Greenpeace para tratar los temas de la cadena de recursos para obtener proveedores sustentables.

En este caso con el CRM Social la identificación de los influenciadores, el alcance e impacto de su mensaje y realizar una correcta respuesta guiada de acuerdo a las reglas de negocio hacia ellos para detener la crisis podría ser posible.

4.3 La integración social de Skyline Boston

Skyline Boston es una empresa proveedora de exhibidores desde hace veinticinco años. Ellos han utilizado el CRM Goldmine, a pesar de ser un CRM completo, el equipo solamente lo utilizaba como rolodex. No utilizaban para ninguna de sus funciones de pronósticos o seguimiento de conversaciones debido a que su arquitectura legada le impedía sincronizar datos, sin mencionar problemas de compatibilidad con nuevos sistemas operativos. Los datos de los clientes estaban esparcidos en diferentes sistemas y archivos. Ellos estaban utilizando su CRM de la forma que no quería, decidieron cambiar al CRM Social Nimble, un CRM que posee funciones sociales integradas. Desde la configuración hasta la ejecución Nimble ha reforzado las relaciones con los clientes de SkylineBoston, creando un único flujo de notificaciones de diferentes redes sociales, gestionando cuentas de diversas redes sociales, sincronizando e importando las tareas y calendarios así como la vista y asignación de tareas, toda esta información siendo almacenada en la nube.

4.4 Conclusión del Capítulo

El CRM Social debe ser tratado como una extensión del CRM, no un reemplazo de este (Behboudi. M, 2011), la integración social en todo el CRM es posible, sin embargo, la formulación de una estrategia del exterior hacia el interior es fundamental para lograrlo, es decir, formular una estrategia que comienza con el cliente y termina con la empresa (Kumar. D, 2011). Los medios sociales pueden representar grandes fuentes de información y oportunidades para la empresa pero no se deben dejar a un lado los medios de comunicación tradicionales ya que representan medios por donde los clientes se pueden comunicar con la empresa, ignorar estos medios equivaldría a ignorar las opiniones de un grupo de clientes.

5 Integración de Redes Sociales en Diversos CRM

En el capítulo anterior se describió el CRM, sus componentes y paradigmas. A continuación se realizará un análisis de integración de redes sociales en CRM. Los CRM serán seleccionados con base al ranking de G2Crowd, una plataforma web cuya comunidad de usuarios califica y evalúa diferentes tipos de software para empresas.

G2 Crowd Grid® for CRM

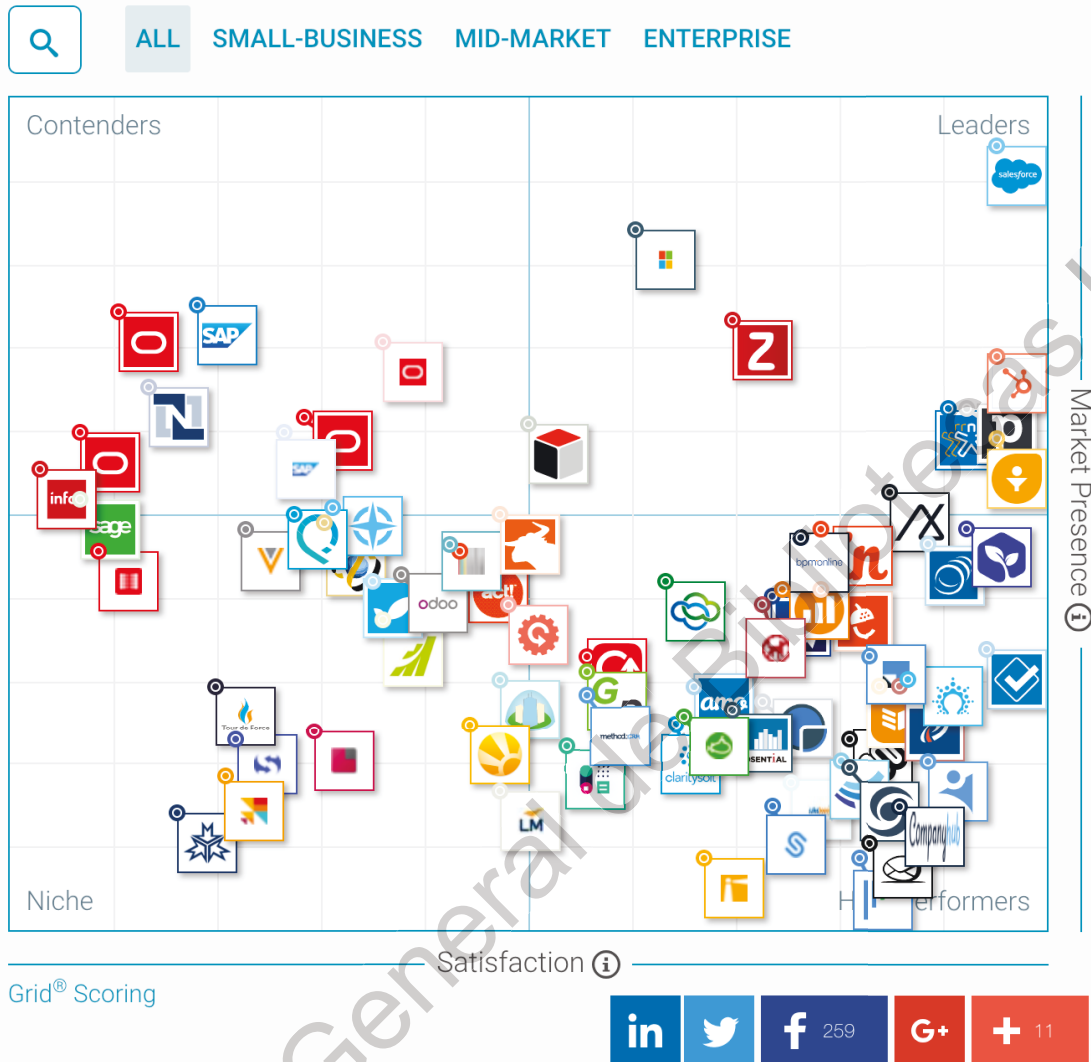


Figura. 5-1 Matriz de CRM. Extraído de

https://www.g2crowd.com/categories/crm?utf8=%E2%9C%93&order=top_shelf

De la matriz anterior se probó la versión de prueba de los siguientes CRM:

5.1 Nimble

El CRM Nimble combina completamente el ecosistema de un CRM a las redes sociales enfocándose únicamente al rastreo de actividad de contactos. Permite enlazar una cuentas de diferentes redes sociales para utilizar los contactos como clientes, visualizar la información de cada contacto así como la información de las cuentas enlazadas al CRM.

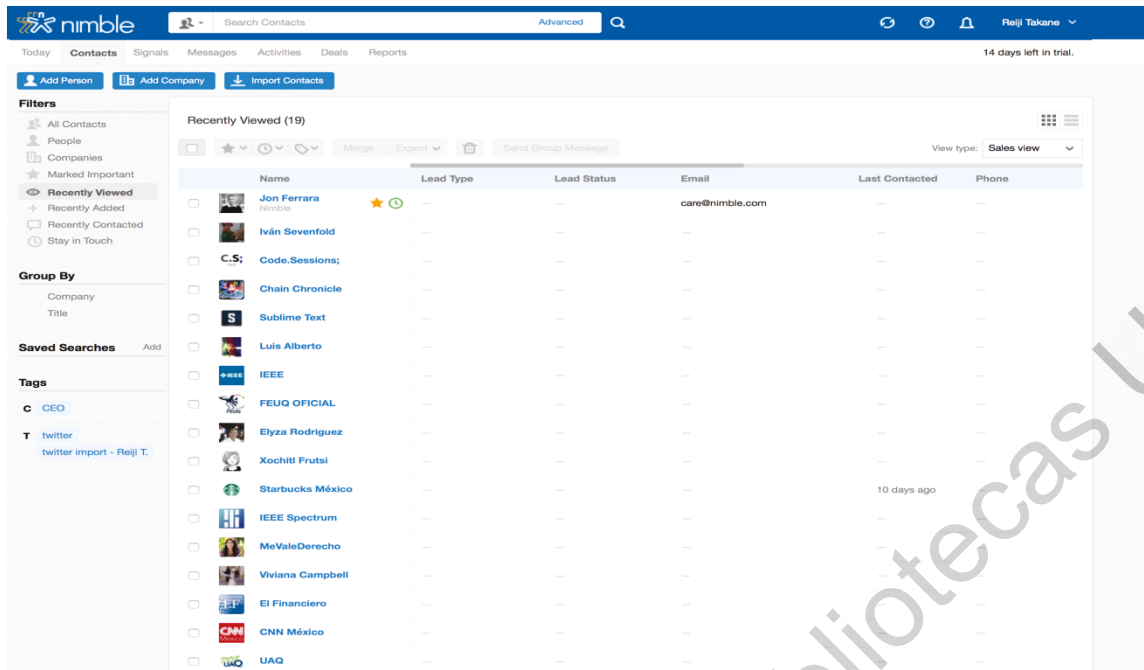


Figura. 5-2 Visualización de followers en el CRM.

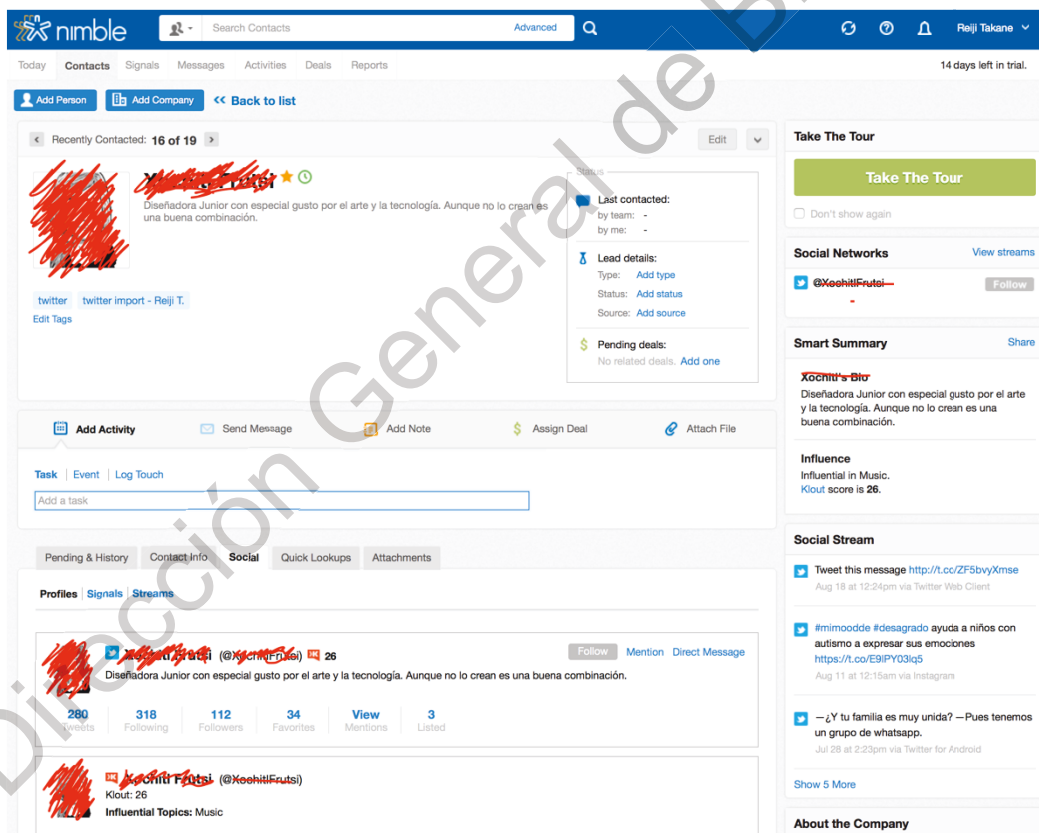


Figura. 5-3 Visualización de Información de contacto

The image shows a screenshot of a Twitter profile page. At the top, there is a profile picture (redacted with a red scribble), a name (redacted), and a bio: "Diseñadora Junior con especial gusto por el arte y la tecnología. Aunque no lo crean es una buena combinación." Below the bio are statistics: 280 Tweets, 318 Following, 112 Followers, 34 Favorites, view Mentions, and 3 Listed. A "Follow" button is visible. The main content is a "Timeline" of tweets. The first tweet is from Aug 18 at 12:24pm via Twitter Web Client. The second is from Aug 11 at 12:15am via Instagram. The third is a retweet from Jul 28 at 2:23pm via Twitter for Android. The fourth is a retweet from Jul 15 at 10:40am via Twitter Web Client. The fifth is from Jul 15 at 9:36am via Twitter Web Client. The sixth is a retweet from Jun 7 at 10:16pm via Twitter for Android. The seventh is from Jun 7 at 10:14pm via Twitter for Android. A large watermark "Dirección General de Bibliotecas UAQ" is overlaid diagonally across the page.

Figura. 5-4 Visualización de la línea de tiempo.

The image shows a Twitter profile page for a user whose name and handle are redacted with a red scribble. The profile bio reads: "Diseñadora Junior con especial gusto por el arte y la tecnología. Aunque no lo crean es una buena combinación." The statistics are: 280 Tweets, 318 Following, 112 Followers, 34 Favorites, view Mentions, and 3 Listed. A "Follow" button is visible in the top right.

Below the profile information, there is a "Twitter" tab and a navigation bar with options: Timeline, Mentions, Favorites, Following (selected), and Followers.

The list of followed accounts includes:

- Boogapp (@boogapp)**: Desarrollo de sistemas personalizados, Marketing Digital, páginas web, apps, juegos y más para tu negocio o empresa. Contacto: ventas@boogapp.com
- Mood Face (@Mood_Face)**: (Profile picture shows a red speech bubble with "MOOD FACE" text)
- Tráfico Querétaro (@Trafico1120)**: Información de la vialidad y el clima en Querétaro. Cuenta oficial de Grupo ACIR Querétaro.
- Alfonso Agúndez (@Alf_agu)**: Social Media Strategist | Comunicación | Coach | Política | Marketing Online | Social Media | #Geek | #Runner. Soy más de vino tinto.

Figura. 5-5 Lista de contactos que sigue el usuario.

[Redacted Name] (@[Redacted]) N/A Follow

Diseñadora Junior con especial gusto por el arte y la tecnología. Aunque no lo crean es una buena combinación.

280 Tweets | 318 Following | 112 Followers | 34 Favorites | [view](#) Mentions | 3 Listed

Twitter

Timeline | Mentions | Favorites | Following | Followers

キラ (@QuiuboKira)
No soy buena persona, no lo puedo evitar. Consejera, amiga, heroína y psicopata.

Boogapp (@boogapp)
Desarrollo de sistemas personalizados, Marketing Digital, páginas web, apps, juegos y más para tu negocio o empresa. Contacto: ventas@boogapp.com

ECcen Community (@ECcenSocial)

Daniel Otero (@DANIELdoiser)
Fundador de <http://t.co/USva5VzOWF> y Profesor UPC. Emprendiendo en tiempos de crisis (<http://t.co/G4GbSGZSfx>). Viajero cuando puedo. #VasoSiempreLleno

watch movie (@online_movies_5)
free movies, series everyday & followback

Figura. 5-6 Lista de seguidores del contacto.

nimble Search Social

Today Contacts **Signals** Messages Activities Deals Reports

Refresh Status Update New Message

- Notifications
- Signals (Beta)
- All Streams
- Profile**
- Pending

Reiji Takane

Twitter

- Whitey** (@WhiteyCat94) ~~El gato siempre adopta una postura fotogénica #CatsOfTwitter #Sábado #TengoHambre #Fotos~~
Aug 29 at 11:01am via Nimble
- Whitey** (@WhiteyCat94) ~~El gato siempre adopta una postura fotogénica #CatsOfTwitter #Sábado #TengoHambre #Fotos~~
Aug 29 at 11:00am via Nimble
- Whitey** (@WhiteyCat94) El gato siempre adopta una postura fotogénica #CatsOfTwitter #Sábado #TengoHambre #Fotos
Aug 29 at 10:35am via Zoho TwitterApp
- Whitey** (@WhiteyCat94) Tengo mucha hambre #hambre #MeTeniasConElPendiente #comida
Aug 25 at 1:52pm via Vtiger CRM Twitter Connector
- Whitey** (@WhiteyCat94) @StarbucksMex ¿Por qué cambiaron de queso Philadelphia a queso crema genérico? No está tan bueno este queso :(
Aug 20 at 12:19am via Twitter for iPhone
- ~~Whitey~~ RT @~~Whitey~~: #CPMX7 El mejor momento durante mi estancia en #CPMX6 ! :) <http://t.co/6gk3QLzo7U>
Jul 26 at 3:44pm via Twitter Web Client
- Whitey** (@WhiteyCat94) Alguien dígale a mi humano que no vaya a clases mañana!!! #CatsOfTwitter #gatos #InicioDeClases
Jul 26 at 3:34pm via Twitter Web Client
- Whitey** (@WhiteyCat94) Que divertido es usar Vtiger #vTiger

Figura. 5-7 Lista de actividad enlazada al CRM.

5.2 Zoho

El CRM Zoho permite enlazar cuentas de Facebook y Twitter de modo que se puede visualizar el contenido de ambas redes sociales de una forma ordenada, permite otorgar roles y permisos a los usuarios, este módulo está principalmente enfocado a la interacción entre el negocio y el cliente.

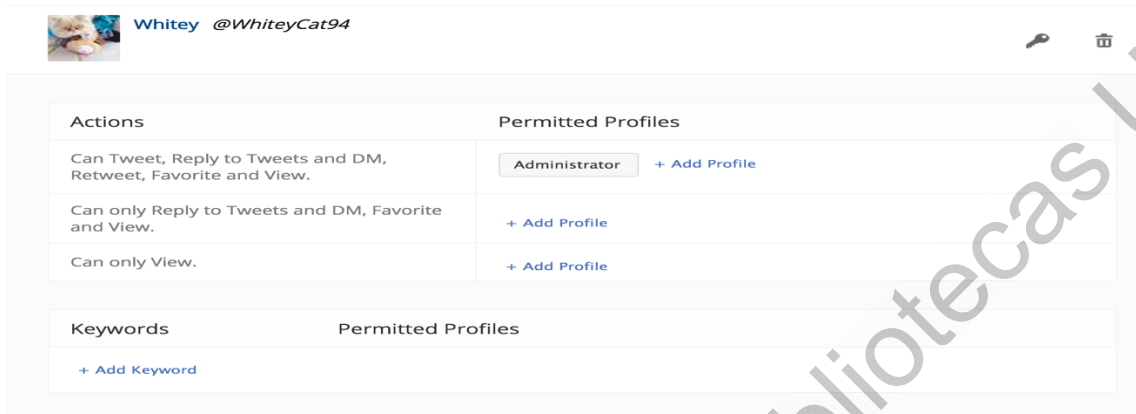


Figura. 5-8 Permisos del usuario.

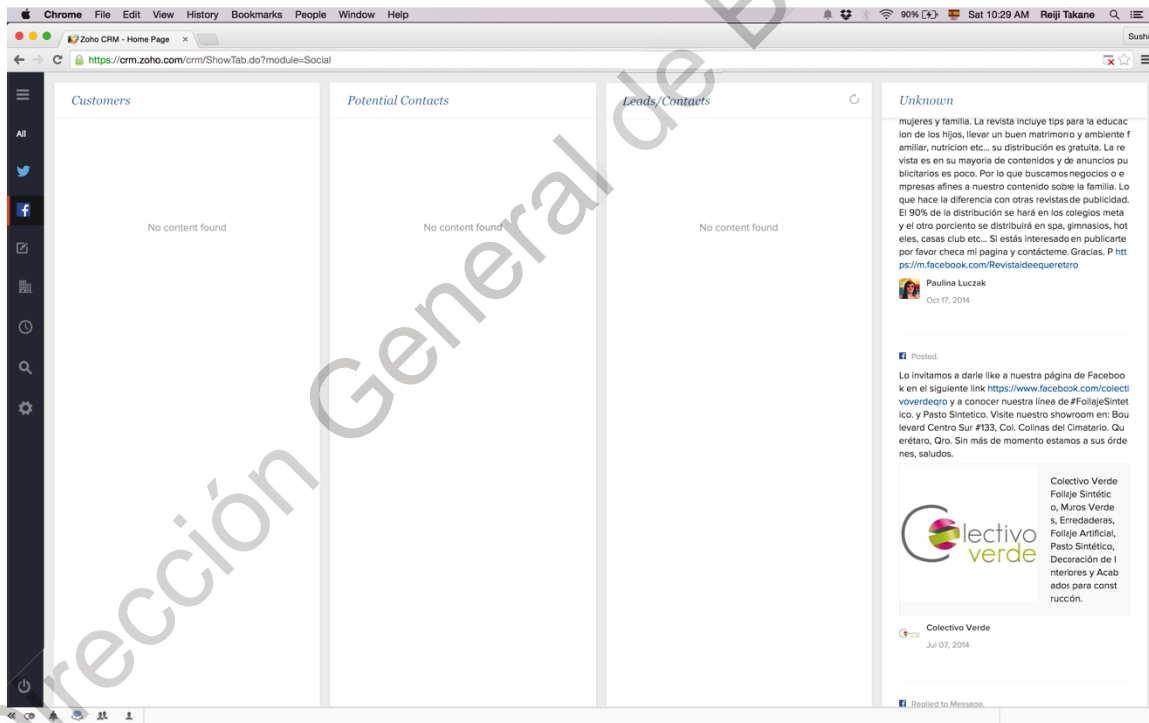


Figura. 5-9 Tabla de contenidos.

5.3 vTiger

El CRM vTiger contiene un módulo que permite publicar tweets así como realizar una búsqueda que nos permite descargar tweets y agregar usuarios dentro de los tweets a la lista de contactos, también contiene una gráfica que describe la actividad de la cuenta en función del tiempo y mensajes.

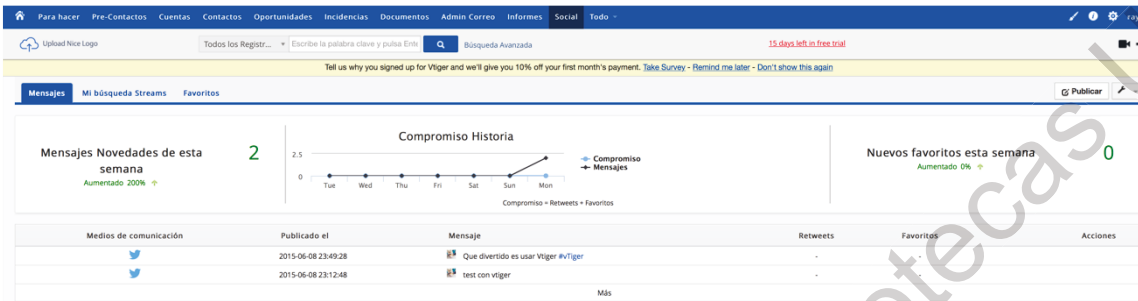


Figura. 5-10 Tabla de mensajes.

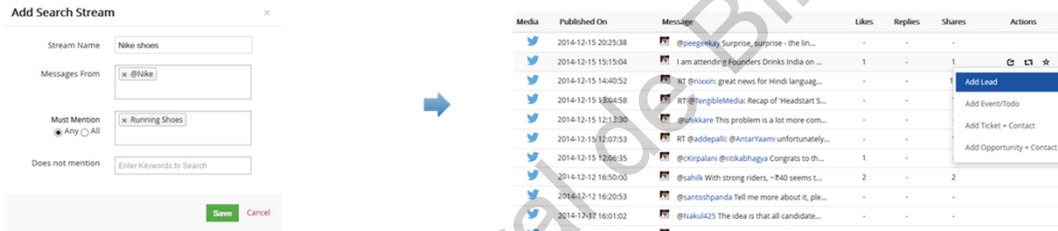


Figura. 5-11 Búsqueda de streams.

Media	Published On	Message	Retweets	Favourites	Actions
	14-05-2015 16:28:19	RT @jerrysaltz: How sexually sick are co...	1967	1134	
	25-03-2015 07:43:33	200 000 étudiants réclament un meilleur ...	14	6	
	19-03-2015 10:52:50	RT @SOQUIJ: L'affaire Proulx c. Québec (...)	3	1	
	19-03-2015 07:44:00	Laurent Proulx et la @Fondation1625 perd...	10	-	
	18-03-2015 16:37:42	@Rob_S_Pierre Le discours de @cerepqa ...	-	-	
	18-03-2015 16:30:24	@Rob_S_Pierre Le discours de @cerepqa ...	-	-	
	18-03-2015 16:20:45	@Rob_S_Pierre Le discours de @cerepqa ...	-	-	
	18-03-2015 16:06:11	@Rob_S_Pierre Le discours de @cerepqa éta...	-	-	
	18-03-2015 16:02:38	Pour l'instant, les propositions les plu...	24	3	
	14-03-2015 06:56:35	RT @Emilie_Ni: 8 garçons en ligne au mic...	3	6	
	20-09-2014 20:54:57	Such a great day to be alive!	-	-	
	13-05-2014 16:01:18	@axilcoffee Is the Flinder's lane shop o...	-	-	
	20-09-2013 00:12:59	@mykimate why do we not have a temporary...	-	-	
	09-06-2013 22:51:05	Just caught someone filming me and posti...	-	3	
	27-05-2013 05:40:08	The gentleman at the tram stop was so ni...	-	1	
	09-05-2013 03:12:48	@lizasmiz like a savoury cupcake. Blegh	-	-	
	09-05-2013 03:12:19	If I saw someone do what I just did on m...	-	-	
	09-05-2013 02:58:43	@tranterpants it's the best. Guarantees ...	-	-	
	01-05-2013 08:00:15	@ghostmeow billy corgan experienced the ...	-	1	
	30-04-2013 17:46:55	@nicholsonr I just ate a banana and can ...	-	-	
	13-10-2014 15:55:42	RT @ADNQro: 3.- Confirmar entrega de pli...	2	-	
	13-10-2014 15:55:25	RT @Enrique_Zamudio: Convoca @FEUQ_UAQ a...	1	-	
	13-10-2014 15:55:23	RT @ADNQro: 1.- Descartada huelga inmedl...	1	-	
	13-10-2014 15:55:20	RT @Quadratin_Qtro: #QUERÉTARO Exhorta e...	1	-	

Figura. 5-12 Tabla de streams.

5.4 Conclusión del capítulo

Los CRM ya están implementando módulos que integran las redes sociales a sus canales de comunicación, sin embargo, su enfoque principal es unificar diversas redes sociales mediante una sola plataforma de comunicación, las herramientas para analizar y medir el contenido están en fases tempranas o inexistentes.

6 Extracción y Minería de texto

En el capítulo anterior se mencionó una serie de métodos para el análisis de grupos sociales y su rastreo evolutivo con el fin de predecir las tendencias que pueden adquirir con base en Wu (2009). El objeto del presente capítulo consiste en investigar los métodos de extracción y análisis de texto con el fin de comprender las técnicas de análisis de texto.

Para realizar el análisis de texto primero se deben obtener los datos, estos datos pueden ser extraídos de diversas fuentes (aplicaciones web, bases de datos, repositorios, CRM, email, etc.), después el contenido extraído debe ser tratado y transformado a un formato adecuado para ser almacenado en un Data Warehouse con el fin de ser analizado y utilizar esos datos para generar gráficos y reportes que generarán conocimiento (Figura. 6-1). Para fines de la presente investigación se revisarán los métodos de web scraping y la API de la plataforma social Twitter.

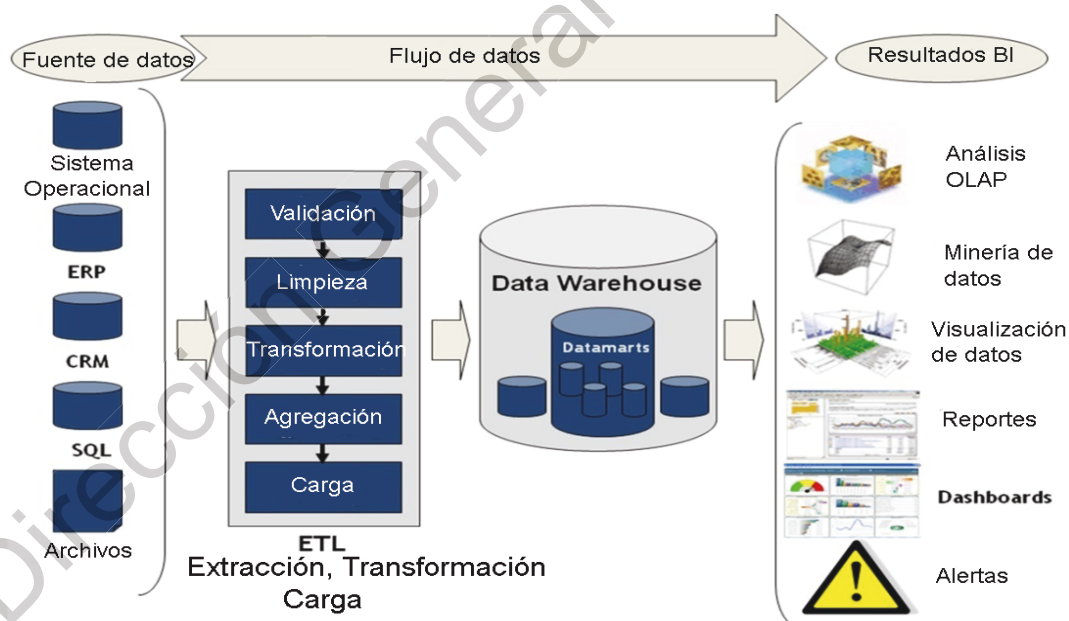


Figura. 6-1 Proceso Business Intelligence. Extraído de <https://www.intechopen.com/source/html/18527/media/image8.jpeg>

La REST API de Twitter son un conjunto de instrucciones en un lenguaje de programación que permiten leer y escribir datos de Twitter, los principales objetivos de esta API son realizar búsquedas individuales, leer información de un usuario o publicar Tweets. La REST API identifica aplicaciones de Twitter utilizando un método de autenticación seguro, esta llave es enviada en cada consulta lo que previene abusos de uso en la API; provee respuestas en formato entendible para el software CRM. (Twitter, 2017).

Mediante la REST API se puede acceder a información detallada de los tweets que cumplen con los criterios de búsqueda requeridos, la REST API cuenta con restricciones de consulta de datos con el fin de aliviar la carga en los servidores de Twitter, sin embargo, Twitter ofrece la Streaming API cuyo fin es monitorear y procesar tweets en tiempo real, esta ofrece diferentes tipos de flujo para cada caso.

Flujo Público: Comprende los flujos de datos públicos en twitter, ideal para extraer datos de un usuario o tema en específico.

Flujo de Usuarios: Flujos de un solo usuario, contiene toda la información correspondiente a un usuario de Twitter.

La diferencia entre la REST y la Streaming API es la forma de comunicación entre el usuario y el servidor. En una consulta con la REST API una aplicación web realiza una o varias peticiones a la API de Twitter, después realiza el formato e impresión de los resultados al usuario como respuesta a la petición inicial.(Twitter, 2017).

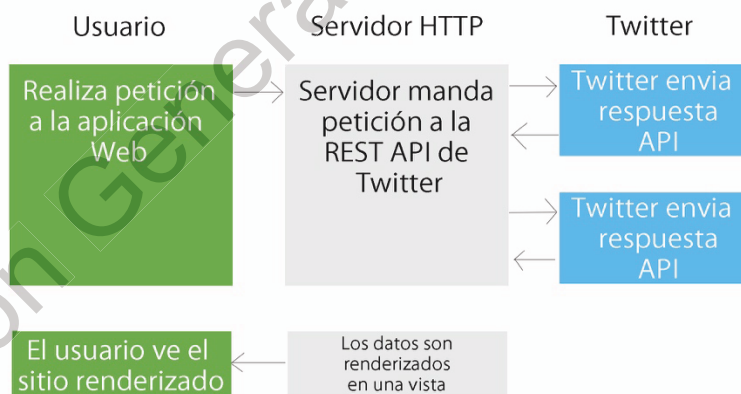


Figura. 6-2 Modelo de comunicación REST extraído de <https://dev.twitter.com/streaming/overview>

En una aplicación web que se conecta con la Streaming API se debe implementar código para mantener un flujo continuo separada del proceso que maneja las peticiones HTTP. El proceso de flujo obtiene los tweets y realiza el proceso de análisis, filtrado y agregación necesaria antes de almacenar los datos. El proceso del manejo de HTTP consulta los datos almacenados de acuerdo a los requisitos de búsqueda (Twitter, 2017).

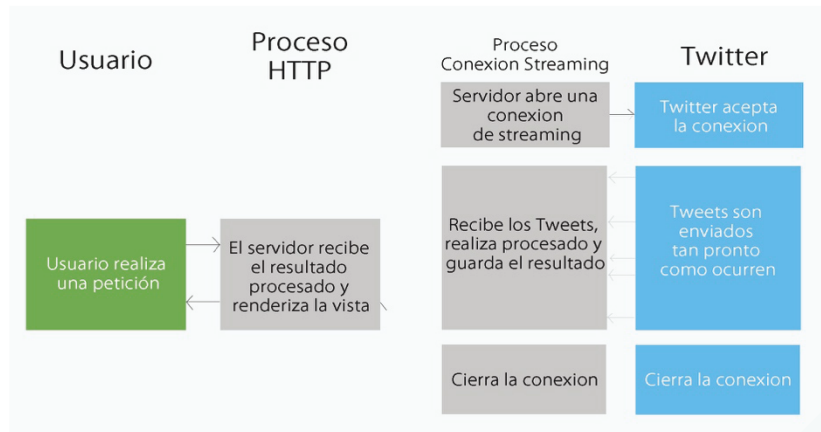


Figura. 6-3 Modelo de comunicación Streaming extraído de <https://dev.twitter.com/streaming/overview>

Como alternativa a las API el Web Scraping es una técnica que consiste en extraer información o “escrubar” una página web de forma automatizada, para realizar esto un crawler visita una dirección URL de la cual extraerá el contenido que se encuentre en las etiquetas que cumplan con los criterios de búsqueda requeridos. El Web Scraping no cuenta con retardos temporales, sin embargo se debe realizar navegación en el sitio web para conocer la estructura de la que está compuesta. Una desventaja de utilizar esta técnica es que a causa de que los crawlers generan tráfico y perjudican la eficiencia del sitio web, algunos administradores utilizan diferentes mecanismos para bloquear el acceso a estos. Es recomendable consultar los términos y condiciones de uso antes de scrapear una página web.

6.1 Minería de texto

La escritura es un sistema de representación gráfica de un idioma por medio de signos trazados con el fin de transmitir una idea. Partiendo de esta definición, los datos de texto son generados por humanos a partir de la percepción del mundo real para consumo humano, haciendo que resulte complicado para las computadoras entender el lenguaje humano, estos datos contienen información acerca de las preferencias y conocimiento del autor. La minería de textos o text mining es el proceso de transformar datos de texto en información o conocimiento para la toma de decisiones. Para responder a la incapacidad de las computadoras de entender el lenguaje humano, el NLP (Natural Language Processing) es la rama de la inteligencia artificial dedicada al procesamiento del lenguaje natural, es decir, enseñar a las computadoras a entender y generar lenguaje humano (Dragomir R).

En la actualidad diversas empresas de T.I. utilizan el NLP para el desarrollo de aplicaciones (Dragomir R). Algunas de las aplicaciones del NLP abarcan:

- Motores de Búsqueda
- Asistentes de lenguaje natural

- Servicios de traducción de idiomas
- Sinopsis de información

Uno de los principales problemas en el análisis de texto es la ambigüedad del mensaje de cual existen diferentes tipos:

Léxica y Polisemia: Consiste en múltiples significados que tiene una palabra que en función del contexto pueden dar un sentido u otro a la frase. Ej. Tengo un hijo bueno

Sintáctica: Se refiere a una frase u oración compleja que se puede analizar de varias formas. Ejemplo: Arregló el camión rápido

Semántica: Ocurre cuando una palabra o concepto tiene un significado de por sí difuso que se basa en el uso informal o generalizado. Ejemplo: Se busca personal de ambos sexos.

6.2 Minería de Opiniones

Son el conjunto de técnicas que se utilizan para el procesamiento del lenguaje natural, teniendo como objetivo la extracción de información subjetiva contenida en un texto. Para representar una opinión se debe tomar en cuenta los siguientes factores:

¿De quién es la opinión?(Emisor)

¿Acerca de qué es la opinión?(Objetivo)

¿Cuál es la opinión?(Contenido)

¿Bajo qué situación fue la opinión expresada?(Contexto)

¿Qué nos dice la opinión acerca de los sentimientos del emisor?(Positivo/Negativo)

Para llevar a cabo esta tarea, el proceso de análisis del texto, comprende el análisis en cuatro partes: a nivel morfológico, sintáctico, semántico y discursivo.

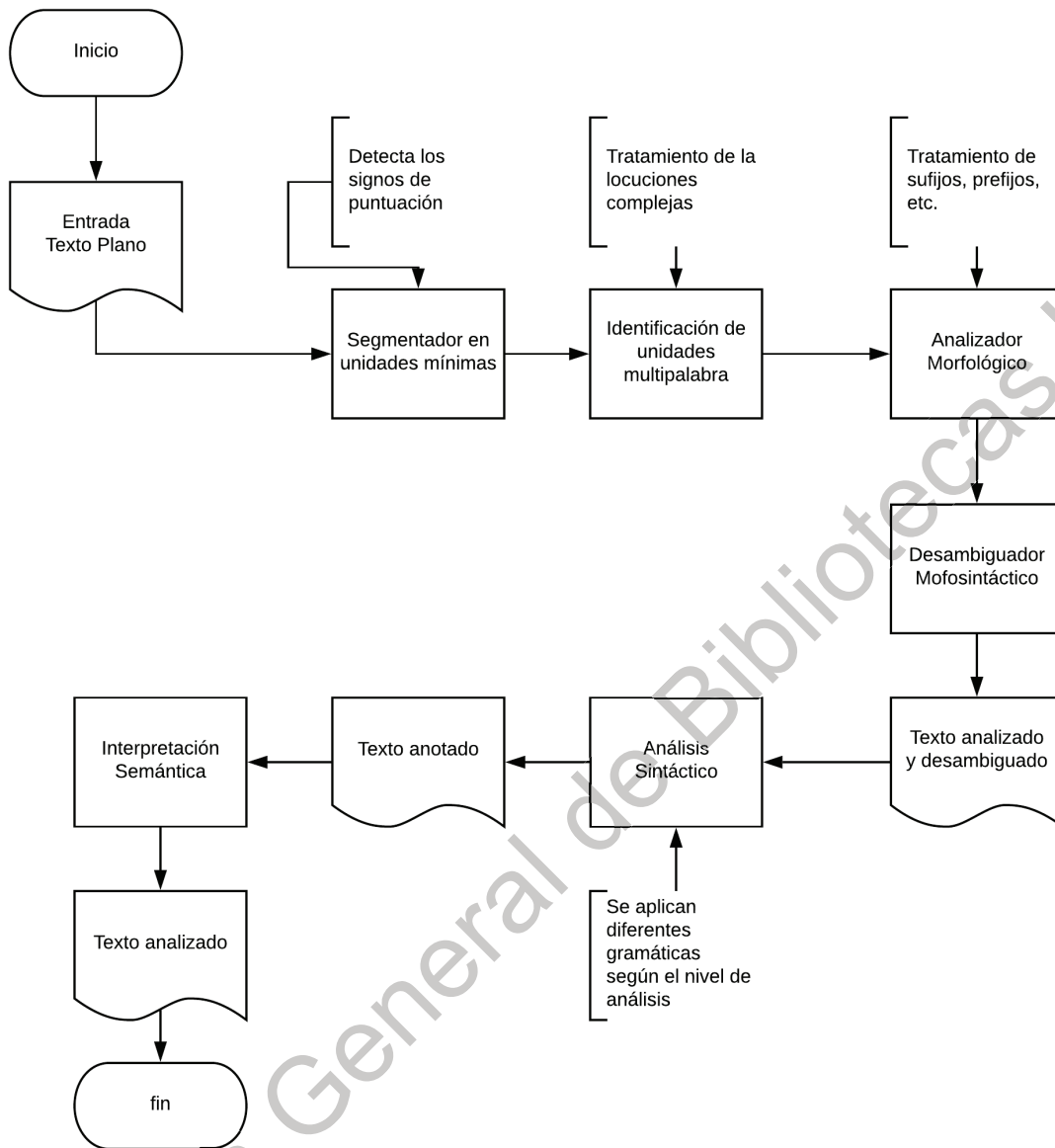


Figura. 6-4 Proceso de análisis de texto.

La primera parte del proceso, el análisis morfológico, consiste en segmentar las palabras en el texto plano para determinar las categorías léxicas de una palabra, decidir cuál categoría es la adecuada en el contexto que aparece la palabra y asignar a cada palabra analizada la información morfológica que le corresponde.

Estos procesos en que se subdivide el análisis morfológico se pueden realizar en tres funciones.

- Incorporación de la información morfológica: Requiere la definición de etiquetas consistentes en categorías o en rasgos que permitan describir los fenómenos

morfológicos propios de cada lengua, es decir, el conjunto de etiquetas que se utilizarán para la categorización de las palabras Figura 6-5.

Grampal		Análisis de: palabras oraciones textos Generación de formas Etiquetario Autores	
Palabra : Etiquetario:		Analiza	
Etiquetario:			
CATEGORÍAS SINTÁCTICAS:			
Sustantivo	N	nombre	<i>gato</i>
	NPR	nombre propio	<i>María</i>
Adjetivo	ADJ		<i>azul</i>
Verbo	V		<i>cantar</i>
Auxiliar	AUX		<i>habría</i>
Pronombre	P		<i>yo, ahora</i>
	REL	relativo	<i>que</i>
	PINT	interrogativo	<i>qué, cómo</i>
Determinante	ART	artículo	<i>el</i>
	POSS	posesivo	<i>su</i>
	DEM	demonstrativo	<i>ese</i>
Cuantificador	Q		<i>uno</i> cardinal
			<i>primer</i> ordinal
			<i>muchos</i> indefinido
Preposición	PREP		<i>en</i>
Adverbio	ADV		<i>así</i>
Conjunción	C		<i>y</i>
Interjección	INTJ		<i>ole</i>
Marcador discursivo	MD		<i>oye</i>
Desconocida	UNKN		<i>xx</i>
RASGOS MORFOLÓGICOS:			
Genero	masc	masculino	<i>gato</i>
	fem	femenino	<i>gata</i>
Número	sing	singular	<i>lápiz</i>
	plu	plural	<i>lápices</i>
Persona	1	primera	<i>amo</i>
	2	segunda	<i>amas</i>
	3	tercera	<i>ama</i>
Tiempo	inf	infinitivo	<i>amar</i>
	ger	gerundio	<i>amando</i>
	part	participio	<i>amado</i>
	pres_ind	presente indicativo	<i>amo</i>
	indf_ind	pretérito perfecto	<i>amó</i>
	impf_ind	pretérito imperfecto	<i>amaba</i>
	fut_ind	futuro indicativo	<i>amará</i>
	pres_subj	presente subjuntivo	<i>ame</i>
	impf_subj	imperfecto subjuntivo	<i>amara</i>
	cond	condicional	<i>amaría</i>
	imper	imperativo	<i>ama</i>

Figura. 6-5 Etiquetario empleado en Grampal. Consultado el 21 de mayo 2017. Extraído de <http://cartago.llf.uam.es/grampal/grampal.cgi?m=etiquetario>

- **Búsqueda en el Léxico:** realiza en un diccionario una búsqueda con todas las formas expandidas de las palabras y con la información correspondiente a la categoría léxica.
- **Desambiguación:** Decisión sobre la categoría léxica de la palabra en función de su contexto de aparición.

Cada	categoria	Q	lema	CADA	
noche	categoria	N	lema	NOCHE	rasgos femenino singular
que	categoria	REL	lema	QUE	
bajo	categoria	V	lema	BAJAR	rasgos singular 1 presente indicativo
a	categoria	PREP	lema	A	
la	categoria	P	lema	LO	rasgos singular 3 femenino
calle	categoria	V	lema	CALLAR	rasgos singular
el	categoria	ART	lema	EL	rasgos masculino singular
músico	categoria	ADJ	lema	MÚSICO	rasgos masculino singular
bajo	categoria	PREP	lema	BAJO	
toca	categoria	V	lema	TOCAR	rasgos singular
el	categoria	ART	lema	EL	rasgos masculino singular
bajo	categoria	N	lema	BAJO	rasgos masculino singular
bajo	categoria	PREP	lema	BAJO	
la	categoria	ART	lema	EL	rasgos femenino singular
luna	categoria	N	lema	LUNA	rasgos femenino singular

Figura. 6-6 Categorización léxica de la oración. Extraído de http://liceu.uab.cat/~joaquim/language_technology/NLP/PLN_analisis.html#Tokenizaci_n

La segunda parte del análisis, el análisis sintáctico consiste en segmentar el párrafo en frases para determinar la validez de una expresión respecto a una gramática que da como resultado un árbol jerárquico, que representa la estructura de los elementos de una frase para utilizarse en el análisis semántico. Consiste de dos procesos, el análisis de dependencias y el análisis para la generación de la estructura de árbol.

El análisis para la generación del árbol jerárquico se puede realizar de dos formas: superficial y profundo como se muestra en la Figura 6-7 y Figura 6-8 respectivamente. El análisis profundo construye un árbol entero para una oración y el análisis superficial construye un juego de árboles para una oración.

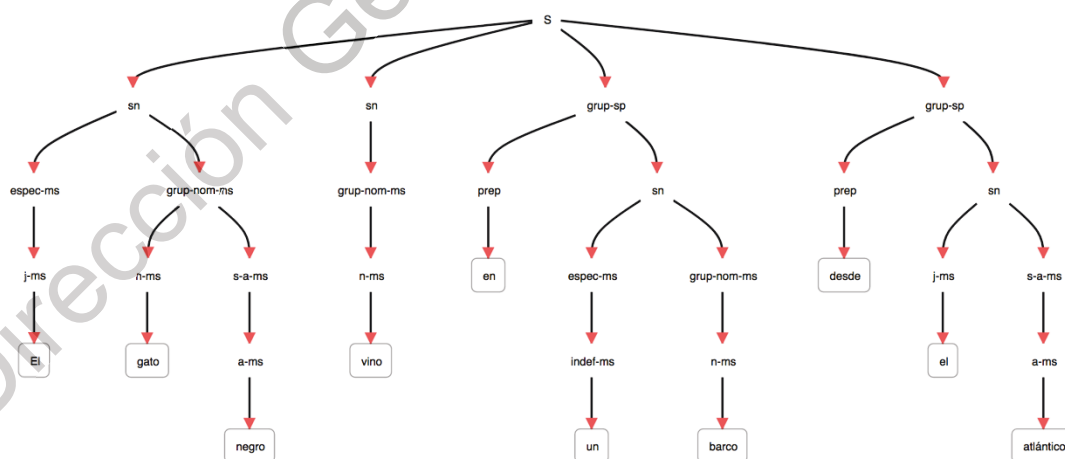


Figura. 6-7 Estructura de un árbol con análisis superficial. Padró, L. (2016).

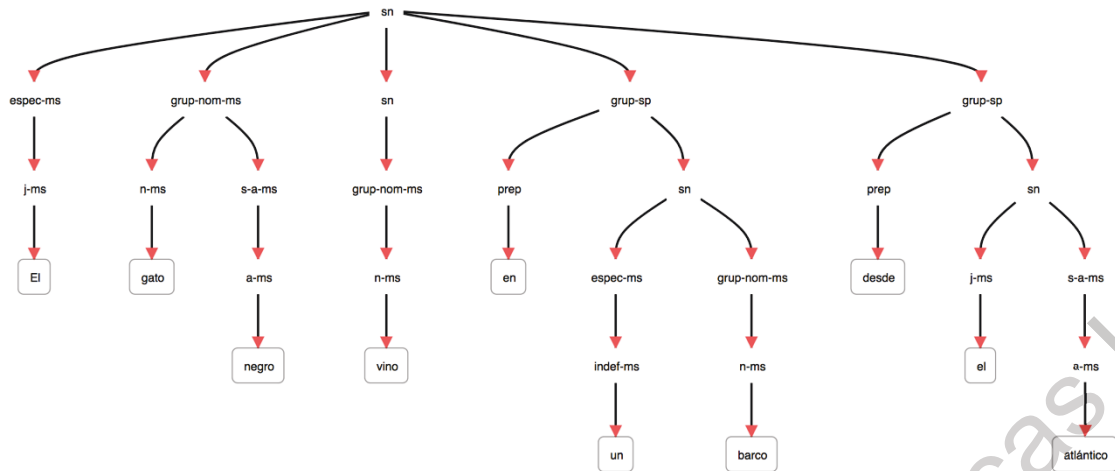


Figura. 6-8 Estructura de un árbol con análisis profundo. Padró, L. (2016).

El sub proceso del análisis de dependencias (Figura 6-9), consiste en analizar la estructura gramatical de la oración, estableciendo relaciones entre palabras “principales” y palabras que modifican a esos principales.

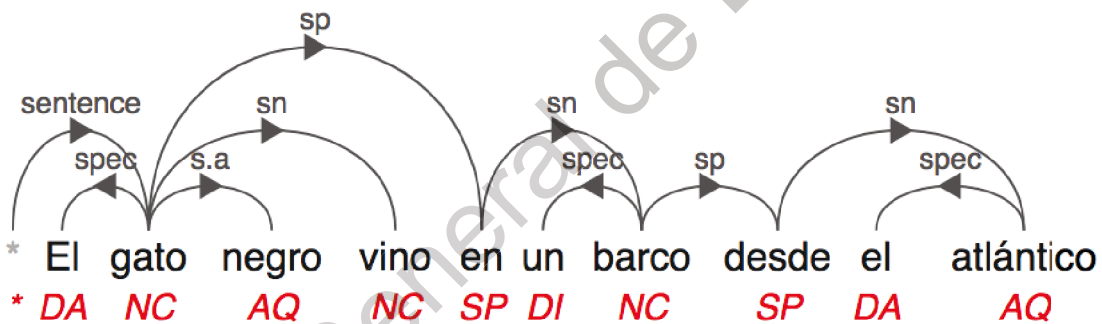


Figura. 6-9 Análisis de dependencias.

Capturando tendencias distribucionales en los datos, se pueden realizar modelos estadísticos para analizar oraciones, facilitando el proceso de desambiguación y en consecuencia reducir las restricciones de lenguajes incrementando la robustez. También ya que los modelos pueden ser generados a partir de datos, estos pueden ser modificados para entender nuevos lenguajes siempre que exista una muestra etiquetada del lenguaje (Nivre. J, 2010).

La tercera parte del proceso es el análisis semántico encargado de asignar sentido a las palabras polisémicas en función del contexto en el que se encuentran. El principal problema en el análisis semántico es la resolución de ambigüedad. Desde el punto de vista de una computadora, muchas palabras del léxico humano están abiertas a múltiples interpretaciones, ya que estas palabras pueden tener más de un significado, o porque ciertas palabras como contadores, operadores negativos pueden aplicar de diferente

forma en algunos textos, o incluso en la referencia de pronombres u otras expresiones para referirse a una persona pueden ser poco claras. (Nivre. J, 2010).

Los roles semánticos sirven para enlazar y clasificar argumentos de una oración en un numero de roles a los cuales se le otorga un estatus con propiedades.

Rol	Descripción
Agente	El participante que deliberadamente realiza la acción
Experimentador	La entidad que recibe sensorial o emocionalmente el estímulo
Estímulo	La entidad que produce no deliberadamente una sensación sensorial o anímica
Paciente	El participante que se somete a la acción y cambia su estado al final de tal acción
Tema/Participe	El participante en el que se lleva a cabo la acción pero que no cambia su estado a consecuencia de ello
Instrumento	El participante que es usado para llevar a cabo la acción
Fuerza/Causa	El participante no animado ni volitivo que realiza la acción
Ubicación	El lugar donde se produce la acción
Dirección	El lugar adonde se dirige la acción hacia
Destinatario	Un tipo especial de participe asociado a verbos que expresan un cambio en la posesión
Fuente	El lugar donde se origina la acción
Tempo	El momento en que sucede la acción o estado
Beneficiario	La entidad en cuyo beneficio se produce la acción
Modo	La forma en que una acción se lleva a cabo
Causa	Un ente que preexiste a la acción y que se considera que da lugar
Finalidad	Un ente que no preexiste a la acción y que se considera el objetivo por el que el agente la realiza
Compañía	Un ente animado que participa de la acción o el estado con el agente, paciente o causa del mismo

Figura. 6-10 Roles semánticos.

La última parte del proceso es a nivel discursivo, consiste en segmentar el texto para señalar temas y subtemas. Usualmente los textos en lenguaje natural consisten en un conjunto de oraciones que forman un discurso (Mitkov. R, 2010). De acuerdo a la RAE la definición de discurso es 1.- “Escrito o tratado, generalmente de no mucha extensión, en que se discurre sobre una materia determinada.” 2.- “Serie de las palabras y frases empleadas para manifestar lo que se piensa o se siente.” Con base a estas definiciones podemos inferir que un discurso manifiesta cohesión, el cual permite que las palabras estén relacionadas para formar oraciones con sentido, y de que las oraciones manifiesten un significado en el discurso (Mitkov. R, 2010). Como ejemplo:

- (1) Pedro tenía hambre. Él comió salmón.
- (2) Pedro tenía hambre. Él se fue a su casa.

La segmentación de texto para la identificación de temas en este proceso es crucial. El algoritmo TextTiling (Hearst. M, 1994), en lugar de capturar relaciones jerárquicas y dependencias que existen entre segmentos, separa documentos en una secuencia lineal de párrafos, cada uno se distingue de tener un subtema dentro del tema principal del discurso. Hearst ilustra este algoritmo utilizando un artículo científico de 21 párrafos llamado Stargazers, cuyo tema principal es la existencia de vida en la tierra y en otros planetas, el artículo consiste de 9 subtemas descritos a continuación indicando con números los párrafos.

- 1-3 Introducción- la búsqueda de vida extraterrestre
- 4-5 Estructura química de la luna
- 6-8 Cómo fue creada la luna
- 9-12 Cómo la luna propició la vida en la tierra

- 13 Improbabilidades de sistemas como la tierra y luna
14-16 ¿Por qué los sistemas solares binarios/trinarios no son viables para el desarrollo de vida?
17-18 Probabilidad de un Sistema solar de una estrella.
19-20 Propiedades del sol que promueven la vida.
21 Resumen.

Hearst asume que un juego de unidades léxicas se encuentra presente en lo largo de un subtema y cuando el subtema cambia una porción significativa del vocabulario cambia. Un problema en la segmentación de subtemas es identificar cuando se aborda un subtema y cuando termina, Hearst implementa la identificación del rango de subtemas mediante el algoritmo de comparación de bloques. Pares de bloques son comparados con el promedio de sus similitudes léxicas y este resultado es asignado a la separación entre cada bloque, si la separación tiene un bajo resultado léxico y es precedido por separaciones con altos valores, entonces el resultado menor indica que un hay un cambio en el vocabulario.

Dirección General de Bibliotecas

7 Propuesta de módulo

El prototipo del módulo social analítico será desarrollado sobre una plataforma CRM, un servicio web dedicado a la comunicación B2B y B2C al que sólo pueden acceder personas de una empresa. Con esto se definió que el usuario final contempla a agentes en el departamento de marketing y en el caso de PYME, un emprendedor, ambos con conocimiento acerca del manejo de una computadora. El usuario final espera obtener un reporte estadístico con base a una colección de tweets previamente extraídos. Se clasificaron los objetivos del sistema en dos categorías descritas a continuación:

Objetivos alto nivel:

- Identificar Influencers, seguidores y entidades
- Identificar palabras que describan oportunidades en un servicio o producto

Objetivos secundarios:

- Visualizar el historial de actividad de los tweets

Para realizar el prototipo se utilizó como base Vtiger CRM, por su simplicidad en desarrollo, prueba e implementación de módulos y su documentación.

Se utilizó MAMP para el desarrollo del módulo del prototipo dentro del CRM, para el análisis de datos se utilizará un nodo Hadoop, una plataforma de software que provee la infraestructura para desarrollar y ejecutar software orientado al análisis de información. Este estaría compuesto de una maquina virtual CentOS7 con 4Gb RAM actuando como el nodo principal (Namenode) en modo pseudodistribuido. La infraestructura ideal de un cluster Hadoop se compone por un número grande de computadoras cada una actuando con un rol distinto, por limitaciones de infraestructura y costos administración se utilizó un nodo en modo pseudodistribuido.

A continuación se describirá la función del nodo.

Namenode: Su función principal es delegar trabajos a los nodos esclavo en el clúster, también se utilizó este nodo como un nodo esclavo para procesar la información.

Slavenode: La función de este nodo será analizar la información.

Para extraer Tweets, se contruyó un extractor utilizando la librería de Java llamada Twitter4J (Yamamoto, s.f), la cual permite obtener un flujo de tweets que cumple un criterio de extracción y después entrarán al proceso de normalización de tweets en Hadoop.

Previo al análisis de tweets es necesario normalizar el texto para realizar un etiquetado adecuado:

1.- Del conjunto de tweets extraídos se reemplazó todos los saltos de línea “\n” por espacios.

- 2.- Se eliminaron “RT @Usuario:” del los tweets que son retweeteados.
- 3.- Se eliminaron los puntos suspensivos “...” que generó la Twitter API como resultado de un tweet largo.
- 4.- Se reemplazaron los hipervínculos por “URL”
- 5.- Se reemplazaron los emoji por su equivalente en texto.

1.1. Vectorización del Texto

Para realizar un análisis cuantitativo en los tweets se deben cuantificar los términos, primero se realizará un índice invertido (Stanford, 2008) de todas las palabras que existen en la colección de tweets. Durante la construcción del índice simplemente se asignará a cada nueva palabra el ID del tweet en el que aparece por primera vez, obteniendo un par, una palabra y ID de Tweet. Posteriormente se ordenarán las palabras en orden alfabético y se agruparán las instancias de cada palabra para obtener un diccionario y los identificadores donde aparece cada palabra. (Figura.7-1).

Doc 1		Doc 2			
term	docID	term	docID	term	doc. freq.
I	1	ambitious	2	ambitious	1
did	1	be	2	be	1
enact	1	brutus	1	brutus	2
julius	1	brutus	2	capitol	1
caesar	1	capitol	1	caesar	2
I	1	caesar	1	caesar	2
was	1	caesar	2	caesar	2
killed	1	caesar	2	did	1
i'	1	did	1	enact	1
the	1	enact	1	hath	1
capitol	1	hath	1	I	1
brutus	1	I	1	i'	1
killed	1	I	1	it	1
me	1	i'	1	it	1
so	2	it	2	julius	1
let	2	julius	1	killed	1
it	2	killed	1	let	1
be	2	killed	1	me	1
with	2	let	2	noble	1
caesar	2	me	1	so	1
the	2	noble	2	the	2
noble	2	so	2	told	1
brutus	2	the	1	you	1
hath	2	the	2	was	2
told	2	told	2	with	1
you	2	you	2		
caesar	2	was	1		
was	2	was	2		
ambitious	2	with	2		

term	doc. freq.	→	postings lists
ambitious	1	→	2
be	1	→	2
brutus	2	→	1 → 2
capitol	1	→	1
caesar	2	→	1 → 2
did	1	→	1
enact	1	→	1
hath	1	→	2
I	1	→	1
i'	1	→	1
it	1	→	2
julius	1	→	1
killed	1	→	1
let	1	→	2
me	1	→	1
noble	1	→	2
so	1	→	2
the	2	→	1 → 2
told	1	→	2
you	1	→	2
was	2	→	1 → 2
with	1	→	2

Figura. 7-1 Descripción Gráfica del Índice Invertido. (Stanford, 2008).

El índice invertido permitirá obtener información estadística como la frecuencia de términos y la longitud de los tweets, que se utilizarán para el algoritmo TF-IDF.

Se utilizará el algoritmo TF-IDF el cual consiste en asignar un peso a cada término dentro de una colección de documentos expresada como el producto de la frecuencia del término por la frecuencia inversa del documento:

$$TFIDF = TF(t, D) * IDF(t)$$

El algoritmo TF-IDF es el producto de dos estadísticas, la frecuencia de términos (TF) y la frecuencia inversa del documento (IDF). La frecuencia de términos se refiere al número de veces que un término (t) se observa en un documento (D). Es posible que un término se repita con mayor frecuencia en documentos que tienen una mayor longitud, por este motivo se divide el número de observaciones sobre la longitud total del documento. De esta forma TF se expresa como:

$$TF(t, D) = \frac{count(t,D)+1}{D}$$

Donde:

t término

D Documento (Tweet)

Durante el cálculo de TF todos los términos en la colección de documentos son tratados por igual, sin embargo, existen términos que se siempre se utilizan y pueden tener poca importancia, por lo tanto se necesita penalizar términos muy recurrentes y compensar los términos raros con la siguiente expresión:

$$IDF(t) = \log\left(\frac{|D|+1}{DF(t,D)+1}\right)$$

Debido a que un término t tiene un peso TFIDF distinto en cada documento (tweet), se realizara el promedio del término en la colección de tweets mediante:

$$PromedioTFIDF(TFIDF_t) = \frac{\sum TFIDF_t}{|t|}$$

Con el promedio TFIDF de cada término se puede calcular el peso de cada tweet mediante el promedio de los términos que están en el tweet.

$$TweetTFIDF = \frac{t_1 + t_2 + \dots + t_n}{|d|}$$

Se filtrarán los términos en la colección en tres categorías, hashtags, usuarios, y términos, en los términos se decidió remover aquellos menores a tres caracteres ya que son palabras utilizadas gramaticalmente en el idioma español y por lo tanto, no serán de utilidad.

Los tweets analizados serán presentados al usuario de la siguiente forma:

Área con gráfica de barras: Se mostrarán los usuarios, hashtags y términos en sus respectivas graficas por separado mostrando los primeros 30 elementos más altos, en el caso de los términos se mostrarán los primeros 90 más altos.

Área con nube de términos: Como complemento al área de grafica de barras se mostrarán en forma de nube de términos los primeros 50 términos más altos de cada categoría, asignándoles colores a cada término de acuerdo a su peso TFIDF.

Área de tweets con mayor peso TFIDF: Esta parte indicará los tweets con el peso TFIDF más alto, junto al tweet aparecerá su peso.

Área con historial de actividad: Esta área mostrará el comportamiento de la actividad de los usuarios durante el transcurso del tiempo utilizando la cantidad de tweets publicados en relación al tiempo.

7.1 Requisitos

Se han dividido los requisitos en funcionales y no funcionales, estas dos especifican qué características debe tener el prototipo y cómo se deben hacer las cosas en cada parte del sistema para que funcione tal y como se ha especificado.

7.1.1 Requisitos Funcionales

Identificador	F-001
Nombre	Introducir Palabras de búsqueda
Descripción	El usuario introduce las palabras a buscar
Necesidad	Alta
Dependencia	No aplica

Tabla. 7-1 Requisito funcional 1

Identificador	F-002
Nombre	Extraer Tweets
Descripción	A través del API de twitter extraerá los tweets que contengan las palabras a buscar
Necesidad	Alta
Dependencia	F-001

Tabla. 7-2 Requisito funcional 2

Identificador	F-003
Nombre	Normalización de tweets
Descripción	Acomodará los datos para su análisis posterior
Necesidad	Alta
Dependencia	F-002

Tabla. 7-3 Requisito funcional 3

Identificador	F-004
Nombre	Índice de documentos
Descripción	Crea un índice de tweets y qué términos contiene
Necesidad	Alta
Dependencia	F-003

Tabla. 7-4 Requisito funcional 4

Identificador	F-005
Nombre	Contador de tweets
Descripción	Contará cuántos tweets iguales hay
Necesidad	media
Dependencia	F-003

Tabla. 7-5 Requisito funcional 5

Identificador	F-006
Nombre	Contador de localización
Descripción	Obtiene la geolocalización de los tweets si es disponible.
Necesidad	baja
Dependencia	F-004

Tabla. 7-6 Requisito funcional 6

Identificador	F-007
Nombre	Contador de fechas
Descripción	Obtiene cuántos tweets fueron publicados por día y hora
Necesidad	media
Dependencia	F-004

Tabla. 7-7 Requisito funcional 7

Identificador	F-008
Nombre	Actividad de usuario
Descripción	Obtiene cuántos tweets fueron retweeteados y su autor original
Necesidad	media
Dependencia	F-004

Tabla. 7-8 Requisito funcional 8

Identificador	F-009
Nombre	Frecuencia de término
Descripción	Obtiene la frecuencia de todos los términos en la colección de documentos.
Necesidad	alta

Dependencia	F-004
-------------	-------

Tabla. 7-9 Requisito funcional 9

Identificador	F-010
Nombre	Frecuencia Inversa de término
Descripción	Obtiene la frecuencia inversa de todos los términos en la colección de documentos.
Necesidad	alta
Dependencia	F-004

Tabla. 7-10 Requisito funcional 10

Identificador	F-011
Nombre	TFIDF
Descripción	Multiplica la frecuencia de término con la frecuencia inversa de término.
Necesidad	alta
Dependencia	F-009, F-010

Tabla. 7-11 Requisito funcional 11

Identificador	F-012
Nombre	Promedio TFIDF por término
Descripción	Obtiene el promedio TFIDF por cada término en la colección
Necesidad	alta
Dependencia	F-011

Tabla. 7-12 Requisito funcional 12

Identificador	F-013
Nombre	Eliminación de términos
Descripción	Elimina los términos menores a tres caracteres.
Necesidad	media
Dependencia	F-012

Tabla. 7-13 Requisito funcional 13

Identificador	F-014
Nombre	Vectorización de tweets
Descripción	Crea vectores de tweets con los pesos TFIDF de los términos que lo conforman.
Necesidad	alta
Dependencia	F-013

Tabla. 7-14 Requisito funcional 14

Identificador	F-015
Nombre	Promedio TFIDF por tweet
Descripción	Obtiene el promedio TFIDF de cada tweet.
Necesidad	alta
Dependencia	F-014

Tabla. 7-15 Requisito funcional 15

Identificador	F-016
Nombre	Filtrado de Hashtags
Descripción	Obtiene el promedio TFIDF de cada tweet.
Necesidad	alta
Dependencia	F-015

Tabla. 7-16 Requisito funcional 16

Identificador	F-017
Nombre	Filtrado de usuarios
Descripción	Obtiene el promedio TFIDF de cada tweet.
Necesidad	alta
Dependencia	F-015

Tabla. 7-17 Requisito funcional 17

Identificador	F-018
Nombre	Filtrado de palabras
Descripción	Obtiene el promedio TFIDF de cada tweet.
Necesidad	alta
Dependencia	F-015

Tabla. 7-18 Requisito funcional 18

Identificador	F-019
Nombre	Ploteo de información
Descripción	Muestra la información al usuario en forma de: Gráficas de barras. Nube de términos. Gráficas lineales.
Necesidad	alta
Dependencia	F-016, F-017, F-018

Tabla. 7-19 Requisito funcional 19

Identificador	F-018
Nombre	Filtrado de palabras

Descripción	Obtiene el promedio TFIDF de cada tweet.
Necesidad	alta
Dependencia	F-015

Tabla. 7-20 Requisito funcional 20

Identificador	F-019
Nombre	CRM vTiger
Descripción	Es la plataforma donde el prototipo del módulo social estará.
Necesidad	Alta
Dependencia	N/A

Tabla. 7-21 Requisito funcional 21

7.1.2 Requisitos no funcionales, Interfaces externas

Identificador	RNFIE-001
Nombre	API Twitter
Descripción	Se encarga de establecer la conexión del extractor con el servidor de Twitter
Necesidad	alta
Dependencia	N/A

Tabla. 7-22 Requisito no funcional 1

Identificador	RNFIE-002
Nombre	API Google Graphs
Descripción	Se encarga de mostrar la información analizada en forma de graficas lineales.
Necesidad	media
Dependencia	N/A

Tabla. 7-23 Requisito no funcional 2

Identificador	RNFIE-003
Nombre	API D3
Descripción	Se encarga de mostrar la información analizada en forma de graficas de barras y nube de términos.
Necesidad	media
Dependencia	N/A

Tabla. 7-24 Requisito no funcional 3

Identificador	RNFIE-004
---------------	-----------

Nombre	Librería Twitter 4J
Descripción	Es la interfaz entre el extractor y el API de twitter.
Necesidad	alta
Dependencia	RNFIE-001

Tabla. 7-25 Requisito no funcional 3

Identificador	RNFIE-005
Nombre	Librería Vdurmont emoji-java
Descripción	Se encarga de transformar los emojis a texto.
Necesidad	media
Dependencia	N/A

Tabla. 7-26 Requisito no funcional 4

Identificador	RNFIE-006
Nombre	Librería Hadoop
Descripción	Es necesario para analizar la información.
Necesidad	alta
Dependencia	N/A

Tabla. 7-27 Requisito no funcional 5

7.1.3 Requisitos no Funcionales Interfaz de Usuario

Identificador	RNFUI-001
Nombre	Conexión a internet
Descripción	Es necesario que el equipo esté conectado a internet para extraer los tweets
Necesidad	alta
Dependencia	N/A

Tabla. 7-28 Requisito no funcional de interfaz de usuario 1

Identificador	RNFUI-002
Nombre	Java Virtual Machine
Descripción	Es necesario que el equipo tenga instalado JVM para ejecutar el extractor.
Necesidad	alta
Dependencia	N/A

Tabla. 7-29 Requisito no funcional de interfaz de usuario 2

7.1.4 Requisitos de rendimiento

Identificador	RNR-001
Nombre	Número máximo de tweets
Descripción	Se desplegarán en la gráfica máximo 30 tweets
Necesidad	media
Dependencia	N/A

Tabla. 7-30 Requisito de rendimiento 1

Identificador	RNR-002
Nombre	Número máximo de términos
Descripción	Se desplegarán en la gráfica máximo 80 términos
Necesidad	media
Dependencia	N/A

Tabla. 7-31 Requisito de rendimiento 2

Identificador	RNR-003
Nombre	Número máximo de hashtags
Descripción	Se desplegarán en la gráfica máximo 30 hashtags
Necesidad	media
Dependencia	N/A

Tabla. 7-32 Requisito de rendimiento 3

Identificador	RNR-004
Nombre	Número máximo de usuarios
Descripción	Se desplegarán en la gráfica máximo 30 usuarios
Necesidad	media
Dependencia	N/A

Tabla. 7-33 Requisito de rendimiento 4

7.2 Diagrama de clases

En este apartado se describirán de qué forma están relacionadas las clases del servidor Hadoop, sus atributos y métodos que se pueden utilizar en cada una de ellas.

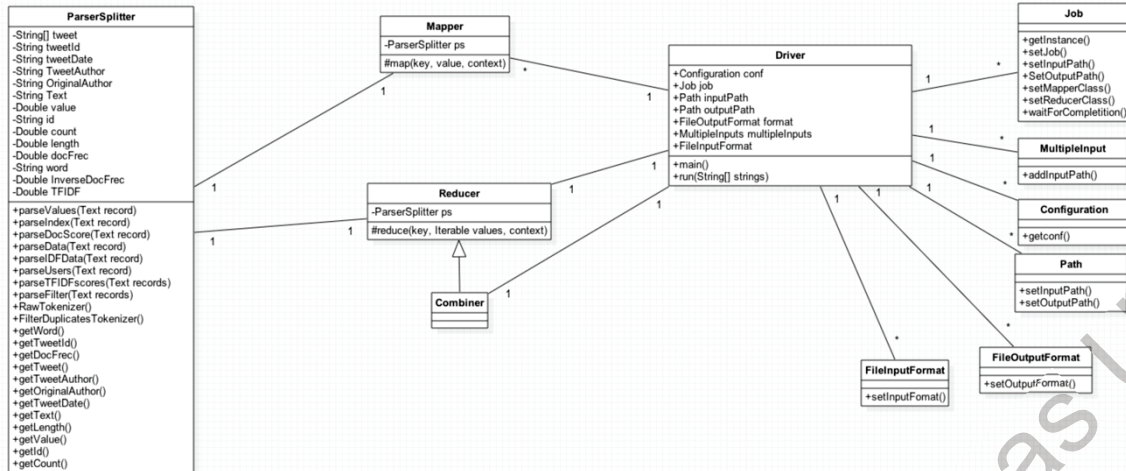


Figura. 7-2 Diagrama de clases utilizadas en Hadoop

Para entender el diagrama de clases en la figura 7-2, se iniciará describiendo el diagrama en dos partes, con el fin de ofrecer una mejor visualización del diagrama de clases se decidió resumir las veintitrés clases que heredan de la clase Mapper y Reducer.

En la figura 7-3, corresponde a la clase Mapper y Reducer, componentes principales del proceso Map-Reduce de Hadoop. Es importante mencionar que cada proceso Map-Reduce requiere de una entrada y produce una salida que se guarda en una dirección dentro del sistema de archivos HDFS (Hadoop Distributed File System), cuya finalidad es almacenar archivos de forma distribuida, es decir, almacenarlos a través de diferentes computadoras. La clase Mapper es la encargada de procesar una entrada de datos en el sistema de archivos HDFS para enviarlos al Reducer en forma de pares de datos (id, valor). La finalidad de la clase Reducer es agrupar todos los valores con el mismo id para ejecutar un proceso en ellos, la clase Combiner es utilizada para reducir el volumen de transferencia de datos entre el Mapper y Reducer, actuando como un Semi-Reducer. Finalmente la clase ParserSplitter será de utilidad para separar los datos de entrada según el caso.

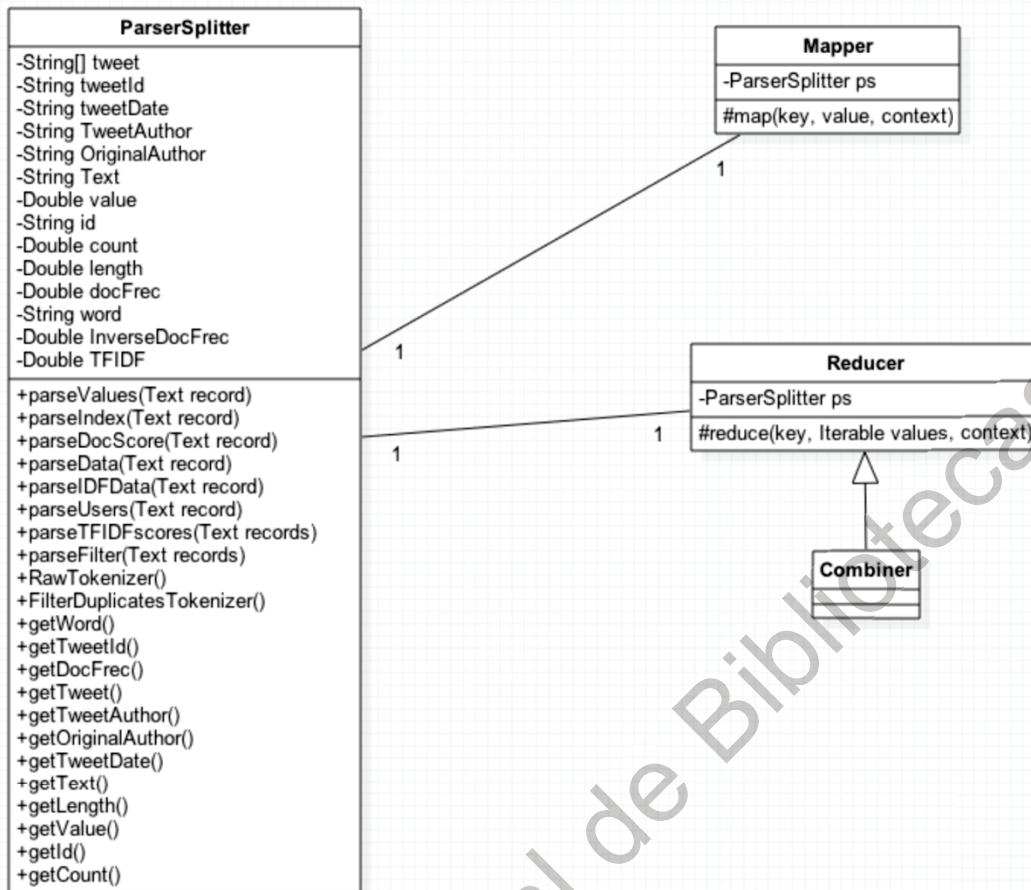


Figura. 7-3 Diagrama de clases Map-Reduce.

En la figura 7-4, es la que corresponde a la parte de la clase Driver, esta es la encargada de controlar el flujo, configuración y seguimiento de los trabajos ejecutados en Hadoop. En la clase Driver se asignarán las clases Mapper y Reducer a un objeto Job, este objeto tendrá información acerca del tipo de formato de entrada, directorio de entrada, directorio de salida y formato de salida. En algunos casos se utilizará más de una entrada, en este caso la clase MultipleInput permite utilizar dos clases Mapper y una clase Reducer.

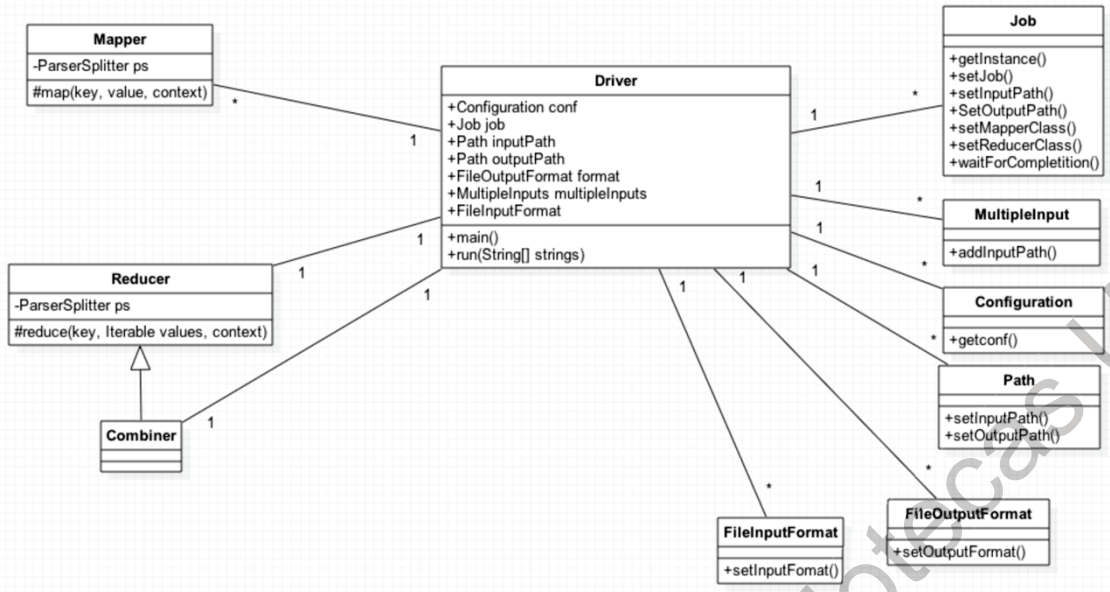


Figura. 7-4 Diagrama de clases correspondiente a componentes del Driver.

7.3 Interfaz de usuario

La información analizada será presentada al usuario de esta forma:

Dirección General de Bibliotecas UAQ

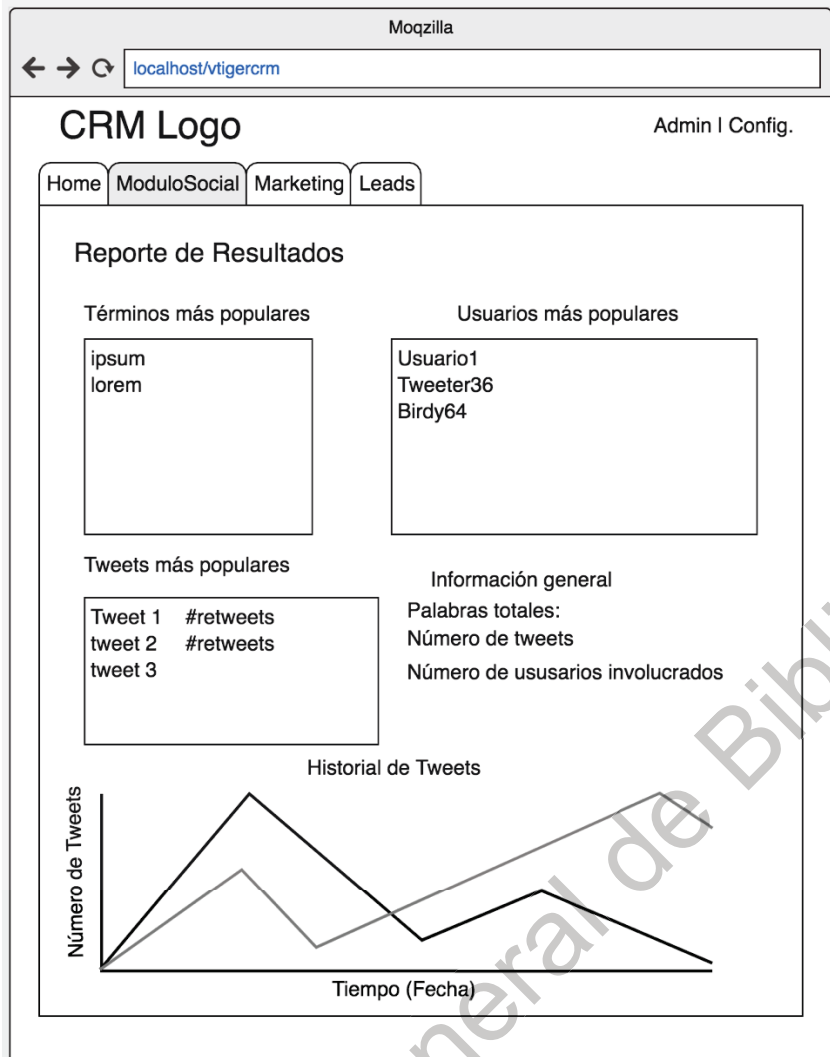


Figura. 7-5 Interfaz de usuario.

8 Caso Práctico

Con la herramienta de extracción de tweets, fueron guardados en formato JSON, un formato utilizado para el intercambio de datos de fácil lectura y escritura (Crockford, D, s.f.). aquellos tweets que incluyeran el hashtag #uber y #cabify, durante la tercera y cuarta semana de septiembre del año 2017 en periodos de una hora cada día. Cada vez que el extractor terminaba, este generó un archivo con los tweets extraídos como se muestra en la figura 8-1.

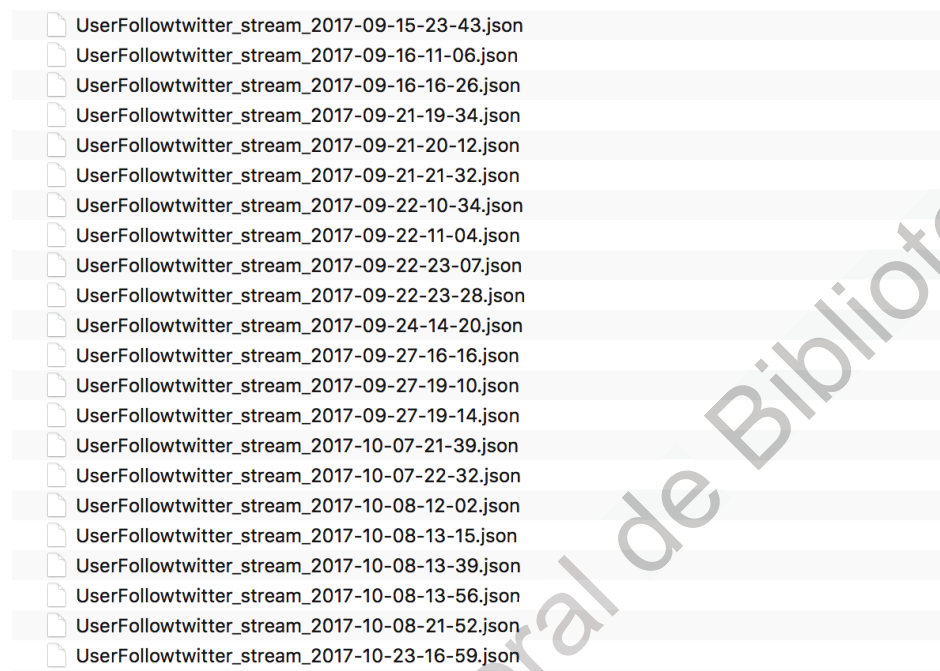


Figura. 8-1 Archivos generados por el extractor.

Posteriormente se procedió a juntar todos los archivos en uno solo con el siguiente diagrama de flujo en la figura 8-2, cuya función consiste en imprimir en un archivo nuevo todas las líneas de los archivos con terminación JSON contenidas en un directorio:

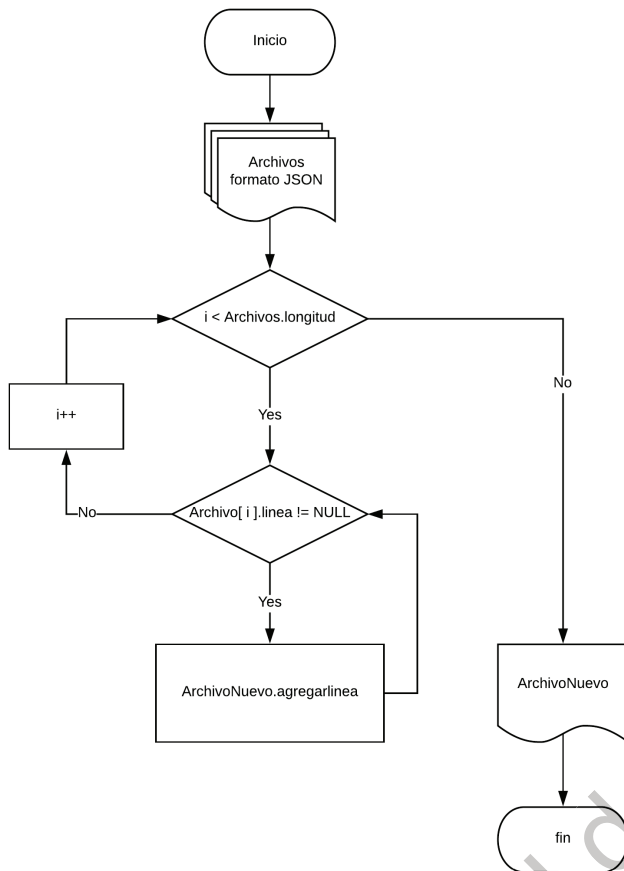


Figura. 8-2 Diagrama de flujo para unir archivos JSON.

En el siguiente paso se identificaron los valores que contiene un tweet (Figura 8-3) y se determinó que los valores se utilizarían en la fase de análisis serían: `createdAt`, `id`, `text`, `user.screenName`, `retweetedStatus.id`, `retweetedStatus.user.screenName`. Se decidió separar los valores de la colección de tweets en diferentes pares de datos con el fin de que fueran legibles para el servidor Hadoop, los pares de datos estarían compuestos de la siguiente forma: `id`, `createdAt` en el archivo de nombre `mapped-id-date`; `id`, `retweetedStatus.id` en el archivo de nombre `mapped-id-id`; `id`, `user.screenName`, `retweetedStatus.user.screenName` en el archivo de nombre `mapped-id-RTU-OTU`; `id`, `text` en el archivo de nombre `mapped-id-text`. A su vez, se realizó la normalización del valor `text` para cambiar emoji por una representación textual y eliminar caracteres especiales.

```
JSON
  createdAt : "Oct 7, 2017 9:46:40 PM"
  id : 916857348504870900
  text : "RT @PretoLG222: Uber tocando pagode antigo ❤️❤️❤️❤️"
  source : "Twitter for Android"
  isTruncated : false
  inReplyToStatusId : -1
  inReplyToUserId : -1
  isFavorited : false
  isRetweeted : false
  favoriteCount : 0
  retweetCount : 0
  isPossiblySensitive : false
  lang : "es"
  contributorsIDs
  retweetedStatus
  userMentionEntities
  uriEntities
  hashtagEntities
  mediaEntities
  extendedMediaEntities
  symbolEntities
  currentUserRetweetId : -1
  user
  quotedStatusId : -1
```

Figura. 8-3 Estructura de un Tweet.

Dirección General de Bibliotecas UAAQ

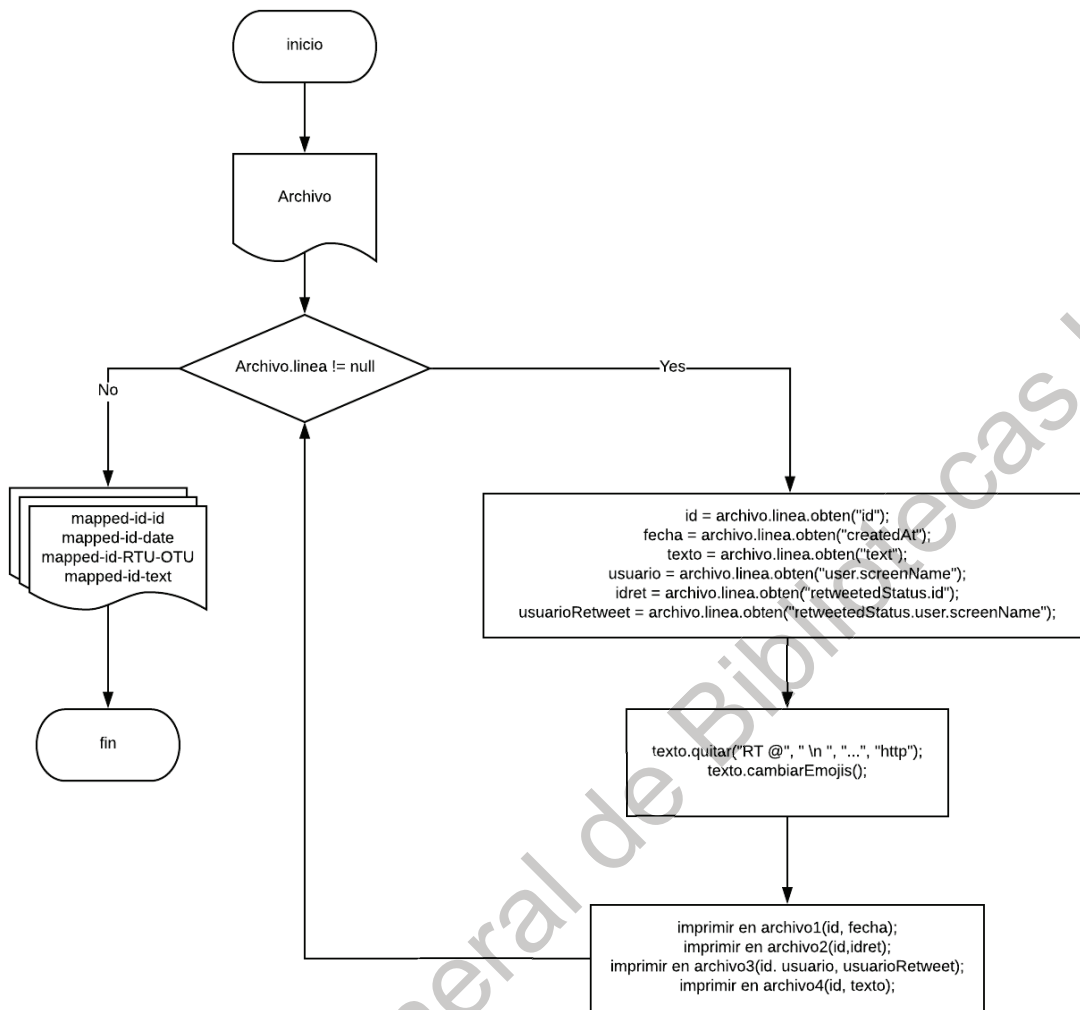


Figura. 8-4 Diagrama de flujo para separar los valores en diferentes archivos.

Una vez separados los valores en diferentes archivos, se transfirieron al servidor Hadoop y se subieron al sistema de archivos HDFS en los siguientes directorios, se les asignará un nombre a cada dirección para evitar confusión:

- /input/TweetCount/mapped-id-id.tsv; como dirección1
- /input/TweetDate/mapped-id-date.tsv; como dirección2
- /input/TweetUsers/mapped-id-RTU-OTU.tsv; como dirección3
- /input/TweetWordCount/mapped-id-text.tsv; como dirección4

En la propuesta de módulo se mencionaron las clases Mapper y Reducer así como el sistema de archivos HDFS que utiliza Hadoop para analizar la información, a continuación, se describirán los procesos que se llevaron a cabo para el cálculo TFIDF que consiste en 5 pasos.

En el primer paso, utilizando el archivo en la dirección 4, en se realizó un índice conteniendo las veces que aparece cada palabra en toda la colección de tweets

utilizando cada palabra como llave y un valor de 1 en el proceso Mapper para que en el Reducer se realice la sumatoria de todos los valores que tengan la misma llave, el índice generado fue guardado en el directorio /output/RawWordCount, en la figura 8-5 se muestra un fragmento de la entrada y salida de datos de este proceso.

/output/TweetWordCount

```

908914275460145153      #Mujeres #Mexicanas si usan #Taxi #Uber #Cab
908914323875131393      Sabes que valió madre cuando te dicen que ni
908914338869768192      Lo hizo porque sabía que podía hacerlo, porq
908914345039532034      Consejo #Mujeres :mx: :taxi: URL
908914361174863872      A Mara la mataron porque el criminal sabía q
908914397795495936      La @FiscaliaPuebla apunta que Ricardo Alexis
908914411791843328      El miedo no anda en burro, anda en Cabify.
908914423762268160      En Cabify, en taxi, en transporte público, a
908914426882985985      Información que cura. #uber #cabify #Cabifyr
908914439113625600      ¿Y Cabify? ¿Qué del servicio que llegó como '
  
```

/output/RawWordCount

```

#          2
#2doinformejmcarreras  1
#a          1
#adnmotor          1
#afp          3
#ahora          1
#ald          3
#alertadegeneropuebla  1
#argentina          1
#aviso          1
#ayudacdmx          5
  
```

Figura. 8-5 Entrada y salida del primer paso.

Seguido de este proceso, se generó un índice de tweets, utilizando el mismo archivo en la dirección 4, pasando como llave el texto de los tweets y como valor el índice que genera Hadoop en la clase Mapper, este índice de tweets fue guardado en el directorio /output/DocIndex, este índice será esencial para la vectorización de tweets, en la figura 8-5 se muestra un fragmento de la entrada y salida de datos de este proceso.

/output/TweetWordCount

```
908914275460145153 #Mujeres #Mexicanas si usan #Taxi #Uber #Cab:
908914323875131393 Sabes que valió madre cuando te dicen que ni
908914338869768192 Lo hizo porque sabía que podía hacerlo, porqu
908914345039532034 Consejo #Mujeres :mx: :taxi: URL
908914361174863872 A Mara la mataron porque el criminal sabía q
908914397795495936 La @FiscaliaPuebla apunta que Ricardo Alexis
908914411791843328 El miedo no anda en burro, anda en Cabify.
908914423762268160 En Cabify, en taxi, en transporte público, a
908914426882985985 Información que cura. #uber #cabify #Cabifyr
908914439113625600 ¿Y Cabify? ¿Qué del servicio que llegó como '
```

/output/DocIndex

```
Ayer pedí un uber para regresar a mi casa después de ayudar un poco en donde se necesitaba, y cuando me bajé el señor c
. 68940
Ayúdame a difundir y conseguir mas recursos. #frida #FridaRescatista #FridaStickers @CulturaColectiv 64811
Ayúdame a difundir y conseguir mas recursos. #frida #FridaRescatista #FridaStickers @CulturaColectiv 62892
Ayúdame a difundir y conseguir mas recursos. #frida #FridaRescatista #FridaStickers @CulturaColectiv 63755
Ayúdame a difundir y conseguir mas recursos. #frida #FridaRescatista #FridaStickers @CulturaColectiv 64520
Ayúdame a difundir y conseguir mas recursos. #frida #FridaRescatista #FridaStickers @CulturaColectiv 68666
Ayúdenos a encontrarla, ayer a las 10 tomé un uber y no ha llegado a su casa. No se ha localizado al chófer. 55527
Ayúdenos a encontrarla, ayer a las 10 tomé un uber y no ha llegado a su casa. No se ha localizado al chófer. 52356
Ayúdenos a encontrarla, ayer a las 10 tomé un uber y no ha llegado a su casa. No se ha localizado al chófer. 61509
Ayúdenos a encontrarla, ayer a las 10 tomé un uber y no ha llegado a su casa. No se ha localizado al chófer. 60123
.....
```

Figura. 8-6 Entrada y salida del segundo paso.

En el tercer paso, se utilizó el índice generado en el paso anterior para tokenizar cada tweet, es decir, segmentar el texto en unidades de palabras, para lograr esta salida, se utilizó como llave la concatenación del id del índice con cada palabra y como valor de salida se utiliza el valor de 1, para realizar posteriormente en el Reducer la sumatoria de los valores que coinciden con la llave, esta sumatoria representa la frecuencia de veces que aparece una palabra en un tweet. Este índice fue almacenado en el directorio /output/VectorizedDocuments posteriormente se procedió a eliminar del índice las palabras menores a tres caracteres con el fin de producir resultados más precisos, el nuevo índice filtrado fue generando un índice filtrado en el directorio /output/FilteredVectDocs, como se muestra en la figura 8-7, en la columna izquierda se muestra un índice sin filtrar, en la columna derecha se muestra un índice después de pasar por el proceso de filtro, la cantidad de palabras que fueron eliminadas de un solo tweet es significativo y será de ayuda para obtener una mejor visualización del contexto.

0	#cabify	1							
0	#mexicanas		1						
0	#mujeres		1						
0	#taxi	1							
0	#uber	1							
0	a	1							
0	amistad	1							
0	c	1							
0	compartan		1						
0	del	1							
0	detalles		1						
0	familiar		1						
0	llamen	1							
0	los	1							
0	no	1							
0	o	1							
0	si	1							
0	una	1							
0	usan	1							
0	viaje	1							
0	y	1							

0	#cabify					1			
0	#mexicanas					1			
0	#mujeres					1			
0	#taxi	1							
0	#uber	1							
0	amistad					1			
0	compartan					1			
0	detalles					1			
0	familiar					1			
0	llamen	1							
0	usan	1							
0	viaje	1							

Figura. 8-7 Índice de Frecuencia de términos antes y después de la fase de filtrado.

Para calcular el TFIDF, se necesita calcular la frecuencia de término denominada como TF y la frecuencia inversa del documento denominada como IDF. Para el Mapper del archivo en /output/FilteredVectDocs, se obtuvo el valor TF mediante el logaritmo natural de la frecuencia de términos mas 1, después se utilizó como llave la palabra y como valor la concatenación de una etiqueta "TF" con el valor obtenido en este proceso. Así mismo en el Mapper para el archivo en /output/RawWordCount, se calculó el valor IDF mediante el logaritmo natural del cociente del total de tweets entre las veces que aparece una palabra en el tweet a los resultados de este proceso se les concatenará la etiqueta "IDF" y se utilizará la palabra como llave para enviarlas al Reducer. El Reducer de este proceso recibirá los valores de los Mappers descritos previamente, de acuerdo a la etiqueta que reciba, realizará un proceso diferente sobre ese valor. Si el valor contiene la etiqueta "IDF", este será guardada en una variable llamada idf, si el valor contiene la etiqueta "TF" se asignará el valor en una variable llamada tf y se procederá a el cálculo TFIDF el cual consiste en el producto de las variables tf e idf los resultados producidos fueron almacenados en /output/SubTFIDFValues. Este resultado representa el valor TFIDF por palabras en cada documento, es decir, cada palabra tendrá un valor TFIDF diferente en cada tweet, con base a estos valores se calculó el valor TFIDF por documento mediante el promedio de los valores TFIDF de las palabras que componen al tweet y se almacenó en el directorio /output/DocumentScores, adicionalmente se realizó un promedio de valores TFIDF por palabra con el fin de obtener un único valor por palabra almacenado en el directorio /output/IndWordTFIDFAvg. Se decidió mostrar los valores TFIDF en tres categorías: hashtags, usuarios y palabras, para esto se filtraron las palabras contenidas en el archivo del directorio /output/IndWordTFIDFAvg, las palabras que empezaran con el símbolo # fueron enviadas a un archivo localizado en /output/IndWords/Hashtags, las

palabras que empezaran con el símbolo @ fueron enviadas a un archivo localizado en /output/IndWords/Users y las palabras que no empezaran con los valores anteriores serían enviadas a un archivo localizado en /output/IndWords/Words.

Desde el servidor donde se alojó el CRM, se extrajeron los archivos del servidor Hadoop para mostrar una representación gráfica de los datos, se decidió representar los datos en forma de nube de términos y gráficas de barras, también se decidió graficar el número de veces que se publicaron tweets por hora y el número de tweets publicados por usuario.

De los tweets extraídos obtuvimos la siguiente información:

Número total de tweets extraídos	3104 tweets
Número de términos únicos (hashtags, usuarios y palabras) antes del filtrado	5046 términos
Número de términos únicos después del filtrado	3858 términos
Número de hashtags únicos	190 hashtags
Número de usuarios mencionados	483 usuarios
Número de usuarios retweeteados	43 usuarios

Tabla. 8-1 Información de resultados general.

Lista de Tweets más altos.

id	TFIDF	Tweet
199756	3.6000111111111104	No quiero usar Uber con miedo. No quiero usar Cabify. No quiero usar un taxi. No quiero salir con miedo de no regresar. #Mara
199902	3.6000111111111104	Los tuits machistas que culpabilizan a la joven asesinada de 19 años en un Cabify en México URL
260942	3.078071428571428	@Miriposita Abrazo grande grande grande para allá. Y la oferta del uber no caduca
261044	3.078071428571428	@martchoka @juancristonomo Uber. Como tienen bloqueado el aeropuerto, pones el pin en el río y lo pedís. Cuando est URL
114765	3.02934	@Cabify_Mexico Ni un viaje más en @Cabify_Mexico no útilices el servicio de cabify URL
114874	3.02934	Por seguridad, Mara Castilla tomó un Cabify para regresar a casa; nunca llegó :pensive:#NiUnaMás URL

104133	2.995809090909091	@LuisBanck VIAJA EN UN SERVICIO SEGURO UBER TÉ LLEBA A SU DESTINÓ SEGURO UBER EN EL TURISMO MEXICANO VISITA PUEBLA
10417	2.995809090909091	No es Cabify, no es el antro ni la hora en que quiso divertirse. A #MaraCastilla la mató el machismo, el odio y la impunidad.
8981	2.995809090909091	Convocan a marcha este domingo en la CDMX para pedir justicia por el feminicidio de #MaraCastilla. #NiUnaMenos.
9113	2.995809090909091	"Quiten el ""llegando a casa a salvo"", más respeto por #MaraFernanda URL"
9330	2.995809090909091	Consejo #Mujeres :mx: :taxi: URL
9509	2.995809090909091	#Mujeres #Mexicanas si usan #Taxi #Uber #Cabify llamen a una amistad o familiar y compartan los detalles del viaje. No c
95897	2.995809090909091	"Buen trato? HP, te hago un favor a"
96002	2.6605	Si es cierto ya sea Uber o Taxistas por que el afán de hacer platica!? A veces uno no está de humor!
96174	2.6605	ojalá mueran todos, y los trague la tecnología (?) URL
96250	2.6604999999999994	El presunto asesino del conductor Uber en la colonia Real del Carmen sólo tiene 18 años de edad... :arrow_right:
96389	2.6604999999999994	Si tienes que salir utiliza mi código de Uber :stars:ANGELADELRIO:stars: para 3 viajes gratis *válido en todo el mundo URL
9649	2.6605	@Cabify_Mexico ALV cabify! Nunca más usaré su app! #fueraCabifydemexico
96535	2.6605	Mientras en el ""Dubái de las Américas"" los taxis y buses piratas están de paro, acosan a Uber a diario y cobran lo
96673	2.6604999999999994	@elmundoes O quizá los que despachan periódicos o los

		conductores de buses o los choferes de Uber o....
96798	2.6604999999999994	Uber pierde su licencia para operar en Londres URL
96868	2.6605	@AgustinaaCaab Veni q te pago el uber
97075	2.6604999999999994	Uber pierde su licencia para operar en Londres, afirman que apelarán URL
97169	2.6604999999999994	@el_shura Snapchat y el servicio de Uber sin límite, MB libres y Claro video gratis desde \$399.00 al mes, además pu URL
97410	2.6604999999999994	- Papá: ¿quieres que te lleve a clase? - Yo: ¡ay sí! - Papá: ¿o te pido un Uber? - Yo: :unamused: tranquilo, yo me voy en bicicleta
9742	2.6604999999999994	#MaraCastilla fue asesinada en un motel por conductor de Cabify; es feminicidio 58: @FiscaliaPuebla Ángulo 7 URL
97567	2.6605	"No, joven, si yo era rescatista. Pero ya ve, ahora aquí and
97701	2.6604999999999994	Que parte de NO QUIERO PLATICAR no entiende el chofer de Uber !
97784	2.6604999999999994	Nuevas reglas para #Cabify y #Uber en la #CDMX, para restringir sus operaciones y tener nuevas obligaciones URL

Tabla. 8-2 Lista de tweets con TFIDF alto

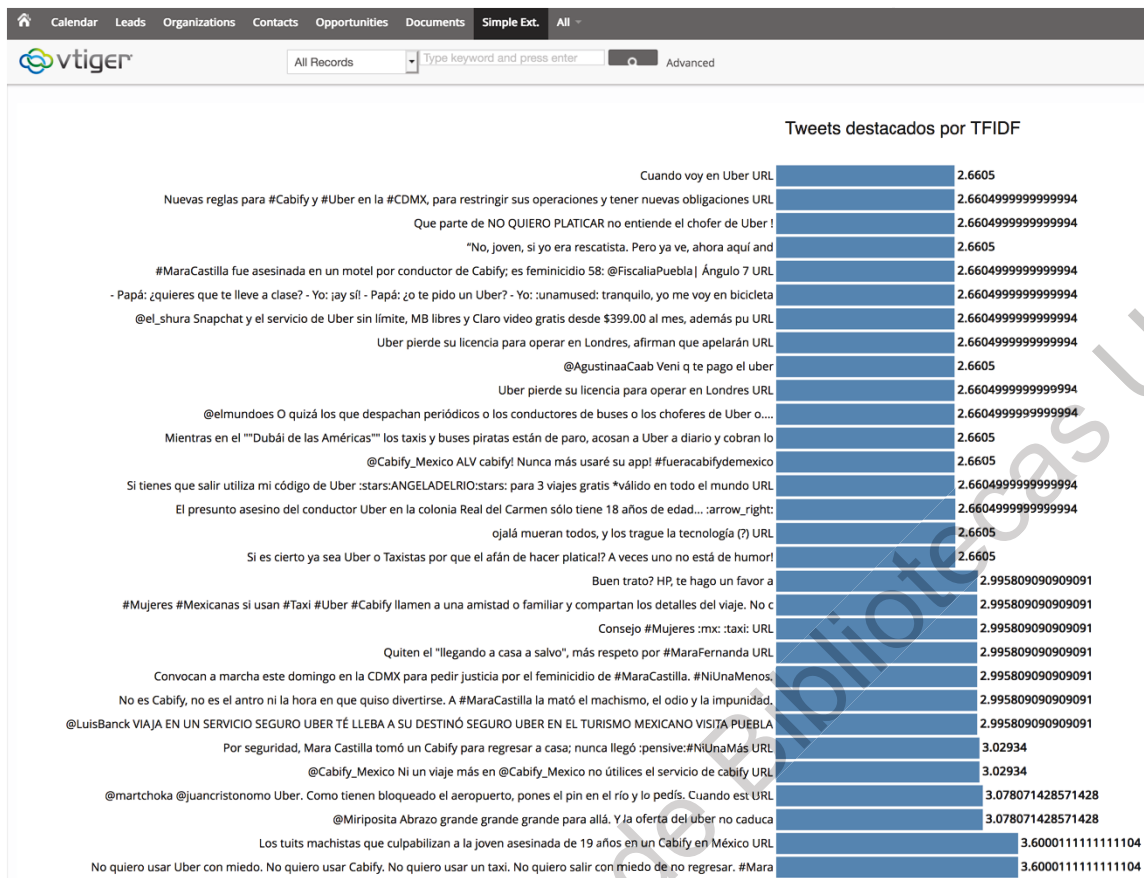


Figura. 8-8 Grafica de tweets destacados por TFIDF.

Lista de hashtags más altos.

Hashtag	TFIDF
#uber	2.6762623931624003
#maracastilla	2.6605000000000006
#cabify	2.6605000000000005
#niunam	2.6605000000000025
#2doinformejmcarreas	2.6605
#adnmotor	2.6605
#afp	2.6605
#ahora	2.6605
#ald	2.6605
#alertadegeneropuebla	2.6605
#argentina	2.6605
#aviso	2.6605
#ayudacdmx	2.6605
#ayudasismo	2.6605
#bares	2.6605
#betaajudabeta	2.6605

#betaquerlab	2.6605
#blogs	2.6605
#btc	2.6605
#buenviernes	2.6605
#bye	2.6605
#cabifyesservicio	2.6605
#cabifymexico	2.6605

Tabla. 8-3 Hashtags con TFIDF alto

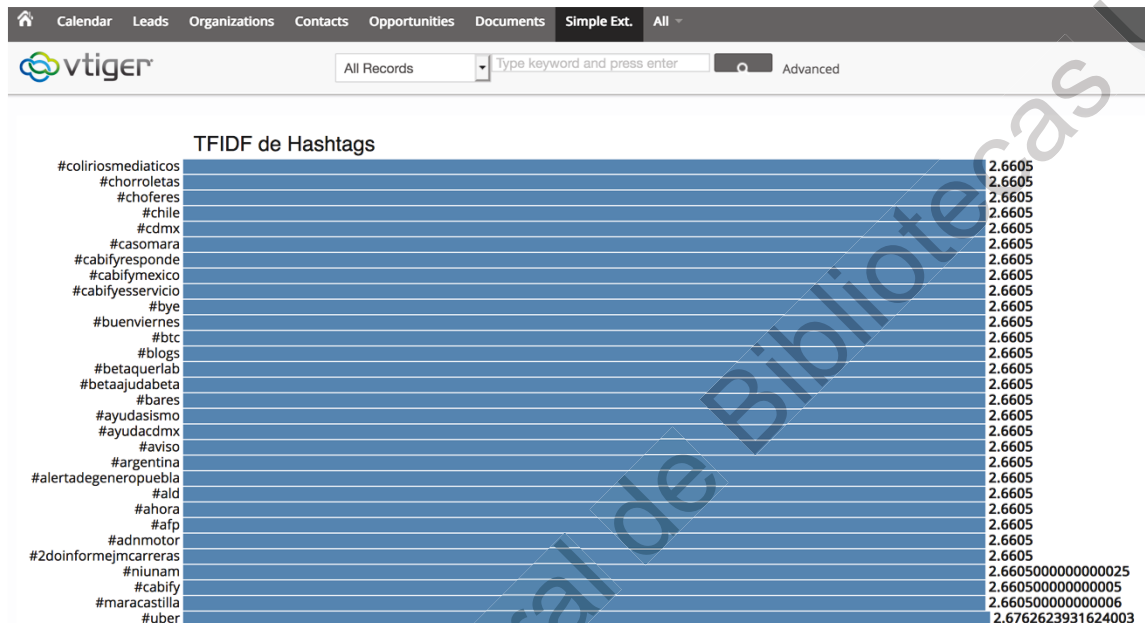


Figura. 8-9 Gráfica de hashtags destacados.

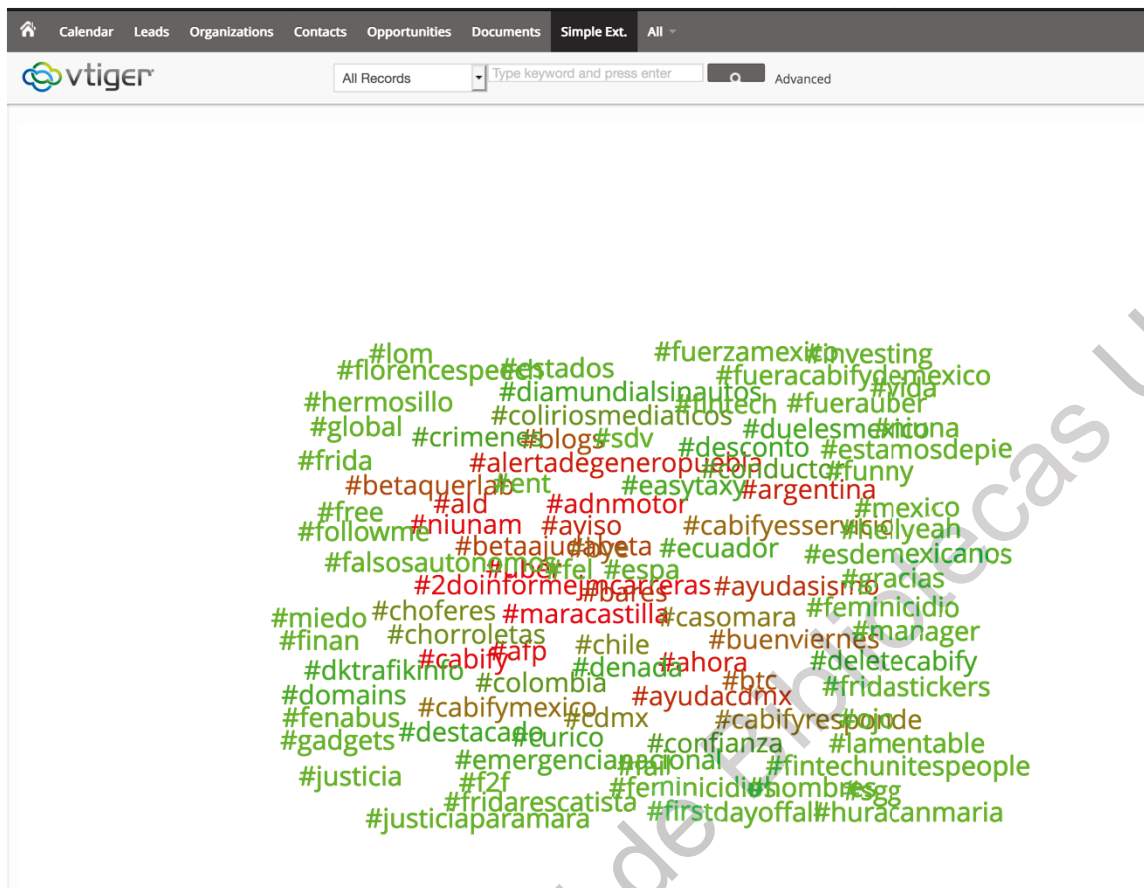


Figura. 8-10 Nube de hashtags.

Lista de Usuarios más altos

Usuario	TFIDF
@uber_mex	2.713191428571428
@cabify_mexico	2.7124492957746513
@uber	2.6605000000000002
@26monique	2.6605
@4lezka	2.6605
@_cesartang14	2.6605
@_roberttg	2.6605
@_zavandija	2.6605
@a_esparzaoteo	2.6605
@abbu25	2.6605
@aearcos	2.6605
@afpespanol	2.6605
@agustinaacaab	2.6605
@ahope71	2.6605
@albaniaoc	2.6605
@aleakbal	2.6605
@alepinillos	2.6605

@alex_081987	2.6605
@alexhsillo	2.6605
@alfonsodasis	2.6605
@alfredelmazo	2.6605
@alfredoalifer_	2.6605
@alfstardust	2.6605
@alvaro_escobar_	2.6605
@alynmon	2.6605
@amicsonora	2.6605
@anacelialara	2.6605
@anapatriciatv	2.6605
@anapruis	2.6605

Tabla. 8-4 Usuarios con TFIDF alto

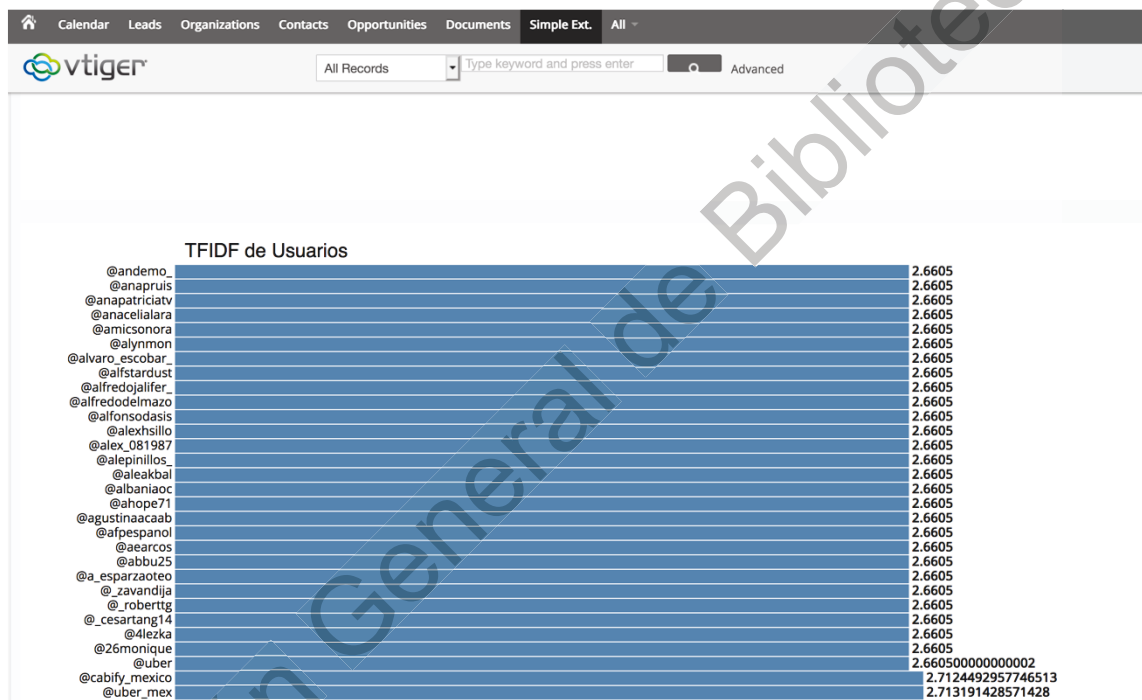


Figura. 8-11 Gráfica de Usuarios destacados.

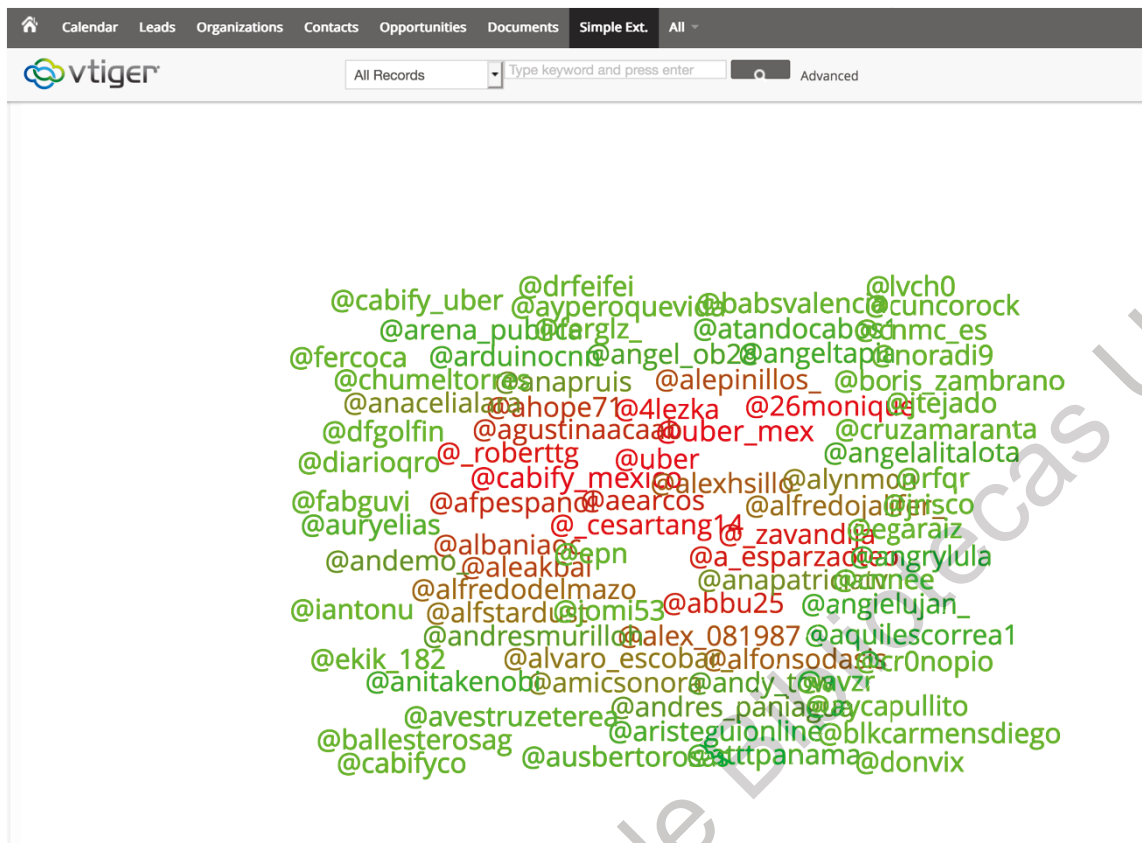


Figura. 8-12 Nube de Usuarios mencionados.

Lista de términos más altos

Término	TFIDF
angry	6.9425
star	6.9425
point_down	5.991958064516134
drooling_face	5.8878
clap	5.5835
guatemala	5.5835
triumph	5.08928
wink	4.8643
beep	4.5047
callar	4.5047
campera	4.5047
chatas	4.5047
cnet	4.5047
coches	4.5047
confirno	4.5047
demandar	4.5047
emisora	4.5047
gurita	4.5047

hablamos	4.5047
heartpulse	4.5047
holder	4.5047
ignorante	4.5047
manden	4.5047
meno	4.5047
nice	4.5047
nulo	4.5047
numeros	4.5047
pelotudo	4.5047
persevere	4.5047
poblaci	4.5047
purple_heart	4.5047
quej	4.5047
spades	4.5047
stars	4.5047
trata	4.5047
vyf3sqxcue	4.5047
warning	4.5047
enza	4.3510166666666666
verg	4.3510166666666666
sleeping	4.122
star2	4.0878333333333333
viaja	4.079115384615384
rage	3.9479749999999996
anda	3.653530769230769
busca	3.5826
conoce	3.5826
fire	3.5826
hombre	3.5826
new_moon_with_face	3.5826

Tabla. 8-5 Términos con TFIDF alto

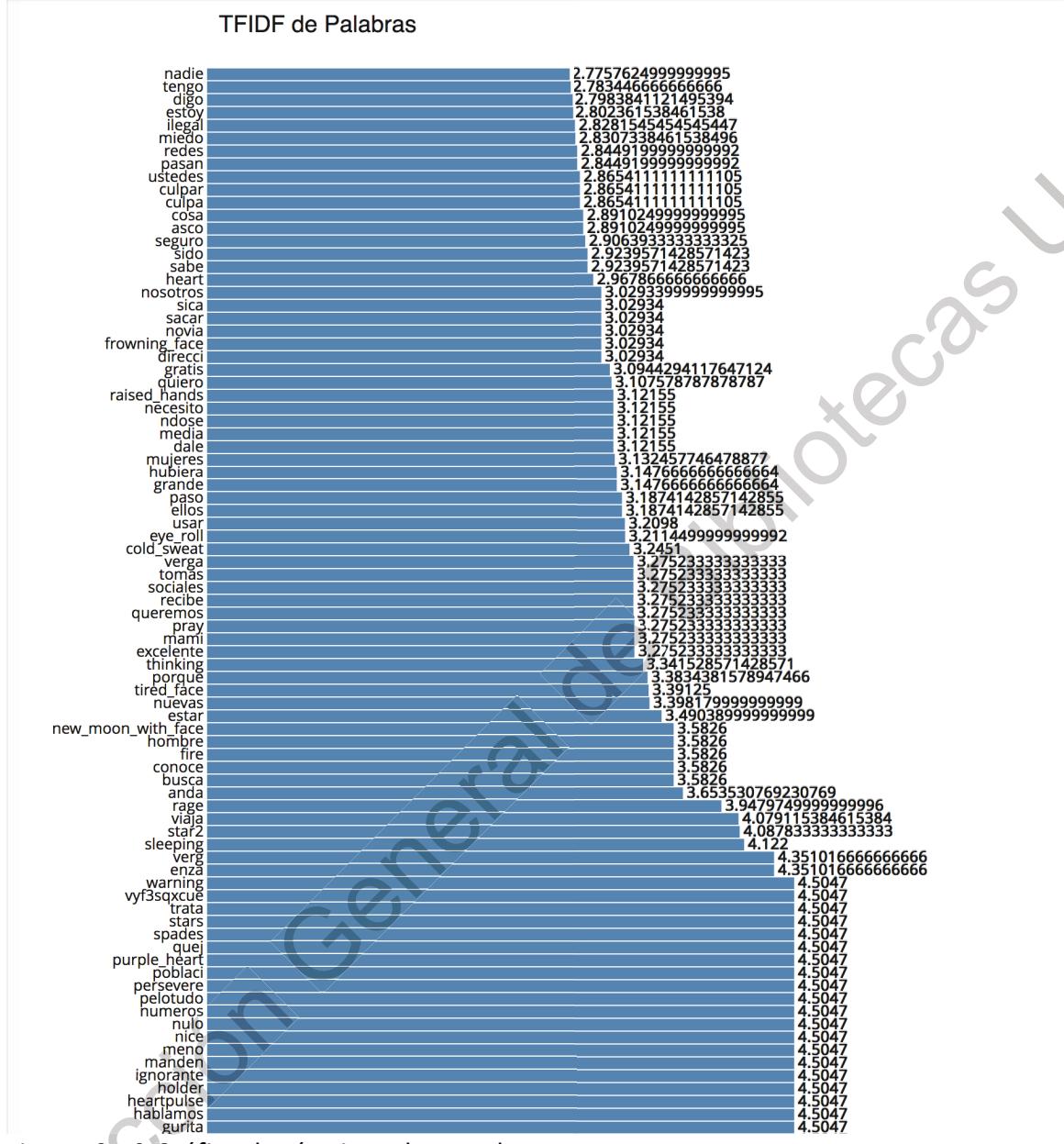


Figura. 8-13 Gráfica de términos destacados.

9 Conclusiones

El desarrollo de un módulo social es de gran ayuda para conocer si un segmento de las redes sociales está hablando de un servicio. El análisis de la frecuencia de términos acompañado de la interpretación de un analista, permite conocer el contexto en el que se encuentra el servicio, sin embargo, este tipo de análisis sólo es recomendado para productos y servicios de consumo popular. También, el desarrollo de un módulo de CRM debe llevarse a cabo si la organización lo requiere y cuenta con los recursos humanos e infraestructura tecnológica para implementarlo.

En relación a las técnicas de análisis de texto, hay que revisar las técnicas de detección de spam, esto puede ser de gran ayuda en la fase de extracción de tweets ya que así podemos enfocar el análisis de texto a tweets que realmente pueden ser relevantes. Existen analizadores gramaticales que soportan el lenguaje español, sin embargo, no tienen compatibilidad con la jerga informal que se utiliza comúnmente.

Finalmente, este proyecto puede ayudar a gente que desea empezar en el análisis de texto, así como a empresas dedicadas al análisis de información.

10 BIBLIOGRAFÍA

Aguirre C. E. (2014). Metodología de desarrollo para sistemas de información basados en web. México: Universidad Autónoma de Querétaro.

Alexa. (2015). Top sites in Mexico. Alexa. Recuperado de <http://www.alexa.com/topsites/countries;0/MX>

Alan. R. H. Salvatore T. M. Jinsoo P. Sudha R. (2004) Design Science in information Systems Research. MIS Quarterly Volumen 28. No 1.

Altwegri. R. (2015). Hybrid CRM Deployment Model. 2015 International Conference on Cloud Computing.

AMIPCI. (2016). 12º Estudio sobre los Hábitos de los Usuarios de Internet en México 2016. Recuperado de https://amipci.org.mx/images/Estudio_Habitosdel_Usuario_2016.pdf

AMIPCI (2015). 11º estudio sobre hábitos de los usuarios de internet en México 2015.

APIC. (2015). APICS OMBOK Framework 310. Recuperado de <http://www.apics.org/industry-content-research/publications/ombok/apics-ombok-framework-table-of-contents/apics-ombok-framework-3.10>

Behboudi. M. (2011). Using social network paradigm for developing a conceptual framework in CRM. Australian Journal of Business and Management Research Volumen 1. No 4. Julio 2011.

Bose. R. (2002). Customer relationship management: key components for IT success. Industrial Managment & Data Systems. Emerald

Billy. L. (2015). A Brief History of Customer Relationship Management. Recuperado de <http://www.business2community.com/business-innovation/brief-history-customer-relationship-management-01245936#4cXSPcs6jglObUxb.97>

Brodia. R. (2010). Nestle's Facebook page: How a Company Can Really Screw Up Social Media. Recuperado de <http://www.cbsnews.com/news/nestles-facebook-page-how-a-company-can-really-screw-up-social-media/>

Candón. M. (2013). Movimientos por la democratización de la comunicación: los casos del 15M y #YOSOY132. Razón y palabra. Departamento de Comunicación Audiovisual, Publicidad y Literatura. Universidad de Sevilla.

CRM Switch Staff. (2013). A Brief History of Customer Relationship Management. Recuperado de <https://www.crmswitch.com/crm-industry/crm-industry-history/>

Corcoran. S. (2009). Defining Earned, Owned and Paid media. Recuperado de http://blogs.forrester.com/interactive_marketing/2009/12/defining-earned-owned-and-paid-media.html

Cortés. L. (2017). A cinco años del #YoSoy132. Recuperado de <http://www.eluniversal.com.mx/articulo/metropoli/cdmx/2017/05/11/cinco-anos-del-yosoy132>

Crockford. D. (consultado el 29 de noviembre del 2018). Introducing JSON. Recuperado de <https://www.json.org>

Danah M. B. Nicole B. E. (2008). Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication. Volumen(13) 210-230. Recuperado de <http://www.home-electronics.ch/HomeElectronic/Files/Web/Dossier/VL%203%20SNS%20History.pdf>

DiablosRojosMX. (9 de noviembre 2016). @HawkersMX Nuestra #PasiónPorMÉXICO. Recuperado de https://twitter.com/DiablosRojosMX/status/796511534180696065?ref_src=twsrc%5Etfw

Dragomir R. Introduction to Natural Language Processing. Coursera. Extraído de <https://www.coursera.org/learn/natural-language-processing/lecture/oX0lw/01-01-introduction-8-38>

Facebook. (2010). Automated Data Collection Terms. Recuperado de https://www.facebook.com/apps/site_scraping_tos_terms.php

Fox. E. (2010). Nestlé hit by Facebook “anti-social” media surge. Recuperado de <https://www.theguardian.com/sustainable-business/nestle-facebook>

Florido. M. (2016). Qué es LinkedIn, Cómo funciona y para que sirve. Consigue empleo con LinkedIn. Recuperado de <http://www.marketingandweb.es/marketing/linkedin-que-es-como-funciona/>

Freeman.C. (1974). *La teoría económica de la innovación industrial*, Madrid: Alianza Universidad.

Galeano. S. (2016). HAWKERS MÉXICO SUFRE UNA GRAVE CRISIS DE REPUTACIÓN TRAS UN TUIT SOBRE TRUMP. Recuperado de <http://marketing4ecommerce.net/polemica-de-hawkers-en-mexico/>

Gomez. D. (2014). Google+: Qué es, cómo funciona y para qué sirve. Recuperado de <http://bienpensado.com/google-plus-que-es-y-para-que-sirve/>

Greenberg. P. (2010). CRM at the speed of light: social CRM strategies, tolos, and techniques for engaging your customers. Cuarta Edición. MacGraw-Hill.

Greenpeace. (2010). Sweet success for Kit Kat campaign: you asked, Nestlé has answered. Recuperado de <http://www.greenpeace.org/international/en/news/features/Sweet-success-for-Kit-Kat-campaign/>

Hanna. R. (2011). We're all connected: The power of the social media ecosystem. U.S.A. Kelly School of Business, Indiana University.

Hawkers. Querido Sergio. Recuperado de <https://www.hawkersmexico.com/pages/querido-sergio>

Marti. A. Hearst (1994). Multi-Paragraph Segmentation of Expository Text. University of California. Berkley.

HTML - HyperText Markup Language, Webopedia. Recuperado de <http://www.webopedia.com/TERM/H/HTML.html>.

ITU (International Telecommunication Union). (2010). The rise of social networking. Recuperado de <http://www.itu.int/net/itunews/issues/2010/06/35.aspx>

José A del Moral. (2014). Recuperado de <http://blogs.alianza.com/redessociales/2014/05/29/para-que-sirve-cada-red-social/>

Kaplan. A. Haenlein. M. (2009). Users of the world, unite! The challenges and opportunities of Social Media. U.S.A. Kelly School of Business, Indiana University.

Kellen V. (2007). CRM Measurement Frameworks. U.S.A. Blue Wolf.

Kotler. P. Armstrong. G. (2014). Principles of Marketing. Edición 15. Pearson.

Kumar. D. (2011). A Social Network Perspective by CRM. IEEE.

LinkedIn. (2015). A brief history of LinkedIn. Recuperado de <https://ourstory.linkedin.com>

McDonald. D. (2004). Alcohol and Other Drug Peer Education in Schools: A review for the ACT Alcohol, Tobacco and Other Drug Strategy Implementation and Evaluation Group. Social Research and Evaluation. Canberra recuperado de <http://mypeer.org.au/planning/what-are-peer-based-programs/what-is-a-peer/>

Magee. K. (2010). Taking Action – Greenpeace vs Nestle. Recuperado de <http://www.prweek.com/article/992445/taking-action---greenpeace-vs-nestle>

Marjorie. A. (2004). PDMA Foundation New Product Development Report of Initial Findings : Summary of Responses from 2004 CPAS.

Mitkov. R. (2010). Discourse Processing. The Handbook of Computational Linguistics and natural language processing (primera edición, pp. 599-629). Clark. A. Et al. Wiley-Blackwell.

Morgan. J. (2010). The Evolution of the Social CRM Process. Chess Media Group. Recuperado de <https://thefutureorganization.com/evolution-social-crm-process/>

Nivre, J. (2010). Statistical parsing. En N. Indurkha y F. J. Damerau (Eds.), *Handbook of natural language processing* (2nd ed., pp. 237-266). Boca Baton, FL: Chapman & Hall/CRC, Taylor & Francis.

Nimble (19 febrero 2017). Case Study: SkylineBoston. Recuperado de <http://www.nimble.com/blog/category/case-studies/page/3/>

Ngai. E. (2005). Customer relationship management research (1992-2002). Marketing Intelligence & Planning Volumen 23. No 6.

Oxford. (2015). Recuperado de <http://www.oxforddictionaries.com/definition/english/search-engine>

Owyang. J. (2011). Social Media On Rise: Be Prepared by Climbing the Social Business Hierarchy of Needs. Recuperado de <http://www.web-strategist.com/blog/2011/08/>

Owyang. J. Wang. R. (2010). Social CRM: The New Rules of Relationship Management. Altimeter. Recuperado de <http://www.web-strategist.com/blog/2010/03/05/altimeter-report-the-18-use-cases-of-social-crm-the-new-rules-of-relationship-management/>

Padró, L. (2016). Demonstration. *FreeLing 4.0. An open-source suite of language analyzers*. Barcelona: TALP - Tecnologies i Aplicacions del Llenguatge i de la Parla, Universitat Politècnica de Catalunya. Consultado en <http://nlp.lsi.upc.edu/freeling/demo/demo.php>

Payne, A. (2005). Handbook of CRM Achieving Excellence in Customer Management Butterworth-Heinemann publications, Great Britain.

Pijamasurf. (1/3/2014). La estética de la desobediencia civil: acrobacias en #PosMeSalto. Recuperado de <http://m.pijamasurf.com/2014/01/la-estetica-de-la-desobediencia-civil-acrobacias-en-posmesalto/>

Quane. J. Rankin. B. (2006). Does it pay to participate? Neighborhood-based organizations and the social development of urban adolescents. *Children and Youth Services Review*. Volumen(28). Número(10). 1229-1250.

Rubín. R. (2016). Qué es Facebook, cómo funciona y qué te puede aportar esta red social. Recuperado de <https://www.ciudadano2cero.com/facebook-que-es-como-funciona/>

Rubín. R. (2016). Qué es Twitter, cómo funciona y qué te puede aportar esta red social. Recuperado de <https://www.ciudadano2cero.com/twitter-que-es-como-funciona/>

Rubira. F. (2013). (consultado el 10 de enero 2017). ¿Qué es Instagram y para qué sirve? Recuperado de http://www.elconfidencialdigital.com/opinion/tribuna_libre/Instagram-sirve_0_2076992284.html

Sergio Pérez. (9 de noviembre 2016). Que mal comentario. Recuperado de https://twitter.com/SChecoPerez/status/796476888390701056?ref_src=twsrc%5Etfw

Schultz. D. (2007). Snackbyte: A view of the ecosystem. Recuperado de <http://www.deborahschultz.com/deblog/2007/11/snackbyte-a-vie.html>

Shao. Yi. (2010). CRM measurement indicator system of logistics enterprises for data-mining. 2010 International Conference on Logistics Systems and Intelligent Management.

Sopitas. (2014). Aquí están los movimientos sociales más importantes del 2014. Recuperado de <http://www.sopitas.com/420285-aqui-estan-los-movimientos-sociales-mas-importantes-del-2014/>

Singh, S. (2010). *Social Media Marketing for Dummies*. Canada: John Wiley & Sons.

Stanford (2008). A first take at building an inverted index. Cambridge University Press. Recuperado de <https://nlp.stanford.edu/IR-book/html/htmledition/a-first-take-at-building-an-inverted-index-1.html>

T.L.Children. A.R.Rao. (1992). The influence of Familial and Peer-based Reference Groups on Consumer Decisions. *Journal of Consumer Research Inc*.

Twitter. (2015). The Twitter Glossary. Recuperado de <https://support.twitter.com/articles/166337-the-twitter-glossary>.

Twitter. (2015). Twitter Milestones. Recuperado de <https://about.twitter.com/company/press/milestones>

Twitter. (2017). Twitter Limits. Recuperado de <https://support.twitter.com/articles/15364>

Twitter. (2017). REST APIs. Recuperado de <https://dev.twitter.com/rest/public>

Twitter. (2017). Streaming APIs. Recuperado de <https://dev.twitter.com/streaming/overview>

Ueno. K. (2005). The effects of friendship networks on adolescent depressive symptoms. *Social Science Research*. Volumen(34). no. 3, 484-510.

Wikilengua. Abigüedad. Recuperado de <http://www.wikilengua.org/index.php/Ambig%C3%BCedad>

Wu. B., Ye. Q., Yang. S., & Wang. B. (2009). Group CRM: a new telecom CRM framework from social network perspective. *Proceeding of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management*. Recuperado de https://doi.org/10.1007/978-3-642-00528-2_7

Xu. Y. (2002). Adopting customer relationship management technology. *Industrial Management & Data Systems* Volumen 102. No 8.

Yamamoto Y. (2016). Twitter4J. Recuperado de <http://twitter4j.org/en/index.html>

Yuanlei, Z. (2004). A Comparison of Search Engines For Finding Resources. Recuperado de <http://www.yuanlei.com/studies/articles/is567-searchengine/page2.htm>

Youtube. (consultado el 9 de enero del 2017). Acerca de YouTube. Recuperado de <https://www.youtube.com/yt/about/es-419/>

Yeung, K. (2013). LinkedIn is 10 years old today: here's the story of how it changed the way we work. Recuperado de <http://thenextweb.com/insider/2013/05/05/linkedin-10-years-social-network/>