La imagen pública de la química en Twitter

Manuel Guerris Larruy

http://hdl.handle.net/10803/673271

TESIS DOCTORAL

Título	La imagen pública de la química en Twitter
Realizada por	Manuel Guerris Larruy
en el Centro	IQS School of Management
y en el Departamento	Métodos Cuantitativos
Dirigida por	Dr. Jordi Cuadros Margarit y Dr. Lucinio González Sabaté

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

D./Dña. MANUEL GUERRIS LARRUY con DNI 43692609S estudiante del Programa académico de doctorado en Competitividad Empresarial y Territorial, Innovación y Sostenibilidad (CETIS) de IQS School of Management

DECLARO

1. Que soy autor/a del Trabajo que presento para su exposición y defensa titulado: **La imagen pública de la química en Twitter**
2. Que dicho trabajo es original, no ha sido plagiado ni presentado previamente, y he contribuido directamente al contenido intelectual y a la génesis y análisis de los datos que certifico son reales y no han sido falsificados.
3. Que las fuentes utilizadas para su realización y aportes intelectuales de otros autores han sido referenciados en el mismo y no se atenta contra derechos de terceros.

En caso de detectarse fraude, plagio o falsificación asumo las consecuencias y sanciones que se deriven de acuerdo con la normativa vigente de IQS y la Universidad Ramon Llull.

Y para que así conste firmo el presente documento en

Barcelona, a 11 de julio de 2021

Resumen

La imagen pública de la química se ha estudiado mediante el análisis de documentos y encuestas. Twitter es una red social de alcance mundial que, a diferencia de las anteriores fuentes, se basa en opiniones breves y espontáneas. El objetivo de este trabajo es identificar la percepción pública de la química en esta red, es decir, lo que se explica, qué sentimientos se perciben, así como los usuarios más influyentes y sus relaciones.

Para conseguir los objetivos anteriores 256 833 tweets recogidos entre el 1 de enero de 2015 y el 30 de junio de 2015 que contenían las palabras "chemistry", "chemical" o "chem", se depuraron hasta conseguir 50 725 *tweets* con información textual en inglés. Los *tweets* aceptados se agruparon automáticamente utilizando el método *spherical k-means*. Los grupos resultantes fueron clasificados según seis temáticas por 18 expertos en química. Las predominantes fueron el entorno educativo, relacionada con las actividades y tareas de cursos de química, y la actividad humana, referida a hechos y noticias sobre la industria química. La temática conocimiento científico, relativa a la comunicación de los conocimientos de química, solo representó un pequeño porcentaje de los *tweets*.

Se clasificaron los *tweets* de las temáticas más relevantes en base a sus valores de sentimiento y se obtuvieron percepciones más positivas que negativas. No obstante, el análisis de los *wordclouds* de unigramas y bigramas reveló una presencia notable de términos relacionados con la quimiofobia en la temática actividad humana, tanto en los tweets clasificados positivos como en los negativos así como elementos específicos de los cursos de química percibidos negativamente en la temática entorno educativo.

El análisis de los usuarios y sus relaciones entre ellos arrojó que aquellos más relevantes no se corresponden con los presumiblemente más influyentes, empresas y sociedades científicas, en el mundo de la química. Las empresas y sociedades científicas utilizan Twitter como canal de noticias y publicidad sin una finalidad educativa aparente sobre las virtudes y beneficios de la química.

Para finalizar, destacar que, durante este estudio, se ha detectado la existencia de una imagen pública de la química en Twitter centrada en las temáticas de entorno académico y actividad humana, con cierta presencia de quimiofobia pero con el predominio de sentimientos positivos.

Palabras clave

Química; imagen pública; Twitter; análisis de sentimientos; análisis de *influencers*; *clustering*

Resum

La imatge pública de la química s'ha estudiat mitjançant l'anàlisi de documents i enquestes. Twitter és una xarxa social d'abast mundial que, a diferència de les fonts anteriors, es basa en opinions breus i espontànies. L'objectiu d'aquest treball és identificar la percepció pública de la química dins aquesta xarxa, és a dir, el que explica, quins sentiments es perceben així com els usuaris més influents i les seves relacions.

Per aconseguir els objetius anteriors, 256 833 *tweets* recollits entre l'1 de gener de 2015 i el 30 de juny de 2015 que contenien les paraules "chemistry", "chemical" o "chem", es van depurar fins aconseguir 50 725 *tweets* amb informació textual en anglès. Els *tweets* acceptats es van agrupar automàticament utilitzant el mètode *spherical k-means*. Els grups resultants van ser classificats segons sis temàtiques per 18 experts en química. Les predominants van ser l'entorn educatiu, relacionada amb les activitats i tasques de cursos de química, i l'activitat humana, referida a fets i notícies sobre la indústria química. La temàtica coneixement científic, relativa a la comunicació dels coneixements de química, només va representar un petit percentatge dels *tweets*.

Es van classificar els *tweets* de les temàtiques més rellevants en base als seus valors de sentiment i es van obtenir percepcions més positives que negatives. No obstant, l'anàlisi dels *wordclouds* de unigrames i bigrames va revelar una presència notable de termes relacionats amb la quimiofobia en la temàtica activitat humana, tant en els *tweets* classificats positius com en els negatius, així com elements específics dels cursos de química percebuts negativament en la temàtica entorn educatiu.

L'anàlisi dels usuaris i les seves relacions va mostrar que aquells més rellevants no es corresponen amb els presumiblement més influents, empreses i societats científiques, en el món de la química. Les empreses i societats científiques empren Twitter com a canal de notícies i publicitat sense un fi educatiu aparent sobre les virtuts i beneficis de la química.

Per a finalitzar, destacar que, durant aquest estudi, s'ha detectat l'existència d'una imatge pública de la química en Twitter centrada en les temàtiques d'entorn acadèmic i activitat humana, amb una certa presència de quimiofòbia però amb el predomini de sentiments positius.

Paraules clau

Química; imatge pública; Twitter; anàlisis de sentiments; anàlisis de *influencers*; *clustering*

Abstract

The public image of chemistry has been studied through the analysis of documents and surveys. Twitter is a worldwide social network that, unlike the previous sources, is based on brief and spontaneous opinions. The aim of this work is to identify the public perception of chemistry in this network, i.e., what is explained, what feelings are perceived, as well as the most influential users and their relationships.

To achieve the above objectives 256 833 tweets collected between January 1, 2015 and June 30, 2015 containing the words "chemistry", "chemical" or "chem" were purified to 50 725 tweets with textual information in English. Accepted tweets were automatically clustered using the spherical k-means method. The resulting clusters were categorised according to six topics by 18 chemistry experts. The prevailing topics were the Learning Environment topic, related to activities and tasks in chemistry courses, and the Human Activity topic, referring to facts and news about the chemical industry. The Scientific Knowledge topic, concerning communication of chemistry knowledge, only accounted for a small percentage of the tweets.

We classified the tweets of most relevant topics based on their sentiment values and obtained more positive than negative perceptions. Nevertheless, the analysis of the unigrams and bigrams word clouds revealed a notable presence of chemophobia-related terms in the Human Activity topic, both in positive and negative classified tweets as well as negatively perceived chemistry course-specific elements in the Learning Environment topic.

The analysis of the users and their relationships with other users showed that the most relevant users do not correspond to the presumably most influential users, companies and scientific societies, in the chemistry world. Companies and scientific societies use Twitter as a news and advertising channel without an apparent educational purpose about the virtues and benefits of chemistry.

Finally, it should be noted that, during this study, the existence of a public image of chemistry on Twitter centered on the topics of Learning Environment and Human Activity has been detected, with a certain presence of chemophobia but with a predominance of positive feelings.

Keywords

Chemistry; public image; Twitter; sentiment analysis; influencer’s analysis; clustering

Agradecimientos

Quiero agradecer el trabajo, tiempo, dedicación y guía durante todo el peridodo del programa de doctorado realizado por los doctores Jordi Cuadros y Lucinio González que me han transmitido conocimientos, experiencia, valores y competencias que un investigador debe poseer. A mi mujer, Yolanda y a mis dos hijos Biel y Aleix, por el tiempo que no les he podido dedicar y la paciencia que han tenido, proporcionándome periodos de tiempo de calidad en los que he podido desarrollar esta tesis. A mi madre, por los valores que me ha inculcado y que sin ellos no estaría realizando este programa de doctorado. A los revisores y editores de la publicación realizada, parte de esta tesis, sin los cuales ciertas reflexiones no hubiesen sido planteadas y desarrolladas. A los 18 expertos en química de IQS, ya que sin ellos no se podría haber clasificado algunos de los resultados obtenidos. Al grupo de investigación ASISTEMBE y a los miembros que participaron en diferentes pruebas piloto de la metodología. Al departamento TICs por las facilidades proporcionadas para la utilización del servidor ABACO y su gestión, ya que sin este no se podrían haber realizado los cálculos de esta tesis. A la coordinadora del programa de doctorado de IQS Marianna Bosch por su confianza en este tesis, habida cuenta de no ser desarrollada en un programa a tiempo completo. A las diferentes direcciones de IQS por su confianza, y en particular al Dr. Enric Julià exdirector de IQS y a Helena Borbón exdirectora de Programas Ejecutivos de IQS, que me animaron y motivaron a realizar el programa de doctorado. Finalmente, no quiero dejar de agradecer a todos aquellos que han compartido todo este tiempo y confiado de diferentes formas en mi para culminar este programa de doctorado. Muchas gracias a todos.

Índice

1 Introducción

La química se define como la ciencia que estudia la estructura, propiedades y transformaciones de los cuerpos a partir de su composición (RAE, 2020) y sus aplicaciones abarcan múltiples y diversos ámbitos de la vida cotidiana. La química moderna se inicia como ciencia data del siglo XVII y alcanza el rango de ciencia en el siglo XVIII, con múltiples descubrimientos, invenciones y desarrollos a lo largo de los siguientes siglos (ver Figura 1.1).

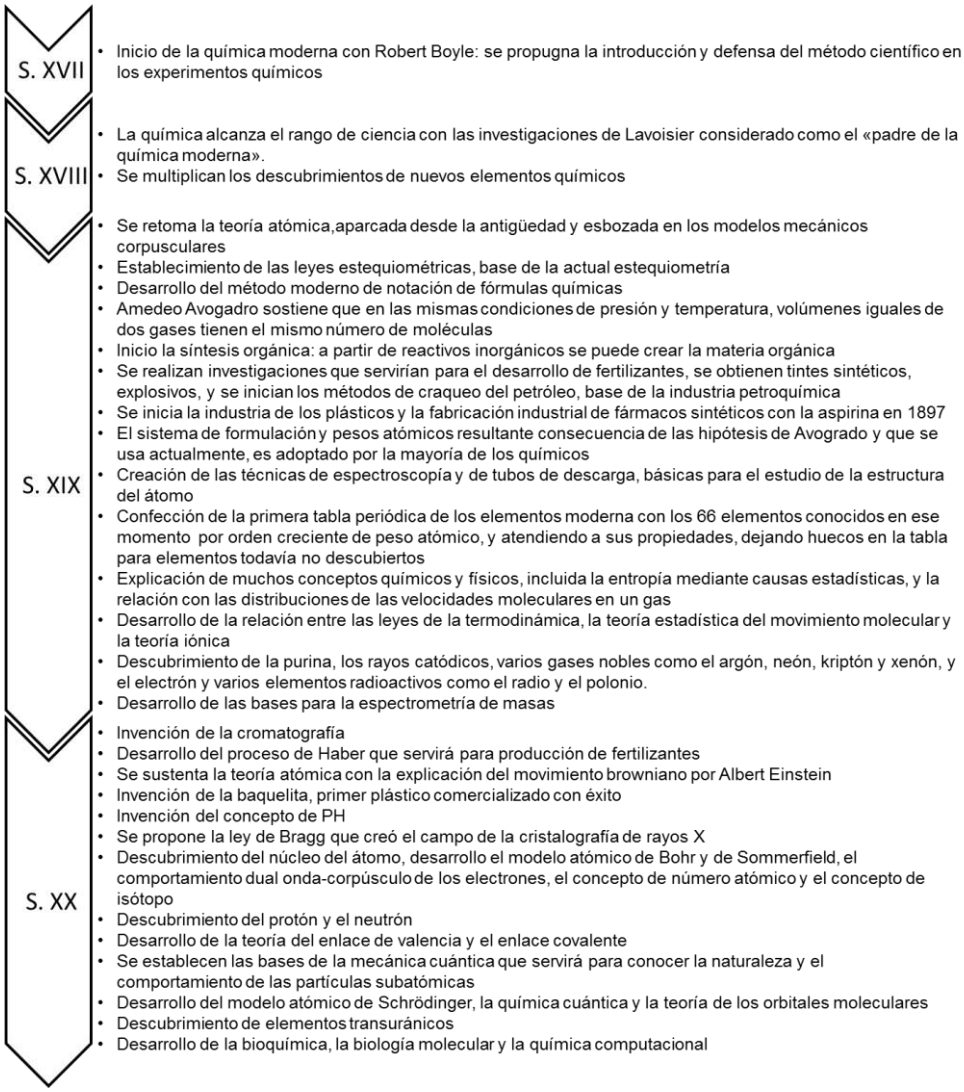


Figura 1.1 Ejemplos de descubrimientos, invenciones y desarrollos en la química (elaboración propia)

Una vez presentada la química, el resto de la introducción se centra en describir su importancia como sector en la economía mundial, en su imagen pública y la relevancia de su estudio por los efectos sobre los recursos para desarrollarla, en las redes sociales *on-line* como nuevos medios de comunicación en los que poder estudiar la

imagen pública de la química y en la organización de esta memoria junto con los procesos desarrollados.

El sector de la química representó el equivalente 7% del PIB mundial (Economics, 2019) con 5 700 000 millones de dólares de valor añadido[1] al PIB con 120 millones de puestos de trabajo, incluyendo el impacto económico directo de sus actividades, el indirecto de su cadena de suministro y el inducido del gasto de los trabajadores directos e indirectos sobre la economía. La distribución geográfica de su valor añadido, liderada por la zona Asia-Pacífico, puede consultarse en la Figura 1.2.

Figura 1.2 Distribución geográfica del valor añadido al PIB y del número de puestos de trabajo en 2017. Extraído de (Economics, 2019)

De forma directa el sector aportó 1 100 000 millones de dólares de valor añadido al PIB mundial con 4 100 000 millones de dólares en ingresos y 15 millones de puestos de trabajo como empleo directo. Este sector es el quinto dentro del sector de producción atendiendo a su contribución anual al PIB mundial industrial (Figura 1.3) e invirtió 51 000 millones de dólares en I+D. El 42% de sus ingresos provienen de ventas de productos dentro del mismo sector (Figura 1.4) siendo el restante proveniente de otros sectores. En este sector, las empresas líderes son grandes tanto en volumen de ingresos (Figura 1.5) como en número de empleados (Figura 1.6).

[1] Valor añadido: calculado como los ingresos menos costes relacionados con la producción de estos ingresos

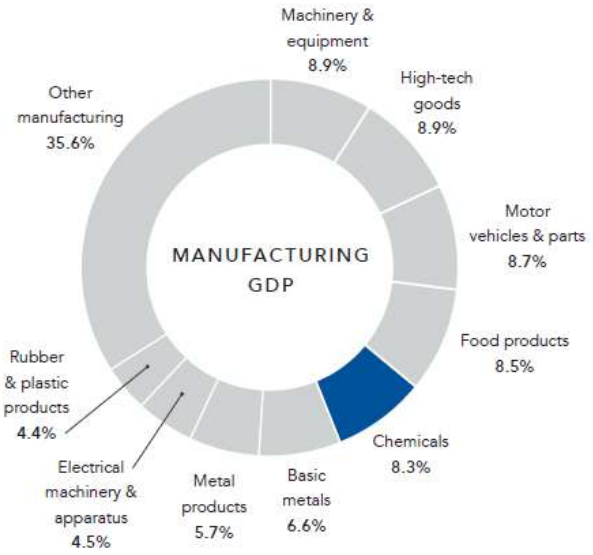


Figura 1.3 Distribución de sectores de producción según el valor añadido al PIB industrial en 2017. Extraído de (Economics, 2019)

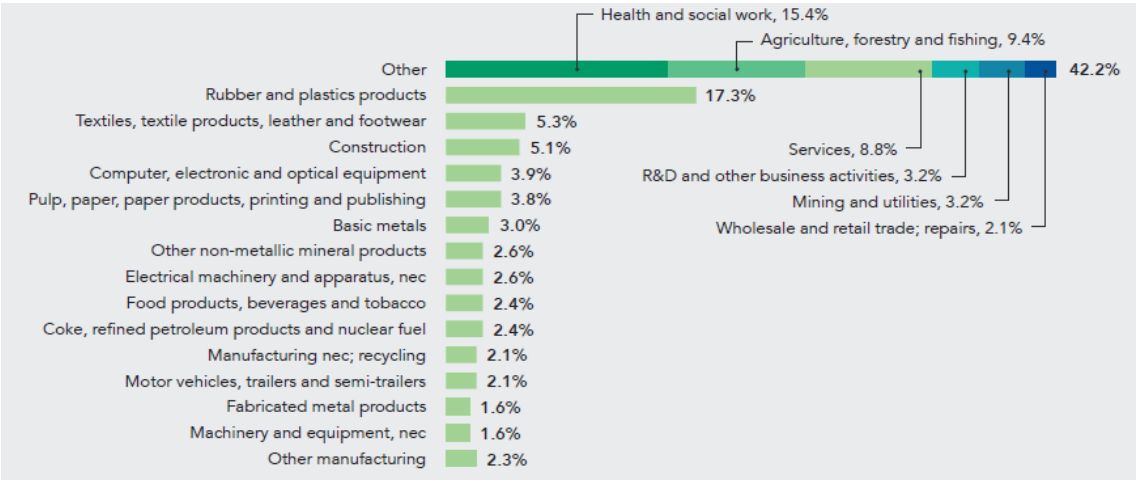


Figura 1.4 Distribución las ventas del sector químico según el sector del cliente (excluidos los fabricantes químicos). Reproducida de (Economics, 2019)

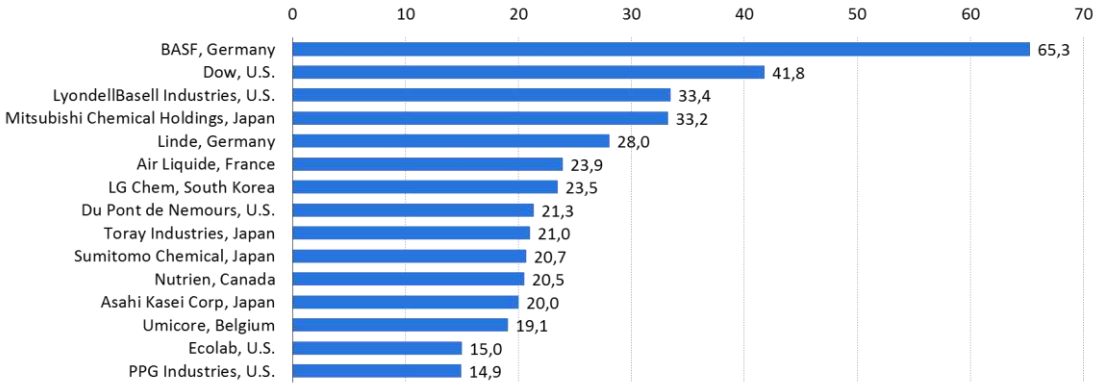


Figura 1.5 Empresas químicas líderes a nivel mundial en función de sus ingresos medidos en billones de dólares (miles de millones) en 2020. Reproducida de (Statista, 2020a)

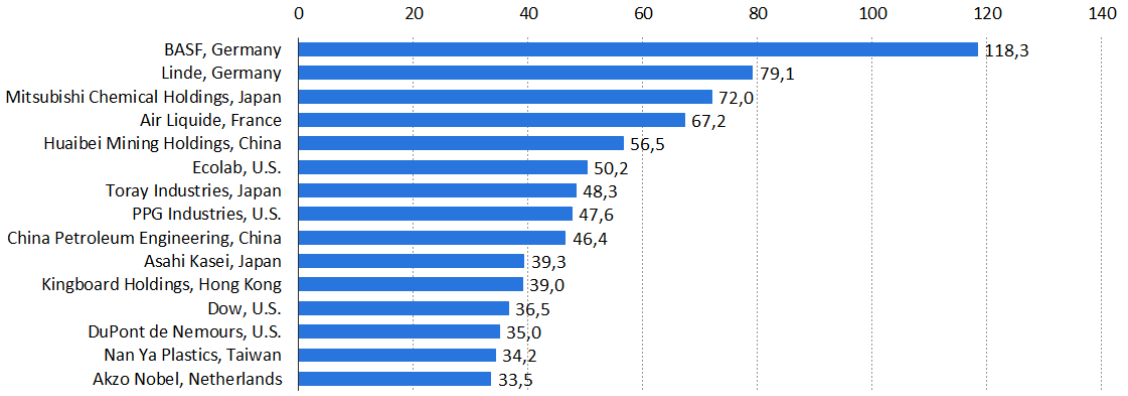


Figura 1.6 Empresas químicas líderes a nivel mundial en función del número de trabajadores directos en 2020 medidos en miles. Reproducida de (Statista, 2020a)

El consumo de productos químicos desde el 2004 ha seguido una tendencia en general creciente (Figura 1.7) y las estimaciones para los próximos años son de crecimientos de la producción en las diferentes regiones mundiales (Figura 1.8) aunque con cierta disminución de este ritmo de crecimiento.

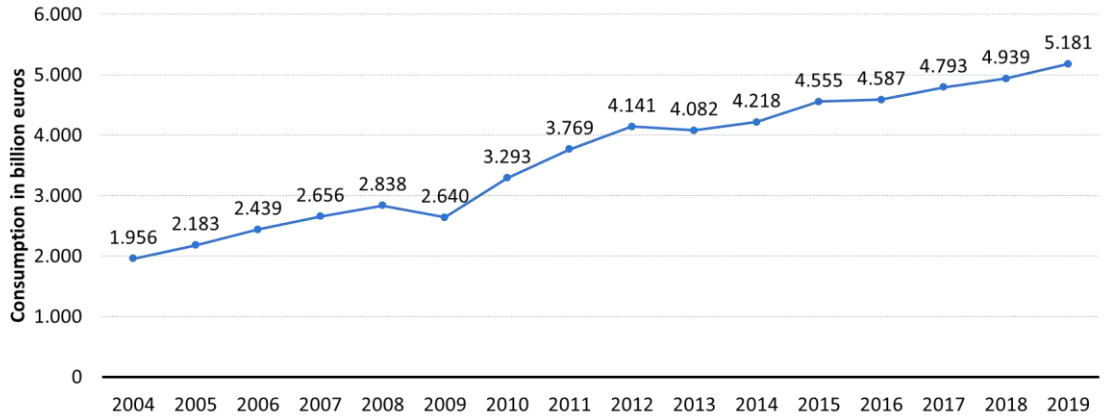


Figura 1.7 Evolución del consumo de productos químicos a nivel mundial expresado en billones de dólares (miles de millones). Reproducida de (Statista, 2020a)

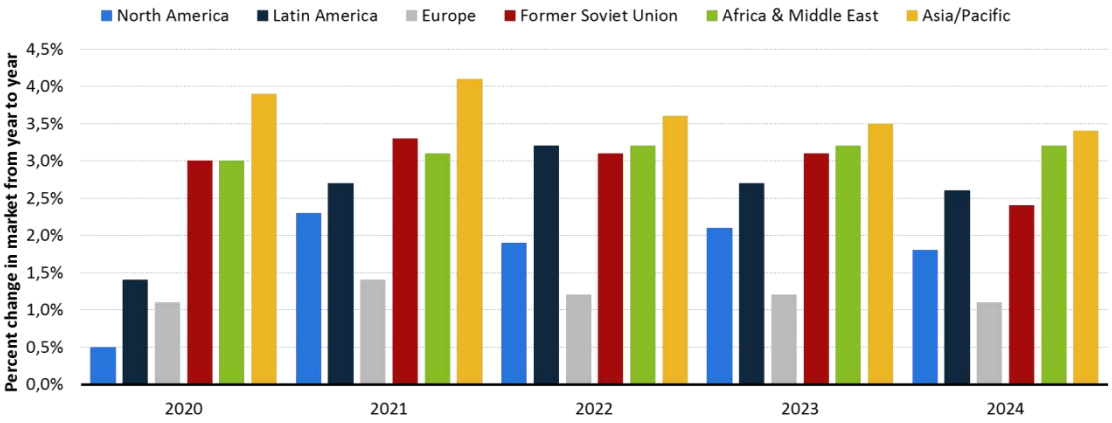


Figura 1.8 Crecimiento estimado de la producción de la industria química a nivel mundial según regiones. Reproducida de (Statista, 2020a)

Una definición de la imagen pública de la química es como la sociedad piensa y siente con respecto acerca de la química, los profesionales químicos y los productos químicos (The Royal Society of Chemistry y TNS BMRB, 2015). Debido al impacto de la química en los cambios de la sociedad, ser considerada la ciencia que en mayor medida ha contribuido a ofrecer respuestas a las necesidades del ser humano (*Declaración de la Química*, 2002) y su importancia en la economía mundial y en el mercado de trabajo, el análisis y comprensión de la imagen pública de la química es relevante ya que su comunicación ayuda a la sociedad a adquirir el conocimiento para participar activamente en el debate científico (European Commission, 2015) y tiene un papel significante en despertar vocaciones científicas (Stekolschik *et al.*, 2010) así como en fomentarlas (Hayden *et al.*, 2011).

En cambio, la falta de vocaciones puede causar insuficientes graduados para poder dar respuesta a la demanda creciente en la Unión Europea y una escasez de profesionales científicos, tecnológicos, de ingeniería y matemáticas (Cedefop, 2016). Adicionalmente, el rol de la química, parece no ser comprendido por los legisladores, financiadores y la misma comunidad química (Palermo, 2018).

Por tanto, el conocimiento de la imagen pública de la química y su comprensión a partir de los contenidos y sentimientos percibidos por la sociedad es de relevancia para los diferentes grupos de influencia de la química. Es adecuado entonces, entender qué se comunica y cómo se percibe para adecuar las políticas en relación a la química y reducir la escasez de profesionales.

Esta imagen se percibe en diferentes campos (Mora Penagos, 1997)., como por ejemplo, el científico, en su estudio para conseguir mejoras a través de la

investigación, el desarrollo y la innovación, el formativo a través de su enseñanza, aplicación y práctica en centros educativos, el industrial en sus aplicaciones en la vida diaria o en la comunicativo a nivel científico, publicitario, literario o cinematográfico.

Asimismo, la imagen ha sido históricamente negativa (Hartings y Fahy, 2011) debido, entre varios motivos, a la creación de productos no deseables con grandes perjuicios para la sociedad, a que ha sido presentada como una ciencia compleja, a que se desarrolla en laboratorios que son entornos desconocidos por el público en general con aparatos tecnológicos difíciles de entender. En cambio, de forma implícita, nos aprovechamos y somos lo que somos como sociedad gracias a ella, como por ejemplo mediante las aplicaciones en el mundo de la salud o simplemente mediante los productos que vestimos y utilizamos diariamente. Aunque ha habido intentos para mejorar esta imagen, la contradicción entre su negatividad histórica y su elevada utilidad social, nos lleva a plantearnos el interés por conocerla actualmente.

En las diversas sociedades y culturas, los seres humanos nos relacionamos y organizamos en redes sociales. Una red social es un conjunto de personas, grupos, comunidades u organizaciones vinculados a través de relaciones sociales. Las redes a distancia en las que no existe un contacto físico y más concretamente las *on-line* o mediante Internet, han proliferado en las últimas décadas debido a la naturaleza social del ser humano y al desarrollo de la red de Internet, las comunicaciones y los dispositivos móviles. Estas redes son de interés porque personas, empresas, asociaciones u organizaciones públicas o privadas las aprovechan como canales de comunicación para transmitir sus mensajes, de forma que influyen en la opinión de los usuarios como por ejemplo en el caso de la política o de contenidos sociales (Statista, 2019c).

Facebook, YouTube, WhatsApp, Twitter o Instagram son ejemplos de redes sociales *on-line* que están en las primeras posiciones de rankings según el número de usuarios activos (Statista, 2019b). Twitter, una red que permite enviar mensajes cortos entre usuarios de 280 caracteres como máximo y con 326 millones de usuarios activos a Enero de 2019 (Statista, 2019b) es una red relevante a nivel mundial, dentro de las cinco primeras más populares medido por el grado de utilización de sus usuarios (Statista, 2019a) y con la posibilidad que ofrece para poder extraer sus mensajes públicos.

El contenido de los mensajes en estas redes está escrito en el lenguaje natural de cada usuario. Gracias al desarrollo de las ciencias y la tecnología, disponemos de maquinaria y algoritmos con los que una computadora puede clasificar y analizar grandes volúmenes de estos contenidos, compararlos entre ellos y valorar los sentimientos y emociones que transmiten, de forma que un experto pueda visualizar, analizar e interpretar los resultados obtenidos.

Con la contradicción entre la importancia social de la química y su imagen histórica negativa, con redes sociales *on-line* como nuevos medios de comunicación, con Twitter una red social *on-line* significativa que permite extraer los contenidos de mensajes públicos, con la capacidad técnica y de conocimiento para clasificar contenidos y los sentimientos que transmiten, y como veremos en el marco teórico de este trabajo, sin haber encontrado ningún estudio que haya analizado la imagen pública de la química en Twitter, en este trabajo nos planteamos la investigación de la imagen pública de la química. Creemos que los resultados que puede aportar servirán a grupos de influencia como científicos, educadores, profesionales químicos, organizaciones públicas y privadas y la sociedad en general para entender la imagen en Twitter y de esta forma cada uno de los miembros de los grupos de influencia poder desarrollar estrategias y acciones más efectivas para conseguir sus objetivos.

Por tanto, el propósito de esta investigación consiste principalmente en descubrir si existe o no una imagen pública de la química en Twitter, y si existe, analizar qué contenidos incluye, valorar qué sentimientos y emociones transmite, identificar la presencia de usuarios influyentes y analizar sus perfiles y publicaciones.

Una vez revisada la bibliografía existente sobre el tema y para conseguir este propósito, captaremos una muestra de los mensajes públicos en Twitter o *tweets* relacionados con la química que estén escritos en inglés. Un *tweet* es un mensaje corto de hasta 140 caracteres hasta el 26-09-2017 (Rosen y IKuhiro, 2017) y desde entonces, con un máximo de 280 caracteres para la mayoría de idiomas. Con técnicas de minería de textos los limpiaremos de palabras y símbolos no útiles para la investigación y eliminaremos *tweets* duplicados.

Agruparemos estos *tweets* en conjuntos con cierto grado de homogeneidad o clústeres de forma que si existen diferentes temáticas en los contenidos de los *tweets*, los clústeres obtenidos contengan el máximo número de términos relacionados con una

temática. Interpretaremos los clústeres mediante la representación gráfica de nubes de palabras o *wordclouds*.

Mediante un diseño de análisis de experimentos, repartiremos el total de clústeres en grupos y proporcionaremos cada grupo a diferentes expertos químicos para que los clasifiquen en diferentes temáticas. De esta forma, podremos analizar si las temáticas se corresponden de forma cualitativa con algún aspecto de la imagen pública de la química encontrado en la bibliografía consultada.

A partir de la clasificación obtenida y considerando solo los *tweets* relacionados con las temáticas de la imagen pública de la química, analizaremos el valor de los sentimientos positivos o negativos que transmiten para compararlos con los sentimientos recopilados de la bibliografía. Adicionalmente, detectaremos las emociones relacionadas con estos *tweet* y analizaremos su predominancia y su valor positivo o negativo.

Por otro lado, buscaremos aquellas organizaciones públicas o privadas que podrían tener un mayor nivel de importancia dentro de la química. Con todos los *tweets* recopilados, detectaremos los usuarios que tienen más influencia y su correspondencia con estas organizaciones para analizar si son las que más influyen en Twitter.

Finalmente, analizaremos el contenido de los *tweets* de los usuarios más influyentes para entender qué temáticas transmiten y mediante redes los representaremos gráficamente junto con sus relaciones con otros usuarios para comprender sus vínculos y su capacidad de transmisión de contenidos.

De acuerdo con el propósito de esta investigación, este trabajo se organiza de forma que estudiaremos en primer lugar la imagen pública de la química para entender en qué consiste, los ámbitos que comprende y qué valoración positiva o negativa transmite. Describiremos también las redes sociales *on-line* para centrarnos en Twitter, red en la que nos focalizaremos para analizar la imagen pública de la química. Para poder analizar los contenidos de los *tweets* de Twitter necesitaremos entender y aplicar alguna de las técnicas más utilizadas y adecuadas en la minería de textos, las cuales revisaremos. Junto con las técnicas revisadas utilizadas en redes sociales y en particular en Twitter para valorar los sentimientos y las emociones de textos, podremos entonces analizar la imagen pública de la química. El estudio de cómo considerar la

influencia y reputación en Twitter atendiendo a las métricas más relevantes y el estudio de la relación entre usuarios nos permitirá analizar los usuarios más influyentes, compararlos con los que son presuntamente más influyentes, analizar los contenidos de sus mensajes y conocer la capacidad de transmisión de estos.

Mediante la discusión de los resultados obtenidos, teniendo en cuenta las limitaciones de la investigación, resumiremos las principales conclusiones a las que hemos llegado tras la investigación realizada y apuntaremos futuras líneas de investigación.

2 Marco teórico

En el marco teórico de esta investigación se revisa la bibliografía sobre la imagen pública de la química, la descripción y uso de la redes sociales *on-line* y en particular Twitter, para finalmente analizar el estado de la imagen pública de la química en esta red social.

2.1 Imagen pública de la química

Se ha definido la imagen pública de la química como la forma en que la sociedad piensa y siente acerca de la química, los profesionales químicos y los productos químicos (The Royal Society of Chemistry y TNS BMRB, 2015). Otra forma de definirla es mediante la imagen manifiesta de la química, que comprende su práctica diaria y la interpretación de las personas (Vivas-Reyes, 2009). Esta imagen se desarrolla desde diversos puntos de vista como la académica por científicos, educadores y filósofos, la escolar en los materiales didácticos y en el trabajo académico de los profesores en el aula y la popular por los medios de comunicación en interacción con la población en general (Mora Penagos, 1997).

La imagen puede ser positiva o negativa. Aunque la química tiene éxitos y aplicaciones en la vida cotidiana, la imagen está relacionada con problemas como contaminantes, calentamiento global o dioxinas entre otros (Vivas-Reyes, 2009). A veces, es tan negativa que puede desarrollar quimiofobia, término definido por la "International Union of Pure and Applied Chemistry" (IUPAC) como "miedo irracional a los productos químicos" (Duffus, 1993).

Los profesionales químicos conviven históricamente con ella sufriendo hostilidad pública debido a múltiples factores, como por ejemplo la guerra química durante la primera guerra mundial o el desastre de Bophal de 1984 (Laszlo, 2006). Actualmente, por ejemplo, parece que dentro del público americano y en el ámbito de la salud la quimiofobia está creciendo debido a una supuesta alta relación entre los productos químicos y el cáncer (Entine, 2011).

Los efectos de la imagen pública de la química hacen que el rol de la química dentro de los planes de mejora de la educación científica sea pequeño en comparación con otras ciencias, aunque los avances dentro de estas como la medicina, la agricultura, la farmacia o la física sean consecuencia de los avances en la química (Breslow, 1993).

Para disminuir la quimiofobia y que con el tiempo desaparezca, se sugiere, por ejemplo, formar a los profesores para presentar el conocimiento de la química como deseable y base de la sociedad moderna (Michaelis, 1996).

La quimiofobia también hace que la difusión de la química al público en general sea difícil y se sugiere a los profesionales químicos mejorar su comunicación (Hartings y Fahy, 2011). Casos históricos que existen, como el evitar mencionar la palabra química en títulos de libros para poder ser populares, demuestran no obstante, que es difícil.

Existe incluso un tipo de marketing quimiofóbico (Moreno, 2013) que hace que los compuestos químicos no sean ni bien vistos ni entendidos. Por un lado se comercializan artículos peligrosos para la salud con la etiqueta “sin productos químicos” y por otro productos químicos que ayudan a mejorarla. Estas acciones favorecen que la sociedad no conozca la realidad de la química a través de hechos. Para contrarrestarlo se sugiere la difusión de la química por los profesionales químicos para que sea cercana y cotidiana, por ejemplo, a través de las redes sociales a los jóvenes y más concretamente, a través de Twitter, para utilizarlo como un escaparate de la química.

La quimiofobia sigue siendo objeto de estudio (Rollini, 2020) con limitaciones debido a ser una temática dispersa, basada en opiniones y con la necesidad de involucrar a campos diversos de las ciencias, entre otros, la psicología o la antropología. Sigue existiendo, y parece no afectar a la química que se percibe de forma positiva o neutral así como a los expertos en química que se consideran personas buenas e inteligentes.

La imagen de la química se ha estudiado de forma científica a partir de dos perspectivas, la académica y la social o popular. A nivel académico, los estudios realizados revelan que la imagen de la química es negativa debido a una desfavorable imagen preestablecida, a unos contenidos alejados de las motivaciones de los estudiantes y a su forma de producción del conocimiento en las aulas.

Ya en el año 1975 se detecta una imagen desfavorable preestablecida de la química (Trozzolo, 1975), que junto con las actitudes negativas de los estudiantes (Yager y Penick, 1983; McDermott, 1984; Schibeci, 1986; Furió Más, 2006) parecen ser uno de los factores determinantes a la hora de elegir estudios de química por parte de

estudiantes (Elías Pérez, 2006; Galiano *et al.*, 2015) que prefieren otras enseñanzas ya sean o no científicas.

Los estudiantes tienen una falta de motivación con respecto las ciencias y más concretamente con la química (Galagovsky, 2005, 2007; Rodríguez Gómez, 2009) debido a su dificultad (Mammino, 2001). Asimismo, la falta de contexto y perspectiva histórica y social adecuada en los planes docentes (Jiménez y Criado García-Legaz, 2005; Muñoz y Nardi, 2011) o el obviar algún tema inicial en las clases que ayuda a poner en contexto la química (Nicolas, 2006), da como resultado una imagen deficiente transmitida a los alumnos y en consecuencia su desinterés (Solbes y Traver, 1996).

La falta de introducción de relaciones entre ciencia, tecnología y sociedad debido a su ausencia en los libros de texto científicos, es otro motivo por el que los estudiantes tengan una imagen de la ciencia alejada del mundo real, del desarrollo tecnológico y de sus implicaciones en el medio ambiente y en la sociedad (Solbes y Vilches, 1992; Ribelles *et al.*, 1995; Malaver *et al.*, 2004; Furió Más, 2006; Nicolas, 2006).

Otro aspecto curricular es la falta de atención suficiente a los campos social y humanístico de la química, en particular de su filosofía, historia, sociología, psicología, y didáctica. Parece ser una deficiencia que ha contribuido a la generación y persistencia de estereotipos negativos que afectan no solo la motivación para estudiar y formarse en las distintas áreas de la química sino también en la formación y el ejercicio docente (Penagos y Lozano, 2009).

La poca claridad de las formas de producción del conocimiento en las aulas también parece ser uno de los factores que ayudan a esta imagen negativa (Penagos y Lozano, 2009) y que, en consecuencia, influye en la dificultad de comprensión por parte de los estudiantes y por tanto, en su motivación. Por ejemplo, una encuesta realizada a 840 estudiantes de bachillerato (Nicolas, 2006), concluyó que los docentes debían esforzarse en explicar más claramente la naturaleza de la química y su importancia en la sociedad del conocimiento.

Esta producción de contenidos afecta a la imagen de la química que los estudiantes construyen y que reflejan en las ideas previas que tienen sobre ésta, siendo la de una disciplina no continua sino a trozos desde el punto de vista de los temas, con conceptos limitados y con descripciones más que explicaciones (Chamizo *et al.*, 2005).

Adicionalmente la visión deformada de la química que a veces se transmite afecta también a la motivación de los estudiantes. Por ejemplo se observa en la enseñanza secundaria (Lacolla *et al.*, 2013) que la idea que los alumnos tienen sobre una reacción química es del concepto de explosión, así como es lo que esperan al realizar un experimento en el laboratorio. Otros investigadores (Fernández *et al.*, 2002; Chamizo, 2011), también han detectado la visión deformada que se transmite en la escuela.

A nivel social, la imagen de la química es poco favorable en comparación con la biología y la física (Bensaude-Vincent y Simon, 2012) y alimentándola (Stocklmayer y K. Gilbert, 2002; Penagos y Lozano, 2009; Hill y Kumar, 2013) tanto a nivel general como de la industria química en particular (Moreau, 2005).

La imagen no ha sido siempre la misma. Históricamente y antes del nacimiento de la química moderna (Partington, 1951) parece percibida como negativa (Hartings y Fahy, 2011). Durante la primera guerra mundial, denominada también Guerra Química, y en la que se utilizó dinamita, explosivos y gas venenoso (Hartings y Fahy, 2011) fue negativa. Cambió a positiva en los años treinta gracias por ejemplo al desarrollo de antibióticos y con intentos en los años treinta y cuarenta de mejorarla en USA (Schummer *et al.*, 2007) volviendo a ser negativa en la segunda guerra mundial. Hubo intentos de cambio después de la segunda guerra mundial por parte de la industria en USA, donde los fabricantes de plástico utilizaron estrategias para ganar aceptación pública del plástico en ambivalencia con las dudas con respecto a sus efectos en el entorno, manteniéndose en los años setenta debido a la contaminación global, y aparentemente no recuperándose aún con los nuevos desarrollos científicos en los años ochenta (Allen, 2004), como por ejemplo con la nanotecnología.

Asimismo parece que esta imagen negativa de la química no es territorialmente homogénea. Por ejemplo, dentro de Europa las opiniones son diferentes en función del país (Schummer, 2004) con un 80% de la población con una imagen negativa de la industria química en Suecia, un 65% en Francia y un 38% en Alemania e Italia.

Esta imagen popular (Penagos y Lozano, 2009) parece también ser negativa debido a propagandas sensacionalistas en los medios de comunicación, utilizando la crítica y vulgarizando y tergiversando las aportaciones científicas, así como a través de opiniones relacionadas con la química en el medio natural y social que hacen que la

industria química y la química se asocien con la contaminación y el deterioro del medio ambiente y percibiéndose como no natural, artificial y potencialmente peligrosa (Mammino, 2001). La química también se ha asociado con venenos, guerras, polución y científicos locos generando una imagen negativa a pesar de las campañas para convencer al público que la química posibilita la salud, el confort y el bienestar (Schummer, 2006; Hill y Kumar, 2013).

No obstante, los estereotipos de la química comunicados a través de imágenes en la historia del arte y la ciencia (Schummer *et al.*, 2007) están cargados tanto con asociaciones históricas negativas como positivas haciendo referencia a ideas clásicas de la belleza. También existen escritores americanos que describen la química como un aspecto de la vida moderna que proporciona texturas, sabores, olores y colores en la vida diaria, ahora bien, también se destaca que la química es difícil de encontrar en novelas a principios del siglo XXI más allá del uso ocasional de venenos en novelas de asesinatos (Schummer, 2004). Adicionalmente, en general, en la literatura aparece el arquetipo del químico como un ser siniestro, peligroso, secreto y loco (Schummer *et al.*, 2007).

También Schummer (2007) aporta que en la ficción y en los filmes se perpetúa este arquetipo que proviene del siglo XIX en la Europa Oeste. En 222 películas entre 1992 y 2001 donde aparecía la temática de las ciencias y de la química, cuando reflejaban una conducta indebida o algo mal hecho, era frecuentemente un químico el culpable siendo en el caso de películas de horror en un 24% el malo de la película un químico, mientras cuando eran otros tipos de científicos el porcentaje era inferior al 10%. Adicionalmente, analiza el caso de cómics en que los científicos locos eran profesionales químicos en el 50% de los casos así como en el caso de imágenes encontradas por Internet, donde los profesionales químicos no eran conocidos, sin afeitarse, viejos y con barba y trabajando en el laboratorio con bata blanca y gafas, mientras que las imágenes de los físicos eran de científicos famosos, reforzando la idea que los profesionales químicos son personas aisladas y no sociables.

El no divulgar los principales desafíos de la química en comparación con la publicación de grandes retos intelectuales, como por ejemplo el funcionamiento del cerebro o la clonación, hace también que se mantengan las imágenes preconcebidas sobre la química (Furió Más, 2006). Adicionalmente, el otorgar descubrimientos y desarrollos científicos químicos a otras ramas de la ciencia como la medicina o la biología molecular (Trozzolo, 1975) no ha ayudado a esta imagen popular, aunque existen

premios y reconocimientos entre otros, el Premio Nobel de Química, que aparentemente deberían ayudar a su mejora.

La percepción social de la industria química preocupa a la propia industria por su impacto social y sus efectos negativos sobre el medio ambiente (Adrian y de Paula, 2013), hasta el punto que entidades relevantes de grupos de influencia de la química como el Foro Química y Sociedad en España que agrupa a representantes de empresas, centros de investigación y sociedades profesionales así como el Responsible Care en Canadá, iniciativa propia de la industria química y que agrupa a las asociaciones nacionales de empresas del país, tienen como objetivo el establecer canales de diálogo con la sociedad.

Aunque estudios manifiestan una imagen negativa (Lazlo y Greenberg, 1991), midiendo incluso que la opinión pública favorable acerca de la industria química del año 1980 al año 1990 se redujo del 30% al 14 %, y su percepción pública desfavorable aumentó del 40% al 58% (King y Lenox, 2000), algunos de los estudios más recientes sugieren un cambio en la imagen pública de la química. En el año 2004, la IUPAC, mediante un estudio (Mahaffy *et al.*, 2008), correlacionó la imagen negativa con un conocimiento limitado de la química, los profesionales químicos, los productos químicos y de las funciones y operaciones de la industria química con respecto al público en general

Asimismo el "Consejo Europeo de la Industria Química" (CEFIC) realizó un estudio en el año 2010 acerca de las percepciones públicas de la industria química en la Unión Europea, en la que los encuestados situaban el sector con una imagen favorable y en el lugar sexto de ocho industrias (telecomunicaciones y electrónica, alimentación, farmacéutica, automovilística, eléctrica, petroquímica y nuclear) comparadas con la química. De forma comparativa la química estaba en el mismo nivel que en el año 2008 habiendo mejorado su percepción favorable desde finales de los años 90 (Hadhri, 2010).

Existen también diversos informes promovidos por la Unión Europea. El eurobarómetro de la Unión Europea del año 2013 (European Commission, 2013) está focalizado en las sustancias químicas y no en la imagen de la química percibida y preguntada directamente a los encuestados. Más concretamente, está focalizado en la percepción de su uso en productos diarios, en las actitudes con respecto a si se

pueden o no eliminar, si pueden ayudar a reducir el uso de recursos naturales, si pueden ayudar a un mejor medio ambiente, si están involucradas en innovaciones industriales así como en su seguridad tanto de consumo como de gobernanza.

También existen los eurobarómetros de los años 2014 (European Commission, 2014) y 2017 (Eurobarometer, 2017) sobre las actitudes de los ciudadanos europeos con respecto al medio ambiente. En éstos se pregunta sobre las implicaciones de los productos químicos en la salud y en el medio ambiente. Un 43% de los encuestados estaban preocupados acerca de su salud y el uso de productos químicos en productos de la vida diaria en 2014, siendo de un 84% en 2017. Asimismo un 90% en 2017 estaban preocupados con respecto al impacto de su uso sobre el medio ambiente. Estos eurobarómetros tampoco están focalizados en el estudio de la imagen pública de la química, aunque de ellos se puede deducir dentro del nivel social parte de esta imagen, siendo no positiva atendiendo a los resultados obtenidos y empeorando.

Un estudio también reciente, pero focalizado en la imagen pública de la química, es el publicado por la “Royal Society of Chemistry” (The Royal Society of Chemistry y TNS BMRB, 2015), realizado en el Reino Unido a mayores de 16 años a través de una encuesta Omnibus a 2104 personas, workshops cualitativos a miembros del público en cuatro localidades (Newcastle, Birmingham, London y Southampton) y a 450 miembros y empleados de la “Royal Society of Chemistry” como representantes los profesionales químicos. En éste se observa (ver Figura 2.1) un sentimiento hacia la química más neutral (51% de los encuestados) o positiva (19% les transmitía una sentimiento de felicidad).

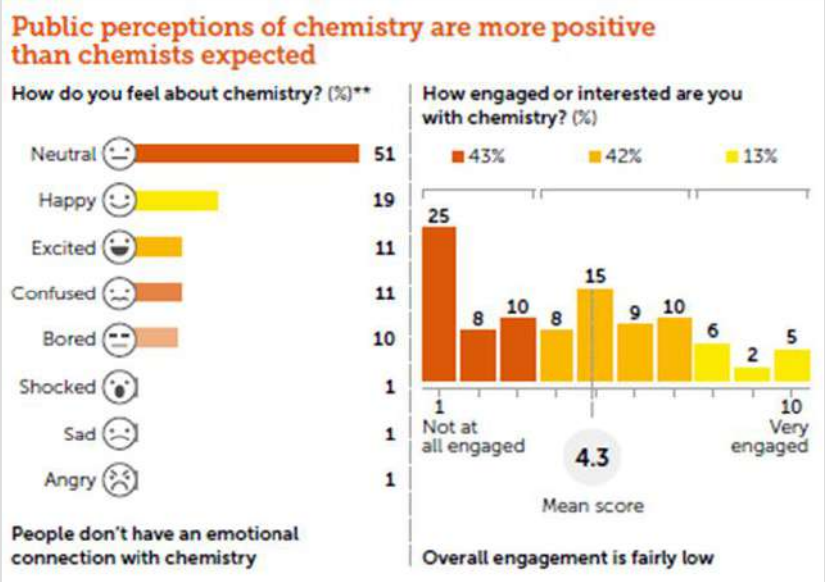


Figura 2.1 Actitud pública de la química. Reproducida de (The Royal Society of Chemistry y TNS BMRB, 2015)

La imagen también era positiva respecto a los impactos de la química en la sociedad debido a sus beneficios (59% de los encuestados respondieron que los beneficios de la química eran superiores que los efectos dañinos), siendo percibida con un impacto positivo mayoritariamente en el bienestar de las personas (75% de los encuestados) y vista mayoritariamente como una solución a grandes desafíos planetarios como la dependencia del petróleo, la escasez de comida, el acceso a agua potable o la contaminación.

Pero por otro lado, las principales asociaciones de los encuestados con la química eran mayoritariamente conceptos relacionados con la educación, como la escuela o profesores (21% de los encuestados), la ciencia (16%) y los elementos químicos (14%). Asimismo sus principales recuerdos estaban relacionados con posibilidades de accidentes en los laboratorios, símbolos de peligro, alarmas, aromas fuertes de amoníaco, azufre o gases y la idea de necesidad de concentración o silencio debido a una noción de dificultad.

Adicionalmente, el 47% de los encuestados estaban de acuerdo con que la escuela les había desanimado con la química o tenían una opinión neutra, un 31% respondían que realmente les servía en su vida diaria y un 52% no tenían confianza para hablar de ella, aunque el 55% respondieron que era importante saber de química en su vida diaria.

Comentarios cualitativos adicionales que la consideraban seria, difícil de entender, inaccesible, dura, metódica e incluso que no estaba implicada en la producción de productos de la vida diaria así como “la química no me interesa”, “nunca la disfruté en la escuela” o “no sé mucho respecto a la química” parecen mostrar un sentimiento o bien negativo o bien neutral de la química.

Resultados cualitativos parecidos circunscritos a la población encuestada en el Reino Unido se encontraron en el “Wellcome Trust Monitor” (Clemence *et al.*, 2013) del año 2013. Los encuestados respondieron que eran poco conscientes con respecto a cómo la química era relevante para ellos, siendo distante en el tiempo con respecto a su educación, con poca visibilidad respecto a lo que significaba, con percepciones de tener una cara peligrosa escondida así como ser difícil y dura, y en resultado, entendiéndola como una disciplina abstracta más que una ciencia aplicada.

Adicionalmente, en el estudio de la “Royal Society of Chemistry” (2015), un 55% de los encuestados respondieron no sentirse informados con respecto a la química, siendo una fuente de desinterés. Este porcentaje variaba con aquellos encuestados que reconocían o no la importancia de la química en su vida diaria (53% vs 24%).

Las fuentes de información que los encuestados describieron que habían escuchado o leído noticias sobre la química son las mostradas en la Tabla 2.1.1:

Tabla 2.1.1. Fuentes de información sobre la química. Extraído de (The Royal Society of Chemistry y TNS BMRB, 2015)

Fuente	**Porcentaje de encuestados**
Televisión	57%
Packaging de productos	27%
Periódicos	23%
Noticias de Internet	22%
Social media	16%
Radio	15%
Revistas o libros	15%
Museos o conferencias	10%

Un 48% de los encuestados que buscaban información de la química en su vida diaria, utilizaron Google o un buscador de Internet disminuyendo a menos de un 20% en el resto de fuentes de información mostradas en la Tabla 2.1.1.

La confianza en los medios de comunicación era superior al 50% excepto con las fuentes de *social media*, Wikipedia y los periódicos sensacionalistas. Los buscadores

de Internet, las páginas web de empresas químicas o farmacéuticas, los periódicos on-line, las páginas web de gobiernos y las de universidades y centros de investigación tenían una confianza superior al 50%. Esta confianza es relevante, debido a que si, por ejemplo, realizamos una búsqueda en Google del término química, aparecen noticias con respecto a la química con contenidos alarmantes ayudando a no crear una imagen positiva de ésta. Los temas de mayor interés eran aquellos relacionados con la medicina, la tecnología de depuración de agua potable y la escasez de alimentación a nivel mundial.

Los encuestados asociaban a los profesionales químicos con los farmacéuticos y los médicos (un 48% de los encuestados) y droguerías (13%), no sabiendo en qué otro tipo de industrias los químicos desarrollan su actividad. Por otro lado, solo un 14% respondió que su percepción de lugar de trabajo de un químico era otras industrias diferentes de farmacias, laboratorios y hospitales. Estos resultados parecían inducir a considerar los profesionales químicos como profesionales que cambian el mundo (95% de los encuestados) y que tienen una serie de cualidades positivas como la honestidad, la cercanía, el entusiasmo y el interés. Por tanto, esta percepción puede ayudar a crear una imagen positiva de la química rompiendo estereotipos o creencias históricas.

Asimismo y con respecto a los productos químicos, los encuestados tienen una percepción neutra (55%) y tienen conocimiento de atributos tanto positivos como negativos de los productos. Pero un 44% de los encuestados tienen una percepción negativa de los productos químicos considerados como sintéticos o perjudiciales, aunque no se sienten suficientemente informados acerca de ellos y tienen la creencia que son beneficiosos para la sociedad. Asimismo, el estudio sugería que las percepciones sobre los productos químicos no estaban asociados con los profesionales químicos o la química.

Es destacable resaltar que los encuestados reconocen que los medios de comunicación diseminan informaciones no positivas de los productos químicos, inconsistencias en el consejo con productos y comidas y que no tienen suficiente información, ni esta es de confianza, en los canales de comunicación de noticias, el packaging y las etiquetas de los productos. Parece que debido a esto, los encuestados no estaban interesados en estos temas.

2.2 Redes sociales en Internet y Twitter

Una red social puede definirse como un servicio basado en web[2] dentro de la red de Internet que permite a sus usuarios tener un espacio social, construir un perfil público compartiendo toda la información que publican, o semipúblico compartiendo parte de su información, generar una lista de usuarios con los que compartir conexiones con aquellos que tienen intereses o conexiones similares y ver y recorrer sus listas de conexiones dentro de la red, pudiendo saber algo acerca del resto de individuos conectados pero no toda su información (Boyd y Ellison, 2007; Dron y Anderson, 2009; Marques *et al.*, 2013; P. C. Lin *et al.*, 2013).

Asimismo, también permiten la experiencia del descubrimiento y el intercambio de contenidos generados por los usuarios, ponen a su disposición las herramientas para poder realizarlo (Gikas y Grant, 2013a) y permiten a sus usuarios la recolección, organización y preservación del conocimiento en forma de contenidos (Virkus, 2008; Madhusudhan, 2012).

Las redes sociales son utilizadas a nivel mundial. Según la web de Statista (2019d) el número de usuarios o personas que están registradas en una red social y que al menos la utiliza una vez al mes a través de cualquier dispositivo electrónico adecuado para su uso, era de 2 480 millones en el año 2017 con unas previsiones de alcanzar los 3 090 millones en el año 2021. Su penetración, definida como el porcentaje de usuarios de redes sociales respecto al total de usuarios de Internet, fue elevada con un valor del 71% en 2017.

Los países o regiones del mundo que lideran las redes sociales son diferentes según el número de usuarios, según el porcentaje de usuarios respecto a la población del país o región o según su penetración como podemos observar en la Tabla 2.2.1, en la Tabla 2.2.2 y en la Tabla 2.2.3.

Tabla 2.2.1. Cinco países del mundo con mayor número de usuarios en redes sociales en 2018 (Statista, 2019d)

País	Millones de usuarios
China	673,5
India	326,1
Estados Unidos	243,6
Brasil	95,2
Indonesia	81

[2] Web: red informàtica. Relativo a sitio web, conjunto de páginas web agrupadas bajo un mismo dominio de internet. (https://dle.rae.es/sitio#Rt2Ilqu consultado el 21/04/2021)

Tabla 2.2.2. Cinco regiones del mundo con mayor porcentaje de la población que es usuario de redes sociales a Junio de 2019 (Statista, 2019d)

País	Porcentaje
Este de Asia	70%
América del Norte	70%
Europa del Norte	67%
América del Sur	66%
Centroamérica	62%

Tabla 2.2.3. Cinco países del mundo con mayor porcentaje de cuentas activas de la red social con mayor número de usuarios en comparación con la población en Junio de 2019 (Statista, 2019d)

País	Porcentaje
Emiratos Árabes Unidos	99%
Taiwán	89%
Corea del Sur	85%
Singapur	79%
Hong Kong	78%

Una de las formas para clasificar las redes sociales es mediante su popularidad, medida por el número de usuarios activos mensuales o usuarios únicos que visitaron la red social durante el último mes. Las tres redes sociales más populares a Octubre de 2019 eran Facebook con 2 414 millones de usuarios activos mensuales o usuarios únicos que visitarion la red durante el último mes, YouTube con 2 000 millones y WhatsApp con 1 600 millones. Su uso, medido como el tiempo que un usuario pasa en ella, está aumentando con un incremento a nivel mundial de 90 minutos al día en 2012 a 136 en 2018.

El rol que un usuario puede tener en una red social puede ser diverso, pudiendo relacionarse con amigos, familia, empleadores o clientes entre algunos potenciales roles como podemos observar en la Tabla 2.2.4.

Tabla 2.2.4. Principales roles de un usuario en una red social durante el último trimestre de 2018 (Statista, 2019d)

Rol de un usuario	Porcentaje
Estar conectado con amigos	40%
Encontrar contenidos de entretenimiento	37%
Buscar o encontrar productos para comprar	30%
Compartir su opinión	29%
Hacer *networking* laboral	24%

Adicionalmente, la percepción del impacto de las redes sociales en aspectos de la vida diaria es diversa. Según la misma fuente de Statista (2019d) a Febrero de 2019, el 57% de los usuarios de Internet encuestados percibían un incremento de este impacto

tanto gracias al acceso a la información como a la facilidad de comunicación y un 50% gracias a la libertad de expresión.

Las organizaciones y empresas utilizan las redes sociales para buscar contenidos y productos (Kietzmann *et al.*, 2011; Zaglia, 2013) para buscar información de los usuarios y aumentar su número (Skeels y Grudin, 2009; Waters *et al.*, 2009; Curtis *et al.*, 2010; Reid y Ostashewski, 2010), para poder aumentar la colaboración entre usuarios, para poder compartir contenidos y para construir comunidad entre los usuarios (Jansen, Sobel, *et al.*, 2009; Jansen, Zhang, *et al.*, 2009; Michaelidou *et al.*, 2011).

Estos contenidos son una fuente de información que puede ser utilizada por organizaciones y empresas, que pueden interactuar de forma directa o mediante publicidad con sus grupos de clientes objetivos (Weller *et al.*, 2014), permiten despertar el interés de usuarios sobre sus nuevos productos y campañas (Larson y Watson, 2011) y recoger quejas y recomendaciones que pueden influir en sus decisiones de compra (Parveen, 2012; Reinhold y Alt, 2012).

Las redes sociales también se están utilizando dentro del contexto educativo y como un medio de comunicación adicional al tradicional dentro de las aulas (Judd, 2010). Están creciendo en popularidad en la educación formal e informal gracias a la posibilidad de crear comunidades de conocimiento (Mason y Rennie, 2008) que promueven el potencial de las redes para ser una conexión entre la educación formal e informal (Greenhow y Lewin, 2016), y están alineadas con el uso de los dispositivos móviles que permiten un acceso instantáneo a los contenidos (Du *et al.*, 2013; Gikas y Grant, 2013b). Adicionalmente, las funcionalidades de las redes sociales y el hecho de poder compartir el conocimiento de una forma simple, permiten una rápida diseminación de la información (Lovejoy *et al.*, 2012).

La valoración del uso de las redes sociales para objetivos docentes parece ser positiva. De una revisión de 662 estudios de tesis y disertaciones sobre *social media* (Piotrowski, 2015), de los que 29 estaban focalizados en el uso de redes sociales en educación, sólo en dos se reportaba aspectos negativos por parte de los estudiantes o de los docentes. No obstante, es un ámbito del conocimiento que evoluciona muy rápidamente debido a la evolución tecnológica y su asimilación, y por tanto, pendiente de explorar. En el caso un estudio en la educación universitaria en Italia (Manca y Ranieri, 2016), los resultados sugerían que el uso de las redes sociales era limitado y

que los docentes no parecían estar dispuestos a integrar su uso en las aulas debido a la resistencia cultural, a aspectos pedagógicos y a limitaciones institucionales, aunque sí que reconocían sus beneficios.

Parece que los académicos tienen activas sus cuentas en las redes sociales por largos periodos (Forkosh-Baruch y Hershkovitz, 2012) y su uso permite compartir el conocimiento en la comunidad docente y un aprendizaje informal mayor. Las utilizan para hacer *networking*, mantener una imagen profesional dentro de la comunidad académica y buscar información académica (Dermentzi *et al.*, 2016).

Por otro lado, permiten a los estudiantes un mayor soporte al aprendizaje (Greenhow, 2011) y un sistema colaborativo de aprendizaje fuera de la clase tradicional (Marques *et al.*, 2013; P.-C. Lin *et al.*, 2013), siendo sus percepciones de uso dentro de los cursos así como las de profesores e instituciones educativas positivas (Neier y Zayer, 2015), aunque con discrepancias en el caso de cursos masivos abiertos on-line (MOOC) (Salmon *et al.*, 2015) siendo percibidos por algunos participantes como una oportunidad de networking y de compartir conocimientos, pero por otros como una pérdida de tiempo.

Los estudiantes, no obstante, no hacen una distinción entre los medios físicos tradicionales y los virtuales (Bicen y Cavus, 2011), e incluso a nivel universitario esperan el uso de redes sociales como soporte al trabajo en el aula (Roblyer *et al.*, 2010). De forma espontánea las utilizan para interaccionar entre ellos dentro del ámbito educativo (Selwyn, 2007; Trinder *et al.*, 2008; Madge *et al.*, 2009) y pueden servir como un puente de aprendizaje informal entre estudiantes nativos digitales y profesorado inmigrante digital (Bull *et al.*, 2008), así como tener implicaciones positivas percibidas por los estudianes en su aprendizaje académico y en algún tipo de evaluación como los ensayos (Arquero y Romero-Frías, 2013).

Dentro de las redes sociales, Twitter se define como una plataforma de comunicación basada en Internet, con servicio de *microblogging* y características de red social (Stevens, 2008; Veletsianos, 2012; Davenport *et al.*, 2014). Un *microblog* es un servicio que permite a los usuarios escribir pequeños textos desde dispositivos móviles y computadoras personales para publicarlos en Internet (Oulasvirta *et al.*, 2010).

Twitter fue diseñado en 2006 por Evan Williams and Biz Stone, que habían trabajado en Google para posteriormente intentar lanzar Odeo, una nueva empresa de

podcasting (The Editors of Encyclopaedia Britannica, 2020). Williams, había estado experimentando con uno de los proyectos colaterales de Odeo, un servicio de mensajes cortos que denominaron Twttr. Previendo un alto potencial de negocio a este producto, Williams compró las acciones de Odeo y lanzó Obvious Corp. para desarrollar este servicio. Jack Dorsey se incorporó al equipo de gestión con el que lanzaron su primera versión del servicio en la conferencia musical Southwest en Austin, Texas en Marzo de 2007. Twitter, Inc., fue creada como empresa al mes siguiente gracias a los fondos proporcionados por un *venture capital*.

Los inicios de Twitter no dejan de tener cierta polémica (Carlson, 2011) ya que Noah Glass, co-fundador de Odeo, no es considerado uno de los fundadores de Twitter. Es sabido que Jack Dorsey tenía la idea de Twitter antes de entrar en Odeo, que Noah Glass era el líder del proyecto Twttr en Odeo y que Biz Stone ayudaba puntualmente, mientras que Evan Williams, al parecer, era inicialmente escéptico. No obstante, Williams era percibido por algunos inversores y empleados de Odeo como una persona calculadora. Williams compró las acciones de Odeo, despidió a Glass y posteriormente ofreció entrar de nuevo a antiguos inversores de Odeo en Twitter, con una valoración de empresa superior y, en consecuencia, con un precio de compra de las acciones de Twitter superior a las que vendieron en Odeo. Aunque, por otro lado, el momento de la entrada de los inversores es posterior al de la venta de sus acciones en Odeo, habiendo la empresa demostrado que el servicio era técnicamente viable y con miles de usuarios inscritos.

Los usuarios de Twitter pueden comunicarse intercambiando mensajes cortos de hasta 140 caracteres hasta el 26-09-2017 (Rosen y IKuhiro, 2017) y desde entonces, con un máximo de 280 caracteres para la mayoría de idiomas. Estos mensajes se denominan *tweets* y se pueden publicar mediante diferentes medios, como por ejemplo la propia web de Twitter o su aplicación móvil.

El modelo de negocio de Twitter consta de tres diferentes tipos de segmentos a los que se dirige, usuarios, empresas y desarrolladores. A los usuarios se les ofrece servicios de *microblogging* al instante a través de diversos canales como su app para smartphones, su página web, y las APIs que permiten integrar Twitter en otras webs. A las empresas, y gracias a la gran cantidad de usuarios y la información que Twitter posee sobre ellos, se les ofrece servicios de publicidad enfocada a aquellos usuarios con mayor probabilidad de comprar los productos de esas empresas. A los desarrolladores, Twitter permite la posibilidad de conectarse a Twitter para generar

herramientas relacionadas con analítica web u otras aplicaciones móviles que ayuden a hacer crecer la masa crítica de usuarios que utilizan Twitter.

Twitter actúa como intermediario, como una plataforma de publicidad, permitiendo a sus usuarios utilizar los servicios de *microblogging* de forma gratuita, consiguiendo ingresos directos por la venta de sus servicios publicidad a las empresas e indirectos gracias al incremento de los usuarios mediante las herramientas de los desarrolladores. Los servicios de publicidad se basan en *tweets* promocionados donde el anunciante paga por mostrar el *tweet* a un segmento de usuarios definido, cuentas promocionadas donde el anunciante paga por adquirir seguidores y tendencias promocionadas donde el anunciante paga por tener más visibilidad de sus *tweets*. Adicionalmente, Twitter consigue ingresos por la licencia de datos o venta del acceso a datos públicos para su análisis histórico o en tiempo real y por las comisiones generadas por la intermediación de compra y venta de anuncios móviles.

Podemos observar en la Figura 2.2 que sus ingresos han seguido una tendencia creciente desde el 2010 proviniendo la mayor parte de los servicios de publicidad aunque con una ligera disminución en 2017 en comparación al año anterior. El número de empleados anuales también ha seguido una tendencia parecida a la evolución de estos servicios, como podemos apreciar en la Figura 2.3. Sus beneficios netos anuales mostrados en la Figura 2.4 consiguieron ser positivos a partir del año 2018, aunque en el segundo trimestre del año 2020 y probablemente debido a la crisis provocada por el Covid y su efecto sobre las empresas reduciendo algunos de sus servicios externos como la publicidad, sus beneficios netos trimestrales han sido negativos como podemos observar en la Figura 2.5.

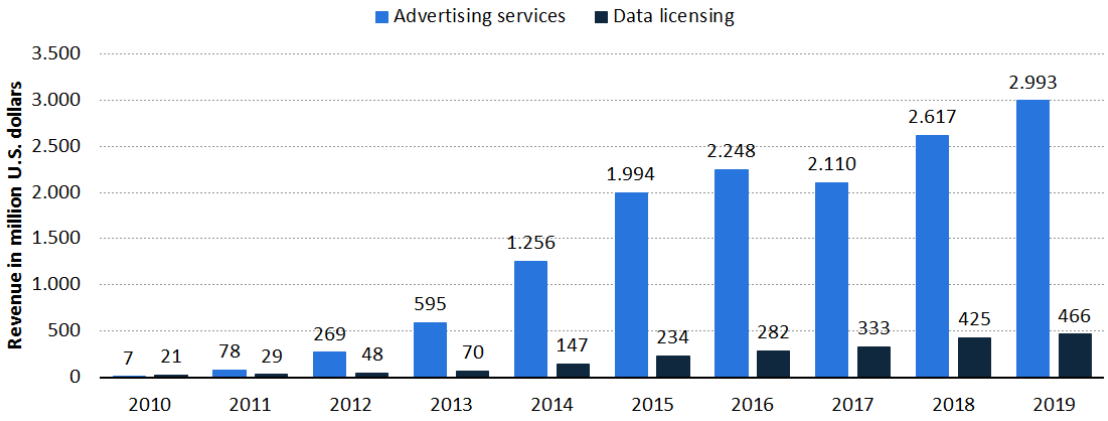


Figura 2.2 Ingresos de Twitter en millones de dólares en función de los servicios de publicidad ("Advertising") y el resto de servicios. Extraída de (Statista, 2020d)

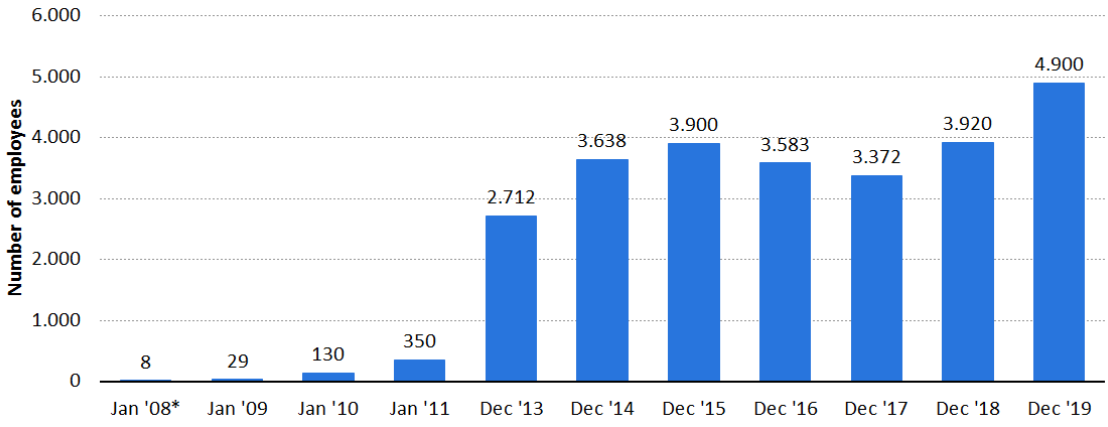


Figura 2.3 Número anual de empleados de Twitter. Extraída de (Statista, 2020d)

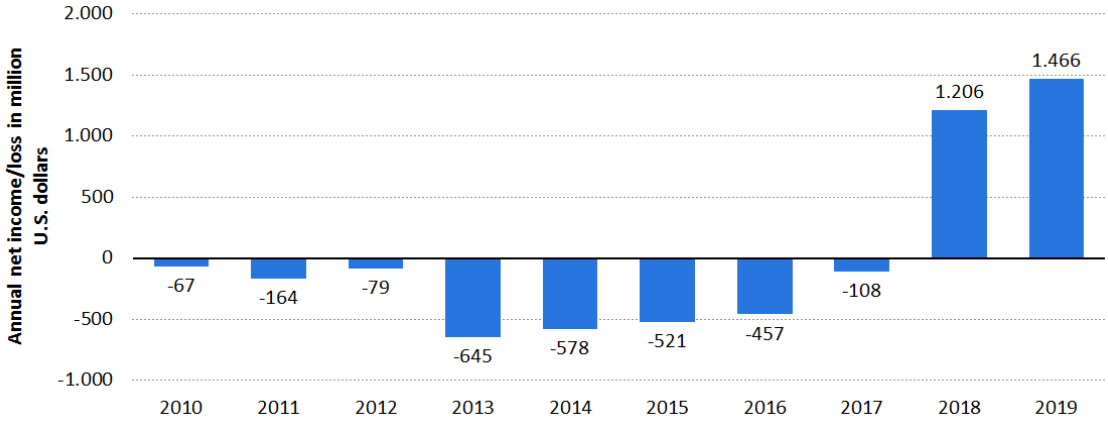


Figura 2.4 Beneficios anuales de Twitter en miilones de dólares. Extraída de (Statista, 2020d)

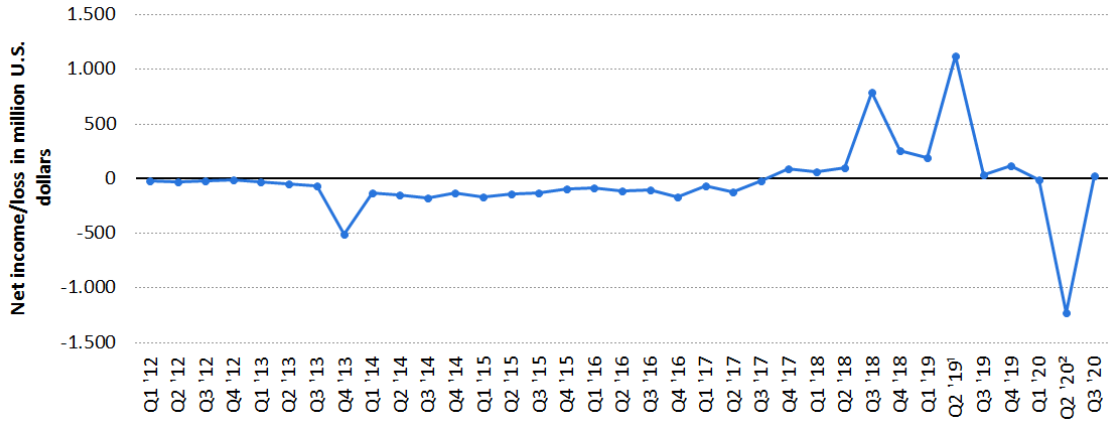


Figura 2.5 Beneficios trimestrales de Twitter en millones de dólares. Extraída de (Statista, 2020d)

Twitter como red social está en decimoséptima posición (Statista, 2020b) en Octubre de 2020 con 353 millones de usuarios considerados como audiencia a la que se puede publicitar (ver Tabla 2.2.5), siendo Facebook la líder.

Tabla 2.2.5. Redes sociales más populares a nivel mundial ordenadas según el número de usuarios activos. Extraída de (Statista, 2020b)

Red Social	**Millones de usuarios activos**
Facebook	2 701
YouTube	2 000
WhatsApp	2 000
Facebook Messenger	1 300
Weixin / WeChat	1 206
Instagram	1 158
TikTok	689
QQ	648
Douyin	600
Sina Weibo	523
QZone	517
Snapchat	433
Reddit	430
Kuaishou	430
Pinterest	416
Telegram	400
Twitter	**353**
Quora	300

El número de usuarios mensuales activos de Twitter (Statista, 2020c) a nivel mundial según el número de usuarios o personas que están registradas en una red social y que al menos la utiliza una vez al mes a través de cualquier dispositivo electrónico adecuado para su uso, ha seguido una tendencia creciente desde 2010 como podemos apreciar en la Figura 2.6, con una ligera disminución a partir del segundo

trimestre de 2018. No obstante esta disminución, el número de usuarios activos que se pueden monetizar (Statista, 2020d) presenta una tendencia creciente hasta el periodo de los datos más recientes disponibles como podemos observar en la Figura 2.7.

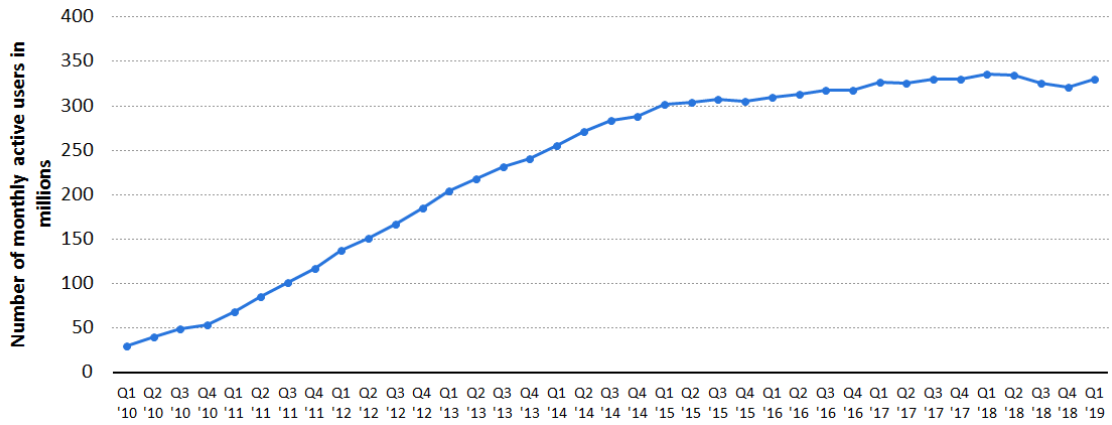


Figura 2.6 Número de millones de usuarios activos mensuales en Twitter. Extraída de (Statista, 2020c)

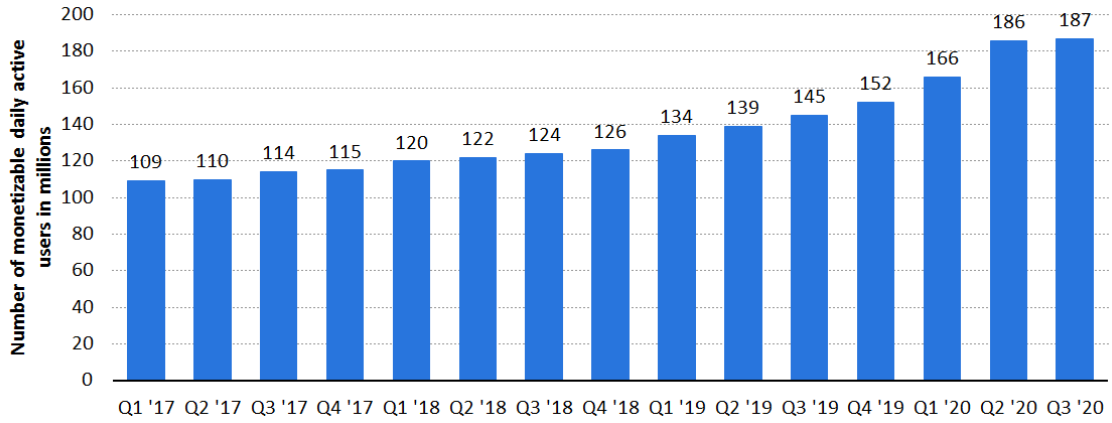


Figura 2.7 Número de millones de usuarios diarios activos monetizables en Twitter desde el primer trimestre de 2017 hasta el tercer trimestre de 2020. Extraída de (Statista, 2020d)

Twitter es sexta por uso (Statista, 2020b) con un 23% de usuarios según una encuesta realizada durante Febrero de 2020 en Reino Unido, USA, Alemania, Francia, España, Italia, Irlanda, Dinamarca, Finlandia, Japón, Australia y Brasil, y en la que se preguntaba qué red social se había utilizado para cualquier tipo de propósito durante la semana previa a la encuesta (ver Tabla 2.2.6), y entre las diez primeras en USA en Septiembre de 2019 (Statista, 2020d) según los minutos pasados en la red social por mes y por usuario mayor de 18 años (ver Tabla 2.2.7).

Tabla 2.2.6. Ranking de redes sociales según su uso. Extraída de (Statista, 2020b)

Red Social	Porcentaje de usuarios
Facebook	63%
YouTube	61%
WhatsApp	48%
Facebook Messenger	38%
Instagram	36%
Twitter	**23%**
Snapchat	13%

Tabla 2.2.7. Ranking de redes sociales según su uso en USA medido por minutos al mes por usuario mayor de 18 años en Septiembre de 2019. Extraída de (Statista, 2020d)

Red Social	Minutos por mes
Facebook (main)	769,2
TikTok	498,1
Messenger by Google	335,5
WhatsApp	292,4
Telegram	274,3
Discord - Chat for Games	253,2
Instagram (main)	202,9
Snapchat	199,9
Kik	164,7
Twitter (main)	**158,2**

La mediana de edad de los usuarios de Twitter (Statista, 2020d) en Octubre de 2020 es de entre 25 y 34 años con más del 75% de los usuarios menores de 49 años y que siguen una distribución mostrada en la Figura 2.8.

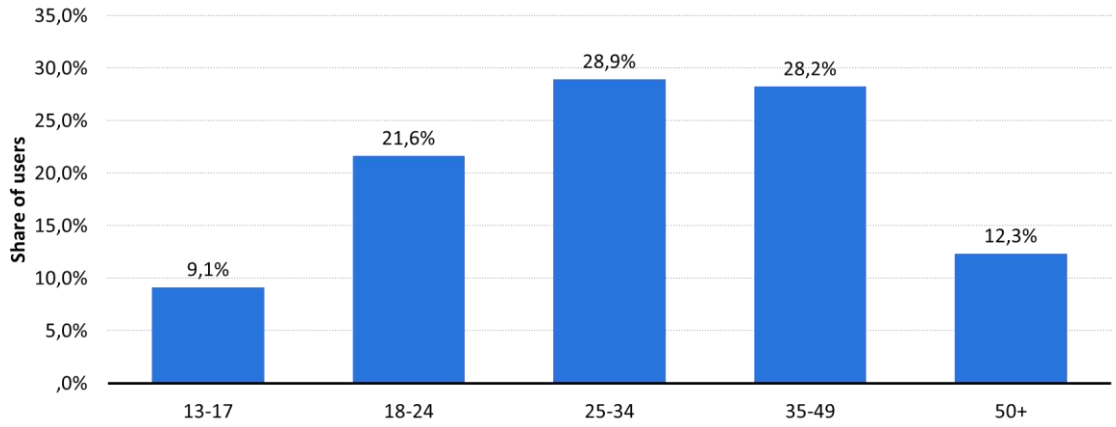


Figura 2.8 Distribución de las edades de los usuarios de Twitter en Octubre 2020. Extraída de (Statista, 2020d)

Respecto a los países con mayor número de usuarios (Statista, 2020d) en Octubre de 2020, USA y Japón concentran una gran parte de ellos con una diferencia elevado con respecto al resto de países (ver Tabla 2.2.8).

Tabla 2.2.8. Ranking de países con mayor número de usuarios en Octubre de 2020. Extraída de (Statista, 2020d)

País	Millones de usuarios
Estados Unidos de América (USA)	68,7
Japón	51,9
India	18,9
Brasil	16,6
Reino Unido	16,6
Turquía	13,4
Indonesia	13,2
Arabia Saudí	12,3
Méjico	10,6
Francia	7,9
Filipinas	7,8
España	7,4
Tailandia	7,3
Canadá	6,2
Alemania	5,4
Argentina	5,2
Corea del Sur	5,2
Egipto	3,7
Colombia	3,4
Malasia	3,1

Twitter a Febrero de 2019 en USA (Statista, 2020d), es utilizado por sus usuarios principalmente para el consumo de noticias, aspecto diferencial sobre el resto de redes. Adicionalmente, es la segunda después de Facebook para realizar networking, y se utiliza como las redes sociales de Instagram, Facebook, Snapchat y Pinterest para ver fotos y vídeos y compartir contenidos (ver Tabla 2.2.9).

Tabla 2.2.9. Uso de las principales redes sociales en USA a Febrero de 2019. Extraída de (Statista, 2020d)

Uso de las redes sociales	Instagram	Facebook	Snapchat	Pinterest	Twitter
Ver fotografías	77%	65%	64%	59%	42%
Ver videos	51%	46%	50%	21%	32%
Compartir contenidos con el resto de usuarios	45%	57%	46%	21%	32%
Compartir contenidos usuario a usuario	31%	43%	45%	12%	20%
Networking	23%	33%	21%	10%	26%
Noticias	18%	38%	17%	9%	56%
Encontrar o comprar productos	11%	15%	5%	47%	7%
Promocionar un negocio	9%	7%	6%	5%	7%

Los idiomas más utilizados en Twitter en 2013 (Mocanu *et al.*, 2013; Richter, 2013) son el inglés con un 34% de *tweets* escritos en inglés, el japonés con un 16% y el castellano con un 12%.

Twitter permite la propagación de la información en tiempo real a los usuarios que pertenecen a la red y hace que sea un entorno para la diseminación de ésta. Los *tweets* parecen seguir una distribución de Pareto (Heil y Piskorski, 2009): investigando una muestra de 300 000 *tweets* aleatorios, descubrieron que el 10% de los usuarios publicaron más del 90% de los *tweets*.

Twitter difiere de otras redes sociales en que en principio un usuario puede seguir a cualquier otro sin la aceptación del primero, los usuarios siguen a otros en función del contenido que publican y todo el contenido dentro de Twitter es público y se puede reenviar a otros usuarios (Brooks, 2011). Permite también publicar fotografías y *Uniform Resource Locators* (URLs) además de texto escrito, emoticonos[3] y *emojis*[4], y actualmente permite que un usuario bloquee a otros para que no puedan acceder a sus contenidos. De esta forma, podemos clasificar los contenidos en públicos y privados, siendo los primeros accesibles y pudiendo ser recogidos mediante la API (*Application Programming Interface)* de "Twitter Search" para ser analizados (Sailunaz y Alhajj, 2019).

Twitter, en comparación con otras redes sociales, ofrece mayor anonimidad y se focaliza menos en quién eres y en tus círculos sociales y más en lo que piensas y dices (Huberman *et al.*, 2009) a través de los *tweets*. Estos permiten mejorar los desafíos que presentan las encuestas (Choi y Pak, 2005; Tourangeau y Yan, 2007; Krumpal, 2013), como los diversos sesgos debidos al diseño de las preguntas de las encuestas y de los cuestionarios, a la forma de realizar las preguntas por un entrevistador y a las respuestas de los entrevistados según la temática de las preguntas, y los desafíos del análisis de documentos (Casadevall y Fang, 2009; Antilla, 2010) como la auto-censura de los analistas según sus opiniones. Los *tweets* son conversaciones entre usuarios de Twitter (Boyd *et al.*, 2010; Huang *et al.*, 2010; Smith *et al.*, 2014) que les ofrecen la capacidad de expresar sus pensamientos, opiniones (Kanavos *et al.*, 2014) y emociones (Tago y Jin, 2018) de forma espontánea, aspectos que son inherentes del comportamiento social humano (Aarts *et al.*, 2012; Ye y Wu, 2013).

[3] Emoticono: secuencia de caracteres ASCII que expresa una emoción
[4] Emoji: ideograma que puede ser utilizado como un emoticono en una conversación o mensaje

No obstante, parece que Twitter no es tan "social". En un estudio de 41 millones de usuarios y 106 millones de *tweets* en 2009 (Kwak *et al.*, 2010), los resultados mostraron que hay una baja reciprocidad en el seguimiento de *tweets* por los usuarios, asemejándose más a un medio informativo donde los usuarios difunden contenidos a sus seguidores. Asimismo, parece existir la existencia de dos redes (Huberman *et al.*, 2009), una primera muy densa con usuarios que siguen y usuarios que son seguidos y otra red dispersa de amigos reales, con una frecuencia de publicación de contenidos menor en la primera red que en la segunda.

El uso de Twitter por empresas y organizaciones dentro del marketing está orientado a difundir información de las empresas y sus productos (Boyd *et al.*, 2010; Stieglitz y Krüger, 2011; Stieglitz y Dang-Xuan, 2013) y parece que la mayoría lo utilizan sobre todo para la distribución de noticias de su ámbito de negocio (Case y King, 2010, 2011). También lo utilizan para monitorizar clientes, observar a los competidores, analizar marcas, productos y la imagen de la empresa (Shannon B. Rinaldo *et al.*, 2011), para comunicarse con sus clientes (Popescu y Jain, 2011; Ioanid y Scarlat, 2017), informar tanto a sus clientes presentes como potenciales acerca de eventos como de nuevos productos (Berinato y Clark, 2010) y promocionar la marca (Popescu y Jain, 2011) siendo raramente utilizada para proporcionar opiniones o consejos así como para la venta de productos o servicios (Swani *et al.*, 2014).

En Twitter existen usuarios aparentemente más importante que otros y que tienen algún tipo de influencia sobre los demás, aunque no existe un acuerdo en la definición de lo que es un usuario más influyente (Riquelme y González-Cantergiani, 2016) así como su forma de medirlo y de denominarlo. Entre otras denominaciones se utiliza líder de opinión, *influencer* o experto de un ámbito. En la química, existen algunos químicos a seguir e *influencers* sugeridos por el *Chemical & Engineering News*[5] (ver Tabla 2.2.10) y *Feedspot*[6] (ver Tabla 2.2.11).

Tabla 2.2.10. Químicos a seguir en Twitter en 2017 según el *Chemical & Engineering News*

Nombre	Descripción	Cuenta de Twitter	Enlace a la cuenta de Twitter
Carolyn Bertozzi	professor of radiology and of chemical and systems biology at Stanford University	@carolynbertozzi	https://twitter.com/CarolynBertozzi
Christopher J. Cramer	professor of chemistry and associate dean at the University of Minnesota, Twin Cities	@ChemProfCramer	https://twitter.com/chemprofcramer

[5] Influencers en la química en 2017: https://cen.acs.org/articles/95/web/2017/11/25-Chemists-should-follow-Twitter.html, consultado el 22-04-2021

[6] Influencers en la química 2021: https://blog.feedspot.com/chemistry_websites/, consultado el 22-04-2021

Nombre	Descripción	Cuenta de Twitter	Enlace a la cuenta de Twitter
Christine Le	postdoc in synthetic organic chemistry at the University of California, Berkeley	@christine_m_le	https://twitter.com/christine_m_le
Brian Wagner	professor of physical chemistry at the University of Prince Edward Island	@DrummerBoy2112	https://twitter.com/drummerboy2112
Ed Sherer	predictive sciences process analyst and chemist at Merck	@edsherer	https://twitter.com/edsherer
Sujata Kundu	teaching fellow at Imperial College London, presenter atDiscovery, and writer at Forbes Science and Standard Issue	@funsizesuze	https://twitter.com/funsizesuze
Henrik Pedersen	professor of materials chemistry at LinkÃ¶ping University	@hacp81	https://twitter.com/hacp81
Debbie Gale	assistant teaching professor of chemistry at theUniversity of Denver	@heydebigale	https://twitter.com/heydebigale
James Batteas	professor of nanochemistry, materials, and biology atTexas A&M University	@jamesbatteas	https://twitter.com/jamesbatteas
Lars Öhrström	professor of inorganic chemistry at Chalmers Universityof Technology	@Larsohrstrom	https://twitter.com/larsohrstrom
Lee Cronin	Regius Chair of Chemistry and professor of chemistry,nanoscience, and chemical complexity at the University of Glasgow	@leecronin	https://twitter.com/leecronin
Luke Gamon	Marie Curie Fellow in Protein Oxidation at the Universityof Copenhagen	@lgamon	https://twitter.com/lgamon
Madison Fletcher	postdoc in chemical biology at the University ofCalifornia, Irvine	@madihfletch	https://twitter.com/madihfletch
Nadine Borduas	postdoc in atmospheric chemistry at ETH Zurich	@nadineborduas	https://twitter.com/nadineborduas
David K. Smith	professor of supramolecular and nanochemistry,medicine, and materials at the University of York and YouTube chemist	@professor_dave	https://twitter.com/professor_dave
Saiful Islam	professor of battery and solar cell materials chemistryat the University of Bath	@SaifulChemistry	https://twitter.com/SaifulChemistry
Sarah Reisman	professor of natural product synthetic chemistry atCalifornia Institute of Technology	@sarah_reisman	https://twitter.com/sarah_reisman
Sarah Cady	associate scientist of NMR at the Chemical InstrumentationFacility of Iowa State University	@sarahdcady	https://twitter.com/sarahdcady
Matthew Hartings	professor of inorganic and food chemistry atAmerican University and author of the book "Chemistry in Your Kitchen	@sciencegeist	https://twitter.com/sciencegeist
Andrea Sella	professor of inorganic chemistry at University CollegeLondon, television and radio broadcaster at the BBC, and author of Sella the Chemist	@SellaTheChemist	https://twitter.com/sellathechemist
Fraser Stoddart	professor of mechanostereochemistry at NorthwesternUniversity, 2016 #ChemNobel	@sirfrasersays	https://twitter.com/sirfrasersays
Stephani Page	postdoc in pharmacology at the University of NorthCarolina, Chapel Hill	@ThePurplePage	https://twitter.com/thepurplepage
Toria Stafford	Ph.D. student in chemistry at the University of Manchester	@ToriaStafford	https://twitter.com/toriastafford
Vittorio Saggiomo	professor of microfluidics, microfabrication andsensors at Wageningen University & Research and author of Labsolutely	@V_Saggiomo	https://twitter.com/v_saggiomo
Vy Dong	professor of organic chemistry at the University ofCalifornia, Irvine, and a host of #VMDchats with scientists	@Vy_Dong_Group	https://twitter.com/Vy_Dong_Group

Tabla 2.2.11. *Influencers* en Twitter en 2021 según *Feedspot*

Nombre	Descripción	Cuenta de Twitter	Enlace a la cuenta de Twitter
Ash Jogalekar	Ashutosh (Ash) Jogalekar is a chemist doing research in biotechnology and is passionate about the history and philosophy of science. He blog's at the Curious Wavefunction where He write about chemistry, drug discovery, physics and history.	@curiouswavefn	https://twitter.com/curiouswavefn
Dr. Jay	Hi my real name is Jason and welcome to my blog. My favourite area is chemistry and my expertise lies in photo chemistry. The blog is official feed for the community project RealTimeChem,	@Doctor_Galactic	https://twitter.com/doctor_galactic

Nombre	Descripción	Cuenta de Twitter	Enlace a la cuenta de Twitter
	connecting chemists.		
Dr Kat Day	My name is DR Kat Day and I have a PhD in chemistry and over a decade's experience as a chemistry teacher. This blog is about Tales of interesting chemical tidbits and chemistry coursework.	@chronicleflask	https://twitter.com/chronicleflask
Egon L. Willighagen	This blog deals with Chemblaics in the broader sense. Chemblaics is the science that uses computers to solve problems in chemistry, biochemistry and related fields.	@egonwillighagen	https://twitter.com/egonwillighagen
Dr. Umesh Laddi	A chemistry blog by Mr. Umesh - Ph.D(Organic chemistry). who has worked worked for various Industries like, Agrochemicals, Pharmaceuticals, Speciality chemicals, and drug intermediates. DR. Umesh Laddi is a Professor and Chair at Channabasaveshwara Institute of Technology.	@dr_umesh	https://twitter.com/dr_umesh
Dr. Anthony Melvin Crasto	Organic chemistry international by DR Anthony Melvin Crasto. Anthony is helping organic chemists with websites, trying to get information at one place, easy picks for users. million hits on google, purely academic, non commercial and free from ads.	@amcrasto	https://twitter.com/amcrasto
Andrei Yudin	Amphoteros is a science blog maintained by Andrei Yudin. The purpose of this blog is to illuminate synthetic and chemical biology efforts in our lab at the University of Toronto. It is also a forum to discuss advances in science, past and present.	@amphoteros	https://mobile.twitter.com/amphoteros
James Ashenhurst	Hi, I'm James Ashenhurst. I founded Master Organic Chemistry to help understand the factors that make learning organic chemistry difficult. The mission of Master Organic Chemistry is: to provide student-centered articles, students success in organic chemistry, address the problems people have in learning organic chemistry.	@jamesashchem	https://twitter.com/jamesashchem?lang=en
Alex M. Clark	This blog by Alex M. Clark is about chemical information software for next generation computing environments. Alex is a Scientist & software engineer: founder of Molecular Materials Informatics, which specialises in next-generation cheminformatics tools, particularly for mobile.	@aclarkxyz	https://twitter.com/aclarkxyz
Antony John Williams, PhD, FRSC	My passion is connecting people to chemistry and I am known as the ChemConnector in the social network. recently posted topics open drug discovery for the Zika virus, Chemical Education, Open Chemistry Platform Update and Learnings.	@ChemConnector	https://twitter.com/chemconnector?lang=es
Scott Milam	Chemistry materials from Scott Milam, International Baccalaureate Chemistry HL teacher from Plymouth, MI IB blogging.	@Ibchemmilam	https://twitter.com/ibchemmilam?lang=es
Katherine Haxton	Katherine Haxton is a lecturer in Physical and Inorganic Chemistry at Keele University in the United Kingdom.	@kjhaxton	https://twitter.com/kjhaxton?lang=es

A nivel empresarial, en Junio de 2020 el 53% de los anunciantes a nivel mundial la utilizaron para marketing (Statista, 2020b), en comparación con un 76% que utilizaron Instagram o un 94% que utilizaron Facebook. Pero Twitter está lejos de otras compañías medido por sus ingresos por publicidad. En 2019 sus ingresos por anuncios eran de 2 990 millones de dólares en comparación con Google con 113 260 millones o Facebook con 69 670 millones.

En el ámbito educativo, el uso de Twitter tiene potencial para la creación de comunidades de aprendizaje formal e informal, el desarrollo del aprendizaje colaborativo y móvil y el pensamiento reflexivo (Ahmad Kharman Shah *et al.*, 2016), el poder ser una herramienta de facilitación del aprendizaje (Carpenter, 2015), de promoción de la participación (Hunter y Caraway, 2014), facilitar el aprendizaje de habilidades de aprendizaje como la autodisciplina y la autoexploración (Luo y Franklin, 2015) y la meta-cognición (Blaschke, 2014), y de herramienta para la evaluación del proceso de enseñanza en el ámbito universitario (Suárez *et al.*, 2015).

Su uso en este ámbito es difícil de ser medido y dependiendo del estudio, del momento y de la muestra estadística utilizada, varía desde un porcentaje pequeño hasta una tercera parte de los docentes (Haustein, 2019). Los docentes lo utilizan en conferencias para referirse a presentaciones y discusiones de las sesiones (Mishori *et al.*, 2014) y como red de intercambio de información y de contactos (Quan-Haase *et al.*, 2015; Mohammadi *et al.*, 2018). Asimismo, publican contenidos relacionados con su trabajo, buscan a otros docentes trabajando en líneas de investigación similares, siguen conversaciones relacionadas con su investigación y consiguen recomendaciones de sus trabajos (Noorden, 2014). En diversos casos concretos parece que los docentes lo están utilizando para su desarrollo profesional (Carpenter y Krutka, 2014; Krutka y Carpenter, 2016) más que para interaccionar con estudiantes.

Su uso es limitado por los docentes (Knight y Kaye, 2016), debido a ser un canal no oficial de comunicación académico y es poco utilizado para comunicaciones académicas (Noorden, 2014), con un gran porcentaje de sus tweets no relacionados con sus trabajos o con la academia (Bowman, 2015), ya que prefieren los canales de comunicación académicos tradicionales para la difusión de sus trabajos (Wilkinson y Weitkamp, 2013).

Algunas de las barreras detectadas en el uso de redes sociales parecen ser su percepción negativa y la falta de tiempo y habilidades para utilizarlas (Donelan, 2016) y además, en el caso de Twitter, el daño a su reputación académica (Sammer y Back, 2011), la reticencia a utilizar otra red social más si ya están utilizando otras (Shannon B Rinaldo *et al.*, 2011), sus dudas respecto a su potencial pedagógico (Krutka, 2014), su preocupación acerca del respeto de la privacidad de los contenidos por parte de los estudiantes (Seaman y Tinti-kane, 2013) y su preocupación con potenciales conflictos con estudiantes o familias (Burden, 2014).

Dentro de un curso educativo, los docentes utilizan Twitter para comunicarse con los estudiantes e informar sobre sus objetivos de evaluación, y en particular enviar información sobre el curso, sus deberes, actividades y fechas claves para a los alumnos (Tang y Hew, 2017). También se ha utilizado para proporcionar retroalimentación por el docente (Kassens-Noor, 2012) y para aumentar la experiencia universitaria conectando dentro y fuera del aula con sus compañeros y profesores (West *et al.*, 2015) con percepciones positivas por parte de los estudiantes sintiéndose más conectados con sus profesores.

En algunos casos los estudiantes hacen uso de Twitter para la recepción pasiva de información y no tanto para la participación en actividades educativas (Knight y Kaye, 2016), aunque en otros casos cuando los estudiantes iniciaban conversaciones soportadas por el docente, los estudiantes participaban de forma activa (Prestridge, 2014). Asimismo, si los docentes no animan a la interacción (Chen y Chen, 2012) o integran a los estudiantes en las actividades docentes planteadas con Twitter (Lowe y Laffey, 2011), la efectividad en el aprendizaje es limitada. Adicionalmente el convencimiento de lo que pueden extraer los estudiantes, su afinidad con la tecnología y la tolerancia al riesgo son variables que afectan a la adopción de Twitter en el aula (Eze *et al.*, 2013).

Aunque se ha explorado la mejora del aprendizaje informal y colaborativo con Twitter (Tur y Marín, 2014; Mercier *et al.*, 2015), se ha sugerido que esta colaboración puede alcanzar el nivel de comunidad de práctica para el aprendizaje (Paoli y Rooy, 2015), tiene impacto en el aprendizaje y la implicación de los estudiantes (Junco *et al.*, 2013) y permite superar los obstáculos de la participación en el contexto de aulas grandes (West *et al.*, 2015). Adicionalmente, existen estudios que concluyen que Twitter tiene un impacto positivo en la implicación y en las notas de los estudiantes (Junco *et al.*, 2011) y en la mejora de sus resultados (Junco *et al.*, 2013), y cuando se utiliza de forma guiada por el docente parece mejorar el aprendizaje en la focalización de tareas y habilidades de pensamiento (Luo, 2015), aunque existen discrepancias sobre la mejora en el rendimiento del aprendizaje de los estudiantes (Welch y Bonnan-White, 2012; Tang y Hew, 2017).

2.3 Síntesis del marco teórico

La imagen pública de la química se ha estudiado desde la perspectiva académica y social. En la académica, los estudios revelan que la percepción de los estudiantes es

generalmente negativa y alejada de la realidad debido a su percepción distorsionada, a una falta de claridad en su comunicación, a unos contenidos académicos alejados de sus motivaciones, a una falta de contexto social e histórico en los contenidos académicos y a la ausencia de relaciones entre ciencia, tecnología y sociedad dentro de su comunicación en las aulas. En la perspectiva social, la imagen pública parece haber heredado una percepción negativa debido a sus asociaciones históricas negativas y a una falta de comunicación eficiente por parte de los profesionales químicos, aún haberse realizado esfuerzos comunicativos para transmitir a la sociedad sobre los aspectos positivos en la calidad de vida de las personas. Los estudios más recientes sugieren un cambio hacia una percepción neutra o positiva, lo cual hace interesante hace interesante estudiar la imagen pública de la química para entender si este cambio se consolida, hecho que puede ayudar a incrementar el número de futuros estudiantes y reducir la falta de vocaciones científicas y de profesionales en el sector química.

Los métodos utilizados hasta ahora para estudiar la imagen pública de la química están basados en encuestas y análisis de documentos, métodos con potenciales sesgos en su diseño y auto-censura y no están diseñados para capturar opiniones e ideas espontáneas. Esto no es así en el caso de las redes sociales y en particular en Twitter donde sus usuarios pueden comunicarse intercambiando mensajes cortos o *tweets* en tiempo real, generando conversaciones, siguiendo los mensajes publicados por otros usuarios y por tanto, pudiendo expresar sus pensamientos, opiniones y emociones. Adicionalmente, Twitter es una red social relevante a nivel mundial por número de usuarios activos y por su penetración en la sociedad, utilizada en de forma transversal tanto por ciudadanos como por empresas y organizaciones y en la que sus contenidos públicos pueden ser recogidos y analizados para intentar captar los sentimientos que transmiten.

Por tanto, teniendo en cuenta todo lo anterior, parece interesante y adecuado para conocer mejor la imagen pública actual de la química estudiarla en Twitter, hecho que complementará la literatura científica existente, ya que no se ha encontrado ninguna publicación relativa a este tema. Con ello se espera aportar una nueva faceta a su estudio y comprensión que sea útil para científicos como para profesionales y otros grupos de influencia del sector.

3 Objetivos

Consecuencia de lo descrito en el marco teórico y enlazando con el último párrafo del capítulo anterior, los objetivos detallados que planteamos en esta investigación sobre la imagen pública de la química en Twitter son los siguientes:

1. ¿A qué se refieren los usuarios cuándo hablan de la química en Twitter?

   Para contestar a esta pregunta hemos considerado los *tweets* escritos en inglés que contienen las palabras clave “chemical”, “chem*” o “chemistry”, publicados en el periodo entre 01/01/2015 y 30/06/2015, con un máximo de 500 *tweets* por día.

2. ¿Qué sentimientos o estado del ánimo y emociones o alteración del ánimo se detectan en los contenidos de los *tweets* aceptados como relevantes en la imagen pública de la química?

   Para contestar a esta pregunta hemos considerado los *tweets* que presumiblemente contienen contenidos relacionados con la imagen pública de la química descrita en el marco teórico.

3. ¿Qué usuarios son relevantes en el conjunto de *tweets* aceptado y qué coincidencia tienen con organizaciones presumiblemente relevantes en la química?

   Para contestar a esta pregunta hemos considerado como válidos todos los *tweets* aceptados para el estudio del del objetivo primero.

Teniendo presente la limitación del número de tweets en inglés recogidos en un periodo acotado de tiempo, para dar respuesta a las tres preguntas formuladas hemos desarrollado una metodología propia, selección de técnicas ya existentes, que se explica en el capítulo siguiente.

4 Metodología

Presentamos una visión general de la metodología utilizada, representada en la Figura 4.1, que describiremos con detalle en los apartados de este capítulo. El resultado del análisis de sentimientos junto con sus procesos relacionados, que se muestran en la metodología, han sido publicados en el *Chemistry Education Research and Practice* (Guerris *et al.*, 2020) y citado en un trabajo de revisión de la quimiofobia (Rollini, 2020). El artículo completo puede consultarse en el Anexo 1. Resultados de pruebas de parte de la metodología fueron presentados en la comunicación "¿Qué significa la química hoy?. Una mirada a partir de Twitter" en las "VI Jornades sobre l'Ensenyament de la Física i la Química" celebradas en Barcelona los días 22, 23, 24 y 25 de Octubre de 2015 por la "Societat Catalana de Química".

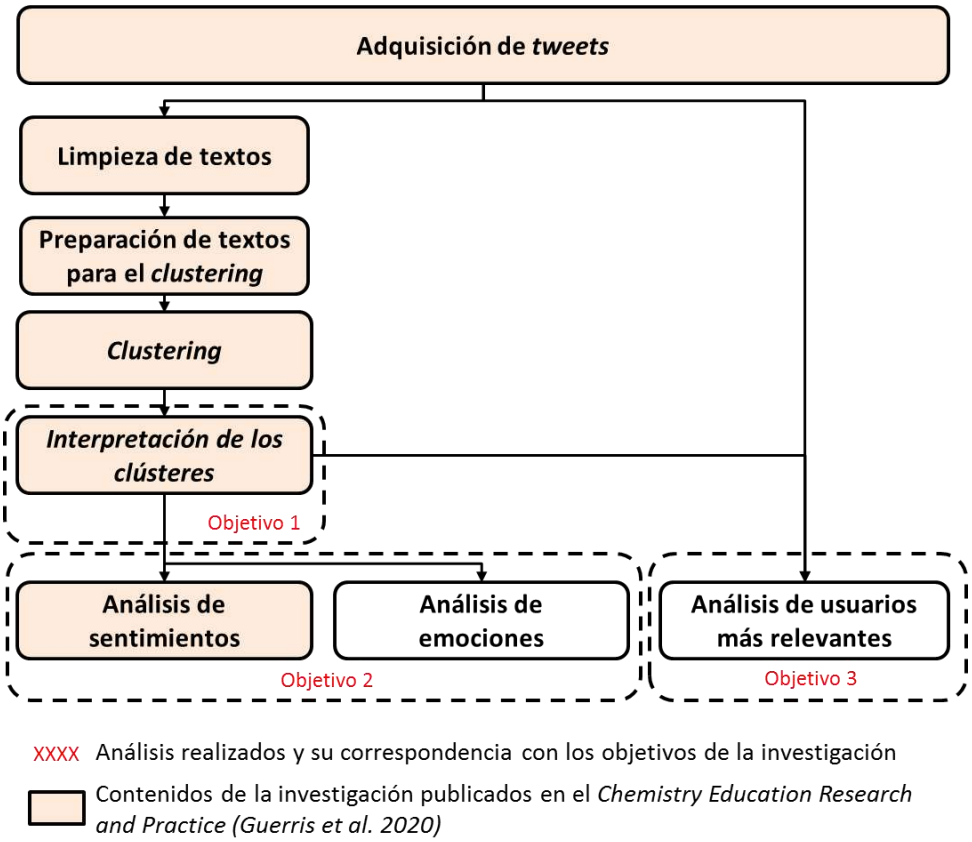


Figura 4.1 Esquema de la metodología utilizada en la investigación

A partir de la adquisición de *tweets* en Twitter mediante palabras clave, los procesos de limpieza y preparación de textos para posteriormente agruparlos (*clustering*), permiten su interpretación y de esta forma intentar dar respuesta al primer objetivo de la investigación sobre a qué se refieren los usuarios cuándo hablan de la química en Twitter. El análisis de sentimientos junto con el de emociones de los *tweets*

relacionados con la imagen pública de la química pretenda dar respuesta al segundo objetivo. El análisis de los usuarios e influencers de los *tweets* recopilados quiere dar respuesta al objetivo tercero de la investigación, sobre los usuarios relevantes en la producción y difusión de la imagen pública de la química. El código R desarrollado y utilizado en este estudio así como los ficheros de datos puede consultarse como documentación electrónica en la dirección web https://github.com/mguerris/Tesis-Doctoral.git y su descripción y organización puede consultarse en el Anexo 17.

4.1 Adquisición de *tweets*

El objetivo de la adquisición es la recopilación de los *tweets* escritos en inglés publicados en el periodo comprendido entre 01/01/2015 y 30/06/2015 y que contienen las palabras clave "chemical", "chem*" o "chemistry". Estas palabras han sido escogidas por considerarse no sesgadas con respecto a la imagen pública de la química siendo conscientes que otros *tweets* relacionados con la imagen pública no se recopilarán.

En el momento de la investigación, un *tweet* está formado por un máximo de 140 caracteres, correspondientes al texto del *tweet* más información adicional relacionada, como por ejemplo, el creador del *tweet*, si es un *retweet*[7] o la fecha y hora que se creó, entre otras. La información detallada de los campos de un *tweet* puede consultarse en la documentación pública (Twitter Inc., 2020b) proporcionada por Twitter.

La API (*Application Programming Interface*) de búsqueda de Twitter (Twitter Search API) versión 1.1 y el paquete `twitteR` (Gentry, 2015) de R (Team y others, 2013), como interfase a la API, permiten la recopilación de *tweets* y son los utilizados en esta investigación.

La API da acceso a la información que los usuarios han decidido compartir, y por tanto pública, así como a la información que no es pública pero que los usuarios han decidido autorizar a desarrolladores (Twitter Inc., 2016, 2020a). Para utilizar la API de Twitter es necesario ser usuario de Twitter. En esta investigación se utiliza la API para recoger los *tweets* que contienen las palabras clave mediante un usuario que no sigue

[7] *Retweet*: *tweet* propio o de otro usuario publicado nuevamente por un usuario para compartirlo con todos sus seguidores (https://help.twitter.com/es/using-twitter/retweet-faqs)

a otro ni que es seguido por otros ni de forma pública como autorizada. Por ese motivo, los *tweets* recopilados son solo *tweets* públicos.

La API devuelve los *tweets* publicados los últimos siete días si dentro del texto del *tweet* o bien dentro del nombre del usuario está la palabra clave de búsqueda. Por este motivo, la adquisición se ha realizado de forma periódica para obtener los *tweets* del día anterior entre los días 01/01/2015 y 30/06/2015 y se ha limitado como máximo a 500 *tweets* por día debido a los problemas técnicos surgidos en la captación cuando se supera este límite.

Un `dataframe` de R (ver ejemplo en Tabla 4.1.1) es una estructura de datos de dos dimensiones o tabla donde las filas corresponden a los diferentes casos, observaciones o individuos y las columnas a variables donde cada columna puede contener un tipo de dato diferente, por ejemplo, texto o números enteros o reales. El tipo de datos de una columna tiene que ser igual en toda la columna así como el número de datos en todas las columnas.

Tabla 4.1.1. Ejemplo de `dataframe`

Observación	Nombre	Edad	Fecha de ingreso
1	María	27	01/01/1985
2	José	30	23/06/1990
3	Ana	18	04/03/2005
4	Lucía	45	15/07/2009
5	Jordi	31	11/12/2019

Los *tweets* recogidos se han incluido dentro de un `dataframe` de R. El `dataframe` resultante contiene los campos descritos en la Tabla 4.1.2, donde cada campo es una columna del `dataframe` y cada fila corresponde a un *tweet*. Adicionalmente se presenta en esta tabla un ejemplo de *tweet.*

Tabla 4.1.2. Descripción del `dataframe` de *tweets*

Nombre del campo	Descripción	Ejemplo
Text	Texto del *tweet*	@chem_cake Ð• Ðµ.
screenName	Nombre del usuario en pantalla de Twitter que subió el *tweet*	_born_wild_
Id	Identificador del *tweet*	550803497651019000
replyToSN	Nombre del usuario en pantalla al que el *tweet* responde	chem_cake
replyToUID	Identificador del usuario al que el *tweet* responde	1872560862
replyToSID	Identificador del *tweet* al que el *tweet* responde	550802825799008000

Nombre del campo	Descripción	Ejemplo
statusSource	Dirección web del tipo de instancia de Twitter dónde se creó el *tweet* (por ejemplo, un dispositivo iphone, Android, un web client, …)	Twitter Web Client
Created	Fecha y hora que el *tweet* se creó con el formato "AAAA-MM-DD HH:MM:SS"	01/01/2015 23:59
Truncated	Valor lógico de si el *tweet* está truncado o no resultado de exceder el texto los 140 caracteres	FALSE
favorited	Valor lógico de si el *tweet* ha sido marcado como favorito	FALSE
favoriteCount	El número de veces que el *tweet* ha sido marcado como favorito	0
is*Retweet*	Valor lógico de si el *tweet* es un *Retweet*	FALSE
*retweet*ed	Valor lógico de si el *tweet* ha sido retuiteado	FALSE
*retweet*Count	El número de veces que el *tweet* ha sido retuiteado	0
longitude	Valor de longitud de la posición GPS del dispositivo donde se creó el *tweet*	0
latitude	Valor de latitud de la posición GPS del dispositivo donde se creó el *tweet*	0

Los *tweets* recogidos, que son públicos, siguen las políticas de privacidad de Twitter que no considera una violación de información privada (Twitter Inc., 2016). Adicionalmente no existe ninguna interacción con los usuarios y por tanto, no consideramos que esta sea una investigación en sujetos humanos atendiendo a las guías más recientes (University of California, 2019; University of Wisconsin, 2019)**.** Por esta razón no consideramos que sea necesaria la revisión y confirmación por parte de un comité de ética para la realización de la investigación.

De cara a facilitar la lectura de esta investigación y en los posteriores apartados y capítulos del documento, vamos a referirnos a un *tweet* como la cadena de caracteres que contiene el texto del *tweet*.

4.2 Limpieza de textos

Un *tweet* está escrito en el lenguaje natural del usuario. Puede contener palabras en diversos idiomas o alfabetos, abreviaturas formales e informales, signos de puntuación, *stopwords*[8], caracteres que no estén entre las letras "a" a la "z" o sus respectivas mayúsculas y símbolos (ver Tabla 4.2.1). Adicionalmente los *tweets* pueden estar duplicados o haber sido retuiteados o reenviados.

[8] *Stopwords*: palabras vacías que son términos frecuentes de un idioma y que no aportan significado por sí solos como por ejemplo adverbios o pronombres

Tabla 4.2.1. Descripción de símbolos en los *tweets*

Símbolos	Descripción	Ejemplo
Marcas HMTL[9]	Contenido entre los símbolos "<" y ">"	Triptych Large wall art Painting on canvas TC02 <U+043E><U+0442>
Menciones	Palabras que contienen el símbolo "@" y que incluyen nombres de cuentas de Twitter y correos electrónicos	@hannah_molitor: Ap chem got me feeling some type of way
Hashtags	Palabras que empiezan con el símbolo #	Todoist task completed #finished - meet with new chem student
Emoticonos	Secuencia de caracteres ASCII que expresa una emoción	it's good to be back :-)
Emojis	Ideograma que puede ser utilizado como un emoticono en una conversación o mensaje	Chem is easy <U+263X>
Direcciones URL ("Uniform Resource Locator")	Palabras que empiezan con la sintaxis "http(s)://…"	Today stats: One follower, No unfollowers via http://t.co/57SzYVgY8Y

El objetivo de la limpieza consiste en eliminar de los *tweets* información no relevante para su posterior análisis, así como excluir los *retweets*, los duplicados y los vacíos. La información no relevante incluye las expresiones, símbolos y palabras que no son de interés en el estudio y los *tweets* vacíos. Para conseguir uno de los objetivos de esta investigación acerca de a qué se refieren sobre la química los usuarios de Twitter, se descartan los *retweets* y los duplicados para evitar que *tweets* de una temática determinada con mayor frecuencia absoluta con respecto al volumen total de *tweets* enmascarasen a *tweets* de otras temáticas con menor frecuencia.

De forma esquemática el proceso de limpieza de textos sigue el representado en la Figura 4.2 y descrito a lo largo de este apartado.

[9] HTML: siglas en inglés de *HyperText Markup Language* ('lenguaje de marcas de hipertexto'), hace referencia al lenguaje de marcado para la elaboración de páginas web (https://dictionary.cambridge.org/dictionary/english/html)

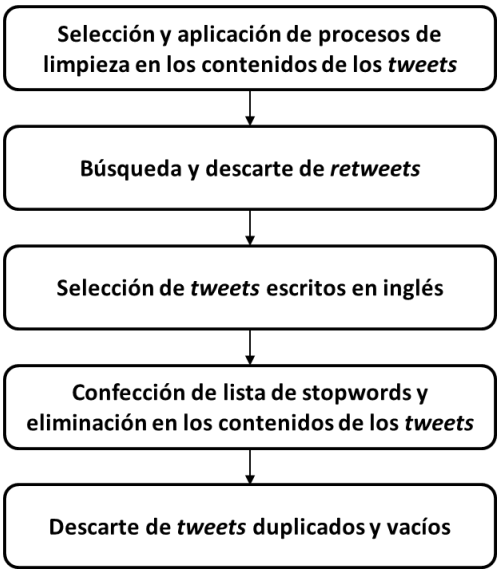


Figura 4.2 Esquema descriptivo del proceso de limpieza de textos utilizado en la investigación

Para decidir las operaciones de limpieza, realizamos un análisis descriptivo de los símbolos en los *tweets* e inspeccionamos un subconjunto de los *tweets* con *hashtags* y menciones. Las marcas HTML, menciones, *hashtags* y direcciones URL en los *tweets* los localizamos mediante el paquete `stringr` (Wickham, 2019) de R, que permite acceder a funciones de manejo de cadenas de texto. Los emoticonos mediante la lista proporcionada por el paquete `qdapDictionaries` (Rinker, 2013) de R, que provee de diccionarios y listas de palabras, como por ejemplo, una lista de adverbios o de verbos de acción. Los emojis los detectamos mediante la lista proporcionada por el paquete `lexicon` (Wickham, 2019) de R, que proporciona diccionarios para análisis de textos.

Las operaciones de limpieza escogidas sobre todos los *tweets* que no interesaba estudiar, así como un ejemplo de cada operación en un *tweet*, se describen en la Tabla 4.2.2.

Tabla 4.2.2. Descripción de las operaciones de limpieza sobre los *tweets*

Operación de limpieza	*Tweet* original	*Tweet* resultante de la operación
Reemplazo por un espacio en blanco del contenido dentro de las marcas HMTL	"ap chem; ap bio test tmr I need blessings. <ed><U+00A0><U+00BD><ed><U+00B8><U+00AD>"	"ap chem & ap bio test tmr I need blessings. < >< >< >< >< >< >"
Reemplazo de las menciones por espacios en blanco	"@bigsullyt what Chem labs? Lol"	" what Chem labs? Lol"
Reemplazo por un espacio en blanco de las direcciones URL	"I could spend ages on this virtual tour of #Oxford http://t.co/csaS0EGmw"	"I could spend ages on this virtual tour of #Oxford "

Operación de limpieza	*Tweet* original	*Tweet* resultante de la operación
Reemplazo de los signos de puntuación por un espacio en blanco	"inactive bc failing chem!!!!!!!!!!!!!!!!!!!!!!!! luv lyfe"	"inactive bc failing chem luv lyfe"
Reemplazo de los caracteres que no estén entre las letras "a" a la "z" o sus respectivas mayúsculas por un espacio en blanco	"Now gotta do my english work and then chem and im done í ½í¸´í ½í²• í ½í²†"	"Now gotta do my english work and then chem and im done"
Reemplazo de letras mayúsculas por sus respectivas minúsculas	"@harrybuxton7 Xavi doesn't have chem with di Maria now"	"@harrybuxton7 xavi doesn't have chem with di maria now"
Reemplazo de las palabras de longitud uno o dos por un espacio en blanco	"It would mean a lot if someone could help me or possibly do my chem lab for me"	"would mean lot someone could help possibly chem lab for"
Reemplazo de los espacios múltiples en cualquier posición de un tweet un por un solo espacio en blanco y eliminación de los espacios al principio y al final de un tweet	"inactive bc failing chem luv lyfe"	"inactive bc failing chem luv lyfe"

Algunas de estas operaciones de limpieza pueden presentar pérdida de contenido y en concreto, inconvenientes debido a la modificación del contenido de los *tweets* como puede consultarse en la Tabla 4.2.3.

Tabla 4.2.3. Potenciales inconvenientes de las operaciones de limpieza del contenido de los *tweets*

Operación de limpieza	Potenciales inconvenientes
Reemplazo por un espacio en blanco del contenido dentro de las marcas HMTL	Pérdida de contenido que no sea una marca HTML
Reemplazo de las menciones por espacios en blanco	Pérdida de contenidos de menciones que no se correspondan con nombres de cuentas de Twitter o correos electrónicos
Reemplazo por un espacio en blanco de las direcciones URL	Pérdida de direcciones URL relevantes para diferentes usuarios
Reemplazo de los signos de puntuación por un espacio en blanco	Pérdida de contenido de palabras compuestas o fórmulas químicas
Reemplazo de los caracteres que no estén entre las letras "a" a la "z" o sus respectivas mayúsculas por un espacio en blanco	Pérdida de contenido de símbolos o fórmulas químicas
Reemplazo de letras mayúsculas por sus respectivas minúsculas	
Reemplazo de las palabras de longitud uno o dos por un espacio en blanco	Pérdida de contenido de símbolos de elementos químicos
Reemplazo de los espacios múltiples en cualquier posición de un tweet un por un solo espacio en blanco y eliminación de los espacios al principio y al final de un tweet	

Un *wordcloud* es una representación visual de las palabras que conforman un texto, donde el tamaño es mayor para las palabras que aparecen con mayor frecuencia y es una técnica considerada adecuada para obtener una idea global de los contenidos de

un texto (Kuo *et al.*, 2007; Cidell, 2010) y evaluarlo (Gottron, 2009; Cui *et al.*, 2010; DePaolo y Wilkinson, 2014; Heimerl *et al.*, 2014; Horn *et al.*, 2017; Smith *et al.*, 2017).

Aunque estas operaciones presentan ventajas en la interpretación del contenido del texto de los *tweets.* Si a modo de ejemplo representamos las palabras de los contenidos de los *tweets* de la Tabla 4.2.4, mediante *wordclouds* antes y después de las operaciones de limpieza, obtenemos la Figura 4.3 y Figura 4.4. Podemos observar en estas figuras como la interpretación del contenido de las palabras en los *tweets* antes de las operaciones de limpieza es más complejo que después de estas.

Tabla 4.2.4. Ejemplo de diversos *tweets*

RT @lucas_bohannon: @holly_becker WHEN YOU LOOKED AT DANNY IN CHEM OMG<ed><U+00A0><U+00BD><ed>
<U+6B21><U+306F>…
My grade on the Chem test if we still have school tommorow #closeFCPS http://t.co/Lj8OORQmwc
RT @sernaum: Ð'Ð¾Ð¹Ð½Ð¾Ð²Ð¸Ñ‡: Â«Ð˜Ð½Ð¾Ð³Ð´Ð° Ñ• Ð¼ÐµÑ€Ñ‚Ð²Ñ‹Ðµ Ð¾Ð´Ð½Ð¾Ð³Ð¾ Ñ‡ÐµÐ»Ð¾Ð²ÐµÐºÐ°
RT @kaylajodell: Walking out of o chem.. @colleen_penny95 @nathanepstein17 http://t.co/ZTefjKWg2t
Chem: C-C chemokine receptor type 2 (CCR2) signaling protects neonatal male mice with hypoxic-ischemic
Been working on this chem lab for 3 hours not even close to done :))))

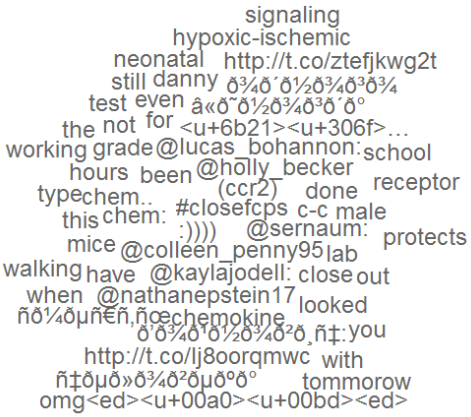


Figura 4.3 *Wordcloud* de palabras de los *tweets* de la Tabla 4.2.4 antes de las operaciones de limpieza

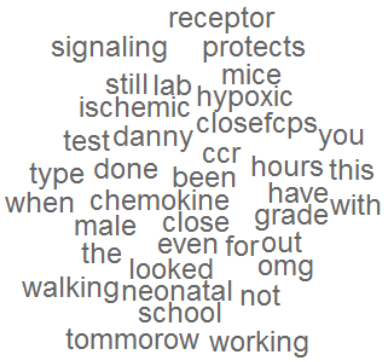


Figura 4.4 *Wordcloud* de palabras de los *tweets* de la Tabla 4.2.4 después de las operaciones de limpieza

Con los *tweets* limpios de contenido, localizamos aquellos considerados como *retweets*. Los *retweets* según Twitter son aquellos que el valor lógico del campo isRetweet es TRUE o bien que empiezan con la expresión “RT” en el campo text. Extraemos los *retweets* del conjunto de *tweets* y se crea un nuevo `dataframe` con el que aplicar nuevas operaciones. Esta operación podría haberse realizado tanto antes de las operaciones de limpieza de contenido del texto de un *tweet* como a posteriori. No afecta al resultado final de todas las operaciones de limpieza.

Con el nuevo conjunto de *tweets* utilizamos el paquete `textcat` (Hornik *et al.*, 2013) de R para detectar los escritos en inglés. Este paquete permite clasificar cadenas de textos en 74 idiomas, obteniendo un idioma para cada cadena de texto. Mediante el uso de este paquete obtenemos para cada *tweet* un idioma y seleccionamos aquellos con el idioma *english*. El resto de *tweets* fueron descartados.

Del conjunto aceptado, se eliminan las *stopwords*. Mostramos las utilizadas en la Tabla 4.2.5.

Tabla 4.2.5. Lista ordenada de *stopwords* utilizada

about	doesn	how	own	very
above	doing	into	same	was
after	don	Isn	shan	wasn
again	down	its	she	were
against	during	itself	should	weren
all	each	*just*	shouldn	what
and	few	Let	some	when
any	for	more	such	where
are	from	most	than	which
aren	further	*much*	that	while
because	*get*	mustn	the	who
been	*got*	myself	their	whom
before	had	nor	theirs	why
being	hadn	Not	them	*will*
below	has	*now*	themselves	with
between	hasn	off	then	won
both	have	once	there	would
but	haven	only	these	wouldn
can	having	other	they	you
cannot	her	ought	this	your
could	here	our	those	yours
couldn	hers	ours	through	yourself
did	herself	ourselves	too	yourselves
didn	him	out	under	
does	his	over	until	

Para confeccionar esta lista, se parte de la lista de *stopwords* proporcionada por el paquete `tm` (Feinerer *et al.*, 2008) de R, que permite realizar operaciones de minería de textos o extracción de información mediante el análisis de textos. A la lista del paquete `tm` se le aplica las mismas operaciones de limpieza que a los contenidos de los *tweets* y se le añade las palabras "just", "now", "got", "will", "get", "much" y "can", que no aportan contenido semántico al texto de un *tweet.*

Estas palabras fueron detectadas en pruebas previas mediante la representación de las palabras de los *tweets* con un *wordcloud* (ver Figura 4.5).

Figura 4.5 *Wordcloud* de palabras de los *tweets* limpios sin las palabras clave "chemistry", "chemical" y "chem"

Finalmente los *tweets* duplicados y vacíos se excluyen del conjunto de *tweets.* Los *tweets* duplicados se detectaron mediante la función `duplicated` del paquete base de R. Los *tweets* vacíos se detectaron mediante la comparación de su contenido con la expresión "".

De cara a facilitar la lectura de esta investigación y en los posteriores apartados y capítulos del documento, vamos a referirnos a un *tweet* como la cadena de caracteres que contiene el texto del *tweet* después de haber aplicado las operaciones de limpieza mencionadas.

4.3 Preparación de textos para el clustering

El objetivo de la preparación de textos consiste en aplicar un conjunto de procesos sobre los *tweets* para obtener una estructura de datos sobre la que se puede aplicar un método de *clustering.*

De forma esquemática el proceso de preparación de textos sigue el representado en la Figura 4.6 y descrito a lo largo de este apartado.

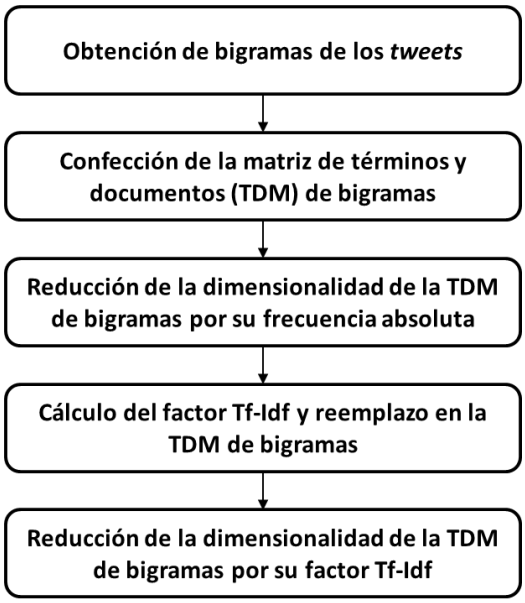


Figura 4.6 Esquema descriptivo del proceso de preparación de textos utilizado en la investigación

Un n-grama es un conjunto de *n* palabras consecutivas en un documento siendo *n* un número natural. La denominación habitual es de unigrama, bigrama y trigrama desde *n* igual a 1 hasta *n* igual a 3 respectivamente. Cuando *n* es igual o superior a cuatro la denominación es n-grama sustituyendo el valor de *n* por su número. Por ejemplo un n-grama cuando *n* es igual a cuatro se denomina 4-grama.

En la clasificación de textos, un documento se representa comúnmente por su lista de palabras (“bag of words”) (Bekkerman y Allan, 2004). Esta lista es un vector en el que cada posición corresponde a una palabra o unigrama. Este vector puede contener en cada posición un bigrama en lugar de un unigrama, denominándose “bag of bigrams”, siendo el uso de unigramas o bigramas extendido en la clasificación de textos.

Los *wordclouds* de bigramas son los que producen mejores resultados en su interpretación (Kaptein *et al.*, 2010) debido a la información adicional de contexto que proporcionan. Asimilando un *tweet* a un documento, en esta investigación decidimos representar un *tweet* mediante su lista de bigramas por su mejor interpretación en los *wordclouds*.

Una matriz de documentos y términos (“Term-document matrix”) de bigramas o TDM de bigramas es una matriz donde cada fila *i* corresponde a uno de los bigramas existentes en los documentos analizados y donde cada columna *j* corresponde a un documento. El contenido de la celda *ij* de la matriz contiene el número de veces o

frecuencia absoluta que el bigrama *i* aparece en el documento *j*. En nuestro caso un documento equivale a un *tweet* con su contenido previamente depurado. Podemos observar el ejemplo de obtención de bigramas y de la TDM de bigramas a partir de varios *tweets* depurados en la Tabla 4.3.1.

Tabla 4.3.1. Ejemplo de *tweets* depurados, sus bigramas y la TDM de bigramas

<table>
<tr><td>Tweets depurados</td><td>Tweet 1: "dreading going back college especially monday chem lab"
Tweet 2: "baking cookies chem lab accidentally used self rising flour omg"</td></tr>
<tr><td>Bigramas de los tweets</td><td>Tweet 1: "dreading going", "going back", "back college", "college especially", "especially monday", "monday chem", "chem lab"
Tweet 2: “baking cookies”, “cookies chem”, “chem lab”, “lab accidentally”, “accidentally used”, “used self”, “self rising”, “rising flour”, “flour omg”</td></tr>
<tr><td>TDM de bigramas</td><td>
<table>
<tr><td></td><td>Tweet 1</td><td>Tweet 2</td></tr>
<tr><td>“dreading going”</td><td>1</td><td>0</td></tr>
<tr><td>“going back”</td><td>1</td><td>0</td></tr>
<tr><td>“back college”</td><td>1</td><td>0</td></tr>
<tr><td>“college especially”</td><td>1</td><td>0</td></tr>
<tr><td>“especially monday”</td><td>1</td><td>0</td></tr>
<tr><td>“monday chem”</td><td>1</td><td>0</td></tr>
<tr><td>“chem lab”</td><td>1</td><td>1</td></tr>
<tr><td>“backing cookies”</td><td>0</td><td>1</td></tr>
<tr><td>“cookies chem”</td><td>0</td><td>1</td></tr>
<tr><td>“lab accidentally”</td><td>0</td><td>1</td></tr>
<tr><td>“accidentally used”</td><td>0</td><td>1</td></tr>
<tr><td>“used self”</td><td>0</td><td>1</td></tr>
<tr><td>“self rising”</td><td>0</td><td>1</td></tr>
<tr><td>“rising flour”</td><td>0</td><td>1</td></tr>
<tr><td>“flour omg2</td><td>0</td><td>1</td></tr>
</table>
</td></tr>
</table>

Los bigramas se han obtenido mediante el paquete `RWeka` (Hall *et al.*, 2009) de R. Este paquete es una interfase a Weka. Weka es una colección de algoritmos de “machine learning” para minería de datos con herramientas de procesamiento de textos entre otras y desarrollados en Java. La TDM de bigramas se ha obtenido mediante el paquete de `tm` (Feinerer *et al.*, 2008) de R.

Una de las dificultades en el clustering de textos es la alta dimensionalidad del espacio de características (Yang y Pedersen, 1997; Feldman y Sanger, 2007) debido a la gran cantidad de términos en relación al número de documentos. Este efecto aumenta en el caso de la utilización de bigramas de forma que este espacio sea superior que en el caso de unigramas. Adicionalmente, como cada documento es un *tweet* y además su contenido de texto es corto, la TDM está repleta de ceros en la mayoría de posiciones, excepto en aquellos bigramas que aparecen en cada uno los *tweets*. La TDM es por tanto una matriz cuasi vacía, hecho que afecta al proceso de cálculo de la similitud entre dos documentos. La alta dimensionalidad junto con la dispersión de la TDM hace que el clustering sea computacionalmente costoso afectando a su rendimiento (Aggarwal y Yu, 2000) y siendo un problema mayor en el caso de los *tweets* (Aggarwal

y Zhai, 2012). La reducción de la dimensionalidad, por tanto, ayuda a mejorar el rendimiento del clustering.

Existen diferentes grupos de métodos para reducir la dimensionalidad, entre ellos los de selección automática de características (Yang y Pedersen, 1997; Alelyani Salem *et al.*, 2013), en los que se eliminan los términos que no aportan información atendiendo a estadísticas relacionadas con los documentos y sus contenidos.

Uno de los métodos y que utilizamos en nuestra investigación es el umbral de frecuencia de un término en el grupo de documentos (Luhn, 1957). Se eliminan los términos que tienen una frecuencia total, suma de la frecuencia de los términos en el conjunto de los documentos, por debajo de un cierto umbral. Se asume que los términos con menor frecuencia no contienen información relevante o no son influyentes en el proceso de clustering (Yang y Pedersen, 1997; L. Liu *et al.*, 2005). Es un método simple y su complejidad computacional es lineal con el número de documentos, aunque puede impactar negativamente en el clustering si los términos menos frecuentes contienen cierto grado de información relevante.

Para aplicar el método de umbral de frecuencia de un término en el grupo de documentos, la dimensionalidad de la TDM de bigramas se reduce mediante la eliminación de los bigramas con una frecuencia absoluta total, suma de la frecuencia absoluta del bigrama en todos los *tweets*, inferior a 30. Se eliminan también de la TDM de bigramas los *tweets* que quedaron vacíos de bigramas.

Adicionalmente reemplazamos las frecuencias absolutas de los bigramas en los *tweets* por el factor Tf-Idf ("Term frequency-Inverse document frequency") (Salton y Buckley, 1988) de forma normalizada. Este factor, aplicable a n-gramas y documentos, tiene las siguientes características:

- Tf-Idf refleja la importancia de un n-grama en un documento dentro de una colección de documentos y se obtiene mediante la multiplicación del factor Tf por el factor Idf. Discrimina los n-gramas más relevantes que podrían quedar relegados frente a los n-gramas más frecuentes pero genéricos si solo se utilizase su frecuencia absoluta.
- Tf o frecuencia de término tiene en cuenta la frecuencia en que un n-grama aparece en un documento. Normalizado, el valor Tf es la frecuencia del n-grama en el documento dividido por el total de frecuencias de los n-gramas que aparecen en un documento.

- Idf o frecuencia inversa de documento tiene en cuenta en cuántos documentos aparece un n-grama siendo mayor este valor en cuantos menos documentos aparezca. El valor Idf pretende discriminar n-gramas genéricos o poco relevantes ya que los n-gramas más genéricos que aparecen en muchos documentos tendrán un valor Idf bajo. Se calcula como el logaritmo decimal o en base 2 del número total de documentos de la colección dividido por el número de documentos en que el n-grama aparece en la colección y a este resultado se le suma el factor correctivo +1.

La fórmula de cálculo del factor Tf-Idf aplicada a nuestra investigación es la expresada en la Ecuación 4.3.1.

$$tfidf_{i,j} = tf_{i,j} \cdot idf_i = \frac{n_{i,j}}{\sum_k n_{k,j}} \cdot \log_2 \frac{|D|}{|\{d \mid t_i \in d\}|}$$

Ecuación 4.3.1 Fórmula de cálculo del factor Tf-Idf

donde:

$tf_{i,j}$ es el factor Tf,

idf_i es el factor Idf.

$n_{i,j}$ es la frecuencia de aparición del bigrama t_i en el *tweet* d_j,

$\sum_k n_{k,j}$ es el total de frecuencias de los bigramas que aparecen en el bigrama d_j,

$|D|$ es el número total de *tweets* y

$|\{d \mid t_i \in d\}|$ es el número de *tweets* donde el bigrama t_i aparece.

Una vez calculado el factor Tf-Idf en cada celda TDM, se eliminaron aquellos bigramas de la matriz en los que el factor era cero en todos los documentos, ya que este factor se utiliza como un método de selección automática de bigramas utilizado para reducir la dimensionalidad de la TDM (L. Liu *et al.*, 2005; Aggarwal y Zhai, 2012) y mejorar el rendimiento del clustering (Aggarwal y Yu, 2000) eliminando los bigramas que no aparecen en ningún documento y aumentando la velocidad de cálculo.

4.4 Clustering

El objetivo del proceso de *clustering* es la obtención de clústeres de *tweets* de forma que cada clúster agrupe a *tweets* parecidos en función de su contenido temático y a la vez cualquier clúster sea lo más diferente respecto a los demás. De esta forma los clústeres podrán ser interpretados más fácilmente en procesos posteriores.

De forma esquemática el clustering sigue el proceso representado en la Figura 4.7 y descrito a lo largo de este apartado

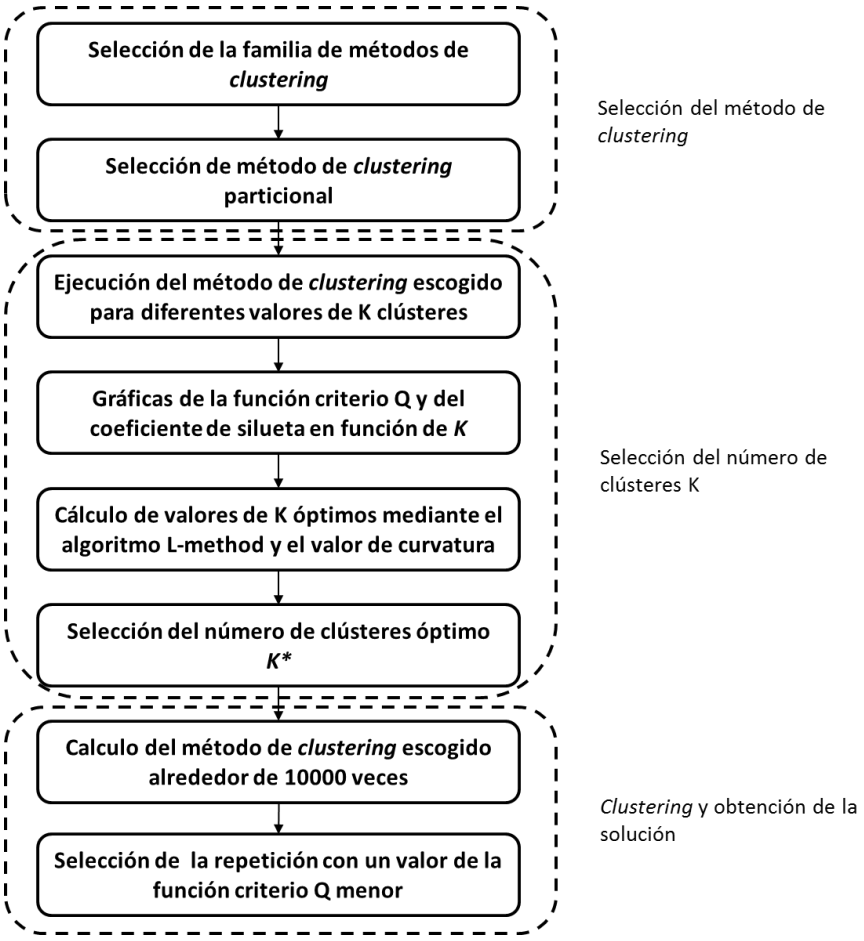


Figura 4.7 Esquema descriptivo del proceso de *clustering* utilizado en la investigación.

Los métodos de clustering se pueden dividir en dos tipos (Jain, 2010), los jerárquicos y los particionales. Los jerárquicos intentan formar clústeres de forma aglomerativa, asignando cada individuo a un clúster y de forma recursiva mezclando los pares de clústeres más similares, o de forma divisiva, partiendo de todos los individuos en un

clúster y recursivamente dividiendo cada clúster en clústeres más pequeños. Los métodos jerárquicos consiguen obtener una jerarquía en los clústeres.

Los métodos particionales se basan en organizar los individuos a agrupar en un número determinado de particiones o clústeres de forma simultánea, sin obtener ningún tipo de jerarquía en los clústeres y su uso se ha extendido gracias a su eficiencia computacional. Los individuos se agrupan en función de su similitud siendo alta dentro un clúster y baja con respecto a los individuos de otros clústeres. Los clústeres pueden diferir en el número de individuos y cada clúster tiene un representante de los individuos que forman aquel clúster denominado prototipo de forma genérica o centroide o medioide en función del método de cálculo utilizado para encontrar el prototipo.

En esta investigación hemos seleccionado como método de *clustering* el *spherical k-means*. Este es uno de los métodos particionales generalmente utilizado en el *clustering* de documentos y una versión de otro método ampliamente utilizado en *clustering*, el *k-means* (Hartigan y Wong, 1979). El *spherical k-means* es adecuado en el *clustering* de textos debido a que los prototipos de los clústeres que se obtienen tienen información semántica representativa de los clústeres, explota la escasez del número de palabras similares en un texto, puede ser paralelizado a nivel computacional, y alcanza rápidamente un mínimo o máximo local (Dhillon y Modha, 2001) siendo uno de los más rápidos para el clustering de textos (Zhong, 2005).

En el caso del *clustering* de textos un individuo equivale a un documento y en nuestra investigación, un documento corresponde a un *tweet* representado por su respectivo vector de bigramas. El vector de bigramas de un *tweet* corresponde a la columna de la TDM de bigramas asociado a ese *tweet* y descrita en el apartado 4.3.

La similitud de dos documentos se calcula utilizando la distancia k-dimensional de los dos vectores que los representan, de forma que, a menor distancia, mayor similitud entre los dos documentos. Existen diferentes definiciones de distancia, siendo la más comúnmente utilizada en la clasificación de documentos el coseno del ángulo que forman sus vectores (Feldman y Sanger, 2007) para evitar la sobrerrepresentación de documentos grandes y mitigar el efecto de diferentes longitudes de documentos (Dhillon y Modha, 2001).

El método estándar del *spherical k-means* utilizado para agrupar documentos tiene como objetivo maximizar el coseno del ángulo que forman los vectores de los documentos asignados a cada clúster y los prototipos de los clústeres, es decir, que el ángulo que formen sea el menor posible. En el caso que un documento solo puede estar asignado a un clúster, este objetivo está representado con la función criterio Q mostrada en la Ecuación 4.4.1.

$$Q = \sum_{i,j} \mu_{ij}\left(1 - \cos(x_i, p_j)\right)$$

$$\mu_{ij} = \begin{cases} 1, si\ c(i) = j \\ 0, si\ c(i) \neq j \end{cases}$$

Ecuación 4.4.1 Función criterio Q del método de *clustering spherical k-means*

donde:

x_i representa el vector de características del documento *i*,

p_j el prototipo del clúster *j*,

c(i) es la función de asignación del documento *i* a uno de los 1 a *K* grupos o clústeres,

μ_{ij} es la función de pertenencia de un documento al clúster tal que su valor es 0 si el documento x_i no pertenece al clúster *j* y 1 si pertenece,

i toma los valores de 1 a *n* documentos y

j de 1 a *K* grupos o clústeres.

El valor de la función criterio Q disminuye a medida que aumenta el número de clústeres *K*, ya que existen más prototipos y por tanto, un mayor número de documentos estarán más cercanos a los prototipos. En el caso que tuviéramos un número de prototipos igual al número de documentos y que coincidiesen con los documentos, el valor de la función criterio Q sería cero.

En nuestro caso un documento equivale a un *tweet* y queremos conseguir que habiendo fijado un número determinado de clústeres *K*, cada *tweet* esté asignado a un clúster y no quede ningún *tweet* sin asignar, de forma que cada clúster contenga

tweets con una mínima distancia y a la vez máxima con los *tweets* del resto de los *K* clústeres.

Representados los *tweets* mediante la TDM de bigramas donde cada columna de la matriz representa un *tweet* y es un vector de longitud igual al número de bigramas de todos los *tweets* en el que cada posición del vector contiene el factor Tf-Idf del bigrama en ese *tweet*, dos *tweets* serán más "parecidos" si el valor del coseno del ángulo de los dos vectores que los representan es más cercano a +1.

La implementación que hemos utilizado es la del paquete `skmeans` (Hornik *et al.*, 2012) de R. Este paquete permite calcular el método de *spherical k-means* e intenta buscar como solución aquellos prototipos y valores de la función de pertenencia μ_{ij} que minimicen la función criterio Q (Ecuación 4.4.1) partiendo de unos valores de la función de pertenencia iniciales aleatorios. Esta optimización la intenta conseguir mediante una técnica propia y mejorada de optimización con un algoritmo genético desarrollado a partir del aplicado (Krishna y Murty, 1999) en la técnica particional del *k-means*.

Un algoritmo genético (Holland, 1992; Whitley, 1994) es una técnica que se utiliza por ejemplo en problemas de optimización. Inspirado en la teoría de la evolución de Darwin, se genera una población inicial, usualmente aleatoria, de soluciones del problema a resolver. Estas soluciones evolucionan de forma similar a lo que sucede en la naturaleza mediante operadores de cruce y mutación con el objetivo de mejorar la siguiente población con respecto a la anterior. Cada paso en el que se aplican los operadores de cruce y mutación se denomina iteración. La bondad de cada solución y de la población se mide con una función de aptitud o salud, que está relacionada con la función a optimizar si el objetivo es buscar un mínimo o máximo de una función. La población inicial se hace evolucionar un número determinado de iteraciones hasta que o bien la función de aptitud global de la población casi no se ve modificada por el paso de las iteraciones o bien hasta un número limitado de iteraciones.

En la implementación del paquete `skmeans` de R fijamos el parámetro "genetic" que es el que por defecto a través del algoritmo genético utilizado permite obtener clústeres no superpuestos, es decir, cada *tweet* solo puede pertenecer a un clúster. En nuestro caso una solución corresponde a la asignación de cada *tweet* de la TDM de bigramas a uno de los *K* clústeres. Siendo los algoritmos genéticos estocásticos, fijamos una población de 50 potenciales soluciones iniciales aleatorias en lugar de las seis que

proporciona por defecto el algoritmo genético implementado en el paquete. Aumentando el número de soluciones que conforman la población se reduce la probabilidad de convergencia rápida de la población a una solución durante las primeras iteraciones del algoritmo.

Otro de los parámetros necesarios en el método *spherical k-means* es la selección del número de clústeres. Buscamos un número de clústeres óptimo K^* que optimice la función criterio (Zhong, 2005), método utilizado comúnmente. Para obtenerlo, calculamos la implementación del método desde K=2, el mínimo número de clústeres, hasta K=285, un número considerado suficientemente alto de clústeres para poder detectar el número de clústeres óptimo K^*. Repetimos el cálculo 50 veces para cada valor de K debido a la naturaleza estocástica de la implementación de este método y para cada K seleccionamos la repetición que minimiza la función criterio Q. El valor de la función criterio Q es proporcionado por la implementación del *spherical k-means.*

Debido a la carga computacional y de memoria necesarias para poder realizar los cálculos, paralelizamos la obtención de los resultados mediantes el paquete `parallel` (R Core Team, 2018) de R y los ejecutamos en el servidor de cálculo ABACO proporcionado por IQS. El paquete `parallel` proporciona un conjunto de funciones para poder ejecutar cálculos en paralelo tanto en procesadores multinúcleo como en computadores diferentes. El servidor ABACO está compuesto por cinco nodos nombrados del c0 al c4. Utilizamos el nodo c4 que tiene una configuración 32 unidades centrales de procesamiento (CPUs) y 96 Gigabytes (GB) de memoria. Los cálculos se realizan entre los meses de Abril y Julio de 2018 e incluyeron ajustes del código R, pruebas del código, pruebas piloto de los cálculos en el nodo y los cálculos.

Con los resultados obtenidos, representamos dos gráficas, el valor de la función criterio Q y el valor del coeficiente de silueta, las dos en función de K. El coeficiente de silueta es una medida independiente del método de *clustering* utilizado que sirve para evaluar la validez de los clústeres obtenidos y para obtener el número de clústeres óptimos K^* (Rousseeuw, 1987) sugerido en la implementación del método *spherical k-means* (Hornik *et al.*, 2012).

El valor de silueta (Rousseeuw, 1987; Kaufman y Rousseeuw, 1990) mide como un individuo de un clúster es similar a los individuos del mismo clúster en comparación a los individuos del resto de clústeres. Este valor está comprendido entre -1 y +1 tal que cuanto más cercano a +1 implica que aquel individuo se parece mucho más a los

individuos del mismo clúster y a la vez menos a los individuos del resto de clústeres. La fórmula para obtener el valor de silueta de un individuo asignado a un clúster se describe en la Ecuación 4.4.2.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$
$$a(i) = d(i, A)$$
$$b(i) = \min_{C \neq A} d(i, C)$$

Ecuación 4.4.2 Fórmula de cálculo del valor de silueta de un individuo asignado a un clúster

donde:

s(i) es el valor de silueta del individuo *i*,
A es clúster donde está asignado el individuo *i*,
C es el conjunto de clústeres,
d(i,A) es el valor promedio de la disimilitud del individuo i con respecto de los individuos del clúster *A*,
a(i) se corresponde con *d(i,A)*,
d(i,C) es el conjunto de valores promedio de disimilitud para cada clúster dentro de *C* exceptuando *A*, del individuo *i* con respecto a los individuos del clúster y
b(i) es el mínimo del conjunto de valores *d(i,C)*.

El coeficiente de silueta se obtiene mediante el promedio de los valores *s(i)* (Rousseeuw, 1987; Zhao, 2012).

Los valores *s(i)* y el coeficiente de silueta se obtienen mediante el paquete `cluster` (Maechler *et al.*, 2019) de R. Este paquete permite realizar procesos de *clustering* con diferentes métodos y acceder a utilidades para analizar resultados de *clustering* como el coeficiente de silueta, entre otras.

Para encontrar el número de clústeres óptimo K^* en la función criterio Q utilizamos el método del codo o *elbow method* (Thorndike, 1953; Madhulatha, 2012; Kodinariya y Makwana, 2013). Es una regla heurística visual consistente en buscar cuál es el número de clústeres K^* en el que el valor de una función que represente la dispersión de las agrupaciones de los elementos de cada uno de los clústeres en función del número de clústeres tenga un cambio de tendencia o codo (Figura 4.8).

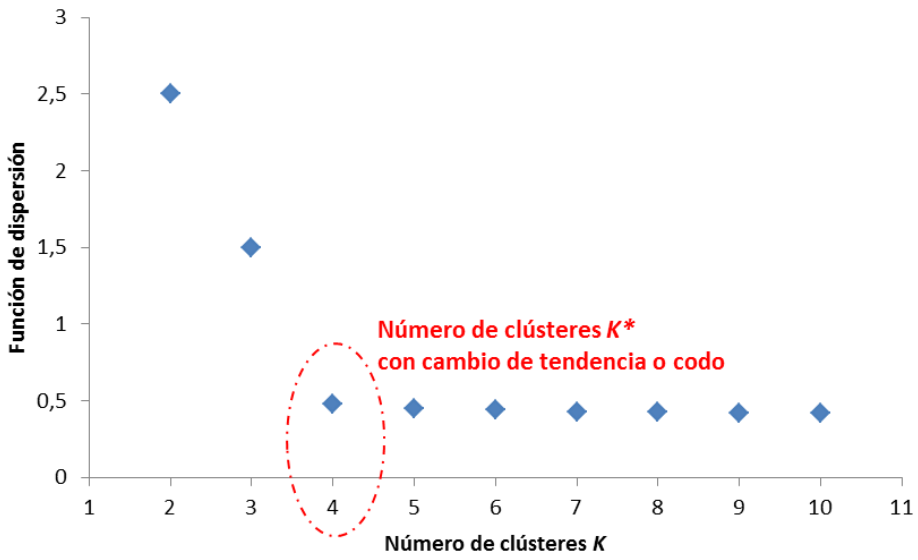


Figura 4.8 Ejemplo indicativo del método del codo

Según como esté definida, una función de dispersión puede ser decreciente o creciente en función del número de clústeres *K*. A medida que aumente el número de clústeres *K* el valor de esta función disminuirá (aumentará), aunque marginalmente existirá un número determinado de clústeres en que esta disminución (aumento) no será tan pronunciada/o generando este cambio de tendencia.

En nuestra investigación, la función criterio Q (Ecuación 4.4.1) calcula la suma para todos clústeres de la suma de las distancia entre los puntos y los centroides de los clústeres, de forma que a menor distancia, menor valor de la función criterio Q. Tal como está definida, la función criterio Q es una función de dispersión y decrece a medida que aumenta el número de clústeres *K*.

En el caso del coeficiente de silueta, el número de clústeres óptimo K^* corresponde al valor del coeficiente de silueta mayor. El coeficiente de silueta aumenta con el número de clústeres K debido a que los datos tienden a estar agrupados en clústeres con mayores valores *s(i)*, pero a partir de un número de clústeres óptimo K^* disminuye ya que algunos clústeres en que los datos eran más similares pueden haber sido divididos disminuyendo los valores *s(i)* de los individuos (Rousseeuw, 1987). De forma gráfica lo podemos observar en la Figura 4.9 .

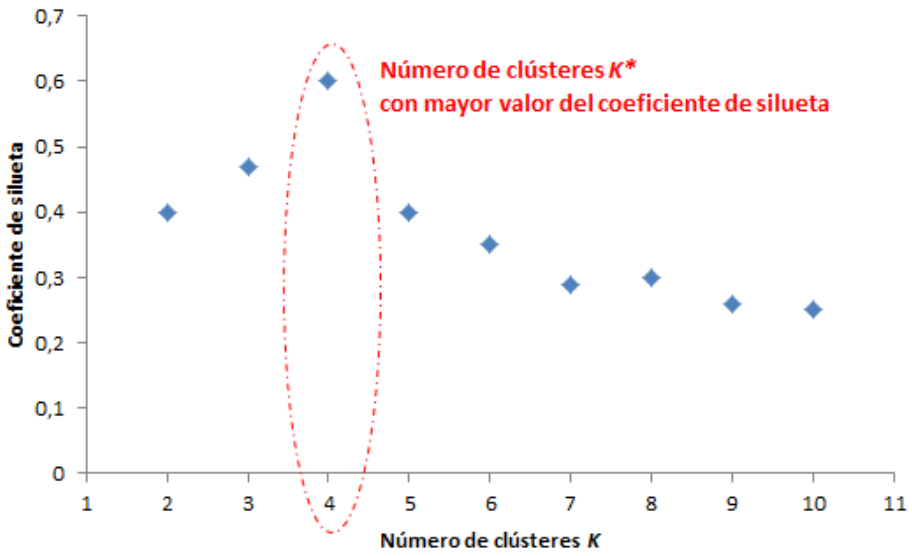


Figura 4.9 Ejemplo indicativo del coeficiente de silueta

No obstante, la gráfica de la función criterio Q puede no presentar un cambio de tendencia nítido y el coeficiente de silueta no presentar un valor máximo claro y por tanto, es difícil el escoger visualmente el número de clústeres óptimo K^*. Podemos apreciar un ejemplo en la Figura 4.10 y Figura 4.11 .

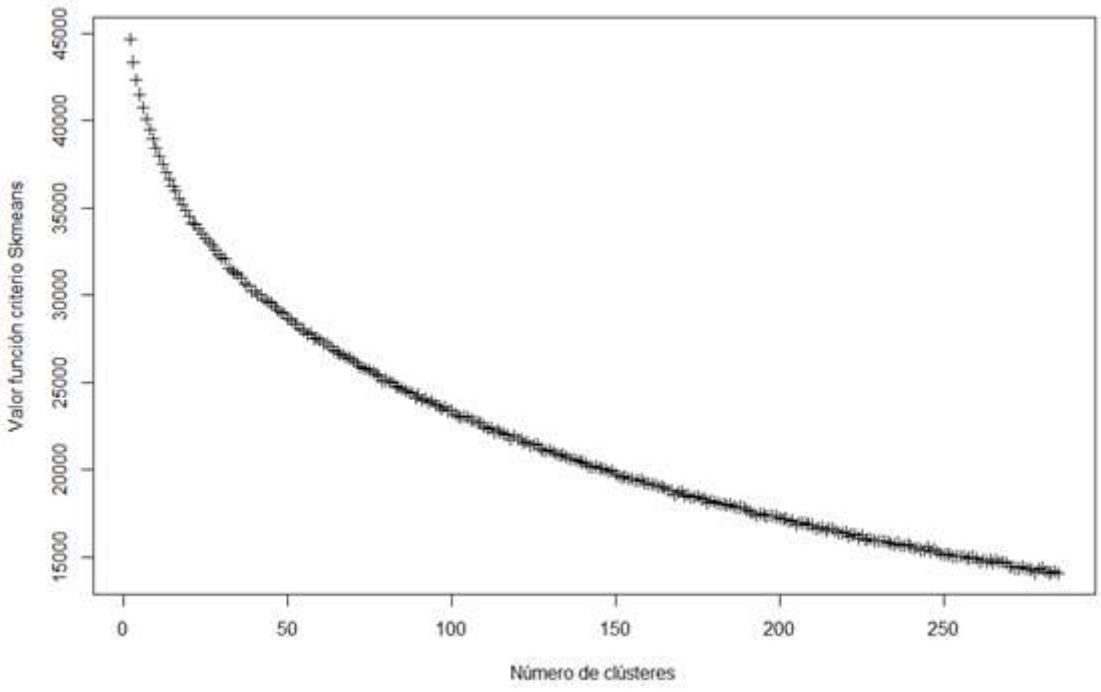


Figura 4.10 Ejemplo indicativo de función criterio con un cambio de tendencia o codo difícil de apreciar visualmente

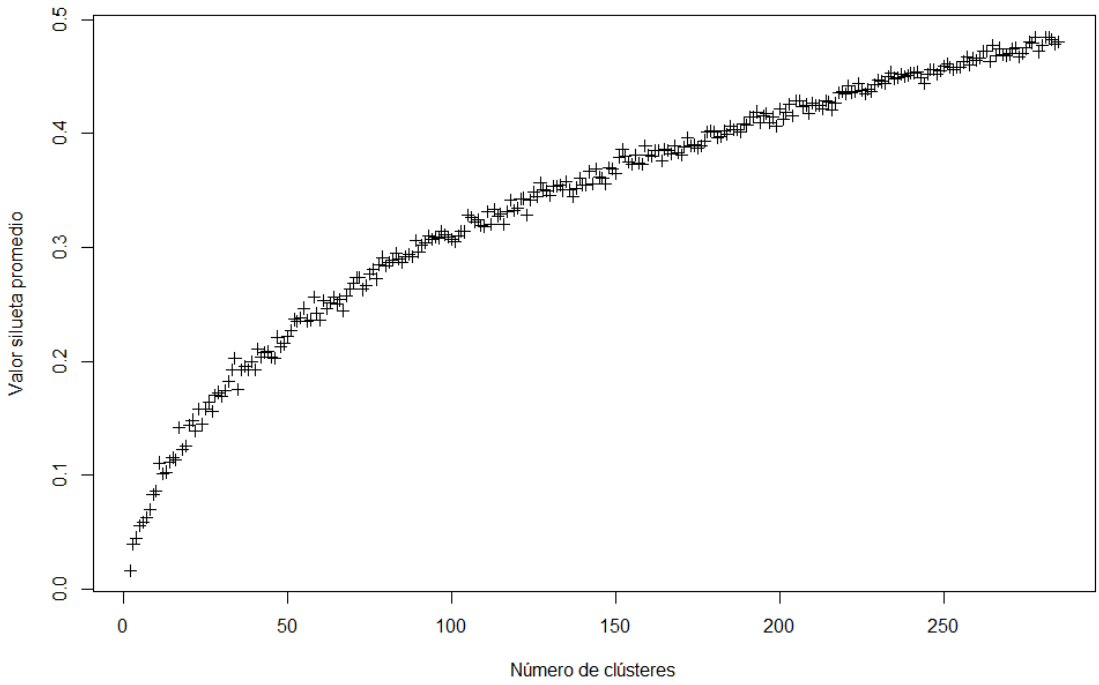


Figura 4.11 Ejemplo indicativo del coeficiente de silueta con un valor máximo difícil de apreciar visualmente

Para intentar objetivar la selección del número de clústeres óptimo K^* decidimos utilizar algún método basado en un índice cuantitativo. Siendo una línea de investigación abierta dentro del *clustering* utilizamos dos métodos, el algoritmo *L-method* (Salvador y Chan, 2004) y el cálculo de la curvatura de una gráfica (Zhang *et al.*, 2017) para intentar encontrar un conjunto de valores aproximados de K^* y posteriormente seleccionar un valor K^* adecuado con los propósitos de la investigación.

El algoritmo *L-method* consiste en dados *n* valores correspondientes a los puntos de una gráfica, calcular el valor RMSE (*Root Mean Square Error*) o error cuadrático medio ponderado de todos los pares de rectas formadas por al menos dos valores empezando por los puntos o bien al inicio o bien al final de la gráfica y seleccionar aquel valor RMSE ponderado tal que sea el mínimo. La fórmula de cálculo del valor RMSE es la mostrada en la Ecuación 4.4.3.

$$RMSE_c = \frac{c-1}{b-1} \cdot RMSE(L_c) + \frac{b-c}{b-1} \cdot RMSE(R_c)$$

Ecuación 4.4.3 Fórmula de cálculo del valor RMSE

donde:

b-1 corresponde al número de puntos en la gráfica, es decir *k*-1,

c es el punto entre 2 y *b* donde se realiza la partición,

L_c y R_c son las particiones izquierda y derecha de los datos en *x=c*,

L_c es la partición izquierda que contiene los datos de *x=2* hasta *x=c*,

R_c contiene los datos de *x=c*+1 hasta *b* y

$RMSE(L_c)$ y $RMSE(R_c)$ son el error cuadrático medio de las ecuaciones de las rectas utilizando las particiones de datos L_c y R_c respectivamente.

Si empezamos por el inicio de los puntos de la gráfica, se crea el primer par de rectas, la primera con el primer y segundo valores y la segunda con el resto de los valores. Se genera el segundo par de rectas, la primera añadiendo el siguiente valor inicial a los dos primeros y la segunda con el resto de valores. Sucesivamente y con el mismo método se generan pares de rectas hasta que la segunda recta es generada con los dos últimos valores de la gráfica. De forma ilustrativa podemos ver la creación de dos pares de rectas en la Figura 4.12.

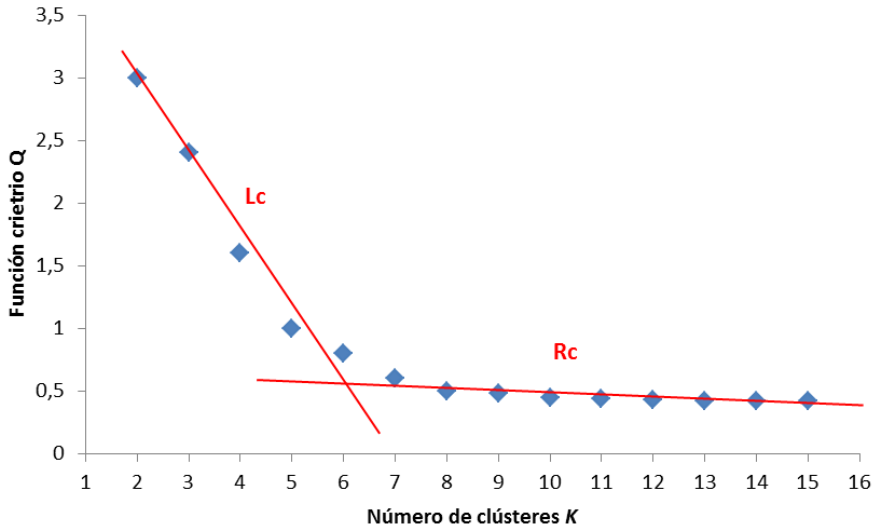


Figura 4.12 Ejemplo indicativo de la creación de pares de rectas con el algoritmo L-method

Para cada par de rectas se calcula el valor RMSE ponderado como la suma del RMSE de la primera recta ponderado por su respectivo número de puntos sobre el total de puntos menos uno y el RMSE de la segunda recta ponderada por su respectivo número de puntos. Si en la gráfica existe un codo, este coincidirá con el menor valor RMSE ponderado (Salvador y Chan, 2004).

Para calcular el módulo de curvatura (κ) de una gráfica se utiliza la fórmula descrita en la Ecuación 4.4.4.

$$\kappa = \frac{|y''|}{(1 + y'^2)^{3/2}}$$

Ecuación 4.4.4 Fórmula de cálculo del módulo de curvatura de una gráfica

donde:

> *y'* e *y''* corresponden a la primera y segunda derivadas de la función que queremos obtener su módulo de curvatura.

Como para la función criterio Q y el coeficiente de silueta no se dispone de una fórmula matemática simbólica, ambas se aproximan mediante splines cúbicos y se calculan sus respectivas primera y segunda derivada mediante el paquete `stats` (R Core Team, 2018) de R. Este paquete proporciona un conjunto de funciones y métodos matemáticos y estadísticos. Si existe cambio de tendencia o codo en la función este corresponde al mayor valor del módulo de curvatura (Zhang *et al.*, 2017).

Aplicando el algoritmo *L-method* y el cálculo del módulo de curvatura a la función criterio Q y al coeficiente de silueta obtenemos cuatro valores de clústeres óptimos K^*. Seleccionamos un valor K^* óptimo de clústeres cercano a estos cuatro valores pero a la vez no demasiado grande para clasificar visualmente los clústeres en función de su temática por expertos de la química intentando evitar su sesgo.

Con el número de clústeres óptimo K^* repetimos el cálculo de *spherical k-means* alrededor de 10000 veces para intentar conseguir una solución que minimizase globalmente la función criterio Q atendiendo al carácter estocástico del método. Con todas las soluciones obtenidas, seleccionamos la mejor de todas las repeticiones correspondiente a aquella con un valor de la función criterio Q menor.

De la misma forma que se ha descrito en este apartado se utiliza la implementación del paquete `skmeans` de R fijando el parámetro “genetic” para obtener clústeres no superpuestos con una población de 50 potenciales soluciones iniciales aleatorias en lugar de las seis que proporciona por defecto el algoritmo genético implementado en el paquete para reducir la probabilidad de convergencia rápida de la población a una solución durante las primeras iteraciones del algoritmo.

Debido a la carga computacional y de memoria necesarias para poder realizar los cálculos, estos se realizaron en el servidor de cálculo ABACO proporcionado por IQS, paralelizamos la obtención de los resultados mediante los paquetes `doParallel` (Microsoft y Weston, 2019a) y `foreach` (Microsoft y Weston, 2019b) de R. Ambos paquetes permiten realizar cálculos en paralelo de cualquier función o proceso mediante bucles `for … next`. Se utilizan estos dos paquetes de R en lugar del paquete `parallel` descrito anteriormente debido a que el paquete `parallel` generaba problemas en la obtención de las soluciones, abortando aleatoriamente la ejecución de los cálculos. El servidor ABACO tiene la misma configuración que la descrita previamente en este apartado. Los cálculos se realizan entre los meses de Abril y Julio de 2018 e incluyeron ajustes del código R, pruebas del código, pruebas piloto de los cálculos en el nodo y los cálculos.

4.5 Interpretación de los clústeres

El objetivo de la interpretación de los clústeres es conseguir clasificar cada uno de los clústeres en función de una temática determinada, ya que aunque el *clustering* es capaz de agrupar los *tweets* en función de su similitud, al ser un método no supervisado, no es capaz de conocer si un clúster corresponde a una temática en función del contenido de los *tweets* y adicionalmente asegurar el agrupamiento perfecto de los *tweets* de una temática en un clúster.

De forma esquemática el proceso de interpretación de los clústeres sigue el representado en la Figura 4.13 y descrito a lo largo de este apartado.

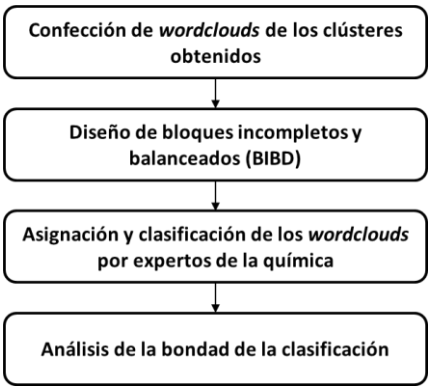


Figura 4.13 Esquema descriptivo del proceso de interpretación de los clústeres utilizado en la investigación

A partir de los clústeres obtenidos en el proceso de *clustering* se generaron dos *wordclouds* por clúster, uno con los 100 unigramas y el segundo con los 100 bigramas más frecuentes de los *tweets* en cada clúster. Los *wordclouds* son considerados una técnica adecuada de visualización para obtener una idea global de los contenidos de textos (Kuo *et al.*, 2007; Cidell, 2010) así como para analizarlos (Gottron, 2009; Cui *et al.*, 2010; DePaolo y Wilkinson, 2014; Heimerl *et al.*, 2014; Horn *et al.*, 2017; Smith *et al.*, 2017). Asimismo proporcionan mejores resultados cuando se representan los bigramas (Kaptein *et al.*, 2010) debido al contenido extra sobre el contexto del texto.

Adicionalmente a las referencias encontradas, se realizaron tests previos para decidir los tipos de *wordclouds* a utilizar. En estos tests, se seleccionó una muestra de los *tweets* de los que se obtuvieron desde sus unigramas *hasta sus* 10-gramas*.* Cada conjunto de estos n-gramas fue representado mediante sus respectivos *wordclouds.* A partir de su análisis se concluyó que con los *wordclouds* de unigramas se mantenían los términos del texto pero podían ser difíciles de interpretar por no tener suficiente contexto y con los *wordclouds* de bigramas no se perdían tantos términos que con el resto de n-gramas, siendo $n \geq 3$, y aportaban contexto del texto suficiente para ser interpretados en comparación con los *tweets* originales..

Para representar los *wordclouds* de unigramas y bigramas se buscan dentro del conjunto total de *tweets* aquellos que pertenecen a cada clúster resultante del *clustering*. Con estos se crean dos matrices TDM, una con los unigramas y la segunda con los bigramas. Los unigramas y su matriz TDM se obtienen mediante el paquete de `tm` (Feinerer *et al.*, 2008) de R. Los bigramas se obtiene mediante el paquete `RWeka` (Hall *et al.*, 2009) de R y la TDM de bigramas mediante el paquete de `tm` (Feinerer *et al.*, 2008) de R. Seleccionando los 100 términos más frecuentes de las respectivas matrices y su frecuencia, se generan los *wordclouds* de unigramas y bigramas mediante el paquete *`wordcloud`* (Fellows, 2018) de R.

Los *wordclouds* obtenidos fueron analizados y clasificados por diversos expertos de la química. Todos ellos tienen el título de doctor en química de los cuales el 75% desde hace 11 años o más y ejercen como investigadores y profesores universitarios en el IQS.

Los dos *wordclouds* permitieron a los expertos analizar e interpretar visualmente los contenidos del clúster y poderlo asignar a una temática. El criterio de asignación a una temática es que la mayoría de términos de los *wordclouds* sugieran al experto aquella

temática concreta y ninguna del resto. Para realizar esta asignación, las temáticas propuestas en este trabajo, a falta de referencias externas en las que se describiesen y después de una prueba piloto en las que se detectaron, son las siguientes:

- **Actividad Humana (AH):** se refiera a la presencia de la química dentro de la actividad que desarrolla la humanidad, como por ejemplo en los ámbitos de producción y de la industria química.
- **Conocimiento Científico (CC):** se refiere a conceptos químicos y entidades abstractas.
- **Entorno Educativo (EE):** se refiere a la química como asignatura o curso enseñado en clase, así como con actividades o ejercicios comunes de los estudiantes.
- **Entretenimiento (E):** se refiere a manifestaciones culturales, como por ejemplo una canción, un grupo musical, el título de una película o el de una serie de televisión.
- **Relación Humana (RH):** se refiere a los sentimientos entre dos o más personas, o bien emociones en general.
- **Indeterminada (I):** se refiere a varias de las temáticas definidas anteriormente de forma que ninguna predomine sobre las demás, o bien a temáticas no definidas en esta lista.

Para asignar los wordclouds de los clústeres a cada experto utilizamos un diseño en bloques incompletos y balanceados (BIBD) (Fleiss, 1981). El BIBD permite que un clúster sea evaluado por un subconjunto de los expertos totales con el objetivo de evitar su fatiga, que cada experto evalúe el mismo número de clústeres y que cada clúster sea evaluado por el mismo número de expertos. Esto implica satisfacer necesariamente el conjunto de condiciones (Fleiss, 1981) descritas en la Ecuación 4.5.1.

$$m * r = n * k$$
$$m \leq n$$
$$\lambda * (m - 1) = r * (k - 1)$$

Ecuación 4.5.1 Condiciones de satisfacción de un diseño en bloques incompletos y balanceados (BIBD)

donde:

m es el número total de expertos del diseño,
n el número total de clústeres,
k el número de expertos que evalúa cada clúster (*k*<*m*),
r el número de clústeres evaluados por cada experto *(r*<*n*),
λ el número de clústeres iguales evaluado por cada par de expertos y
m, k, r, n y λ deben ser números enteros.

Se generaron diferentes posibilidades de diseños fijando el número de clústeres, variando el número de expertos *m* y los expertos que evalúa cada clúster *k* para obtener los valores del número de clústeres evaluados por cada experto *r* y el número de clústeres iguales evaluado por cada par de expertos λ. Los valores obtenidos fueron chequeados para comprobar la condición de si eran números enteros o no. Los BIBD con *r* y λ enteros se guardaron y posteriormente fueron verificados junto con las condiciones de suficiencia para ser considerado diseño BIBD con el paquete `crossdess` (Sailer, 2013) de R. El paquete `crossdess` contiene funciones para la construcción y verificación de diseños de experimentos.

Como ningún BIBD cumplía las condiciones necesarias y suficientes con el número total de expertos, estos fueron divididos en tres grupos para poder cumplir las condiciones de uno de los BIBD. Asimismo, se redujo el número de clústeres a ser evaluados por cada experto para aumentar su calidad de análisis. Los expertos fueron asignados de forma aleatoria a cada uno de los grupos de clústeres definidos en el BIBD así como el orden de presentación los clústeres y sus respectivos *wordclouds*. El diseño detallado del BIBD se desarrolló utilizando el paquete `bibd` (Mandal, 2019) de R. El paquete `bibd` contiene funciones para el diseño en bloques incompletos y balanceados.

A cada experto se le entregó en mano un documento en el que cada página contenía los dos *wordclouds* de unigramas y bigramas correspondientes al clúster a analizar y una tabla con el nombre y definición de cada temática que el experto tenía que marcar según su criterio. Adicionalmente el documento describía en la página inicial las instrucciones de cómo y qué criterio seguir para decidir si un clúster pertenecía a una temática (ver Anexo 2). En la entrega del documento también se explicó y revisó con cada experto el criterio de pertenencia para intentar minimizar la subjetividad debida a la diferente interpretación de las instrucciones.

Como resultado de la interpretación de los clústeres, obtenemos la lista de clústeres votados a cada temática por cada experto. A partir de esta lista calculamos el porcentaje de clústeres y *tweets* asignados a cada temática sumando todos los votos que un clúster había recibido en aquella temática. La temática asignada a un clúster es la que recibe un número mayor de votos. Los *tweets* que pertenecen al clúster heredan la asignación temática del clúster. Si un clúster tiene los mismos votos en diferentes temáticas, se decide asignarlo a la temática Indeterminada (I).

Para evaluar la bondad de la clasificación temática de los clústeres analizamos los resultados mediante el estadístico kappa de Fleiss (Fleiss, 1971). La kappa de Fleiss mide, dado un número de expertos que realizan un conjunto de valoraciones categóricas sobre una serie de observaciones, hasta qué punto el acuerdo entre los expertos es o no fruto de la casualidad. La kappa de Fleiss tiene en cuenta un número fijo de expertos y que diferentes observaciones pueden ser valoradas por diferentes expertos.

En nuestra investigación calculamos la kappa de Fleiss para cada temática y para todo el diseño, su error, el valor z y el p-valor (Fleiss *et al.*, 2003) según las ecuaciones descritas en la Ecuación 4.5.2, Ecuación 4.5.3, Ecuación 4.5.4, Ecuación 4.5.5 y Ecuación 4.5.6.

$$\kappa_j = 1 - \frac{\sum_{i=1}^{n} x_{ij}(m - x_{ij})}{nm(m-1)\bar{p}_j\bar{q}_j}$$

$$\sum_{j=1}^{k} x_{ij} = m$$

$$\bar{p}_j = 1 - \bar{q}_j$$

Ecuación 4.5.2 Cálculo de la kappa de Fleiss de una temática

$$\kappa = 1 - \frac{nm^2 - \sum_{i=1}^{n}\sum_{j=1}^{k} {x_{ij}}^2}{nm(m-1)\sum_{i-1}^{n}\bar{p}_j\bar{q}_j}$$

$$\sum_{j=1}^{k} x_{ij} = m$$

$$\bar{p}_j = 1 - \bar{q}_j$$

Ecuación 4.5.3 Cálculo de la kappa de Fleiss del diseño global

$$se(\kappa_j) = \sqrt{\frac{2}{nm(m-1)}}$$

$$\sum_{j=1}^{k} x_{ij} = m$$

Ecuación 4.5.4 Cálculo del error estándar por temática

$$z_j = \frac{\kappa_j}{se(\kappa_j)}$$

Ecuación 4.5.5 Cálculo del valor Z de las evaluaciones por temática

$$p\text{-}valor_j = Prob(z \geq z_j)$$

Ecuación 4.5.6 Cálculo del p-valor por temática

donde:

n es el número clústeres,

m es el número de expertos que evalúa cada clúster,

k es el número de temáticas,

x_{ij} es el número de evaluaciones del clúster *i* en la temática *j,*

$\bar{p}_j$ es la proporción de evaluaciones en la temática *j*,

κ_j es la kappa de Fleiss de la temática *j*,

κ es la kappa de Fleiss del diseño global,

$se(\kappa_j)$ es el error estándar por temática,

z_j es el valor Z de las evaluaciones por temática siendo Z la distribución normal estándar y

p-valor el área derecha de la distribución normal estándar por encima del valor z_j .

Kappa puede tomar los valores entre -1 y +1. Cuanto más cercano es el valor de kappa a +1, mejor es el acuerdo, en cambio si es cercano o menor a 0 se considera que el acuerdo se ha alcanzado por casualidad. La validez de kappa la evaluamos mediante su respectivo p-valor. Con p-valores inferiores a 0.05, kappa es estadísticamente significativa y por tanto, el acuerdo se ha conseguido porque los expertos comparten alguna razón latente para escoger la temática.

Para evaluar la calidad de los acuerdos obtenidos, se utiliza un criterio habitual (Landis y Koch, 1977) que se puede consultar en la Tabla 4.5.1. Este criterio se aplica a todas las temáticas y en particular, a los *tweets* clasificados en las temáticas de AH y EE, temáticas que se espera contengan la mayoría de términos de la imagen pública de la química e indicios de la quimiofobia.

Tabla 4.5.1. *Benchmark* de Landis and Koch (1977)

Valor de kappa	Bondad del acuerdo
<0.00	*Poor*
0.00-0.20	*Slight*
0.21-0.40	*Fair*
0.41-0.60	*Moderate*
0.61-0.80	*Substantial*
0.81-1.00	*Almost perfect*

Con los procesos descritos en este apartado, se consigue clasificar los *tweets* en temáticas.

4.6 Análisis de sentimientos

El objetivo del análisis de sentimientos y emociones y correspondiente a uno de los objetivos de esta investigación, es la clasificación los *tweets* de las temáticas AH (actividad humana) y EE (entorno educativo) según su valor de sentimiento[10] que transmiten para su posterior análisis cualitativo. Las temáticas AH y EE, por su definición, se ha observado que son las que contienen la mayoría de *tweets* relacionados con la imagen pública de la química en Twitter y que según la bibliografía tiene una dimensión social y académica. Las temáticas E (entretenimiento), RH (relación humana) no contienen *tweets* dentro del ámbito científico y técnico de la química. La temática I (indeterminada) puede contener *tweets* de mezcla de temáticas u otras temáticas no definidas. La temática CC (conocimiento científico), aunque sí puede contener, no se consideró por contener pocos *tweets*.

De forma esquemática el análisis de sentimientos sigue el proceso representado en la Figura 4.14 y descrito a lo largo de este apartado.

[10] Sentimiento: Estado afectivo del ánimo. (https://dle.rae.es/sentimiento consultado el 11/01/2021)

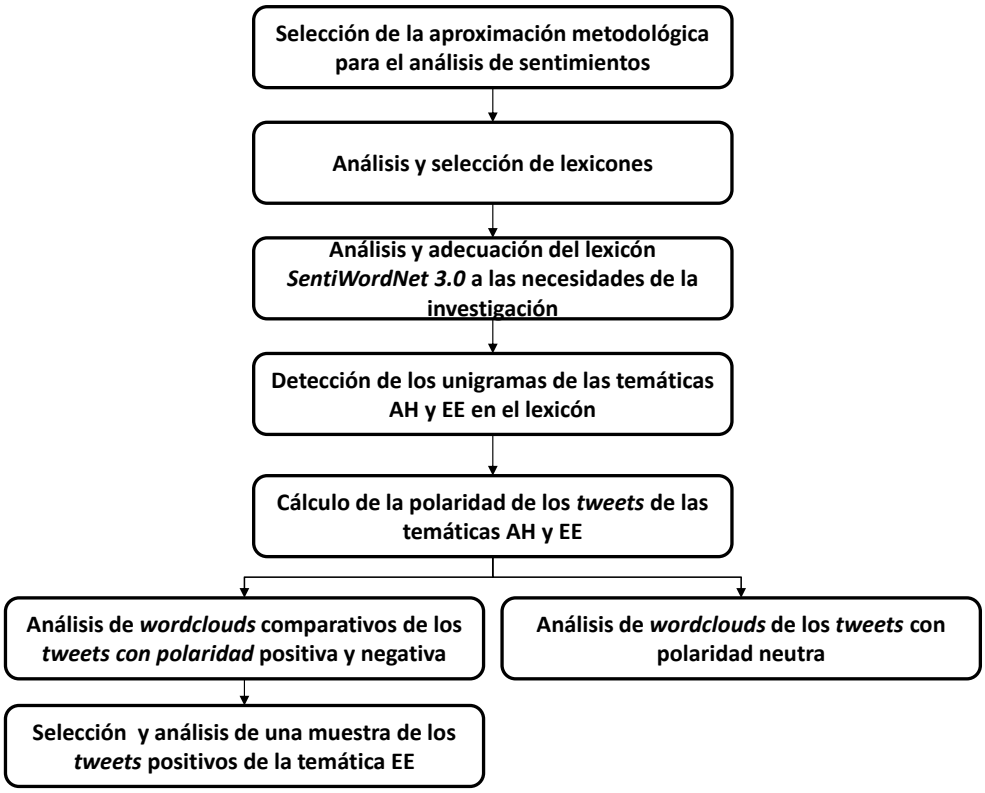


Figura 4.14 Esquema descriptivo del proceso de análisis de sentimientos utilizado en la investigación

En el análisis de sentimientos en Twitter, existen dos grandes líneas de aproximación para su cálculo y análisis (Giachanou y Crestani, 2016). La primera está basada en técnicas de clasificadores y la segunda en lexicones[11].

En ambas líneas, el valor de un sentimiento se mide según su polaridad. La polaridad generalmente es un valor que varía entre -1 y 1, siendo el sentimiento negativo si la polaridad es negativa, neutro si la polaridad es cero y positivo si la polaridad es positiva. También puede tomar otra escala de valores entre un valor negativo y otro positivo, los dos habitualmente iguales en valor absoluto.

Las técnicas de clasificadores pertenecen al ámbito de la inteligencia artificial y en particular, del *machine learning*. Las más extendidas se engloban dentro del aprendizaje supervisado, en el que a partir de datos y sus polaridades asociadas se entrena a un clasificador obteniendo un modelo que sea capaz de calcular las polaridades de nuevos datos.

[11] Lexicón: diccionario. (https://dle.rae.es/lexic%C3%B3n?m=form, consultado el 11/01/2021)

La segunda línea de aproximación está basada en lexicones. Se utiliza una lista de palabras en la que cada una tiene una polaridad asociada. La polaridad de un texto se obtiene con la suma de polaridades de las palabras coincidentes en el lexicón. Estos han sido aplicados en textos como blogs, fórums y revisiones de productos, pero menos explorados en Twitter (Giachanou y Crestani, 2016). en comparación con las técnicas de machine learning debido al contenido cambiante de los *tweets* y las expresiones coloquiales utilizadas.

La ventaja principal de los lexicones es que no requieren de datos de entrenamiento. Su limitación principal es que si una palabra no existe en el lexicón no se considera, no están generalmente especializados en ámbitos de conocimiento concretos y que no pueden adaptar la polaridad a contextos particulares (Martin-Valdivia y Martínez Camara, 2013; Montejo-Ráez *et al.*, 2014).

Al no existir ningún clasificador especializado en el área de la química, ni tener datos clasificados en esta área para poder entrenarlo, se han clasificado los sentimientos de los *tweets* de AH y EE mediante la aproximación basada en lexicones donde la polaridad de un *tweet* se mide como la suma de las polaridades de las palabras del *tweet* encontradas dentro del lexicón (Sun *et al.*, 2017). Un valor de polaridad del *tweet* positivo implica un sentimiento positivo, un valor negativo implica un sentimiento negativo y un valor 0 implica un sentimiento neutro.

Existen diferentes lexicones (Giachanou y Crestani, 2016; Mohey y Hussein, 2016; Soleymani *et al.*, 2017; Zimbra *et al.*, 2018), entre los cuales, el Affin (Nielsen, 2011), el Bing (Hu y Liu, 2004; B. Liu *et al.*, 2005), el NRC (Mohammad y Turney, 2013), el SentiStrength (Thelwall *et al.*, 2010), el SentiWordNet 3.0 (Baccianella *et al.*, 2010), el Syuzhet (Matthew L. Jockers, 2015) o el Vader (Hutto y Gilbert, 2014).

De todos ellos descartamos utilizar el SentiStrength al no tener un paquete de R con el que poder acceder a él y ser eminentemente comercial. El resto, que pueden ser accesibles desde diferentes paquetes de R, tienen las características básicas de número de términos con polaridades negativa, neutra y positiva y número total de términos, mostrados en la Tabla 4.6.1.

Tabla 4.6.1. Descripción de características básicas de lexicones de sentimientos

Lexicón	Número de términos			
	Polaridad negativa	Polaridad neutra	Polaridad positiva	Total
Affin	1 598	1	878	2 477
Bing	4 783	0	2 006	6 789
NRC	3 243	8 709	2 230	14 182
SentiWordNet 3.0	14 726	89 805	13 128	117 659
Syuzhet	7 161	0	3 587	10 748
Vader	4 041	0	3 195	7 236

El lexicón seleccionado en nuestra investigación es el SentiWordNet 3.0 (Baccianella *et al.*, 2010) debido a tener una mayor cantidad de términos tanto positivos como negativos, tener un cierto equilibrio entre el número de términos positivos y negativos, ser accesible en R y ser uno de los habitualmente referenciados en la literatura (Medhat *et al.*, 2014; Mohey y Hussein, 2016; Sun *et al.*, 2017; Yadollahi *et al.*, 2017; Mäntylä *et al.*, 2018).

El lexicón SentiWordNet 3.0 está estructurado en una tabla que contiene los campos mostrados en la Tabla 4.6.2. Puede obtenerse en la dirección web https://github.com/larsmans/sentiwordnet/blob/master/SentiWordNet_3.0.0_20130122.txt y es accesible a través del paquete `syuzhet` (Matthew L. Jockers, 2015) de R. Este paquete contiene funciones para poder extraer los valores de polaridad de textos a partir de lexicones propios o de lexicones externos al paquete que pueden ser introducidos por los usuarios, como es el caso del SentiWordNet 3.0.

Tabla 4.6.2. Descripción del lexicón SentiWordNet 3.0

Nombre del campo	Tipo	Descripción
POS	Carácter	"Part of Speech". Informa que tipo de palabra es desde el punto de vista de utilización en una frase (adjetivo, verbo, pronombre, nombre)
ID	Entero	Identificador único del lema en Wordnet
PosScore	Numérico	Valor positivo (entre 0 y 1)
NegScore	Numérico	Valor negativo (entre 0 y 1)
Terms	Carácter	Palabras simples o compuestas que están relacionadas entre sí a través de su lema y que están en Wordnet. No son palabras sinónimas sino palabras derivadas o compuestas. Están separadas mediante el símbolo # un número y un espacio.
Gloss	Carácter	Descripción del significado de las palabras con ejemplos de utilización

Podemos observar un ejemplo ilustrativo de parte del contenido de este lexicón en la Tabla 4.6.3.

Tabla 4.6.3. Ejemplo del contenido del lexicón SentiWordNet 3.0

POS	ID	PosScore	NegScore	Terms	Gloss
a	19349	0.375	0	reachable#1 approachable#3	easily approached; "a site approachable from a branch of the Niger"
a	19505	0.625	0	getatable#1 get-at-able#1 come-at-able#2	capable of being reached or attained; "a very getatable man"; "both oil and coal are there but not in getatable locations"
a	19731	0.125	0.125	ready_to_hand#1 handy#1	easy to reach; "found a handy spot for the can opener"
a	19874	0.5	0.125	unaccessible#1 inaccessible#1	capable of being reached only with great difficulty or not at all

El lexicón SentiWordNet 3.0 puede contener en el campo Terms más de un término, términos duplicados en diferentes filas y no tiene un valor de polaridad ya calculado.

A partir de un texto y un lexicón, la función `get_sentiment` del paquete `syuzhet` busca la polaridad de las términos del texto que coinciden con las del lexicón, y calcula la polaridad del texto sumando las polaridades asociadas de los términos del texto coincidentes. La función `get_sentiment` con el lexicón SentiWordNet 3.0 busca en cada fila de la tabla del lexicón el primer término dentro del campo Terms, obviando el resto de términos. Adicionalmente, si existen varios términos coincidentes en el lexicón, sólo tiene en cuenta el primer término encontrado.

Por este motivo, construimos un lexicón propio a partir del lexicón SentiWordNet 3.0 en el que cada fila contuviera un término único con su polaridad asociada a partir del campo Terms y de los campos PosScore y NegScore. Las operaciones realizadas para su construcción junto con ejemplos ilustrativos son las siguientes:

- Cálculo de la polaridad de los términos en el campo Terms. La polaridad se calcula como la suma de PosScore y NegScore, debido a que el campo NegScore es un valor negativo.
    - Antes de la operación:

Terms	PosScore	NegScore
reachable#1 approachable#3	0.375	0

    - Después de la operación:

Terms	Polaridad
reachable#1 approachable#3	0,375

- Eliminación de números y el símbolo # del campo Terms

- Antes de la operación:

Terms	Polaridad
reachable#1 approachable#3	0,375

- Después de la operación:

Terms	Polaridad
reachable approachable	0,375

- Detección de cada uno de los términos en el campo Terms y creación de nuevas filas con cada uno de los términos y su polaridad
  - Antes de la operación:

Terms	Polaridad
reachable approachable	0,375

  - Después de la operación:

Terms	Polaridad
reachable	0,375
approachable	0,375

- Cálculo de la polaridad media de los términos repetidos
  - Antes de la operación:

Terms	Polaridad
aberrate	-0,25
aberrate	0

  - Después de la operación:

Terms	Polaridad
aberrate	-0,125
Aberrate	-0,125

- Eliminación de las filas de términos repetidos y su polaridad
  - Antes de la operación:

Terms	Polaridad
aberrate	-0,125
aberrate	-0,125

  - Después de la operación:

Terms	Polaridad
aberrate	-0,125

Como el valor de PosScore puede estar entre 0 y +1 y el de NegScore entre 0 y -1, la polaridad es un valor entre -1 y +1 así como la polaridad media. El cálculo de la polaridad media es uno de los métodos referenciados en la literatura (Gatti y Guerini, 2012; Guerini *et al.*, 2013) y utilizado con buenos resultados en comparación con otros métodos, así como utilizado en diversos estudios y aplicaciones (Devitt y Ahmad, 2007; Denecke, 2009; Thet *et al.*, 2010; Sing *et al.*, 2012).

Con el lexicón resultante se comparan sus términos con los unigramas de los *tweets* obtenidos en el apartado 4.2 Limpieza de textos, y clasificados en las temáticas EE y AH. Calculamos la polaridad del *tweet* como la suma de las polaridades de los términos coincidentes en ambos textos. Como la polaridad de un término de un *tweet* puede estar entre -1 y +1, el valor de la polaridad de un *tweet* puede ser superior a +1 e inferior a -1.

Se separan los *tweets* de cada una de las temáticas EE y AH en dos grupos, los de polaridad con signo positivo y los de signo negativo. Los *tweets* con polaridad con signo positivo son los de sentimiento positivo y los de signo negativo son los de sentimiento negativo.

Para cada temática calculamos la matriz TDM de unigramas y la matriz TDM de bigramas de los dos grupos de *tweets* positivos y negativos. La matriz TDM de unigramas contiene dos documentos, el de los unigramas de los tweets de sentimiento positivo y el de los unigramas de los tweets de sentimiento negativo. De igual forma la matriz TDM de bigramas contiene dos documentos, el de los bigramas de los tweets de sentimiento positivo y el de los bigramas de los tweets de sentimiento negativo.

La matriz TDM de unigramas se obtiene mediante el paquete `tm` (Feinerer *et al.*, 2008) de R. La matriz TDM de bigramas mediante el paquete `RWeka` (Hall *et al.*, 2009) y el paquete `tm` (Feinerer *et al.*, 2008) de R.

Un *wordcloud* comparativo es una representación gráfica de términos o n-gramas de varios documentos en el mismo *wordcloud*. Los términos comunes son asignados al documento donde aquel n-grama tiene su desviación máxima, calculada como su frecuencia en aquel documento menos la frecuencia media en todos los documentos. Permite comparar las frecuencias de los términos de los documentos en el mismos *wordcloud*.

A partir de la matriz de TDM de unigramas y de TDM de bigramas y para cada una de las temáticas EE y AH generamos sus *wordclouds* comparativos mediante el paquete `wordcloud` (Fellows, 2018) de R. Obtenemos dos *wordclouds* comparativos para cada temática, uno de unigramas y otro de bigramas. Cada *wordcloud* comparativo contiene como máximo 200 términos positivos y 200 términos negativos para interpretar de forma más sencilla los contenidos más destacados dentro de los *wordclouds*. Mostramos un ejemplo ilustrativo de un *wordcloud* comparativo en la Figura 4.15.

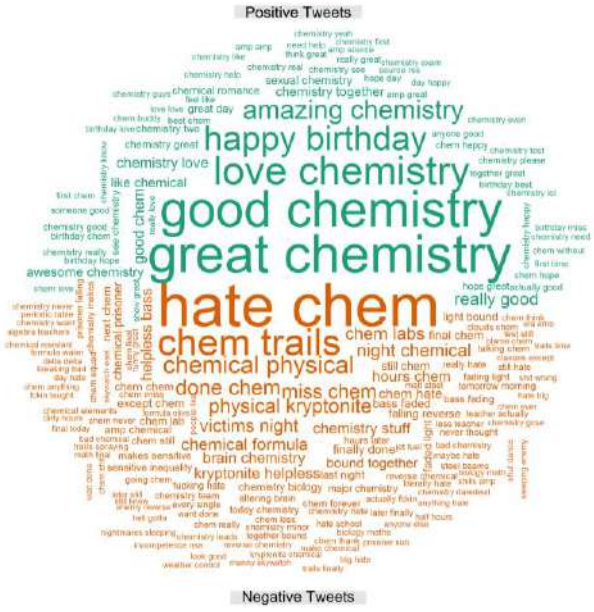


Figura 4.15 Ejemplo ilustrativo de *wordcloud* comparativo de bigramas

Los *wordclouds* comparativos aplicados en esta investigación permiten analizar los unigramas y bigramas de *tweets* con sentimiento positivo y negativo de forma separada y compararlos. Adicionalmente, permite detectar unigramas y bigramas que aparentemente denotan un sentimiento opuesto al sentimiento del *tweet* con el que fueron clasificados..

Según el marco teórico, no parecería que en la temática EE debieran existir muchos *tweets* con sentimiento positivo, aunque el uso de un lexicón puede en el análisis de sentimientos puede hacer que un unigrama o bigrama tenga una polaridad negativa y el *tweet* donde aparece positiva. Para entender con mayor detalle los *tweets* positivos de la temática EE que sus unigramas y bigramas creemos que deberían aparecer en los *tweets* negativos, seleccionamos una muestra aleatoria representativa de aquellos

que contenían los unigramas más frecuentes representados en el *wordcloud* comparativo y que a su vez englobasen también a los bigramas más frecuentes.

El tamaño de las muestras estadísticamente representativas del conjunto de tweets fueron calculadas mediante la Ecuación 4.6.1 para estimar el tamaño de muestra de una proporción con corrección para poblaciones finitas (Morales Vallejo, 2012). El cálculo se realizó mediante el paquete de `samplingbook` (Manitz *et al.*, 2017) de R. Este paquete proporciona diversas funciones para el cálculo de muestreo en poblaciones.

$$n = \frac{z^2 p(1-p)N}{\epsilon^2(N-1) + z^2 p(1-p)}$$

Ecuación 4.6.1 Cálculo de la estimación del tamaño de muestra de una proporción con corrección para poblaciones finitas

donde:

N es el tamaño de la población, que en nuestro caso es el número de *tweets* sobre los que queremos calcular el tamaño de muestra,

p es la proporción esperada de *tweets* con sentimientos negativos o positivos. Sin información previa sobre la proporción esperada, se toma el valor más desfavorable (*p*=0,5),

z es el valor de variable normal estándar asociado al nivel de confianza del intervalo de confianza deseado para la estimación de la proporción y

ϵ es la precisión que corresponde a la mitad de la amplitud del intervalo de confianza.

Se analiza cada uno de los *tweets* de la muestra y clasifica en positivo, neutro, negativo o sin clasificar utilizando los criterios respecto al contenido del *tweet* descritos en la Tabla 4.6.4.

Tabla 4.6.4. Criterios de clasificación de la muestra de *tweets* con sentimiento positivo de la temática EE

Clasificación	Descripción	Ejemplo
Positivo	Contenidos que transmiten sentimientos positivos con respecto a elementos del entorno educativo, como por ejemplo, el sentirse preparado ante un examen, el haber conseguido una nota superior a las expectativas previas o el considerar de forma positiva un docente respecto a su actividad en la clase.	"He's gonna see my first test score and laugh í ½í¸¬í ½í¸¬ I NEED TO GET AN A ON MY CHEM TEST I AM SO DETERMINED TO IMPRESS HIM"
Negativo	Contenidos que transmiten sentimientos negativos con respecto a elementos del entorno educativo, como por ejemplo, el no sentirse preparado o tener miedo ante un examen, el no haber conseguido una nota superior a las expectativas previas, no pasar un examen o el considerar de forma negativa un docente.	"whY THE FUCK WOULD YOU GIVE US HOMEWORK THE DAY BEFORE A TEST THAT ISN'T EVEN RELEVANT TO ANYTHING WE HAVE LEARNED @ CHEM TEACHER"
Neutro	Contenidos que siendo interpretables parecen no transmitir ningún sentimiento.	"Can someone please FaceTime me for like 10 mins and teach me how to do the Chem test"
Sin clasificar	Contenidos que son difícilmente interpretables respecto a la clasificación anterior.	"When you have 6 hours of dance, calculus homework, and a chem test to study for :,) http://t.co/K2qHVRzyxU"

También se detectan los *tweets* con ironías[12] en su contenido, aspecto difícilmente analizable sólo a partir del lexicón utilizado ya que no se tiene en cuenta ni el contexto ni la semántica. Por ejemplo, "That chemistry test was some kinda ridiculous #prayformygrades".

Los resultados de la clasificación se muestran en forma de porcentajes con respecto al total de la muestra y se analizan para entender si existen *tweets* con aparentes sentimientos negativos o neutros dentro de los *tweets* clasificados como positivos en la temática EE.

Para analizar si los sentimientos son fuertes o débiles en las polaridades positivas o negativas, y por tanto, existe un posicionamiento en estos sentimientos polarizado o intermedio, analizamos los *tweets* con polaridad neutra. Por ejemplo, un número elevado de *tweets* con polaridad neutra con respecto al total de *tweets* en una de las temáticas sugeriría que aunque existen sentimientos positivos y negativos con

[12] Ironía: Expresión que da a entender algo contrario o diferente de lo que se dice, generalmente como burla disimulada (https://dle.rae.es/iron%C3%ADa, consultado el 14/01/2021).

respecto aquella temática, estos no estarían polarizados y en cambio el posicionamiento sería intermedio.

Calculamos el porcentaje de *tweets* con polaridad neutra con respecto al total de *tweets* de cada una de las temáticas AH y EE y representamos sus *wordclouds* de unigramas y bigramas para comparar si los términos más frecuentes coinciden con los términos más frecuentes de los *wordclouds* comparativos de unigramas y bigramas de los *tweets* de AH y EE.

Con los procesos descritos en este apartado, se consigue analizar los sentimientos de los *tweets* clasificados en las temáticas AH y EE.

4.7 Análisis de emociones

El objetivo del análisis de emociones es la clasificación los *tweets* de las temáticas AH (actividad humana) y EE (entorno educativo) según el tipo de emoción[13] que transmiten para su posterior análisis cualitativo. Las temáticas AH y EE son las que contienen la mayoría de *tweets* relacionados con la imagen pública de la química en Twitter.

De forma esquemática el análisis de emociones sigue el proceso representado en la Figura 4.16 y descrito a lo largo de este apartado.

[13] Emoción: Alteración del ánimo intensa y pasajera, agradable o penosa, que va acompañada de cierta conmoción somática. (https://dle.rae.es/emoci%C3%B3n consultado el 11/01/2021)

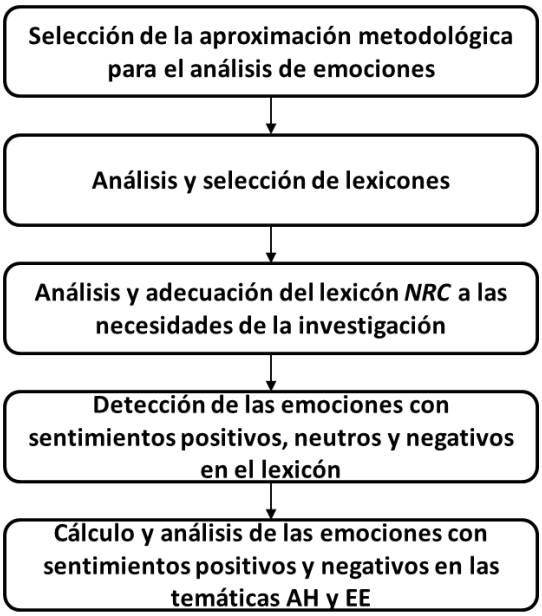


Figura 4.16 Esquema descriptivo del proceso de análisis de emociones utilizado en la investigación

Las técnicas más extendidas para calcular las emociones de textos se basan, de la misma forma que en el caso del análisis de sentimientos, en métodos basados en el procesamiento del lenguaje y métodos de clasificación dentro del área de conocimiento del *machine learning* (Chen *et al.*, 2018).

Los métodos basados en el procesamiento del lenguaje se dividen en cuatro grupos (Sailunaz *et al.*, 2018), los basados en las palabras clave, los basados en lexicones, los basados en *machine learning* y los híbridos, que combinan cualquiera de los anteriores métodos.

En los métodos basados en palabras clave se busca en una frase la palabra que contiene la emoción que se compara con una lista de palabras clave de las que se sabe su emoción. La emoción de la frase corresponde a la emoción de la palabra clave que coincide con la palabra buscada. Los métodos basados en lexicones[14] utilizan un lexicón en lugar de una lista de palabras clave y tienen en cuenta todas las palabras de una frase. Utilizan diversas formas para calcular la emoción de una frase. Los métodos *machine learning* buscan construir un modelo con un texto del que se conocen sus emociones para poder predecir las emociones de un nuevo texto.

[14] Lexicón: diccionario. (https://dle.rae.es/lexic%C3%B3n?m=form, consultado el 11/01/2021)

En nuestra investigación decidimos utilizar los métodos basados en lexicones. Los motivos de esta decisión fueron el no tener disponibilidad ni de un modelo de *machine learning* ni tampoco de un método basado en palabras clave ya desarrollado y de aplicación a nuestro ámbito de estudio en Twitter, y en cambio, sí poder tener acceso a lexicones generalmente utilizados en la literatura, públicos y accesibles a través de un paquete de R.

Algunos de los lexicones más habitualmente citados en la literatura (Yadollahi *et al.*, 2017) son el Wordnet Affect (Strapparava y Valitutti, 2004), el LIWC (Pennebaker *et al.*, 2007), el NRC (Mohammad y Turney, 2013), el NRC hashtag (Mohammad, 2012) o el CBET (Gholipour Shahraki, 2015). Sus características principales respecto al número de términos y emociones valoradas están reflejadas en la Tabla 4.7.1.

Tabla 4.7.1. Descripción características básicas de lexicones de emociones

Lexicón	**Número de términos**	**Emociones valoradas**
Wordnet Affect	4787	Jerarquía de emociones
LIWC	5000	*affective or not, positive, negative, anxiety, anger, sadness*
NRC	14,182	*anger, fear, anticipation, trust, surprise, sadness, joy, disgust, positive and negative sentiment*
NRC hashtag	16,862	*anger, fear, anticipation, trust, surprise, sadness, joy, disgust*
CBET	24,000	*anger, fear, joy, love, sadness, surprise, thankfulness, disgust, guilt*

En esta investigación utilizamos el lexicón de valencia de emociones NRC (Mohammad y Turney, 2013) versión 0.92, debido a ser un lexicón ampliamente referenciado en la literatura (Bravo-Marquez *et al.*, 2013; Kiritchenko *et al.*, 2014; Giachanou y Crestani, 2016; Yadollahi *et al.*, 2017; Chaturvedi *et al.*, 2018), ser accesible mediante varios paquetes de R y tener un lista importante de palabras.

El lexicón NRC está estructurado en una tabla que contiene los campos mostrados en la Tabla 4.7.2.

Tabla 4.7.2. Descripción del lexicón NRC

Nombre del campo	Tipo	Descripción
term	Carácter	Palabra valorada en el lexicón.
AffectCategory	Carácter	Palabra de la lista de emociones (*anger, fear, anticipation, trust, surprise, sadness, joy, disgust)* o de la de sentimientos *(positive, negative).*
AssociationFlag	Numérico	Valor binario (0 o 1).

El campo term contiene el término o unigrama asociado un conjunto de emociones y sentimientos. El campo AffectCategory contiene una de las palabras de la lista de emociones *anger, fear, anticipation, trust, surprise, sadness, joy, disgust* o de la lista de sentimientos *positive, negative*. El campo AssociationFlag contiene el valor de valencia siendo un valor binario (0 o 1). Si el valor de valencia es 0 la palabra de emoción o sentimiento en el campo AffectCategory no está asociada con el unigrama del campo term siendo 1 en caso contrario. Mostramos un ejemplo ilustrativo de parte de este lexicon en la Tabla 4.7.3.

Tabla 4.7.3. Ejemplo ilustrativo del lexicón NRC

Term	AffectCategory	AssociationFlag
abandon	anger	0
abandon	anticipation	0
abandon	disgust	0
abandon	fear	1
abandon	joy	0
abandon	negative	1
abandon	positive	0
abandon	sadness	1
abandon	surprise	0
abandon	trust	0

En la literatura existen diversos modelos de emociones siendo los más frecuentes en la literatura y que expresan la mayoría de emociones humanas (Sailunaz *et al.*, 2018) los mostrados en la Tabla 4.7.4.

Tabla 4.7.4. Modelos de emociones

Modelo	Emociones
Ekman (Ekman, 1992)	*Anger, disgust, fear, joy, sadness, surprise*
Shaver (Shaver *et al.*, 1987)	*Anger, fear, joy, love, sadness, surprise*
Oatley and Johnson-Laird (Oatley y Johnson-Laird, 1987)	*Anger, anxiety, disgust, happiness, sadness*

Modelo	Emociones
Plutchik (Plutchik, 1980)	*Acceptance, admiration, aggressiveness, amazement, anger, annoyance, anticipation, apprehension, awe, boredom, contempt, disapproval, disgust, distraction, ecstasy, fear, grief, interest, joy, loathing, love, optimism, pensiveness, rage, remorse, sadness, serenity, submission, surprise, terror, trust, vigilance*
Circumplex Russell (Russell, 1980)	*Afraid, alarmed, angry, annoyed, aroused, astonished, at ease, bored, calm, content, delighted, depressed, distressed, droopy, excited, frustrated, glad, gloomy, happy, miserable, pleased, relaxed, sad, satisfied, serene, sleepy, tense, tired*
OCC Ortony (Ortony *et al.*, 1988)	*Admiration, anger, appreciation, disappointment, disliking, fear, fears- confirmed, gloating, gratification, gratitude, happy-for, hope, liking, pity, pride, sorry-for, relief, remorse, reproach, resentment, self-reproach, shame*
Lovheim (Lövheim, 2012)	*Anger/rage, contempt/disgust, distress/anguish, enjoyment/joy, fear/terror, interest/excitement, shame/humiliation, surprise/startle*

Las emociones del lexicón NRC corresponden a las del modelo de Plutchik (Plutchik, 1980) que incluye las emociones de Ekman (Ekman, 1992) de ira, asco, miedo, tristeza, sorpresa, y alegría y añade las emociones de confianza y anticipación. Los valores de valencia de cada emoción y sentimiento fueron obtenidos mediante *crowdsourcing* [15] con Amazon's Mechanical Turk (Mohammad y Turney, 2010).

Ekman (1992) clasifica según su sentimiento las emociones de ira, miedo, asco y tristeza como negativas y las de sorpresa y alegría como positivas. Plutchik (1992) clasifica las emociones contrarias entre sí: alegría/tristeza, anticipación/sorpresa, ira/miedo, asco/confianza. El lexicón NRC no proporciona información sobre qué emociones transmiten un sentimiento positivo o negativo.

Para conocer qué emociones transmiten un sentimiento positivo o negativo en el lexicón NRC, calculamos la polaridad de cada palabra como el valor de valencia cuando la palabra del campo AffectCategory es *positive* menos el valor de valencia cuando la palabra del campo AffectCategory es *negative.* La polaridad puede tomar los valores enteros -1, 0 o 1 correspondientes a los sentimientos negativo, neutro y positivo respectivamente.

Seleccionamos las palabras del lexicón NRC con sentimientos positivos y negativos. Representamos las emociones de cada uno de estos conjuntos mediante un gráfico de

[15] Crowdsourcing: colaboración masiva que prestan individuos que no forman parte de una entidad o institución para realizar un conjunto de tareas. (https://definicion.de/crowdsourcing/, consultado el 15/01/2021)

barras, donde cada barra representa una emoción y su altura es el número de palabras que el valor de valencia en aquella emoción es igual a 1. Analizamos los dos gráficos para clasificar qué emociones predominan cuando los sentimientos son positivos o son negativos.

El contenido del lexicón NRC lo obtenemos mediante el paquete `syuzhet` (Matthew L. Jockers, 2015) de R. Este paquete contiene funciones para poder extraer los valores de polaridad de textos a partir de lexicones propios o de lexicones externos al paquete que pueden ser introducidos por los usuarios. Los valores de valencia de un texto los obtenemos con la función `get_nrc_sentiment.` Si existen unigramas repetidos en un texto, la función `get_nrc_sentiment` solo tiene en cuenta uno y no acumula los valores de valencia de los unigramas repetidos en cada emoción.

Para tener en cuenta la existencia unigramas repetidos y calcular los valores de valencia por emoción de un *tweet,* detectamos los unigramas de los *tweets* obtenidos en el apartado 4.2 Limpieza de textos y clasificados en las temáticas EE y AH. Mediante la función `get_nrc_sentiment` obtenemos sus valores de valencia por emoción. Los valores de valencia por emoción de un *tweet* es la suma de los valores de valencia por emoción de cada uno de los unigramas del *tweet.*

Para cada *tweet* obtenemos ocho valores, cada uno de ellos asociado a una emoción. Cada valor corresponde al número total de unigramas asociadas a aquella emoción dentro el *tweet*. A partir de los valores obtenidos, normalizamos el valor de cada emoción dividiendo su valor por la suma del valor de todas las emociones del *tweet*.

Mostramos un ejemplo ilustrativo del cálculo de las emociones del *tweet* “tbt happy birthday old roomie baby gehl chem love happynewyear” en la Tabla 4.7.5 y Tabla 4.7.6.

Tabla 4.7.5. Cálculo de las emociones de los unigramas de un *tweet*

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
tbt	0	0	0	0	0	0	0	0
happy	0	1	0	0	1	0	0	1
birthday	0	1	0	0	1	0	1	0
old	0	0	0	0	0	0	0	0
roomie	0	0	0	0	0	0	0	0
baby	0	0	0	0	1	0	0	0
gehl	0	0	0	0	0	0	0	0

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
chem	0	0	0	0	0	0	0	0
love	0	0	0	0	1	0	0	0
happynewyear	0	0	0	0	0	0	0	0

Tabla 4.7.6. Cálculo de las emociones totales y normalizadas de un *tweet*

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
Emociones totales	0	2	0	0	4	0	1	1
Emociones normalizadas	0	0,25	0	0	0,5	0	0,125	0,125

Con las emociones normalizadas de todos los *tweets* de las temáticas AH y EE, para cada temática se calcula la suma de los valores de sus respectivos *tweets* en cada emoción. Representamos estos valores mediante un gráfico de barras, donde cada barra representa una emoción y su altura es el valor suma calculado. Analizamos los dos gráficos para clasificar qué emociones predominan teniendo también en cuenta los dos grupos de emociones, las predominantes a sentimientos positivos y las predominantes a sentimientos negativos.

Con los procesos descritos en este apartado, se consigue analizar las emociones de los *tweets* clasificados en las temáticas AH y EE.

4.8 Análisis de usuarios más relevantes

El objetivo del análisis de usuarios más relevantes es detectar quienes son, como están relacionados con el resto y si las organizaciones más relevantes de la química se corresponden con los usuarios más relevantes en el conjunto de *tweets,* incluidos los *retweets*. Adicionalmente, también analizamos los contenidos de los *tweets* relacionados con las organizaciones más relevantes de la química para entender las temáticas que comunican.

La detección de usuarios más relevantes en Twitter puede tener diversas acepciones como actividad, popularidad o influencia y existen diversas métricas para medirla (Riquelme y González-Cantergiani, 2016). En nuestra investigación utilizamos las métricas de número de *tweets,* métricas de análisis de redes y contenidos de los *tweets*, habitualmente usadas en Twitter. Estas métricas y el proceso obtenido para obtenerlas son las que describimos a lo largo de este apartado.

De forma esquemática el análisis de usuarios más relevantes y su comparación con las organizaciones más relevantes de la química sigue el proceso representado en la Figura 4.17.

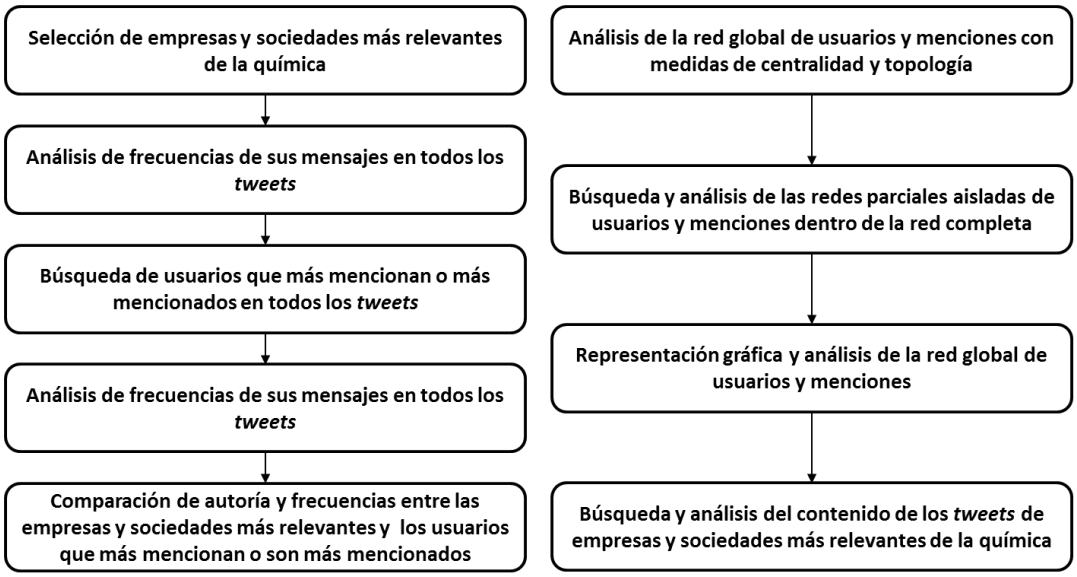


Figura 4.17 Esquema descriptivo del proceso de usuarios más relevantes

Se consideran organizaciones relevantes de la química (Dorronsoro, 2017) las grandes empresas del sector químico y las principales asociaciones y sociedades profesionales de química e ingeniería química que tenían cuentas en Twitter en el periodo entre enero de 2014 y mayo de 2017. Esta lista de empresas y sociedades es la que utilizamos en nuestra investigación y puede consultarse en el Anexo 3.

En el caso de empresas, se escogieron las 50 primeras empresas de la lista de las 100 mayores empresas químicas en 2015 según el volumen de ventas (ICIS Chemical Business, 2016) confeccionada por ICIS (“Independent Commodity Intelligence Services”, https://www.icis.com/explore/). De las 50, solo se consideraron aquellas que tenían alguna cuenta en Twitter. La búsqueda de sus cuentas de Twitter se realizó buscando el nombre de la empresa en Twitter o bien buscando las cuentas de Twitter presentes en la página web de cada empresa.

Si la empresa era una división de una corporación y no poseía cuenta de Twitter, se consideraron las cuentas de Twitter de la corporación. Adicionalmente, la cuenta @DuPont_ability perteneciente a la empresa Dupont y que divulga sus compromisos e

informes de sostenibilidad, también se incorporó a la lista porque estaba vinculada al apartado Sustainability de su página web[16].

En el caso de sociedades profesionales, se generó una lista de asociaciones y federaciones únicas tanto de profesionales como de empresas químicas a partir de la lista de federaciones (Network Science Corp., 2012a), la lista de sociedades profesionales para química analítica (Network Science Corp., 2012b) y la lista de sociedades profesionales para química (Network Science Corp., 2012c) de Network Science Corporation (http://www.netsci.org/) y solo se consideraron aquellas que tenían alguna cuenta en Twitter. La búsqueda de sus cuentas de Twitter se realizó mediante las cuentas de Twitter presentes en la página web de cada sociedad.

En el caso de ACS (American Chemical Society) y RSC (Royal Society of Chemistry) también se consideraron ciertas cuentas adicionales vinculadas a estas asociaciones, ya que son cuentas de divulgación química para públicos específicos. Estas son ACS Green Chemistry (@ACSGCI) y C&EN (@cenmag) para ACS y Chemistry World (@chemistryworld) para RSC.

El nombre de una cuenta en Twitter corresponde con contenido del campo screenName del *tweet*. El contenido de una cuenta empieza con el símbolo "@" seguido de una lista de caracteres. Una mención a un usuario es una cadena de texto de como mínimo un carácter que se encuentra a continuación del símbolo @ y puede contener las letras entre la a y la z, entre la A y la Z, un carácter de barra bajo "_" o cualquier número. El nombre del usuario mencionado corresponde al contenido de la mención sin el símbolo @.

Para cada cuenta de Twitter de las cuentas de empresas y sociedades obtenidas, se buscan los *tweets* del conjunto de *tweets* escritos en inglés con *retweets* en los que el nombre de la cuenta coincide con el contenido del campo screenName del *tweet* o en los que aparece dentro del contenido los *tweets* citados por otros usuarios.

Para medir su relevancia se compara el número de empresas y sociedades tanto que habían enviado *tweets* como que habían sido mencionadas en relación al número respectivo de empresas o sociedades y al número total de usuarios. Asimismo se mide el número de *tweets* con respecto al total de tweets y se confecciona y analiza la

[16] Apartado Sustainability de su pàgina web: http://www2.dupont.com/Sustainability/en_US/ consultado el 23/01/2021

distribución de frecuencias del número de *tweets* para detectar las empresas y sociedades que más *tweets* publican, los usuarios que más mencionan a empresas y sociedades y la relación entre ambos.

Consideramos los usuarios más relevantes como aquellos que más mencionan a otros usuarios o bien que son más mencionados por el resto de usuarios. Para obtenerlos cuantificamos el número de menciones que un usuario realiza a otros y el número de menciones que un usuario tiene en los *tweets* y *retweets* de otros usuarios. Asimismo analizamos las relaciones entre los usuarios que más mencionan y los más mencionados.

Para cuantificar las menciones se analiza el contenido de todos los *tweets* buscando sus menciones y se crea una tabla con tres columnas que contienen la mención del *tweet*, el identificador del *tweet* o campo id del *tweet* y el nombre de usuario que escribió el *tweet* o campo Screename del *tweet.*

Un *tweet* puede tener más de una mención y por tanto más de una fila en la tabla. A partir de los datos obtenidos en la tabla calculamos el número de veces que un usuario menciona a otro usuario, el número de veces que un usuario fue mencionado y el número de menciones totales a ese usuario según quién lo menciona (Casanella, 2019).

Los resultados obtenidos se analizan mediante estadística descriptiva siendo los usuarios más mencionados o que mencionan aquellos que más número de menciones tienen o realizan respectivamente y destacando los cinco primeros en ambos casos. Adicionalmente se busca si existen relaciones entre los cinco usuarios que más mencionan y los cinco más mencionadas.

Finalmente, se compara el nombre de las cuentas de Twitter de los cinco usuarios que más mencionan o más mencionados con la lista de las cuentas de empresas y sociedades profesionales más relevantes de la química para entender si las empresas y sociedades son relevantes dentro de los datos de la investigación.

Para complementar el análisis de usuarios más relevantes y entender su relación con otros usuarios se utilizaron técnicas y procedimientos del área de conocimiento del análisis de redes sociales, en inglés, “Social Network Analysis” (SNA). El análisis de redes sociales incluye la detección, mapeo y medición de las relaciones de fuentes de

información o conocimiento, entre otras, personas, grupos, ordenadores y máquinas, que estén conectadas o relacionadas entre sí (Scott y Carrington, 2011) y en particular, en Twitter permite entender las relaciones entre usuarios (personas u organizaciones) dentro de esta red social (Kwak *et al.*, 2010). En esta investigación nos centramos en el análisis de los usuarios de los *tweets* y sus relaciones, medidas mediante las menciones entre usuarios.

Un sociograma es una forma de representación de redes de individuos u objetos y sus relaciones en la que los individuos se representan mediante nodos o vértices y sus vínculos mediante aristas. Usualmente los sociogramas se dibujan mediante puntos o círculos para los nodos, y líneas o flechas para las aristas si existe una relación entre dos nodos (Sabater y Sierra, 2002; Hansen *et al.*, 2010). La representación gráfica de los sociogramas son útiles para analizar las redes sobre las que están basados (De Laat *et al.*, 2007).

Las redes pueden ser parciales o globales considerando parte o la totalidad de los nodos respectivamente (Scott y Carrington, 2011) y dirigidas o no (Lloyd *et al.*, 1976) según si es relevante o no la dirección de la relación entre dos nodos, representando mediante una flecha la arista entre los nodos si la red es dirigida y mediante una línea si la red es no dirigida.

Una red puede analizarse matemáticamente mediante métricas, y gráficamente mediante su sociograma. Matemáticamente analizamos la red global no dirigida de usuarios de los *tweets* y sus menciones, y las redes parciales formadas por los grupos de usuarios que dentro de la red global solo tienen menciones entre ellos estando aislados del resto de usuarios.

Para construir la red completa de usuarios y menciones consideramos como nodos los usuarios que realizan alguna mención o son mencionados en los *tweets* y *retweets* escritos en inglés. Se considera que existe una arista entre dos nodos si cualquiera de los usuarios que representa cada uno de los nodos realiza una mención a la cuenta del usuario que representa el otro nodo y los usuarios que representan a los dos nodos son diferentes. Si un usuario realiza una mención a sí mismo, esta arista no se considera. Asimismo una arista tendrá más o menos intensidad o peso según el mayor o menor número de menciones entre los nodos.

Para obtener los nodos, calculamos la tabla de usuarios que escriben menciones y la tabla de los usuarios mencionados. Como un usuario puede mencionar y ser mencionado a la vez, unimos las dos tablas y buscamos los usuarios únicos. La lista de usuarios únicos son los nodos de la red. Para obtener las aristas, calculamos la tabla que contiene el número de veces que un usuario que ha escrito un *tweet* realiza menciones a otro usuario y el nombre de los dos usuarios, el que ha escrito el *tweet* y el usuario mencionado. Cada fila de la tabla es una arista y su peso es el número de menciones entre los dos nodos.

A partir de los nodos y las aristas, obtenemos la red global mediante el paquete `igraph` (Nepusz y Csardi, 2006) de R, que permite la definición de redes, operaciones sobre ellas y métricas para su análisis. Con la función `graph_from_data_frame` definimos la red con la lista de nodos y la lista de aristas y con la función `simplify` eliminamos de la red las aristas correspondientes a menciones entre usuarios iguales. Con la función `decompose.graph` de este paquete detectamos y obtenemos las redes parciales de la red global.

Utilizamos dos clases de métricas para analizar las redes resultantes, las que exploran los nodos de la red o métricas de centralidad y las que analizan la topología o forma global de la red. Las métricas de centralidad intentan analizar cuáles son los nodos que ocupan posiciones más céntricas en la red o con características especiales en ésta. En esta investigación se obtuvieron las métricas habitualmente utilizadas (Borgatti, 2005; Aggarwal, 2011) y descritas en la Tabla 4.8.1.

Tabla 4.8.1. Descripción de métricas de centralidad de análisis de redes

Nombre	**Descripción**
Degree centrality **o centralidad de grado o grado**	Medida que calcula el número total de aristas de un nodo ya sea origen o destino. Cuando las redes son dirigidas se puede también diferenciar entre el grado de entrada o *in-degree* en el que solo se consideran las aristas en las que el nodo es destino y el grado de salida o *out-degree* en el que se consideran las aristas en las que el nodo es origen.

Nombre	Descripción
***Eigenvector centrality* o centralidad de vector propio**	Medida para determinar la influencia de un nodo en la red. Amplia el concepto de grado de un nodo y tiene en cuenta no solo el grado de un nodo sino también los grados de los nodos conectados a este. Los nodos que poseen un valor alto de esta medida están conectados a muchos nodos que a su vez están bien conectados al resto de nodos. Como una red se puede expresar mediante la matriz de adyacencia de los nodos de la red, el valor de centralidad de vector propio se calcula como el vector propio de esta matriz (Bonacich, 1972; Newman y Newman, 2010) donde cada elemento del vector corresponde al valor de centralidad de vector propio de un nodo.

Para analizar las redes en su globalidad, atendiendo a su topología o forma se obtuvieron las métricas habitualmente usadas (Scott y Carrington, 2011) descritas en la Tabla 4.8.2.

Tabla 4.8.2. Descripción de métricas de topología de redes

Nombre	Descripción
Densidad de la red	Valor definido como el número de aristas en relación a su máximo número en esa red. La densidad puede tomar cualquier valor entre cero y uno siendo más cercano a uno cuanto más densa sea la red.
Distancia media	Siendo la distancia entre dos nodos la longitud del camino más corto entre ellos medida como el número de vértices mínimo que debe recorrerse para unirlos, la distancia media es el valor medio de la distancia de todos los nodos diferentes de la red.
Diámetro	Medida que representa el tamaño lineal de una red y se calcula como la máxima distancia de todos los pares de nodos de la red.

.

Adicionalmente se analiza el número total de nodos y aristas para entender el tamaño de las redes. Las métricas de centralidad y de topología descritas anteriormente, el número total de nodos y el número total de aristas se obtienen mediante el paquete `igraph` (Nepusz y Csardi, 2006) de R.

Un sociograma y la red en la que está basado pueden representarse gráficamente de formas diferentes según el método de visualización que se utilice y permite la detección, comprensión e identificación de patrones que forman los nodos y sus relaciones en la red. La mayor parte de los métodos de representación de redes en dos dimensiones se pueden categorizar (Gibson *et al.*, 2013) según la dirección de la fuerza o *forced-directed*, en la reducción de la dimensionalidad y en mejoras computacionales. Los *force-directed* representan la red de forma que los nodos se atraen o repelen como en un sistema físico mediante algún tipo de fuerza. Los basados en la reducción de la dimensionalidad se basan en proyectar los datos de la

red de un espacio dimensional alto a uno bajo tratando de mantener el máximo de información posible. Los basados en las mejoras computacionales intentan que los *force-directed* sean más eficientes disminuyendo, por ejemplo, el tiempo computacional necesario que se debe invertir para visualizar redes con un gran número de nodos.

Los *force-directed* suelen ser estéticamente agradables, presentan simetrías, tienden a producir trazados sin cruces entre las aristas (Tamassia, 2013), son los métodos más frecuentemente utilizados y pueden ser aplicados a redes con miles de nodos (Gibson *et al.*, 2013). Dentro de los *forced-directed* existen dos tipos de métodos, los basados en sistemas de muelles y los basados en las soluciones de optimización de problemas. Los basados en sistemas de muelles utilizan un sistema eléctrico o de muelles con el que se intenta buscar un equilibrio global de las fuerzas de cada nodo de la red. Los basados en las soluciones de optimización de problemas intentan minimizar una función global de energía relacionada con la red y sus fuerzas.

En nuestra investigación utilizamos el método force-directed basado en las soluciones de optimización de problemas de Kamada-Kawai (Kamada y Kawai, 1989), inspirado en la tensión de un muelle y que posiciona los nodos en el espacio de forma que la representación gráfica de la red debería aproximarse a la distancia teórica de la red considerada como la longitud más corta de los diferentes caminos entre dos nodos (Gibson *et al.*, 2013). No existen fuerzas de atracción y repulsión entre los nodos. En cambio, si un par de nodos están más cerca (lejos) geométricamente, mediante su distancia euclídea, que su distancia en la red, los nodos se repelen (atraen) entre ellos respectivamente (Kobourov, 2012) para buscar un equilibrio de la posición. El método genera diversas posiciones de los nodos hasta que consigue la mayor estabilidad global de todos los nodos calculada como la diferencia entre la distancia teórica y la de la red. Adicionalmente, prioriza la simetría (Gibson *et al.*, 2013) a través de la función de energía global, que modeliza la falta de simetría de la representación gráfica de la red.

Al estar basado en fuerzas de atracción y repulsión, permite conseguir visualizaciones interpretables en las que las comunidades se pueden detectar visualmente como grupos de nodos (Noack, 2009). Adicionalmente, permite obtener una representación que sea cercana a la definición de una buena representación gráfica (Kobourov, 2012), aunque computacionalmente sea costoso, del orden de entre el cuadrado y el cubo del número de nodos dependiendo del algoritmo de optimización que utilice. Las

bondades del método, así como su acceso público mediante el paquete `ggraph` (Lin Pedersen, 2019) de R que permite la representación gráfica de redes con diversos métodos, nos hizo decantar hacia su uso en esta investigación.

Mediante el método descrito anteriormente, representamos y analizamos la red global a través de diversas visualizaciones con todos los nodos y filtrando un número suficiente priorizando aquellos con mayor número de menciones realizadas por ellos o por el resto para comprender visualmente la forma de cada una de estas redes y la agrupación de sus usuarios.

Finalmente y para acabar el análisis de los usuarios más relevantes, buscamos los *tweets* publicados por las empresas y sociedades de la lista de empresas y sociedades seleccionadas como relevantes en la química, los *tweets* en las que fueron mencionadas por otros usuarios y analizamos la temática de su contenido para comprender qué aspectos comunicaban sobre la química.

En el capítulo siguiente presentaremos los principales resultados obtenidos con estas metodologías.

5 Resultados

A continuación mostramos los principales resultados obtenidos. En el apartado 5.1 recogemos los resultados de los procesos descritos en los apartados 4.1, 4.2, 4.3, 4.4 y 4.5 correspondientes con el primer objetivo de la investigación sobre a qué se refieren los usuarios cuándo hablan de la química en Twitter. En el apartado 5.2 recogemos los resultados de los procesos descritos en los apartados 4.6 y 4.7 correspondientes con el segundo objetivo de la investigación sobre los sentimientos o estados del ánimo y emociones o alteraciones del ánimo que se detectan en los contenidos de los tweets aceptados como relevantes en la imagen pública de la química. En el apartado 5.4 recogemos los resultados los de los procesos descritos en el apartado 4.8 correspondientes con el tercer objetivo de la investigación sobre los usuarios que son relevantes en el conjunto de tweets aceptado y qué coincidencia tienen con organizaciones presumiblemente relevantes en la química. Como se ha descrito en el apartado 4, el código R desarrollado para obtener los resultados de este estudio así como los ficheros de datos puede consultarse como documentación electrónica en la dirección web https://github.com/mguerris/Tesis-Doctoral.git y su descripción y organización puede consultarse en el Anexo 17.

5.1 Resultados de la minería de textos y los clústeres

En este apartado mostramos los resultados de los procesos de adquisición de *tweets*, limpieza de textos de los tweets, su preparación para el *clustering*, el *clustering* y la interpretación los clústeres obtenidos. Con los *tweets* recogidos y siguiendo la metodología descrita en el apartado 4.2 aplicamos los procesos de limpieza descritos, buscamos y descartamos los *retweets*, seleccionamos aquellos escritos en inglés, eliminamos las stopwords de la lista confeccionada y descartamos los *tweets* duplicados y vacíos. Con los *tweets* limpios y con la metodología descrita en el apartado 4.3, obtenemos sus bigramas, confeccionamos la matriz de términos y documentos (TDM) de bigramas, reducimos su dimensionalidad por su frecuencia absoluta, calculamos el factor Tf-Idf (ver Ecuación 4.3.1) de cada documento y bigrama sustituyéndolo por su valor de frecuencia absoluta en la matriz TDM y reducimos la dimensionalidad de la matriz TDM eliminando los documentos que todos sus bigramas tienen un factor Tf-Idf igual a cero.

Con la matriz TDM de bigramas obtenida y teniendo en cuenta los procesos descritos en el apartado 4.4 ejecutamos el método escogido *spherical k-means*, para diferentes valores de *K* para obtener las gráficas de la función criterio Q (ver Ecuación 4.4.1) y del coeficiente de silueta en función de *K* con las que calculamos los valores de *K* óptimos mediante el algoritmo *L-method* y el valor de curvaturas y seleccionamos el valor *K** de clústeres óptimo. Con este valor, calculamos el método *spherical k-means* alrededor de 10 000 veces y seleccionamos la repetición con el valor de la función criterio Q menor.

Con el valor de la función criterio Q menor y aplicando los procesos descritos en el apartado 4.5, confeccionamos los *wordclouds* de los clústeres obtenidos, diseñamos el experimento de bloques incompletos y balanceados (BIBD) para asignar los *wordclouds* a cada uno de los expertos de la química que los clasificaron según las temáticas descritas. La clasificación temáticas de los clústeres, de la que analizamos su bondad mediante el estadístico kappa de Fleiss (Fleiss, 1971) y la calidad de los acuerdos obtenidos mediante el criterio de Landis y Koch (1977), nos permite dar respuesta al primer objetivo de la investigación sobre a qué se refieren los usuarios cuándo hablan de la química en Twitter.

Entre 01/01/2015 y 30/06/2015 adquirimos 256 833 *tweets* y *retweets* con su distribución mostrada en la Figura 5.1 y descritos en las Tabla 5.1.1 según el campo text (ver apartado 4.1) del *tweet*. Este conjunto de datos fue adqurido por el Dr. Jordi Cuadros del departamento de métodos cuantitativos de IQS.

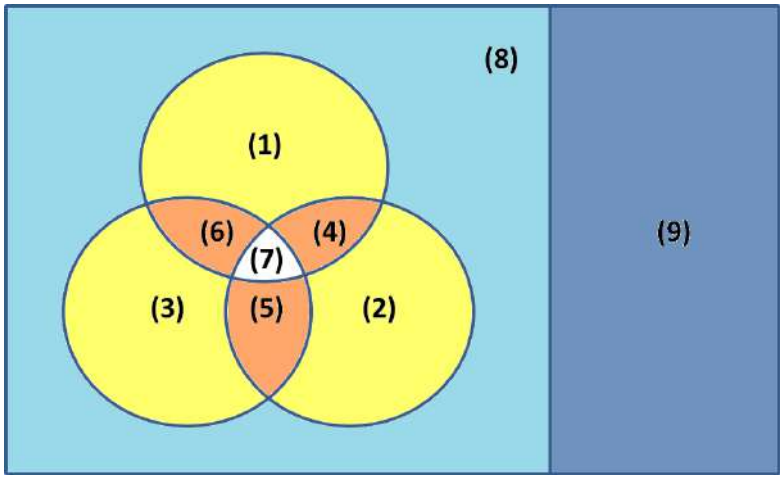


Figura 5.1 Diagrama descriptivo de los *tweets* y *retweets* recogidos

Tabla 5.1.1. Distribución del número de *tweets* en función de las palabras clave de búsqueda

Número	Descripción del texto del *tweet*	Número de *tweets*
(1)	Contiene el unigrama[17] "chemical" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "RT @emojiartworks: MY **CHEMICAL** ROMANCE - THREE CHEERS FOR SWEET REVENGE http://t.co/k0Ajjy3Th0"	66 646
(2)	Contiene el unigrama "chemistry" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "Spending the first day of 2015 doing chemistry....wouldn't wanna spend it doing anything else í ½í¸…í ½í²£"	79 323
(3)	Contiene el unigrama "chem" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "Drowning in AP **Chem** to ring in the new year. Ya nothing has changed..."	54 539
(4)	Contiene los unigramas "chemical" y "chemistry" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "When I was at UCL's chemistry labs we had to checkour facts otherwise the chemical could blow up in our faces https://t.co/IoctyAviia"	328
(5)	Contiene los unigramas "chemistry" y "chem" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "7 years of doing **chemistry** and I managed to get a right answer in a **chem** question in our family pub quiz #praisebetodrmcmurray"	557
(6)	Contienen los unigramas "chemical y "chem" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "Nippon Shokubai files patent infringement lawsuit against LG **Chem** :: **Chemical** Week http://t.co/CoD9pDn9IR"	198
(7)	Contienen los unigramas "chemistry", "chemical" y "chem" en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave Ejemplo: "Last week my **chemistry** teacher asks me what's **chemical** bonding. I say I don't know and she asks me how did I get to grade 12. í ½í¹\u008d I hate **chem**"	4
(8)	Contiene la cadena de texto "chem"[18] en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas y no contiene los unigramas "chemical", "chemistry" o "chem" Ejemplo: ""#TBT and Happy 25th birthday to my old roomie and baby gehl **chem_ee**! Love you! í ½í¸˜í ½í¸˜í ½í¸˜ #HAPPYNEWYEAR http://t.co/1jXdBX6O3B" "**@chem_cake** Ð\u009dÐµ."	13 285
(9)	No contiene[19] la cadena de texto "chem" ni ninguna de los unigramas clave de búsqueda	41 953
	TOTAL = (1)+(2)+(3)+(4)+(5)+(6)+(7)+(8)+(9)	256 833

[17] Unigrama: entendido como palabra delimitada por espacios o por delimitadores de frase o de documento

[18] Cadena de texto "chem": la cadena de texto "chem" puede contener más caracteres antes o después de espacios o delimitadores de frase o documento

[19] No contiene la cadena de texto "chem" en el campo *text* (ver apartado 4.1) del *tweet.* El campo *screenName* (ver apartado 4.1) que contiene el autor del *tweet* contiene la cadena de texto "chem"

El número de tweets con al menos algún *hashtag,* mención, marca HTML[20], dirección URL ("Uniform Resource Locator"), emoticono o emoji contenido en el campo *text* (ver apartado 4.1) de los *tweets* se muestra en la Tabla 5.1.2.

Tabla 5.1.2. Número de *tweets* con al menos algún hashtag, mención, URL, marca HTML, emoticono o emoji

Descripción	Número de *tweets*	Porcentaje de tweets
Hashtags	50 867	19.8%
Menciones	129 354	50.3%
Marcas HMTL	50 388	19.6%
Dirección URL	100 423	39.1%
Emoticonos	6 220	2.4%
Emojis	2 728	1.1%

El análisis *hashtags* arrojó que podían contener información relevante de la química, no siendo así con la menciones, siendo primordialmente menciones a otros usuarios, ni con las marcas HTML correspondientes a códigos sobre como mostrar el texto, ni con las URLs siendo la mayoría de ellas URLs cortas[21]. El número de *tweets* con emoticonos o emojis es muy bajo atendiendo al total de *tweets*. Por estos motivos, el proceso de limpieza implementado mantiene los *hashtags* y descarta el resto.

Aplicamos los procesos ya explicados (ver apartado 4.2) para limpiar los textos de los *tweets* y obtenemos la distribución de *tweets* descrita en la Tabla 5.1.3.

Tabla 5.1.3. Distribución del número de *tweets* en función de los procesos de limpieza de textos

Número	Descripción del proceso	Número de *tweets* (A)	% de *tweets* de (A) sobre los adquiridos (1)
(1)	Adquiridos	256 833	100.0%
(2)	(1) Después de aplicar los procesos de limpieza (ver Tabla 4.2.2)	256 833	100.0%
(3)	(2) Después de buscar y eliminar los *retweets*	175 149	68.2%
(4)	(3) Después de seleccionar los *tweets* escritos en inglés	89 663	34.9%
(5)	(4) Después de eliminar *stopwords* y descartar los *tweets* duplicados y vacíos	76 242	29.7%

En el Anexo 4 presentamos un ejemplo del resultado de la limpieza de los *tweets* así como la detección del idioma de éstos. En la documentación electrónica de la tesis se pueden consultar los ficheros de *dataframes* de los *tweets* adquiridos y los resultantes

[20] HTML: siglas en inglés de *HyperText Markup Language* ('lenguaje de marcas de hipertexto'), hace referencia al lenguaje de marcado para la elaboración de páginas web (https://dictionary.cambridge.org/dictionary/english/html)
[21] URL corta: dirección IP abreviada para dirigir a la misma página que la dirección más larga

después de los procesos de limpieza y preparación, “CyClCm.RData” y “CyClCm_ff.RData” respectivamente así como el fichero del corpus creado con el nombre “Corpus_CyClCm_f.RData”.

Podemos observar que después de los procesos de descarte de *retweets* (3) y de detección de los escritos en inglés (4), el número de *tweets* son aproximadamente un tercio de los *tweets* adquiridos. La distribución de los *tweets* resultantes analizados con los criterios descritos en la Tabla 5.1.1 y según el texto limpio es la descrita en la Tabla 5.1.4.

Tabla 5.1.4. Distribución del número de *tweets* según el el texto limpio aplicados los procesos de limpieza

Número	**Descripción del texto del *tweet***	**Número de *tweets***
(1)	Contiene el unigrama “chemical” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	20 630
(2)	Contiene el unigrama “chemistry” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	28 634
(3)	Contiene el unigrama “chem” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	22 838
(4)	Contiene los unigramas “chemical” y “chemistry” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	175
(5)	Contiene los unigramas “chemistry” y “chem” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	321
(6)	Contienen los unigramas “chemical y “chem” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	56
(7)	Contienen los unigramas “chemistry”, “chemical” y “chem” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas. No contiene el resto de palabras clave	3
(8)	Contiene la cadena de texto “chem” en cualquier parte del texto ya sea en minúsculas o mayúsculas o parte en minúsculas y parte en mayúsculas.y no contiene los unigramas “chemical”, “chemistry” o “chem”	367
(9)	No contiene la cadena de texto “chem” ni ninguna de los unigramas clave de búsqueda	3 218
	TOTAL = (1)+(2)+(3)+(4)+(5)+(6)+(7)+(8)+(9)	76 242

A partir del conjunto de *tweets* depurados obtenemos sus bigramas y confeccionamos la matriz de términos y documentos (TDM) de bigramas (ver Tabla 4.3.1 como ejemplo

de cálculo de la matriz TDM de bigramas). La dimensionalidad de la TDM de bigramas se reduce mediante la eliminación de los bigramas con una frecuencia absoluta total, suma de la frecuencia absoluta del bigrama en todos los *tweets*, inferior a 30. Esta frecuencia representa el percentil 99.72 de la distribución de la frecuencia total de bigramas en todos los *tweets* depurados. Se eliminan también de la TDM de bigramas los *tweets* que quedaron vacíos de bigramas. En la documentación electrónica de la tesis se pueden consultar los ficheros de estas dos matrices, con el nombre "Tdm_f.RData" y "Tdm_ff.RData" respectivamente. La estadística descriptiva de la TDM de bigramas puede consultarse en la Tabla 5.1.5.

Tabla 5.1.5. Estadística descriptiva de las matrices TDM de bigramas de los *tweets* depurados

TDM	Número de *tweets*	Número de bigramas	Frecuencia total de los bigramas en todos los *tweets* depurados				
			Min	Q1	Mediana	Q3	Max
Bigramas	76 242	302 637	1	1	1	1	3 990
Bigramas con frecuencia superior a 29	50 725	864	30*	36	48	81	3 990

*La frecuencia total 30 representa el percentil 99.72 de la distribución de la frecuencia total de bigramas en todos los *tweets* depurados

Como observamos en los resultados, el seleccionar los bigramas con una frecuencia superior a 29 tiene un gran impacto en la matriz reduciéndose el número de bigramas de 302 637 a 864 y el número de *tweets* de 76 242 a 50 725. Está reducción es debida a la baja frecuencia de la mayoría de los bigramas con al menos el 75% de ellos apareciendo solo una única vez en los *tweets* (*Q3*=1).

Calculamos el factor Tf-Idf (ver Ecuación 4.3.1) normalizado de la traspuesta de la matriz TDM de bigramas con frecuencia superior a 29 y se eliminan aquellos bigramas de la matriz en los que el factor es cero en todos los documentos y los documentos vacíos de bigramas. Este proceso no afecta el número de bigramas y documentos de la matriz TDM de bigramas con frecuencia superior a 29. En la documentación electrónica de la tesis se puede consultar el fichero de esta matriz, con el nombre "Dtm_f.RData".

Con la matriz traspuesta de la matriz TDM de bigramas con frecuencia superior a 29 donde cada columna de la matriz representa un *tweet* y es un vector de longitud igual al número de bigramas de todos los *tweets* en el que cada posición del vector contiene el factor Tf-Idf del bigrama en ese *tweet*, repetimos la implementación del método

spherical k-means de *K*=2 a *K*=285 clústeres 50 veces para cada *K*. Los resultados de cada repetición pueden consultarse en la documentación electrónica de la tesis. Seleccionamos la mejor repetición en cada *K* correspondiente al menor valor de la función criterio Q (ver Ecuación 4.4.1) y calculamos su coeficiente de silueta como promedio de los valores de silueta (ver Ecuación 4.4.2). En la documentación electrónica de la tesis se puede consultar el fichero de resultados con el nombre "Best_skm_rep50.RData" y el fichero de los valores de silueta promedio con el nombre "Resum_silouette.RData".

Representamos dos gráficas, el valor de la función criterio Q y el valor del coeficiente de silueta las dos en función de *K* y obtenemos los resultados mostrados en la Figura 5.2 y Figura 5.3.

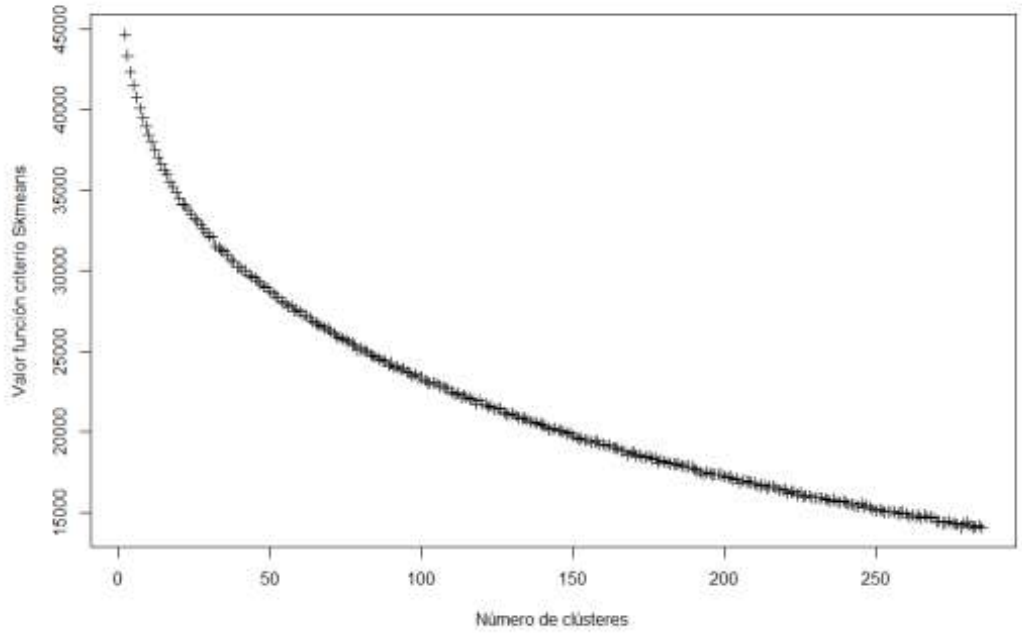


Figura 5.2 Gráfica del valor de la función criterio Q del método spherical k-means en función del número K de clústeres.

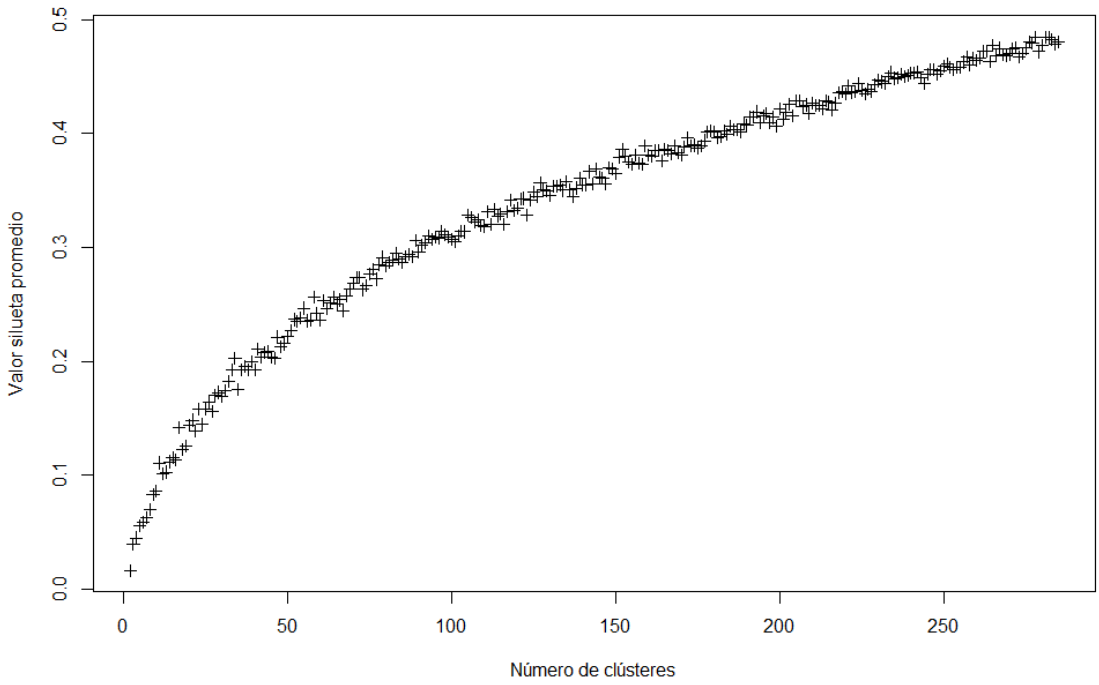


Figura 5.3 Gráfica del coeficiente de silueta en función del número *K* de clústeres

Podemos observar como no existe un codo visualmente claro en ninguna de las dos gráficas anteriores siendo difícil encontrar el mejor número de clústeres. Aplicando el algoritmo *L-method* (ver apartado 4.4) y el cálculo del módulo de curvatura (ver Ecuación 4.4.4) a la función criterio Q y al coeficiente de silueta obtenemos cuatro valores de clústeres óptimos K^*. Los valores (ver Ecuación 4.4.3) del algoritmo *L-method* y los valores del módulo de curvatura en función de K para la función criterio Q y para el coeficiente de silueta los representamos en las gráficas de la Figura 5.4.

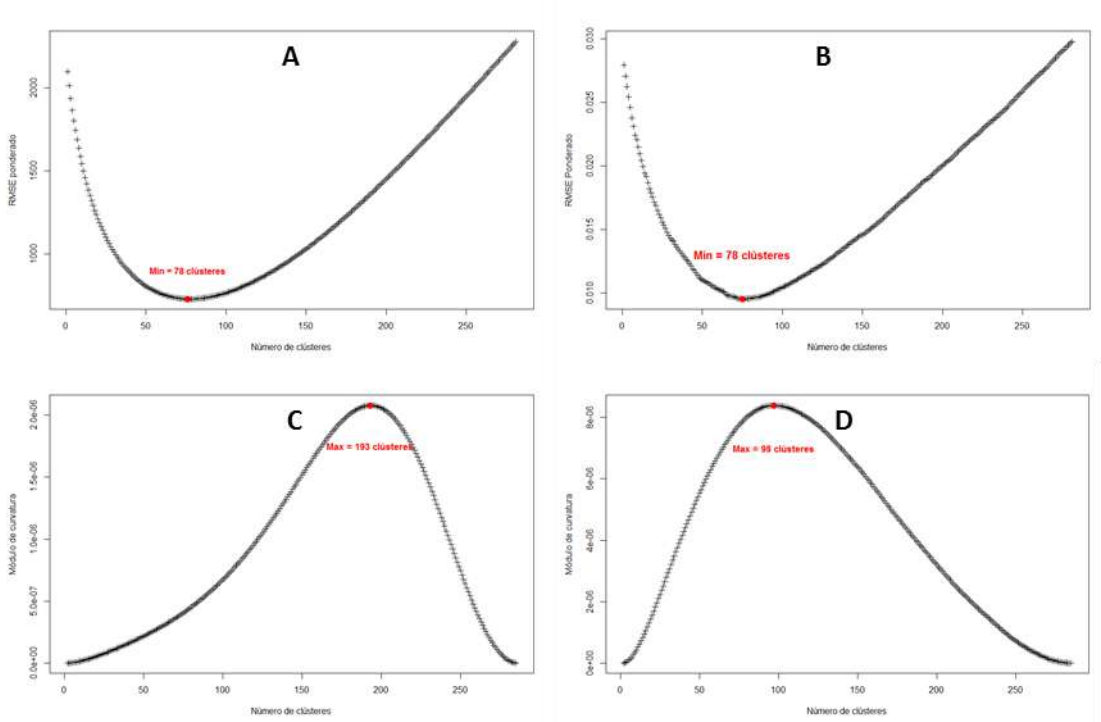


Figura 5.4 Gráficas de resultados: (A) algoritmo *L-method* aplicado a la función criterio Q y (B) al coeficiente de silueta en función del número *K* de clústeres, (C) módulo de curvatura de la función criterio Q y (D) del coeficiente de silueta en función del número *K* de clústeres. Número de clústeres *K óptimos en (A) 78 clústeres, (B) 78 clústeres, (C) 193 clústeres y (D) 98 clústeres**

Observamos en la Figura 5.4 que el número *K** óptimo de clústeres obtenido fue de 78 y 193 para la función criterio Q y de 78 y 98 para el coeficiente de silueta. Como no existe una solución coincidente para el mejor número de clústeres decidimos escoger como mejor valor *K**=100 clústeres. Este es un número cercano a los obtenidos de forma numérica, suficientemente elevado para intentar minimizar los clústeres con temáticas mezcladas y suficientemente pequeño para ser clasificado los expertos de la química y no afecte a la calidad de su clasificación debido a un cansancio excesivo.

Con *K**=100 clústeres repetimos el cálculo del método *spherical k-means* 9 723 veces. Los resultados obtenidos del valor de la función criterio Q para cada repetición en función del número de repetición y la distribución de resultados son los representados gráficamente en la Figura 5.5. Los resultados de cada repetición pueden consultarse en la documentación electrónica de la tesis.

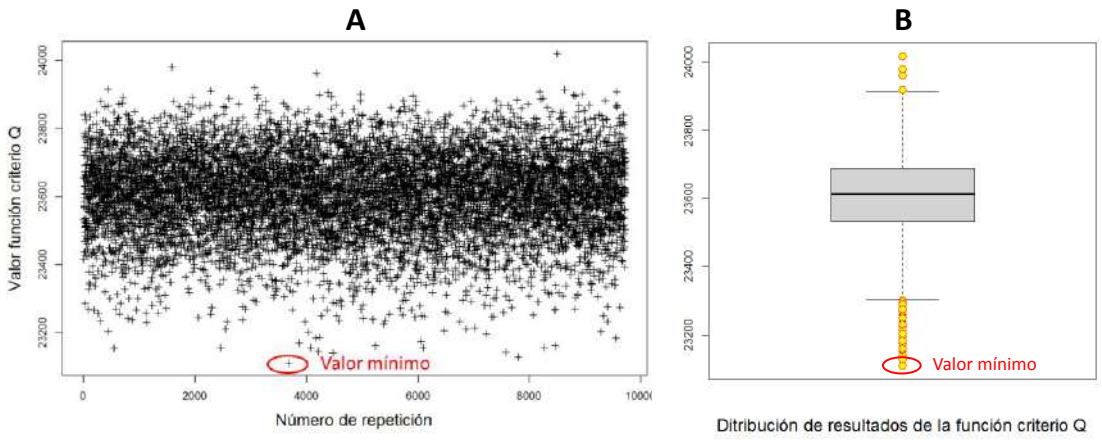


Figura 5.5 (A) Valores de la función criterio Q del método *spherical k-means* para *K=100 clústeres en función del número de repetición. (B) Boxplot de la distribución de los valores de la función criterio Q**

Observamos en la Figura 5.5 que la distribución de los valores de la función criterio Q parece que es simétrica de acuerdo con el boxplot y sugiere estabilidad del método *spherical k-means* con el número de repetición con valores de Q entre 23 000 y 24 000 aproximadamente. No obstante, nuestro interés está en la repetición con un menor valor de la función criterio Q que minimiza la distancia de los *tweets* en los clústeres en función de sus bigramas y los agrupa de forma más compacta, aunque sin poder asegurar que corresponde con un mínimo global de la función criterio Q (ver apartado 4.4). Es por este motivo que repetimos el cálculo del método *spherical k-means* un número elevado de veces. De todas las repeticiones seleccionamos la mejor correspondiente a la repetición con menor valor de la función criterio Q. Esta puede encontrarse en la documentación electrónica de la tesis con el nombre "Bestskm_100clusters.RData". Esta solución tiene las características respecto al número de *tweets* por clúster descritas en la Tabla 5.1.6 y representada en la distribución mostrada en la Figura 5.6.

Tabla 5.1.6. Estadística descriptiva de la distribución del número de *tweets* por clúster de la mejor solución obtenida por el método *spherical k-means*

***Tweets* por clúster**				
Min	**Q1**	**Mediana**	**Q3**	**Max**
95	251	383	647	3476

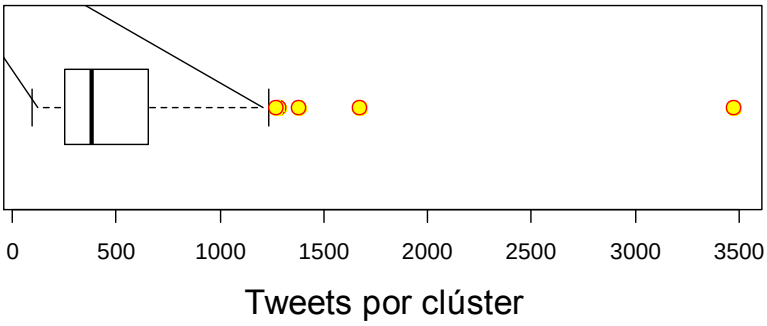


Figura 5.6 Boxplot de la distribución del número de *tweets* por clúster de la mejor solución obtenida por el método *spherical k-means*

Atendiendo a los resultados de la Tabla 5.1.6 y la Figura 5.6 observamos que el 75% de los clústeres (*Q3*) tienen menos de 648 *tweets* y que la mediana es de 383 sobre un total de 50 725 *tweets* clasificados. Asimismo mediante el boxplot observamos que solo cinco clústeres, los marcados en círculos rojos y amarillos, son considerados outliers. Estos cinco clústeres, el 5% de total de clústeres, contienen 1 267, 1 282, 1 377, 1 669 y 3 476 *tweets* que conjuntamente representan el 17.9% del total de *tweets*. La distribución de los *tweets* por clúster sugiere que, aunque el método de *clustering spherical k-means* no garantiza la obtención de clústeres homogéneos, su aplicación conjuntamente con los procesos de limpieza y preparación de textos parece conseguir una mayoría de clústeres de tamaño razonable, disminuyendo el riesgo de obtener clústeres con temáticas mezcladas y ser más fácilmente interpretables.

Para poder clasificar de forma visual en función de las temáticas descritas (ver apartado 4.5) los clústeres resultantes, generamos dos *wordclouds* con los unigramas y los bigramas de los *tweets* asignados a cada uno de los clústeres. Estos *wordclouds* pueden consultarse en el fichero de la documentación electrónica de la tesis. Adicionalmente, en el Anexo 5 puede consultarse a modo de ejemplo los *wordclouds* de los diez primeros clústeres obtenidos. Un ejemplo de éstos es el clúster número 8 con 372 *tweets.* Sus respectivos *wordclouds* de unigramas y bigramas son los representados en la Figura 5.7 y Figura 5.8.

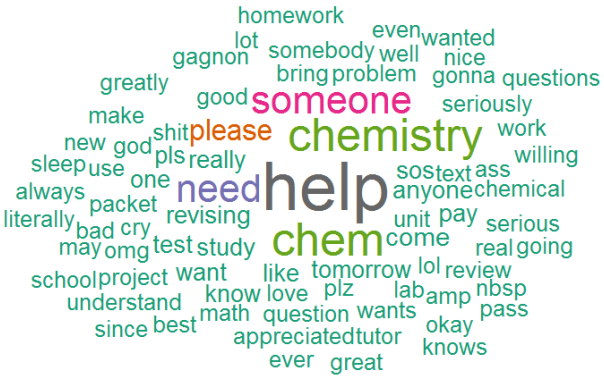


Figura 5.7 *Wordcloud* de unigramas del clúster número 8

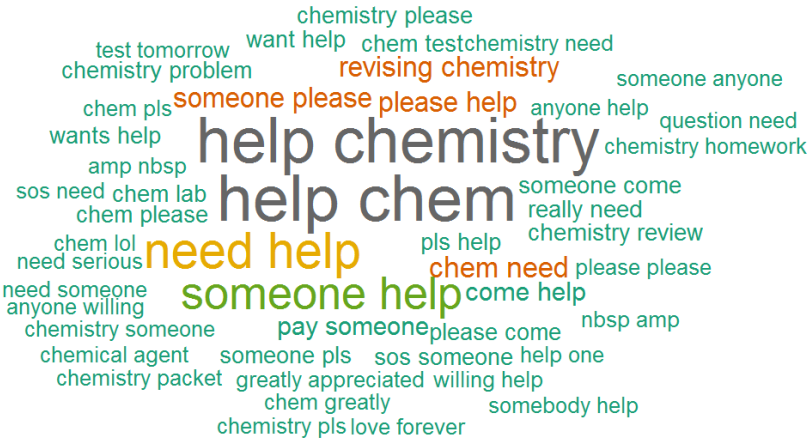


Figura 5.8 *Wordcloud* de bigramas del clúster número 8

Como ya se ha descrito en el apartado 4.2, un *wordcloud* es una representación visual de las palabras que conforman un texto, donde el tamaño es mayor para las palabras que aparecen con mayor frecuencia, es una técnica considerada adecuada para obtener una idea global de los contenidos de un texto y proporcionan mejores resultados (ver apartado 4.5) cuando se representan los bigramas debido al contenido extra sobre el contexto del texto. Como podemos observar en la Figura 5.7 y Figura 5.8, el término “help” en el *wordcloud* de unigramas tiene un tamaño mayor proporcional a su frecuencia de aparición en los *tweets* del clúster pero que sin el *wordcloud* de bigramas donde aparece en los bigramas “help chem” y “help chemistry” más destacados entre otros, su significado es difícil de entender. La paleta de colores fue escogida con colores suficientemente diferentes y el número de términos limitado a

los 100 con mayor frecuencia (ver apartado 4.5) de forma que los *wordclouds* fueran fácilmente interpretables.

Se seleccionaron 18 expertos en función de los criterios definidos en la metodología (ver ver apartado 4.5) para clasificar los clústeres. Los expertos fueron seleccionados atendiendo a su educación similar en química, al alto número de años de experiencia dentro del sector químico y la diversidad de sus conocimientos. La información detallada de cada experto puede consultarse en el Anexo 6.

Para asignar los expertos a los clústeres para clasificar se diseñó un diseño en bloques incompletos y balanceados (BIBD) (ver apartado 4.5). Para conseguir un BIBD factible, generamos diferentes BIBD con el número total de clústeres *n* fijo e igual a 100, variando el número de expertos *m* de 2 a 30 y el número de expertos *k* que evalúan cada clúster de 2 hasta el número de expertos fijado para obtener los valores del número de clústeres evaluados por cada experto *r* y el número de clústeres iguales evaluado por cada par de expertos λ. Posteriormente chequeamos el cumplimiento de las condiciones de cada BIBD (ver Ecuación 4.5.1), obteniendo los resultados descritos en la Tabla 5.1.7 y resaltado en color rojo en el texto y amarillo de fondo el BIBD seleccionado..

Tabla 5.1.7. Lista de potenciales diseños BIBD. Seleccionado el BIBD con color rojo en el texto y amarllo de fondo

Número de expertos (*m*)	Número de clústeres (*n*)	Número de expertos que evalúan cada clúster (*k*)	Número de clústeres evaluados por cada experto (*r*)	Número de clústeres iguales evaluado por cada par de expertos (λ)	Existencia del diseño (V=Verdadero / F=Falso)
2	100	2	100	100	F
3	100	3	100	100	F
4	100	3	75	50	V
4	100	4	100	100	F
5	100	2	40	10	V
5	100	3	60	30	V
5	100	4	80	60	V
5	100	5	100	100	F
6	**100**	**3**	**50**	**20**	**V**
6	100	6	100	100	F
7	100	7	100	100	F
8	100	8	100	100	F
9	100	9	100	100	F
10	100	9	90	80	V
10	100	10	100	100	F

Número de expertos (*m*)	Número de clústeres (*n*)	Número de expertos que evalúan cada clúster (*k*)	Número de clústeres evaluados por cada experto (*r*)	Número de clústeres iguales evaluado por cada par de expertos (*λ*)	Existencia del diseño (V=Verdadero / F=Falso)
11	100	11	100	100	F
12	100	12	100	100	F
13	100	13	100	100	F
14	100	14	100	100	F
15	100	15	100	100	F
16	100	4	25	5	V
16	100	12	75	55	V
16	100	16	100	100	F
17	100	17	100	100	F
18	100	18	100	100	F
19	100	19	100	100	F
20	100	19	95	90	V
20	100	20	100	100	F
21	100	21	100	100	F
22	100	22	100	100	F
23	100	23	100	100	F
24	100	24	100	100	F
25	100	3	12	1	V
25	100	4	16	2	F
25	100	6	24	5	F
25	100	7	28	7	F
25	100	9	36	12	F
25	100	10	40	15	F
25	100	12	48	22	F
25	100	13	52	26	F
25	100	15	60	35	F
25	100	16	64	40	F
25	100	18	72	51	F
25	100	19	76	57	F
25	100	21	84	70	F
25	100	22	88	77	F
25	100	24	96	92	V
25	100	25	100	100	F
26	100	13	50	24	F
26	100	26	100	100	F
27	100	27	100	100	F
28	100	28	100	100	F
29	100	29	100	100	F
30	100	30	100	100	F

Como ninguno de los BIBD con 18 expertos químicos era factible, seleccionamos el BIBD (ver selección en la Tabla 5.1.7) definido por *m* = 6 expertos químicos, *n* = 100

clústeres, k = 3 expertos que evalúan cada clúster, r = 50 clústeres evaluados por cada experto y λ = 20 clústeres iguales evaluado por cada par de expertos.

Se decidió este BIBD por ser factible y para aumentar la calidad de la clasificación de los expertos gracias a la disminución del número de clústeres a clasificar por experto de potencialmente 100 a 50. Se crean tres grupos de expertos, compuestos cada uno ellos por seis expertos, de forma que cada clúster es evaluado nueve veces (tres expertos por clúster multiplicado por tres grupos de expertos). La distribución de clústeres BIBD asignados por experto puede consultarse en la Tabla 5.1.8. Por ejemplo, si observamos el clúster número 1 en la tabla, este será evaluado por tres grupos de expertos, el grupo formado por los expertos 2, 8, 14, el grupo formado por los expertos 4, 10, 16 y el grupo formado por los expertos 6, 12 y 18.

Tabla 5.1.8. Distribución de clústeres asignados por experto (el color gris corresponde al clúster asignado).

Número de Cluster	Experto 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1		■		■		■												
2		■		■		■												
3		■		■		■												
4		■		■		■												
5		■		■		■												
6		■		■		■												
7		■		■		■												
8			■		■	■												
9		■		■		■												
10			■		■	■												
11			■	■		■												
12	■				■	■												
13	■		■			■												
14	■				■	■												
15	■			■		■												
16	■			■		■												
17	■			■		■												
18	■			■		■												
19	■				■	■												
20	■			■		■												
21	■			■		■												
22	■			■		■												
23	■				■	■												
24	■				■	■												
25	■			■		■												
26	■				■	■												
27	■				■	■												
28	■				■	■												
29	■				■	■												
30	■			■		■												
31	■				■	■												
32	■				■	■												
33				■	■	■												
34	■				■	■												
35				■	■	■												
36				■	■	■												
37				■	■	■												
38				■	■	■												
39				■	■	■												
40				■	■	■												
41				■	■	■												
42				■	■	■												
43				■	■	■												
44				■	■	■												
45				■	■	■												
46				■	■	■												
47				■	■													
48				■		■												
49	■			■														
50	■			■														
51	■		■			■												
52	■		■			■												
53	■		■			■												
54	■		■			■												
55	■		■			■												
56		■		■	■	■												
57		■		■	■	■												
58		■	■	■														
59		■	■	■														
60	■	■	■															
61	■	■	■															
62	■	■	■															
63	■		■			■												
64	■		■			■												
65	■	■	■															
66	■		■			■												
67		■	■			■												
68		■	■			■												
69		■		■		■												
70		■	■			■												
71		■		■		■												
72		■	■			■												
73		■		■		■												
74		■		■		■												
75		■				■												
76			■			■												
77			■			■												
78	■	■	■															
79		■	■		■													
80		■	■		■													
81		■	■		■													
82	■	■			■													
83		■	■		■													
84		■	■		■													
85	■	■			■													
86	■	■			■													
87		■	■		■													
88	■	■			■													
89	■	■			■													
90		■	■	■														
91	■	■			■													
92		■	■	■														
93	■	■			■													
94	■	■			■													
95	■	■			■													
96	■	■			■													
97	■	■			■													
98	■	■			■													
99	■	■		■														
100		■	■	■														

Teniendo en cuenta la lista con los nombres de los expertos, cada uno de ellos se asignó de forma aleatoria a cada uno de los números de expertos de esta distribución. Adicionalmente el orden de visualización del número del clúster en el documento que

se entregó a cada uno de los expertos se asignó también aleatoriamente. La tabla final de asignación de clústeres a expertos puede consultarse en el Anexo 7. Los documentos entregados a cada experto pueden consultarse en la documentación electrónica de la tesis.

Una vez que todos los expertos clasificaron sus respectivos documentos, éstos se recogieron para tabular los resultados. Estos resultados pueden consultarse en el Anexo 8. Aplicando el sistema de contaje de las clasificaciones definido en la metodología (ver apartado 4.5) obtenemos los resultados descritos en la Tabla 5.1.9.

Tabla 5.1.9. Resultados de la clasificación porcentual ordenada por clústeres (a la izquierda) y por *tweets* (a la derecha) de todos los expertos en función de la temática asignada

	Porcentaje clasificado			Porcentaje clasificado	
Temática	**Clústeres**	***Tweets***	**Temática**	**Clústeres**	***Tweets***
Entorno Educativo (EE)	41%	35%	Entorno Educativo (EE)	41%	35%
Actividad Humana (AH)	20%	18%	**Indefinido (I)**	18%	21%
Indefinido (I)	18%	21%	Actividad Humana (AH)	20%	18%
Relación Humana (RH)	10%	8%	Entretenimiento (E)	5%	13%
Conocimiento Científico (CC)	6%	5%	Relación Humana (RH)	10%	8%
Entretenimiento (E)	5%	13%	Conocimiento Científico (CC)	6%	5%

Podemos apreciar que el 18% de los clústeres y el 22% de los *tweets* fueron clasificados en la temática Indefinido (I) probablemente y entre diversas causas debido a la dificultad de interpretación de clústeres por la mezcla de temáticas o al criterio de clasificación utilizado. Las temáticas Entorno Educativo (EE) y Actividad Humana (AH) fueron las mayoritariamente clasificadas con un 41% de los clústeres y un 35% de los *tweets*, y un 20% de los clústeres y un 18% de los *tweets* respectivamente. En cambio, las temáticas Conocimiento Científico (CC) y Entretenimiento (E) fueron las menores. Observamos como la temática CC con un 6% de los clústeres y un 5% de los *tweets* sugiere un bajo uso de Twitter para transmitir y comunicar el conocimiento científico más formal.

Para evaluar la bondad de los resultados obtenidos calculamos el estadístico kappa de Fleiss para cada temática, para el experimento global obteniendo los resultados y comparamos los valores de kappa de Fleiss con el *benchmark* de Landis and Koch (1977). Aunque descrita en el apartado 4.5 (ver Tabla 4.5.1), para facilitar la lectura mostramos de nuevo los valores del benchmark de Landis and Koch (1977) en la Tabla 5.1.10. La evaluación de la bondad de los resultados es mostrada en la Tabla 5.1.11.

Tabla 5.1.10. *Benchmark* de Landis and Koch (1977)

Valor de kappa	Bondad del acuerdo
<0.00	*Poor*
0.00-0.20	*Slight*
0.21-0.40	*Fair*
0.41-0.60	*Moderate*
0.61-0.80	*Substantial*
0.81-1.00	*Almost perfect*

Tabla 5.1.11. Resultados del estadístico kappa de Fleiss y comparativa con el benchmark de Landis and Koch (1977) para cada temática y para el experimento global

	Estadístico kappa de Fleiss				
Temática	**Valor**	**Bondad del acuerdo Landis and Koch (1977)**	**Error**	**Z valor**	**p-valor**
Actividad Humana (AH)	0,388	*Fair*	0,017	23,289	$<1x10^{-6}$
Conocimiento Científico (CC)	0,281	*Fair*	0,017	16,887	
Entorno Educativo (EE)	0,517	*Moderate*	0,017	30,994	
Entretenimiento (E)	0,525	*Moderate*	0,017	31,477	
Relación Humana (RH)	0,367	*Fair*	0,017	22,044	
Indefinido (I)	0,103	*Slight*	0,017	6,157	
Global	**0,381**	***Fair***	**0,008**	**44,403**	

A raíz de los resultados obtenidos del estadístico kappa de Fleiss para cada temática y para el experimento global, observamos que todas las temáticas tienen un valor superior a cero y por tanto los resultados del acuerdo entre los expertos analizados por temáticas no son por casualidad, es decir, los expertos muestran un criterio coincidente por alguna razón latente. Adicionalmente, todos los valores de Z tienen un valor de significación p-valor$<1x10^{-6}$ y por tanto los resultados son estadísticamente significativos respecto la hipótesis nula, que el acuerdo entre los expertos sea fruto de la casualidad.

La comparación de los valores de kappa de Fleiss con el benchmark de Landis and Koch (1977) sugiere también que el acuerdo es como mínimo pequeño (*slight)* en todas las temáticas y justo (*fair*) y moderado (*moderate*) en las temáticas AH y EE respectivamente, temáticas que se espera contengan la mayoría de términos de la imagen pública de la química e indicios de la quimiofobia. Esto resultados en estas dos temáticas sugieren que el acuerdo es bastante bueno con un bajo nivel de casualidad entre los expertos.

En resumen, después de adquirir, limpiar y preparar los *tweets* para el proceso de *clustering*, hemos agrupado con el método *spherical k-means* los *tweets* en clústeres que han podido ser mayoritariamente clasificados por 18 expertos químicos en las temáticas de Actividad Humana (AH) y Entorno Educativo (EE), no siendo su resultado por azar y como mínimo justo según el el *benchmark* de Landis and Koch (1977).

5.2 Resultados del análisis de sentimientos

En este apartado y en el siguiente mostramos los resultados del análisis de sentimientos y emociones de los los *tweets* de las temáticas AH (actividad humana) y EE (entorno educativo), aceptados como los que presumiblemente contienen la mayoría de *tweets* relacionados con la imagen pública de la química en Twitter. Con estos *tweets* y aplicando los procesos descritos en el apartado 4.6, detectamos los unigramas de las temáticas AH y EE presentes en el lexicón definido a partir del lexicón SentiWordNet 3.0 y calculamos su polaridad, valor de su sentimiento. Confeccionamos y analizamos los *wordclouds* comparativos de los *tweets* con polaridad positiva y negativa y seleccionamos y analizamos una muestra aleatoria representativa de aquellos *tweets* de la temática EE con sentimiento positivo que contenían los unigramas más frecuentes atendiendo a que las conclusiones del marco teórico no parecen sugerir un sentimiento positivo en esta temática. Asismismo analizamos los *tweets* con polaridad neutra mediante sus porcentajes con respecto al total de tweets y confeccionamos y comparamos los términos más frecuentes de sus *wordclouds* para entender si existe un posicionamiento de sentimientos polarizado o intermedio.

Las emociones de los los *tweets* de las temáticas AH (actividad humana) y EE (entorno educativo) aceptados las obtenemos aplicando procesos descritos en el apartado 4.7. Detectamos los unigramas de los tweets *coincidentes en* lexicón definido a partir del lexicón de emociones NRC y obtenemos las emociones presentes en los tweets así como el sentimiento positivo o negativo que estas emociones transmiten. De esta forma, damos respuesta al segundo objetivo sobre los sentimientos o estados del ánimo y emociones o alteraciones del ánimo que se detectan en los contenidos de los tweets aceptados como relevantes en la imagen pública de la química.

Como se describe en el apartado 4.6 de la metodología, el lexicón seleccionado en nuestra investigación es el SentiWordNet 3.0 debido a tener una mayor cantidad de

términos tanto positivos como negativos que los lexicones mayormente utilizados (ver tabla Tabla 4.6.1), tener un cierto equilibrio entre el número de términos positivos y negativos, ser accesible en R y ser también uno de los habitualmente referenciados en la literatura.

El lexicón SentiWordNet 3.0 permite la obtención la polaridad de un *tweet* como la suma de las polaridades de las palabras del *tweet* encontradas dentro del lexicón. Un valor de polaridad del *tweet* positivo implica un sentimiento positivo, un valor negativo implica un sentimiento negativo y un valor 0 implica un sentimiento neutro.

Para adecuarlo a nuestra investigación, modificamos las 117 659 palabras agrupadas y sus polaridades del lexicón SentiWordNet 3.0 para obtener una lista de 146 842 únicas palabras asociadas cada una a una polaridad. Un extracto del lexicón resultante puede consultarse en el Anexo 9. Tanto el lexicón SentiWordNet 3.0 como el lexicón final completo pueden consultarse en la documentación electrónica de la tesis en los archivos "SentiWordNet_3.0.0_20130122.txt" y "Sentiword_mod.RData" respectivamente. Las características del lexicón resultante pueden consultarse en la Tabla 5.2.1, Tabla 5.2.2 y Figura 5.9.

Tabla 5.2.1. Características del lexicón utilizado en el análisis de sentimientos Sentiword_mod.RData y obtenido a partir del lexicón SentiWordNet 3.0

	Número	**Porcentaje**
Palabras con polaridad negativa	20 288	13,8%
Palabras con polaridad neutra	108 965	74,2%
Palabras con polaridad positiva	17 589	12,0%
Total	**146 842**	**100,0%**

Observamos como la mayoría de las palabras tienen un valor de polaridad neutra. Solo un 25,8%, es decir, 37 877 palabras tienen una polaridad diferente. Si solo consideramos estas últimas, sus características estadísticas según si las palabras tienen polaridad positiva o negativa son las representadas en la Tabla 5.2.2.

Tabla 5.2.2. Estadística descriptiva de las palabras con polaridad positiva o negativa del lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0

	Min	Q1	Mediana	Media	Q3	Max
Palabras con polaridad positiva	0,003	0,125	0,187	0,234	0,350	1,0
Palabras con polaridad negativa	-1,0	-0,375	-0,250	-0,288	-0,125	-0,002
Palabras con polaridad positiva o negativa	-1,0	-0,250	-0,062	-0,0455	0,125	1,0

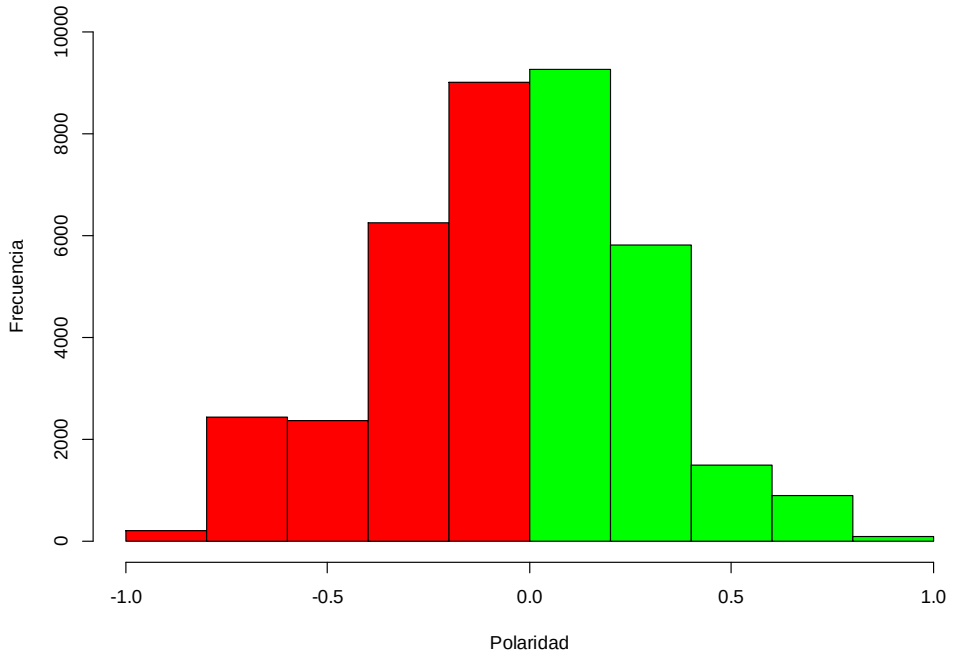


Figura 5.9 Histograma de la distribución de las palabras del lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0 con polaridad positiva (en verde) y negativa (en rojo)

Podemos observar en la Tabla 5.2.2 que como aunque la mediana y la media de las palabras positivas y negativas tienen polaridad negativa estos valores están cercanos al cero y el histograma la distribución de las palabras positvas y negativas representado en la Figura 5.9 es bastante simétrica, parece que el lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0 no está sesgado. Esto nos da una cierta tranquilidad con respecto a la fiabilidad de los resultados.

Detectamos los unigramas de los 9 174 *tweets* de la temática AH y de los 19 387 de la temática EE. Obtenemos 80 085 unigramas en la temática AH y 150 244 en la temática EE. Con el lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0 detectamos las palabras coincidentes con las palabras de los

tweets. Obtenemos 57 620 unigramas detectados en la temática AH y 112 874 en la temática EE que representan un 72% de los unigramas totales de la temática AH y un 75% de los unigramas totales de la temática EE, siendo ambos porcentajes bastante elevados. Adicionalmente solo 15 *tweets* (un 0,16%) de los 9 174 de la temática AH y 33 *tweets* (un 0,17%) de los 19 387 de la temática EE son *tweets* en los que no se detectó ninguna palabra.

Calculamos la polaridad de los *tweets* de las temáticas AH y EE mediante la suma de polaridades de las palabras coincidentes en los *tweets* con las del lexicón y clasificamos los *tweets* en positivos, neutros o negativos según su polaridad superior, igual o inferior a cero respectivamente.

A modo de ejemplo mostramos en la Tabla 5.2.3 los *tweets* de AH con mayor valor de polaridad positiva y negativa que contienen los bigramas "expert killed", "chemical free", "surveys marketresearchreports"[22] y "forecasts marketing", la categorización de palabras según su polaridad y la polaridad del tweets obtenida.

Tabla 5.2.3. Ejemplo de clasificación de los *tweets* en función de su polaridad. Palabras con polaridad positiva en verde, negativa en rojo, neutra en naranja y palabras no encontradas en el lexicón en negro

Bigrama	***Tweet* limpio con la polaridad de cada palabra**	**Polaridad total del *tweet***
expert killed	***chemical*** *weapons* ***expert*** *killed* ***coalition*** *airstrike* ***battle*** ***really*** *begins* ***know***	Positivo
chemical free	***awesome*** ***sustainable*** ***design*** ***pollution*** *busting* ***billboard*** *grows* ***chemical*** ***free*** ***organic*** *vegetables*	Positivo
surveys marketresearchreports	***chemical*** ***phosphoric*** ***acid*** ***mono*** ***alkyl*** *esters surveys marketresearchreports* ***forecast*** *mrx*	Negativo
forecasts marketing	***chemical*** ***acid*** ***mordant*** ***brown*** *surveys marketresearchreports forecasts* ***marketing*** *mrx* ***market*** ***research***	Negativo

Destacar como ejemplo como "free" y "acid" son términos evaluados como negativos por el lexicón debido su definición con polaridad negativa en diferentes contextos. El contexto de "free" incluye enlaces no ligados en una molécula y capaces de cierto movimiento, no estar ocupado o en uso, no estar fijo en una posición y ser capaz de dañar alguien y no estar sometido a servidumbre en la época de la Guerra Civil americana. El contexto de "acid" incluye compuestos solubles en agua que pueden dañarla, tener las características de un ácido y una reacción ácida.

[22] El bigrama "surveys marketresearchreports" corresponde a la unión del unigrama "surveys" con el *hashtag "*#MarketResearchReports"

Un extracto de los resultados puede consultarse en el Anexo 10. Los resultados completos pueden consultarse en la documentación electrónica de la tesis en los archivos con nombre "Polaridad_AH.csv" y "Polaridad_EE.csv". La distribución del número absoluto y en porcentaje de *tweets* sobre el total de *tweets,* 9 174 de la temática AH y 19 387 de la temática EE, según su polaridad está reflejado en la Tabla 5.2.4 y Tabla 5.2.5.

Tabla 5.2.4. Distribución del número de *tweets* de AH y EE según su polaridad calculada con el lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0

	Con alguna palabra incluida en el lexicón			
	Polaridad negativa	**Polaridad neutra**		**Polaridad positiva**
		Contienen palabras con polaridad positiva y negativa	**Solo contienen palabras con polaridad neutra**	
Actividad Humana (AH)	3 326	100	626	5 107
Entorno Educativo (EE)	6 652	103	521	12 528

Tabla 5.2.5. Distribución porcentual[23] de los *tweets* de AH y EE sobre el total de tweets, 9 174 de la temática AH y 19 387 de la temática EE, en cada temática según su polaridad calculada con el lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0

	Con alguna palabra incluida en el lexicón			
	Polaridad negativa	**Polaridad neutra**		**Polaridad positiva**
		Contienen palabras con polaridad positiva y negativa	**Solo contienen palabras con polaridad neutra**	
Actividad Humana (AH)	36,2%	1,1%	6,8%	55,7%
Entorno Educativo (EE)	33,5%	0,5%	2,6%	63,2%

Observamos en la Tabla 5.2.4 y Tabla 5.2.5 que los *tweets* con polaridad neutra son pocos en comparación con el total de *tweets* sólo representando un 7,9% y un 3,1% en las temáticas AH y EE respectivamente. La mayoría de *tweets* son o bien positivos o negativos en ambas temáticas.y predominan los *tweets* con polaridad positiva sobre los *tweets* con polaridad negativa ambas temáticas (55,8% vs 36,3% en AH y 63,2%

[23] La suma de los porcentajes no es igual a 100% debido a que no están contemplados los *tweets* que no contienen ninguna palabra en el lexicón. El porcentaje redondeado en la temática AH y EE es del 0,2%

vs 33,5% en EE) y por tanto con sentimientos positivos predominando sobre los negativos.

La distribución estadística del valor de las polaridades de los *tweets* con alguna palabra detectada es la reflejada en la Tabla 5.2.6.

Tabla 5.2.6. Estadística descriptiva del valor de polaridad de los *tweets* de AH y EE con alguna palabra detectada por el lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0

	Min	**Q1**	**Mediana**	**Media**	**Q3**	**Max**
Actividad Humana (AH)	-2,49	-0,08	0,03	0,07	0,24	2,88
Entorno Educativo (EE)	-2,26	-0,07	0,07	0,11	0,30	2,29

Observamos en la Tabla 5.2.6 que los valores extremos de polaridad de los tweets son son -2,49 y +2,88 en la temática AH y -2,26 y +2,29 en la temática EE. Estos valores son inferiores a -1 y superiores a +1 debido a que calculamos el valor de polaridad del *tweet* como la suma de las polaridades de los términos coincidentes entre el tweet y el lexicón Sentiword_mod.RData obtenido a partir del lexicón SentiWordNet 3.0. Como la polaridad de un término de un *tweet* puede estar entre -1 y +1, el valor de la polaridad de un *tweet* puede ser superior a +1 e inferior a -1. También observamos que ambas temáticas presentan medias y medianas ligeramente superior a cero.

Parece que en ambas temáticas tanto en los porcentajes de *tweets* (ver Tabla 5.2.5) como en los valores de polaridad (ver Tabla 5.2.6), la polaridad positiva es superior a la negativa. Por tanto, parece que la percepción de sentimientos positivos es superior a los negativos.

Para comparar los contenidos de los *tweets* positivos y negativos de cada una de las temáticas AH y EE los representamos mediante sus respectivos *wordclouds* comparativos de unigramas y bigramas. Los *wordclouds* comparativos de la temática AH son los reflejados en la Figura 5.10 y Figura 5.11, y los de la temática EE en la Figura 5.12 y Figura 5.13.

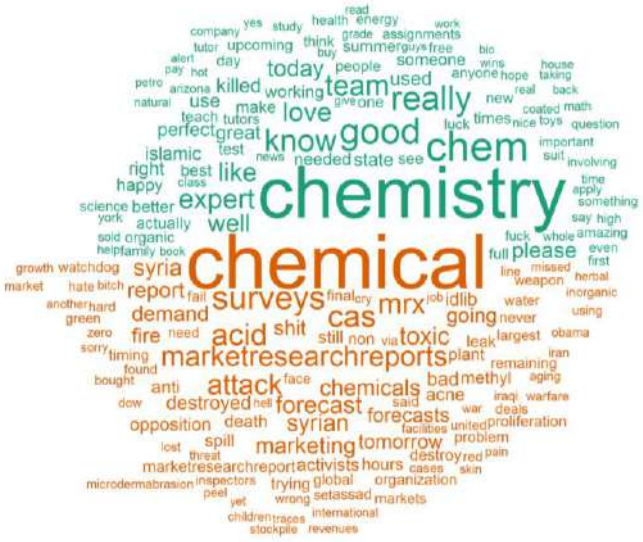


Figura 5.10 *Wordcloud* comparativo de unigramas correspondientes a los *tweets* con polaridad positiva (unigramas en verde) y negativa (unigramas en rojo) de AH

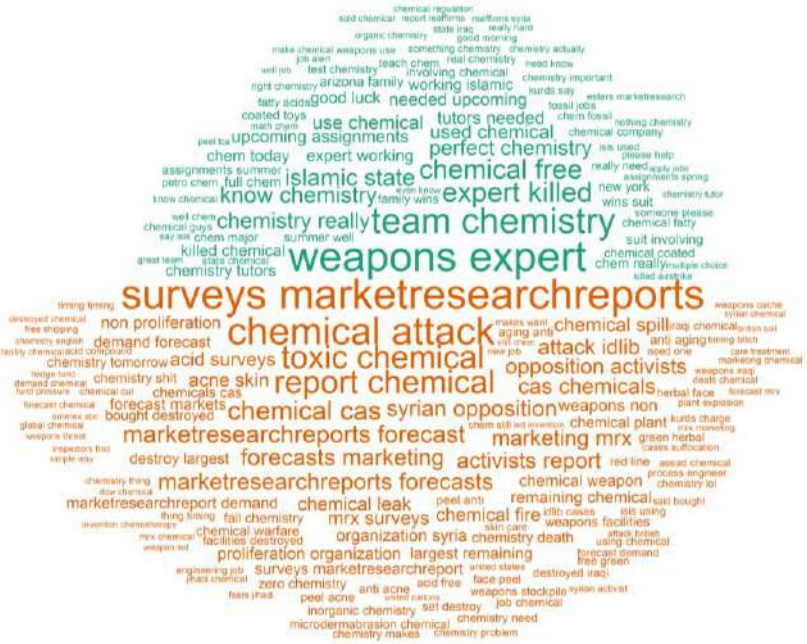


Figura 5.11 *Wordcloud* comparativo de bigramas correspondientes a los *tweets* con polaridad positiva (bigramas en verde) y negativa (bigramas en rojo) de AH

El análisis visual de los *wordclouds* comparativos de AH reflejados en la Figura 5.10 y Figura 5.11 sugiere que los términos como “attack”, “syria”, “chemical attack”, ”syrian opposition”, “chemical warfare” y “toxic”, “toxic chemical”, “chemical fire”, “chemical leak” de los *tweets* clasificados como negativos predominan sobre el resto de los negativos. Estos términos son susceptibles de alimentar actitudes relacionadas con la quimiofobia. La existencia de términos como “chemical free” y “used chemical” en *tweets* clasificados como positivos parecen reforzar este efecto. Observamos también como el término “chemical” resalta dentro de los *tweets* negativos mientras el término “chemistry” destaca en los positivos y como los términos “marketresearch report”, “surveys” o “forecasts” están contenidos dentro de *tweets* negativos cuando aparentemente deberían sugerir un contenido positivo. En cambio, en los *tweets* positivos la aparición de términos como “good”, “expert”, “team”, “team chemistry” y “weapons expert” con un sentido razonablemente favorable y positivo con la química, parecen contraponerse con términos como “expert killed” y “islamic state”

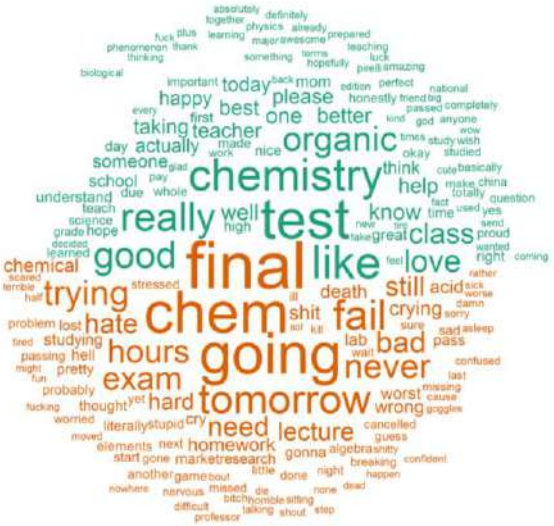


Figura 5.12 *Wordcloud* comparativo de unigramas correspondientes a los *tweets* con polaridad positiva (unigramas en verde) y negativa (unigramas en rojo) de EE

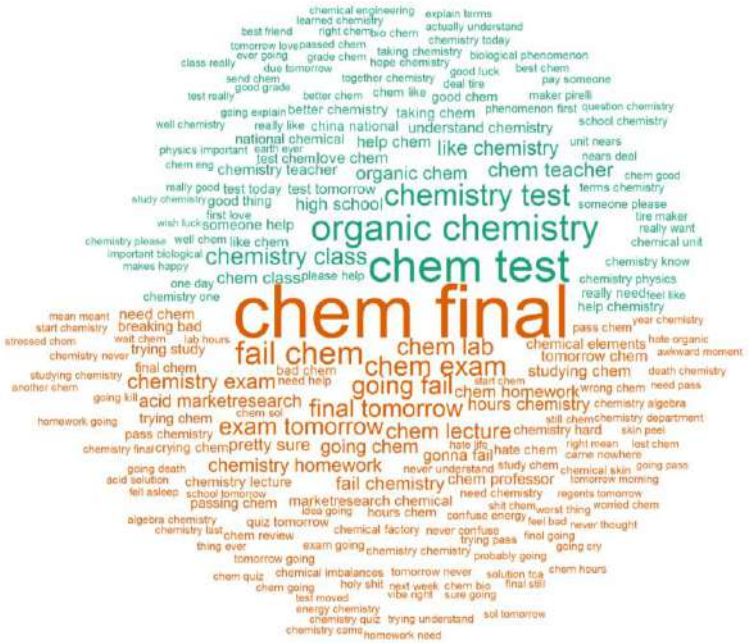


Figura 5.13 *Wordcloud* comparativo de bigramas correspondientes a los *tweets* con polaridad positiva (unigramas en verde) y negativa (unigramas en rojo) de EE

El análisis visual de los *wordclouds* comparativos de la temática EE en la Figura 5.12 y Figura 5.13. sugiere la dificultad de la química como curso académico con términos relacionados con las actividades académicas como “final”, “exam”, “chem final”, “final tomorrow”, “quiz tomorrow”, “lab”, “lecture”, “chem lab”, “chem lecture” y términos relacionados con sentimientos como “hate”, “hard”, “crying“, “need help”, “never understand” y “chemistry hard” en *tweets* clasificados como negativos. Esta dificultad se ve reforzada con la presencia de términos en *tweets* positivos como “someone help” y “help chemistry”. Aunque también términos como “test”, “chem test”, “chemistry test”, “test tomorrow”, “teacher”, “chem teacher”, “chemistry teacher” y “chemistry class” aparecen también en *tweets* positivos siendo contrapuestos a los negativos anteriormente descritos.

Como se describe en el marco teórico, el sentimiento de la química es negativo en la temática EE y no parecería que en esta temática debieran existir muchos *tweets* con sentimiento positivo, aunque como se describe en el apartado 4.6 de la metodología, el uso de un lexicón puede en el análisis de sentimientos puede hacer que un unigrama o bigrama tenga una polaridad negativa y el *tweet* donde aparece positiva.

Para entender con mayor detalle los *tweets* positivos de la temática EE que sus unigramas y bigramas creemos que deberían aparecer en los *tweets* negativos, seleccionamos una muestra representativa de los *tweets* positivos que contenían los términos “test” y “teacher” ya que de esta forma incluíamos a la mayoría de unigramas y bigramas más frecuentes en estos *tweets*.

Teniendo en cuenta los 3 203 y 1 208 *tweets* positivos correspondientes a los términos “test” y “teacher” en la temática EE respectivamente, obtuvimos unas muestras de 344 y 292 *tweets* utilizando la fórmula de cálculo de la Ecuación 4.6.1. correspondientes a la muestra para una proporción con corrección de población finita descrita en la metodología en el apartado 4.6.

Seleccionamos aleatoriamente las muestras del conjunto de los *tweets* originales y las revisamos visualmente clasificando cada *tweet* en positivo, neutro, negativo o sin clasificar (ver Tabla 4.6.4 del apartado 4.6 de la metodología) según el sentimiento que transmitía y en ironía o no según si el contenido era irónico[24] o no (ver apartado 4.6 de la metodología). Un extracto de los resultados de esta clasificación pueden consultarse en el Anexo 11. El resultado detallado puede consultarse en el archivo “Clasificacion muestra tweets EE.csv” de la documentación electrónica de la tesis.

El resumen de la clasificación resultante expresado en porcentaje del total de cada muestra es la reflejada en la Tabla 5.2.7.

Tabla 5.2.7 Resultados de la clasificación visual de las muestras de *tweets* positivos originales de la temática EE que contienen los términos “test” y “teacher"

Clasificación	***Tweets* con el término “test”**	***Tweets* con el término “teacher”**
Positivos	16%	14%
Neutros	36%	41%
Negativos	39%	32%
Sin clasificar	9%	13%
Irónicos	22%	22%
No irónicos	78%	78%

Observamos en los resultados de la Tabla 5.2.7 que existen un porcentaje elevado de *tweets* de las muestras que siendo clasificados como positivos por el lexicón utilizado

[24] Ironía: Expresión que da a entender algo contrario o diferente de lo que se dice, generalmente como burla disimulada (https://dle.rae.es/iron%C3%ADa, consultado el 14/01/2021).

en la investigación, su revisión visual parece que transmiten o bien sentimientos neutros o bien sentimientos negativos.

Si calculamos la tabla cruzando el criterio de clasificación positivo, negativo, neutro, y sin clasificar con el criterio de clasificación irónico y no irónico, obtenemos los resultados reflejados en la Tabla 5.2.8.

Tabla 5.2.8 Resultados de la clasificación visual de las muestras de *tweets* positivos originales de la temática EE que contienen los términos "test" y "teacher" según los dos tipos de clasificación cruzados

Clasificación	***Tweets* con el término "test"**		***Tweets* con el término "teacher"**	
	Irónicos	No irónicos	Irónicos	No irónicos
Positivos	2%	14%	2%	12%
Neutros	1%	35%	6%	35%
Negativos	13%	26%	8%	24%
Sin clasificar	6%	3%	6%	7%

Observamos en los resultados de la Tabla 5.2.8 que en ambas muestras, el mayor porcentaje de *tweets* que parecen contener ironías corresponden a *tweets* con sentimientos aparentemente negativos, siendo las ironías no detectables con el método de análisis de sentimientos utilizado en la investigación.

Adicionalmente y como se describe apartado 4.6 de la metodología, analizamos la información contenida en los *tweets* de AH y EE con polaridad neutra mediante la representación de sus respectivos *wordclouds* de unigramas y bigramas. Los *wordclouds* de unigramas y bigramas con polaridad neutra de la temática AH son los reflejados en la Figura 5.14 y Figura 5.15. Los *wordclouds* de unigramas y bigramas con polaridad neutra de la temática EE son los reflejados en la Figura 5.16 y Figura 5.17.

Figura 5.14 *Wordcloud* de unigramas correspondientes a los *tweets* con polaridad neutra de la temática AH

Figura 5.15 *Wordcloud* de bigramas correspondientes a los *tweets* con polaridad neutra de la temática AH

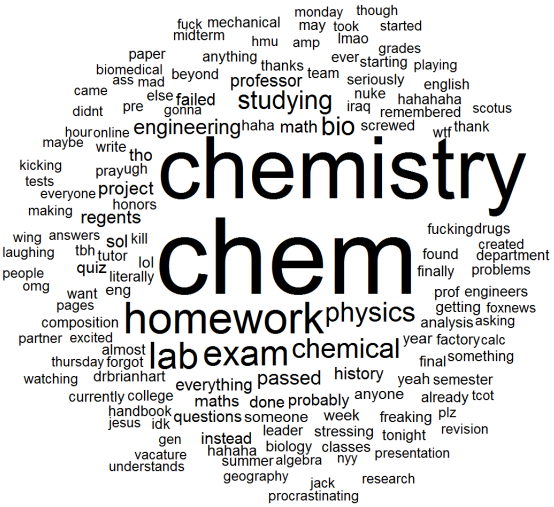


Figura 5.16 *Wordcloud* de unigramas correspondientes a los *tweets* con polaridad neutra de la temática EE

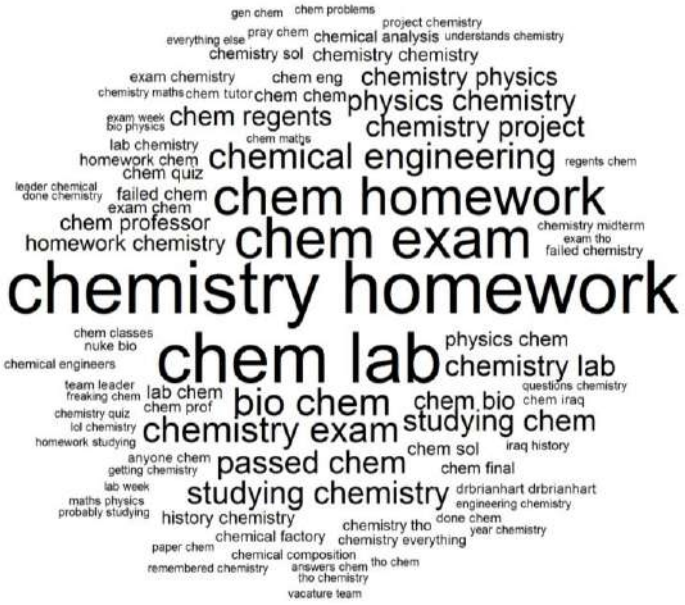


Figura 5.17 *Wordcloud* de bigramas correspondientes a los *tweets* con polaridad neutra de la temática EE

Aunque el número de *tweets* neutros es pequeño en comparación con la suma de los positivos y negativos (726 contra 8 433 en el caso de la temática AH y 624 contra 19 180 en el caso de la temática EE) observamos como términos que se consideraban en el análisis de *tweets* positivos y negativos como “forecasts” o “marketresearchreports” en la temática AH (ver Figura 5.14 y Figura 5.15) y “exam” o “homework” en la temática EE (Figura 5.16 y Figura 5.17) también aparecen en los *tweets* neutros. No

obstante, este bajo número de *tweets* neutros parece sugerir que no existe un posicionamiento intermedio respecto a estos términos sino al contrario polarizado ya sea en positivo, negativo o ambos.

En resumen, adecuando el lexicón *SentiWordNet 3.0* a las necesidades de esta investigación, hemos podido calcular las polaridades de la mayoría de los unigramas de los *tweets* y de los *tweets* clasificados en las temáticas Actividad Humana (AH) y Entorno Educativo (EE), obteniendo pocos *tweets* con polaridad neutra, y mediante su representación gráfica con *wordclouds* comparativos de los unigramas y bigramas de los *tweets* con polaridad positiva y negativa hemos podido detectar términos susceptibles de alimentar actitutdes relacionadas con la quimiofobia en la temática AH y términos que sugieren la dificultad de la química como curso académico en la temática EE. En la temática EE, el análisis de una muestra representativa de los unigramas y bigramas más frecuentes con polaridad positiva sugiere unos sentimientos también negativos.

5.3 Resultados del análisis de emociones

Como se describe en el apartado 4.7 de la metodología, el lexicón seleccionado en nuestra investigación es el lexicón de valencia de emociones NRC versión 0.92 (NRC v0.92), debido a ser un lexicón ampliamente referenciado en la literatura, ser accesible mediante varios paquetes de R y tener un lista importante de palabras.

El lexicón NRC v0.92 proporciona para cada palabra el valor cero o uno de ocho emociones, ira (*anger),* disgusto *(disgust),* miedo *(fear),* tristeza *(sadness),* anticipación *(anticipation),* sorpresa *(surprise),* alegría *(joy),* confianza *(trust),* siendo cero si la palabra no está asociada a la emoción y uno en caso contrario. Nos permite obtener las emociones de un *tweet* localizando las palabras del *tweet* coincidentes en el lexicón y sumando los valores de las emociones de cada palabra coincidente.

Un extracto de este lexicón puede consultarse en el Anexo 12. El lexicón completo puede consultarse en la documentación electrónica de la tesis en el fichero “NRC-Emotion-Lexicon-Wordlevel-v0.92.txt”. La distribución del número y porcentaje de palabras en función de si tienen alguna o ninguna emoción es la reflejada en la Tabla 5.3.1.

Tabla 5.3.1 Distribución de palabras según si tienen algún (suma de valencias de emociones diferente a cero) o ningún (suma de valencias de emociones igual a cero) tipo de emoción en el lexicón NRC v0.92

	Número	Porcentaje
Palabras sin emoción	9 720	68,5%
Palabras con alguna emoción	4 462	31,5%
Total	**14 182**	**100,0%**

Observamos en la Tabla 5.3.1 que existe un número elevado de palabras, un 68,5%, que no tiene ninguna emoción, es decir, que ningún valor de valencia en cada una de las emociones es igual a uno.

La distribución del número de palabras según el número de emociones, su descripción estadística y el número total de palabras por emoción quedan reflejadas en la Tabla 5.3.2, Tabla 5.3.3, y Figura 5.18.

Tabla 5.3.2 Descripción estadística del número de emociones por palabra en el lexicón NRC v0.92

Min	Q1	Mediana	Media	Q3	Max
1	1	1	1,85	2	8

Tabla 5.3.3 Distribución del número de palabras en función del número de emociones en valor absoluto y no acumulado que contiene cada palabra en el lexicón NRC v0.92

Número de emociones absolutas (no acumuladas)	Número de palabras
1	2 344
2	1 021
3	636
4	364
5	76
6	16
7	3
8	2
TOTAL	**4 462**

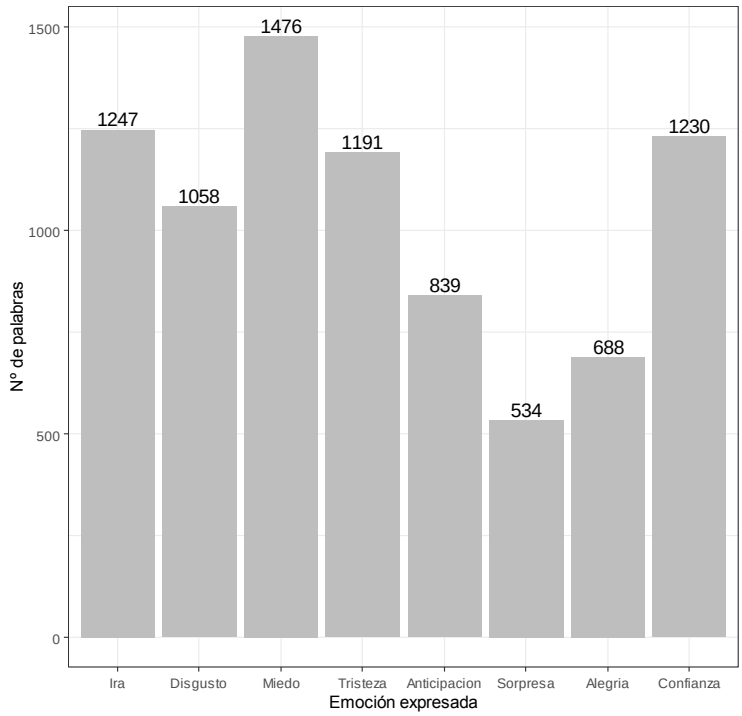


Figura 5.18 Distribución del número total de palabras que tienen alguna emoción (valencia de la emoción igual a uno) según el tipo de emoción del lexicón NRC v0.92

Observamos en la Tabla 5.3.2 que la mayoría de palabras, un 75%, tiene asociada una o dos emociones (*Q3*=2) siendo la mediana igual a una emoción por palabra. En la Tabla 5.3.3 observamos como muy pocas palabras tienen un número de emociones elevado, por ejemplo, sólo 97 palabras (un 2,2%) de las 4 462 con alguna emoción tienen más de cinco emociones. En la Figura 5.18 observamos como el número total de palabras (4 972) que contienen las emociones de ira, disgusto, miedo y tristeza conjuntamente es superior al número total de palabras (3 291) con las emociones de anticipación, sorpresa, alegría y confianza.

Por tanto, el tener el lexicón NRC v0.92 pocas palabras con alguna emoción (un 31,5% de sus palabras), tener una baja diversidad de emociones por palabra (el 75% de las palabras con alguna emoción tiene solo una o dos emociones) y no estar equilibrado con respecto a grupos de emociones (el número de palabras con un grupo de emociones aparentemente con un sentimiento negativo superior al número de palabras del grupo de emociones aparentemente con un sentimiento positivo), puede condicionar los resultados del análisis de las emociones de los *tweets.*

Como se describe en el apartado 4.7 de la metodología, el lexicón NRC v0.92 no proporciona información sobre qué emociones transmiten un sentimiento positivo o

negativo. No obstante, proporciona para cada palabra su sentimiento positivo o negativo. Calculamos el valor de polaridad de cada palabra según la metodología descrita en el apartado 4.7 y el número de palabras por emoción en función de su polaridad positiva, neutra o negativa. La distribución de palabras según su polaridad positiva neutra o negativa es la reflejada en la Tabla 5.3.4. El número de palabras por emoción en función de su polaridad positiva, negativa o neutra es la representada en la Figura 5.19, Figura 5.20 y Figura 5.21.

Tabla 5.3.4 Distribución de palabras según su valor de sentimiento, en positivos (polaridad positiva o mayor que 0), negativos (polaridad negativa o menor que 0) y neutros (polaridad neutra o cerp) en el lexicón NRC v0.92

	Número	**Porcentaje**
Palabras con polaridad negativa	3 243	22,8%
Palabras con polaridad neutra	8 709	61,4%
Palabras con polaridad positiva	2 230	15,8%
Total	**14 182**	**100%**

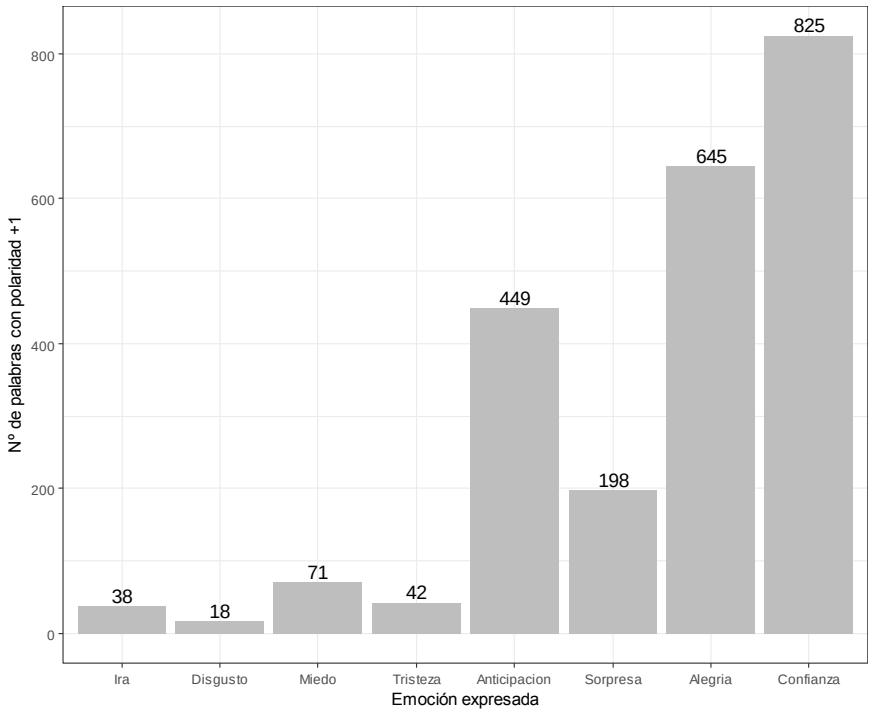


Figura 5.19 Número de palabras según su emoción con polaridad positiva en el lexicón NRC v0.92

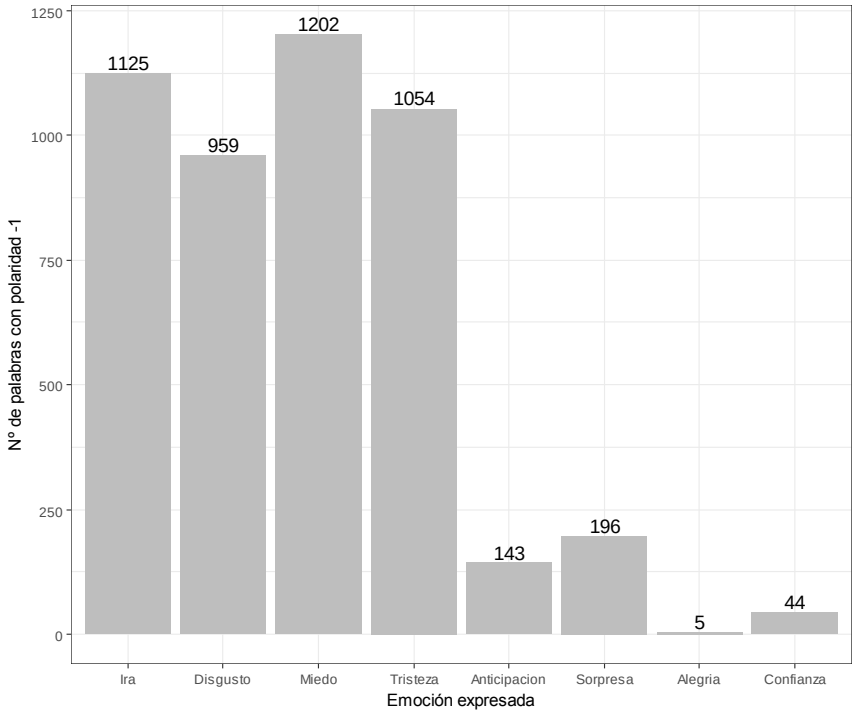


Figura 5.20 Número de palabras según su emoción con polaridad negativa en el lexicón NRC v0.92

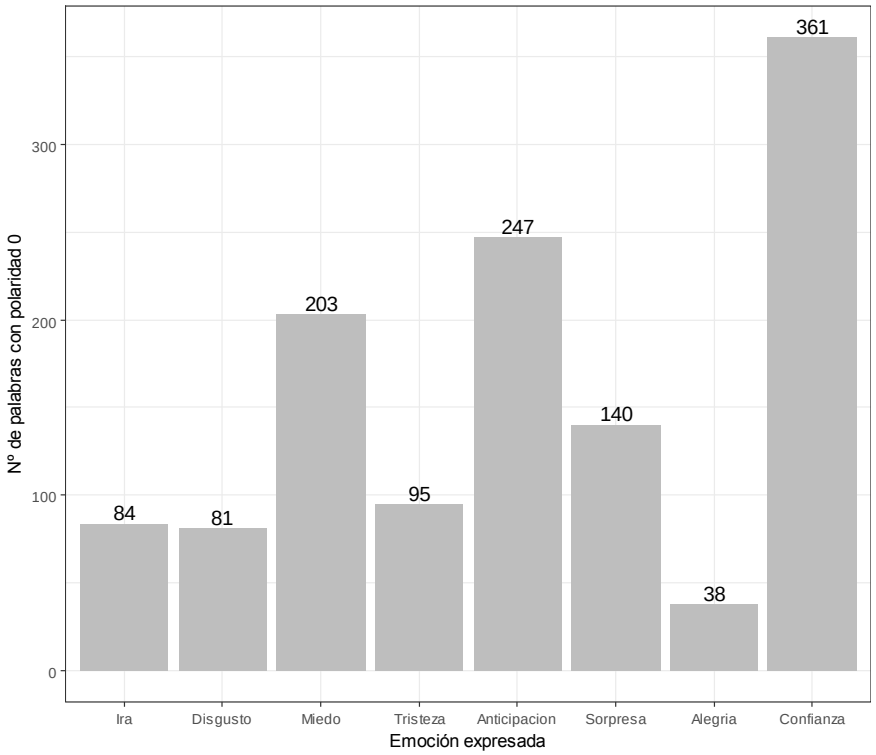


Figura 5.21 Número de palabras según su emoción con polaridad neutra en el lexicón NRC v0.92

Observamos en la Figura 5.19 que las emociones que de forma mayoritaria tienen una polaridad positiva y por tanto un sentimiento positivo son las de anticipación, sorpresa,

alegría y confianza. En la Figura 5.20 las emociones que de forma mayoritaria tienen una polaridad negativa y por tanto un sentimiento negativo son las de ira, disgusto, miedo y tristeza.

En la Figura 5.21 aunque hay palabras con polaridad neutra en todas las emociones, el número de palabras en cada emoción de las emociones de anticipación, sorpresa, alegría y confianza es inferior al número de palabras en cada una de estas emociones en la Figura 5.19. También en la Figura 5.21, el número de palabras en cada emoción de las emociones de ira, disgusto, miedo y tristeza es inferior al número de palabras en cada una de estas emociones en la Figura 5.20. Por tanto, los grupos de emociones con sentimiento positivo y negativo detectados parecen ser adecuados.

En el apartado 5.2 detectamos los unigramas de los 9 174 *tweets* de la temática AH y de los 19 387 de la temática EE. Obtenemos 80 085 unigramas en la temática AH y 150 244 en la temática EE. Con el lexicón NRC v0.92 detectamos las palabras coincidentes con las palabras de los *tweets*. Obtenemos 30 976 unigramas detectados en la temática AH y 70 826 en la temática EE que representan un 39% de los unigramas totales de la temática AH y un 47% de los unigramas totales de la temática EE, siendo ambos porcentajes bajos. Adicionalmente 439 *tweets* (un 4,8%) de los 9 174 de la temática AH y 323 *tweets* (un 1,7%) de los 19 387 de la temática EE son *tweets* en los que no se detectó ninguna palabra.

Como se describe en el apartado 4.7 de la metodología se calcula las emociones de cada *tweet* como la suma del valor de valencia de cada emoción de los unigramas del *tweet* localizados en el lexicón y se normaliza el valor de las emociones en el *tweet* dividiendo cada valor de valencia de por la suma del valor de valencia de todas las emociones del *tweet*. Un extracto del resultado puede consultarse en el Anexo 13. Los resultados completos de los *tweets* de las temáticas AH y EE pueden consultarse en la documentación electrónica anexa en los ficheros “Emociones_AH.RData” y “Emociones_EE.RData” respectivamente.

Representamos el número de unigramas normalizado para cada emoción en la temática AH y en la temática EE en la Figura 5.22 y Figura 5.23.

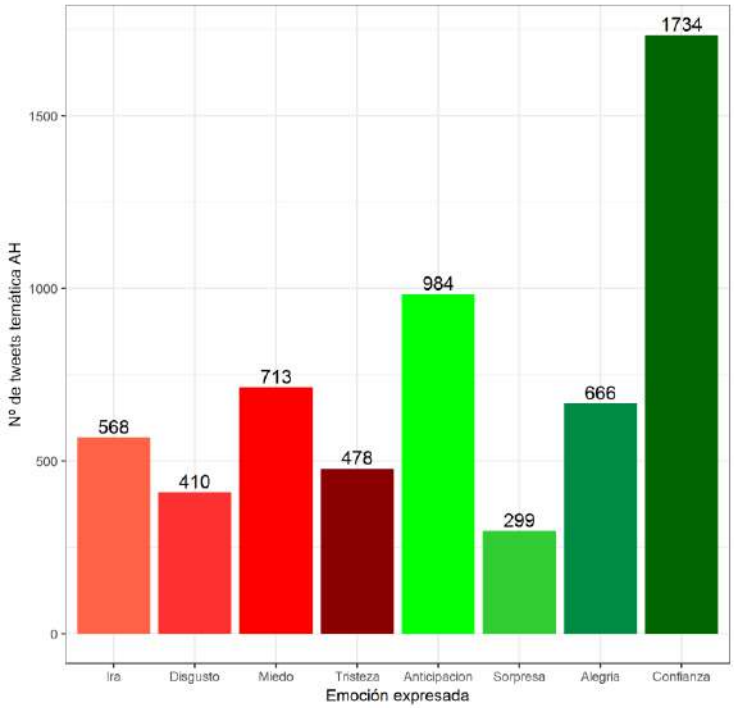


Figura 5.22 Distribución del número de *tweets* normalizados por emoción y sentimiento positivo (en verde) y negativo (en rojo). Temática AH

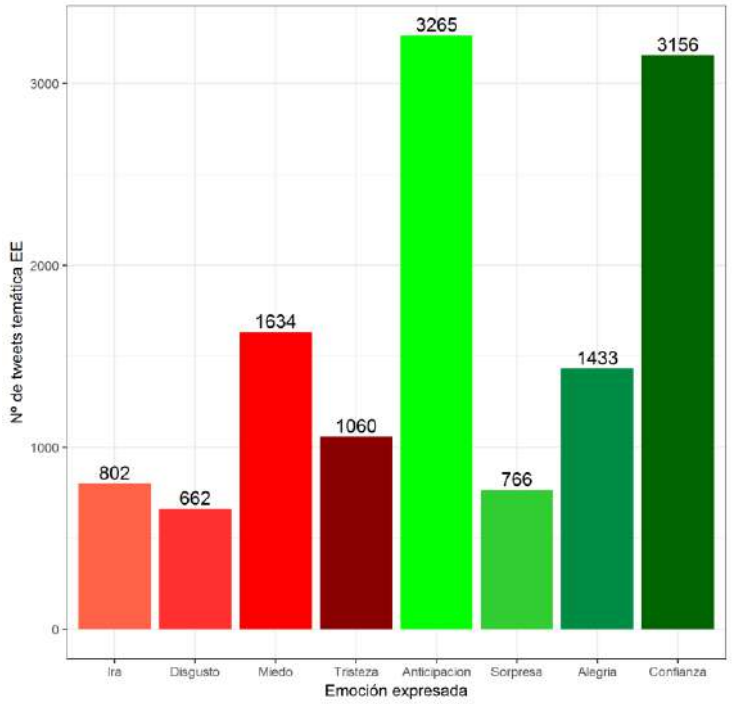


Figura 5.23 Distribución del número de *tweets* normalizados por emoción y sentimiento positivo (en verde) y negativo (en rojo). Temática EE

Observamos en la Figura 5.22 y Figura 5.23 que en ambas temáticas AH y EE las emociones con sentimientos positivos (barras en verde) globalmente predominan sobre las emocions con sentimientos negativos (barras en rojo). Asimismo es destacable en el caso de la temática AH (ver Figura 5.22) el alto valor de la emoción confianza sobre el resto y en la temática EE (ver Figura 5.23) las emociones de anticipación y confianza sobre el resto.

No obstante, cabe no obstante ser prudente con estos resultados ya que el lexicón NRC v0.92 tiene un alto porcentaje de palabras sin emociones, tiene una baja diversidad de emociones por palabra y el porcentaje de las palabras de los *tweets* coincidentes con el lexicón es bajo.

En resumen, adecuando el lexicón *NRC v0.92* a las necesidades de esta investigación, hemos podido calcular las emociones de un porcentaje reducido de los unigramas de los *tweets* y con estos hemos calculado las emociones de los *tweets* clasificados en las temáticas Actividad Humana (AH) y Entorno Educativo (EE), obteniendo pocos *tweets* sin ninguna emoción, y con predominancia de emociones que transmiten sentimientos positivos sobre las emociones que transmiten sentimientos negtivos tanto en la temática AH y como en la temática EE. No obstante, estos resultados deben tomarse con prudencia atendiendo al lexicón utilizado como al reducido porcentaje de los unigramas de los *tweets* incluidos en el lexicón.

5.4 Resultados del análisis de usuarios más relevantes

En este apartado y mediante los procesos descritos en el apartado 4.8 mostramos los resultados del análisis de los usuarios más relevantes de *tweets* y *retweets* y su coincidencia con las organizaciones más relevantes de la química. A partir de la selección de empresas y sociedades más relevantes de la química, comparamos la frecuencia de sus mensajes y con la frecuencia y la autoría de los usuarios que más mencionan o son más mencionados por los demás. Asimismo, analizamos la red social global y las redes aisladas de usuarios y menciones con medidas de centralidad y topología y la red global mediante una forma de representación gráfica para comprender la posición y relación de los usuarios más relevantes. Finalmente, analizamos los contenidos de los *tweets* de empresas y sociedades relevantes de la química para entender así como en los que fueron mencionados por otros usuarios para comprender qué aspectos comunicaban sobre la química. De esta forma damos

respuesta al tercer objetivo de la investigación sobre los usuarios que son relevantes en el conjunto de tweets aceptado y qué coincidencia tienen con organizaciones presumiblemente relevantes en la química.

Como se describe en el apartado 4.8 de la metodología se seleccionan las organizaciones relevantes de la química, grandes empresas del sector químico y las principales asociaciones y sociedades profesionales de química e ingeniería química que tenían cuentas en Twitter (Dorronsoro, 2017). La lista de empresas y sociedades seleccionada puede consultarse en el Anexo 3.

En el caso de empresas se escogieron las empresas con cuenta en Twitter de las 50 primeras empresas en la lista de las 100 mayores empresas químicas en 2015 según el volumen de ventas (ICIS Chemical Business, 2016) confeccionada por ICIS ("Independent Commodity Intelligence Services", https://www.icis.com/explore/). En el caso de sociedades profesionales, se escogieron las asociaciones y federaciones únicas tanto de profesionales como de empresas químicas con cuenta en Twitter a partir de la lista de federaciones (Network Science Corp., 2012a), la lista de sociedades profesionales para química analítica (Network Science Corp., 2012b) y la lista de sociedades profesionales para química (Network Science Corp., 2012c) de Network Science Corporation (http://www.netsci.org/).

Del análisis de la lista de las 50 empresas y 37 sociedades descrita en el Anexo 3, todas las sociedades tienen una cuenta en Twitter. En cambio, de las 50 empresas, 14 (28% del total) no disponían de cuenta, 32 (64% del total) disponían de una y cuatro (8% del total) disponían de dos o más. El hecho que no todas dispongan de una cuenta en Twitter hace suponer que algunas empresas no consideraban Twitter como un medio de comunicación útil para ellas

Seleccionamos los *tweets* con *retweets* escritos en inglés del conjunto de los 256 833 *tweets* recopilados (ver Tabla 5.1.1 del apartado 5.1) y obtenemos 125 188 *tweets* de 103 941 usuarios únicos. De las cuentas de 36 empresas y 37 sociedades con cuenta en Twitter, obtenemos los *tweets* publicados por estas empresas y sociedades y los *tweets* en los que son mencionados. Esta lista puede consultarse en el Anexo 14.

Observamos que solo ocho cuentas de Twitter correspondientes a ocho empresas y sociedades del total de 80 cuentas de empresas y sociedades son activas en el periodo estudiado publicando *tweets*. De las ocho, la mitad son sociedades, "C&EN",

"Chemistry World", "RACI" y "Amer Chem Society", y la otra mitad son empresas, "ChEnected AIChE", "Chevron Phillips", "ExxonMobil" y "Dow". Sus respectivas cuentas representan un 10% del total de cuentas de empresas y sociedades y menos del 0,01% del total de cuentas de usuarios únicos del conjunto de *tweets*. Asimismo, los 15 *tweets* emitidos por estas cuentas representan un 0,01% del total de *tweets*.

Con respecto las menciones a las cuentas de empresas o sociedades y usuarios, 24 cuentas de empresas y sociedades que corresponden a 24 empresas o sociedades son mencionadas por 185 usuarios del periodo de estudio. Las empresas y sociedades mencionadas son "C&EN", "Chemistry World", "ACS Green Chemistry", "American Chemistry", "RACI", "SCI", "Amer Chem Society", "ChEnected AIChE", "ChemHeritage", "ECS", "IChemE", "IUPAC", "Royal Soc. Chemistry", "Cefic", "Chevron Phillips", "Eastman Chemical Co.", "Syngenta", "Yara International", "Evonik" , "ExxonMobil", "BASF North America", "Dow", "DuPont News" y "BASF". De todas ellas 10 son empresas y 14 sociedades. Sus respectivas cuentas representan un 30% del total de cuentas de empresas y sociedades y un 0,02% del total de cuentas de usuarios únicos del conjunto de *tweets*. Asimismo, los 209 *tweets* en los que son mencionados representan un 0,17% del total de *tweets*.

El número de *tweets* mencionados de las empresas y sociedades en función de los usuarios que las mencionan se puede consultar en el Anexo 15. De todos los usuarios solo 185 que representan un 0,17% del total, son los que mencionan a empresas y sociedades con un total de 209 menciones. La distribución estadística del número de *tweets* de menciones en función de empresas y sociedades es la mostrada en la Tabla 5.4.1

Tabla 5.4.1 Distribución del número de *tweets* por usuarios en los que se mencionan empresas y sociedades y del número de menciones por usuario

	Min	**Q1**	**Mediana**	**Media**	**Q3**	**Max**
Empresas y sociedades	1	2	4	8,7	9	42
Usuarios	1	1	1	1,1	1	5

Observamos que en las empresas y sociedades mencionadas, cada una se mencionan de media en casi nueve tweets, así como el 75% de ellas (*Q3*=9) y con un máximo de 42 tweets. Los usuarios que las mencionan, el 75% de ellos (*Q3*=1), las mencionan en un *tweet*, con un máximo de cinco *tweets*.

Si consideramos las cinco empresas y sociedades con mayor número de menciones, “C&EN”, “Amer Chem Society”, “American Chemistry”, “Chemistry World” y “Dow”, sus 143 menciones representan el 68% del total de las 209 menciones de empresas y sociedades y son realizadas por 130 usuarios que representan el 70% del total de usuarios que realizan menciones a empresas o sociedades.

A partir de los 125 188 *tweets* con *retweets* escritos en inglés de los 103 941 usuarios únicos obtenemos (ver apartado 4.8 de la metodología) la tabla compuesta por los usuarios mencionados, el identificador del *tweet* en el que se mencionan y los usuarios que mencionan. Una parte de las tablas, de forma indicativa, se puede consultar en el Anexo 16, así como las tablas completas en los archivos “Tabla de menciones TOTAL.csv” en la documentación electrónica de la tesis.

Del análisis de esta tabla, obtenemos 39 942 menciones únicas de usuarios de un total de 85 195 menciones y 57 636 usuarios únicos que mencionan, que corresponde a un 55,5% de los 103 941 usuarios únicos. Los resultados detallados pueden consultarse en el archivo “Usuarios mencionados TOTAL.csv” en la documentación electrónica de la tesis. La descripción estadística de la frecuencia de menciones es la mostrada en la Tabla 5.4.2.

Tabla 5.4.2 Estadística descriptiva de la frecuencia de menciones de los usuarios mencionados en los *tweets* con *retweets* escritos en inglés

	Min	**Q1**	**Mediana**	**Media**	**Q3**	**Max**
Tweets con *retweets* escritos en inglés	1	1	1	2,12	1	840

Podemos observar como la mayoría de los usuarios mencionados solo se mencionan una vez (*Q3*=1). Asimismo calculamos los percentiles 95%, 99%, 99,5% y 99,9% en los tres conjuntos para entender mejor su distribución obteniendo los valores recogidos en la Tabla 5.4.3.

Tabla 5.4.3 Percentiles 95%, 99%, 99,5% y 99,9% de la frecuencia de menciones de los usuarios mencionados en los *tweets* con *retweets* escritos en inglés

	95%	**99%**	**99,5%**	**99,9%**
Tweets con *retweets* escritos en inglés	5	18	32	105

Los valores de los percentiles obtenidos sugieren como un número reducido de usuarios son los que reciben el mayor número de menciones. Los primeros cinco usuarios más mencionados son el 0,1% del total de usuarios mencionados y sus

menciones representan el 3,6% del total de menciones. Estos usuarios y sus menciones son los descritos en la Tabla 5.4.4.

Tabla 5.4.4 Primeros cinco usuarios más mencionados y sus respectivas menciones en los *tweets* con *retweets* escritos en inglés

Usuario mencionado	Número de menciones
YouTube	840
weyhrauchlaw	814
Learn_Things	520
Michael5SOS	446
VideosOfScience	406

A partir de la información proporcionada por Twitter de cada uno de estos usuarios encontramos su información descriptiva recopilada en la Tabla 5.4.5.

Tabla 5.4.5 Descripción de los cinco usuarios de las cuentas más mencionadas en los *tweets* con *retweets* escritos en inglés

Usuario	Descripción
YouTube	Portal de Internet de compartición y visualización de videos.
weyhrauchlaw	Empresa de consultoría centrada en la aviación, impuestos y leyes sobre patentes con énfasis en la defensa de la aplicación de la legislación de la Administración Federal de Aviación (FAA). Actualmente (consulta realizada el 20/05/2020) esta cuenta ya no existe.
Learn_Things	Cuenta suspendida que Twitter no permite acceder a sus contenidos (23 de noviembre de 2019).
Michael5SOS	Cuenta oficial de Michael Clifford. El usuario es el guitarrista y uno de los contantes de la banda de música australiana de pop y rock 5 Seconds of Summer (5SOS).
VideosOfScience	Cuenta ni activa ni visible en la red de Twitter (24 de enero de 2019).

De los 57 636 usuarios únicos que mencionan analizamos la frecuencia de las menciones que publican y quiénes son los más relevantes. Los resultados detallados de los usuarios pueden consultarse en el archivo “Usuarios que mencionan TOTAL.csv” de la documentación electrónica de la tesis. La descripción estadística de la frecuencia de menciones realizadas es la representada en la Tabla 5.4.6.

Tabla 5.4.6 Estadística descriptiva de la frecuencia con que los usuarios mencionan a otros en los *tweets* con *retweets* escritos en inglés

	Min	Q1	Mediana	Media	Q3	Max
Tweets con *retweets* escritos en inglés	1	1	1	1,48	1	299

Observamos también como en el caso de los usuarios más mencionados en los *tweets*, los usuarios que más mencionan a otros usuarios mencionan con una baja frecuencia de menciones siendo dos (*Q3*=2) o uno (*Q3*=1). Si comparamos los valores

máximos de menciones que ser realizan con los valores de menciones que se reciben en cada uno de los tres conjuntos de datos, observamos que son superiores los valores máximos de menciones que recibe un usuario. Asimismo calculamos los percentiles 95%, 99%, 99,5% y 99,9% en los tres conjuntos para entender mejor su distribución obteniendo los valores recogidos en la Tabla 5.4.7.

Tabla 5.4.7 Percentiles 95%, 99%, 99,5% y 99,9% de la frecuencia con que los usuarios mencionan a otros en los *tweets* con *retweets* escritos en inglés

	95%	**99%**	**99,5%**	**99,9%**
Tweets con *retweets* escritos en inglés	3	6	9	22

Los valores de los percentiles obtenidos sugieren como unos muy pocos usuarios son los más mencionan. Los primeros cinco usuarios que más mencionan y la frecuencia del númro de menciones que realizan son los detallados en la Tabla 5.4.8.

Tabla 5.4.8 Primeros cinco usuarios que con mayor frecuencia mencionan a otros y sus respectivas frecuencias totales de menciones

Usuario que menciona	**Número de menciones que realiza**
BigD_KnowsAll	299
SonalMunot	129
90068San18	126
Chem_Shaw	126
GGflipp	99

Estos cinco usuarios que representan el el 0,009% del total de los 57 636 usuarios únicos realizan el 0,9% del total de las 85 195 menciones. A partir de la información proporcionada por Twitter de cada uno de estos usuarios encontramos su información descriptiva recopilada en la Tabla 5.4.9.

Tabla 5.4.9 Descripción de los cinco usuarios de las cuentas que más mencionan en los *tweets* con *retweets* escritos en inglés

Usuario	**Descripción**
BigD_KnowsAll	Cuenta activa desde octubre de 2010 con último mensaje en 25 de diciembre de 2016 (consulta realizada el 23 de noviembre de 2019).
SonalMunot	Cuenta suspendida.
90068San18	Cuenta activa desde octubre de 2014.
Chem_Shaw	Cuenta inexistente (consulta realizada el 20 de mayo de 2020).
GGflipp	Cuenta activa desde noviembre de 2011 de un usuario que parece seguir noticias mayoritariamente sobre USA.

Comparamos el número de menciones y la autoría de las primeras cinco empresas y sociedades relevantes de la química, los primeros cinco usuarios que más mencionan

a empresas y sociedades, los primeros cinco usuarios más mencionados y los primeros cinco usuarios que más mencionan en la Tabla 5.4.10

Tabla 5.4.10 Tabla comparativa de menciones y la autoría de las primeras cinco empresas y sociedades relevantes de la química, los primeros cinco usuarios que más mencionan a empresas y sociedades, los primeros cinco usuarios más mencionados y los primeros cinco usuarios que más mencionan

Empresa o sociedad mencionada	Número de menciones	Usuario que menciona a empresa o sociedad	Número de menciones	Usuario mencionado	Número de menciones	Usuario que menciona	Número de menciones que realiza
C&EN	42	JSTR_1	5	YouTube	840	BigD_KnowsAll	299
Amer Chem Society	33	khchem	5	weyhrauchlaw	814	SonalMunot	129
American Chemistry	25	silanoldep	5	Learn_Things	520	90068San18	126
Chemistry World	23	234chem	2	Michael5SOS	446	Chem_Shaw	126
Dow	20	AWinnr	2	VideosOfScience	406	GGflipp	99

Observamos que las primeras cinco empresas o sociedades más mencionadas y los primeros cinco usuarios que más mencionan a empresas o sociedades no se corresponden con los usuarios más mencionados o que más mencionan. Adicionalmente, el número de menciones de las primeras cinco empresas y sociedades es menor que el número de menciones de los usuarios más mencionados y que el número de menciones realizadas por los primeros cinco usuarios que más mencionan.

A partir de los usuarios y sus menciones, construimos la red completa no dirigida de los usuarios que realizan alguna mención o son mencionados en los *tweets* y *retweets* escritos en inglés (ver apartado 4.8 de la metodología) y analizamos sus relaciones y estructura. La red obtenida está formada por 91 561 nodos que representan los usuarios que realizan alguna mención o son mencionados y 78 338 aristas que representan si existe una mención entre dos usuarios diferentes con mayor o menor intensidad o peso según el mayor o menor número de menciones entre los dos usuarios. Calculamos las métricas de centralidad y de topología o forma de la red descritas en la Tabla 4.8.1 y Tabla 4.8.2 del apartado 4.8 de la metodología para analizar la red y obtenemos los resultados mostrados en la Tabla 5.4.11 y Tabla 5.4.12.

Tabla 5.4.11 Medidas de centralidad de la red completa no dirigida de los usuarios con menciones y sus menciones de todos los *tweets* y *retweets* escritos en inglés

	Min	Q1	Mediana	Media	Q3	Max
Centralidad de grado	0	1	1	1,71	1	688
Centralidad de vector propio	0	0	0	0,0002	0	1

Los resultados de medidas de centralidad de la Tabla 5.4.11 sugieren que la mayoría usuarios realizan pocas menciones al resto (centralidad de grado media = 1,71 y centralida de grado Q3 = 1) y que existen usuarios sin menciones a otros (centralidad de grado min = 0) debido a que sus menciones eran a sí mismos y en la red sólo tenemos en cuenta las menciones a usuarios diferentes. Los valores de centralidad de valor propio bajos (centralidad de valor propio Q3 = 0) indican que los usuarios están conectados con usuarios poco conectados con el resto de usuarios.

Tabla 5.4.12 Medidas de la topología de la red completa no dirigida de los usuarios con menciones y sus menciones de todos los *tweets* y *retweets* escritos en inglés

Densidad	Diámetro	Distancia media
$1{,}8 \times 10^{-5}$	78	12,2

Los resultados de medidas de topología de la Tabla 5.4.12 sugieren una red muy poco densa debido a su bajo valor de densidad y con los usuarios que realizan menciones a otros aparentemente cercanos atendiendo a los valores de diámetro y distancia media.

Adicionalmente, buscamos los grupos de usuarios conectados entre sí que no estaban conectados al resto de usuarios, es decir las redes aisladas dentro de la red completa no dirigida de los usuarios que realizan alguna mención o son mencionados en los *tweets* y *retweets* escritos en inglés. Obtuvimos 23 077 redes aisladas y calculamos la distribución estadística de su número de nodos, aristas y de las medidas de topología características principales junto con medidas de centralidad y de topología de diámetro, densidad y distancia media. Los resultados son los mostrados en la Tabla 5.4.13.

Tabla 5.4.13 Distribución estadística de las medidas de centralidad y topología de las redes aisladas obtenidas a partir de la red completa no dirigida de los usuarios y sus menciones de todos los *tweets* y *retweets* escritos en inglés

	Min	Q1	Mediana	Media	Q3	Max
Número de nodos	1	2	2	3,97	3	21 048
Número de aristas	0	1	1	3,39	2	26 940
Densidad	0	0,67	1	0,87	1	1

	Min	Q1	Mediana	Media	Q3	Max
Diámetro	0	1	1	1,42	2	78
Distancia media	0	1	1	1,15	1,33	12,2

Los resultados de la Tabla 5.4.13 sugieren la mayoría de redes aisladas con pocos usuarios y menciones (número de nodos Q3 de nodos y aristas ambos con valores inferiores a 3) con unas pocas redes aisladas que concentran un un elevado número de nodos y aristas. Los usuarios en cada red aislada están poco conectados entre ellos (valores de densidad bajos). Asimismo en la mayoría de redes los usuarios están cercanos entre sí (diámetro Q3 = 2 y distancia media Q3 = 1,33).

Estos resultados sugieren que la mayoría de menciones se hacen a unos pocos usuarios y no se propagan al resto de usuarios de la red completa debido a un elevado número de redes aisladas con un número de usuarios pequeño en cada una de ellas.

Representamos gráficamente la red completa de los usuarios con más de una mención con el método de Kamada and Kawai descrito en el apartado 4.8 de la metodología. Para poder entender más fácilmente su representación gráfica, representamos las aristas correspondientes a más de dos menciones y el nombre de los nodos con más de 10 menciones con etiquetas negras. Obtenemos una red con 11 133 nodos y 1 621 aristas que representan el 12,1% de los 91 561 nodos de la red completa global y el 2,1% de las 78 338 aristas de la red completa global. La representación obtenida es la representada en la Figura 5.24.

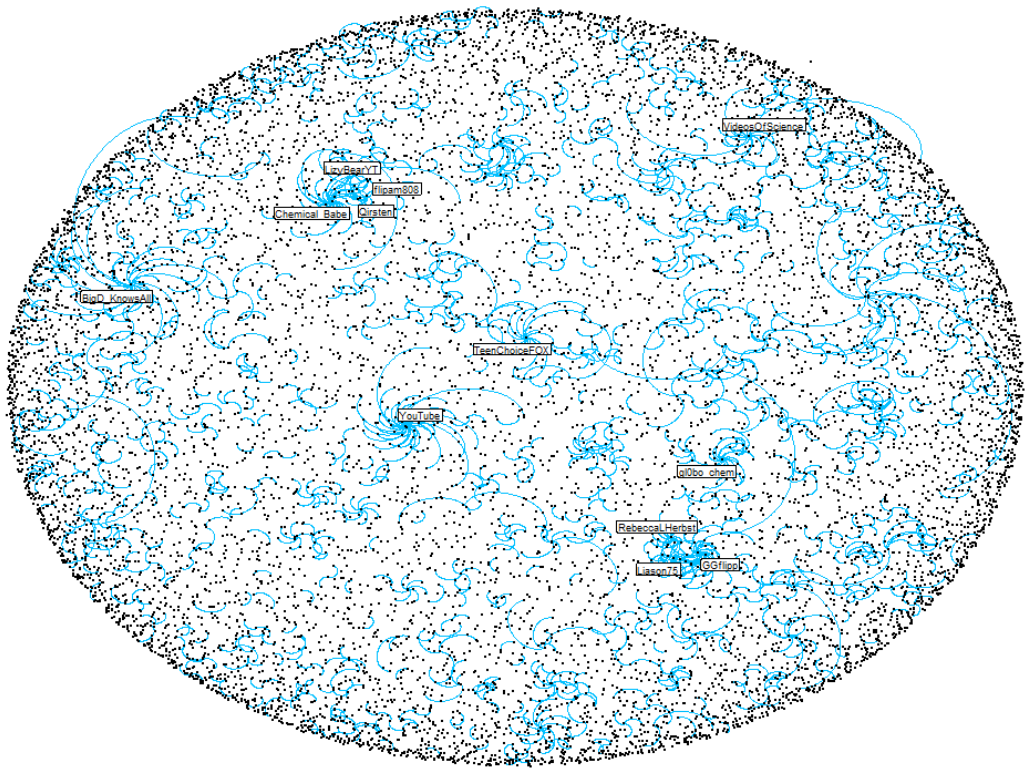


Figura 5.24 Representación gráfica de la red completa no dirigida de los usuarios con menciones y sus menciones de todos los *tweets* y *retweets* escritos en inglés. Selección de los usuarios con un número de menciones superior a dos menciones. Representación de las aristas correspondientes a más de dos menciones con líneas curvas y en color azul, y el nombre de los usuarios con más de 10 menciones con etiquetas negras

Podemos observar en la Figura 5.24 que los usuarios están agrupados en redes con un número pequeño de usuarios y aisladas del resto de redes. Las redes formadas por los usuarios con más de 10 menciones parece que también tienen pocos usuarios dentro de su red. Este efecto es aún mayor si tenemos en cuenta que solo representamos en esta red el 12,1% de los 91 561 nodos de la red completa global y el 2,1% de las 78 338 aristas de la red completa global debido a los valores pequeños de topología de esta red (ver Tabla 5.4.12). La forma de esta red es consecuencia que las menciones no se transmitan de forma lejana a otros usuarios de la red.

También representamos en esta red los usuarios correspondientes a las sociedades y empresas más relevantes de la química que están en la red mediante la etiqueta de su nombre en rojo. La representación obtenida es la representada en la Figura 5.25.

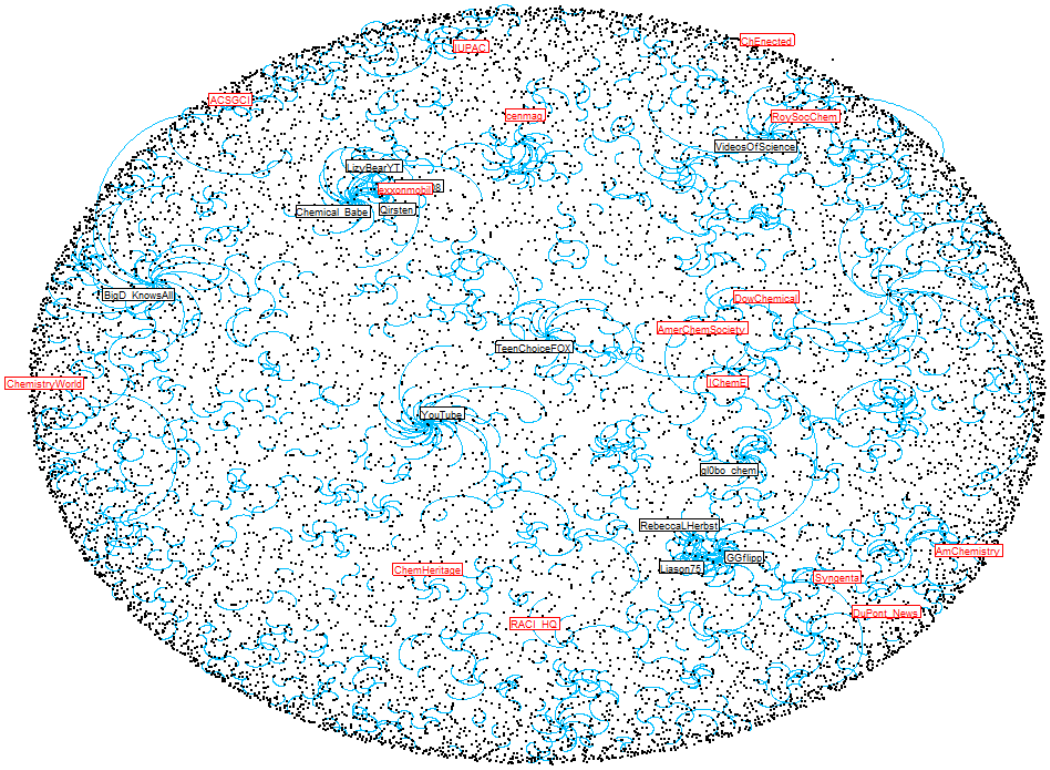


Figura 5.25 Representación gráfica de la red completa no dirigida de los usuarios con menciones y sus menciones de todos los *tweets* y *retweets* escritos en inglés. Selección de los usuarios con un número de menciones superior a dos menciones. Representación de las aristas correspondientes a más de dos menciones con líneas curvas y en color azul, y el nombre de los usuarios con más de 10 menciones con etiquetas negras y el nombre de los usuarios de empresas y sociedades más relevantes de la química con etiquetas rojas

Observamos en la red de la Figura 5.25 que las empresas y sociedades relevantes de la química que aparecen en la red están también dentro de redes aisladas con un pequeño número de usuarios y no tienen conexión con los usuarios que tienen un número de menciones superior a 10.

Mediante el análisis visual de los *tweets* publicados por las empresas y sociedades analizamos sus contenidos así como de las cinco empresas y sociedades más mencionadas por los usuarios. La interpretación de estos tweets puede consultarse en la Tabla 5.4.14 y Tabla 5.4.15.

Tabla 5.4.14 Descripción de la interpretación visual de los contenidos de los *tweets* publicados por las empresas o sociedades

Empresa o sociedad	**Número de *tweets***	**Descripción contenidos de los *tweets***
Amer Chem Society	6	Noticias sobre eventos de la sociedad
C&EN	2	Química en pinturas y en flores y publicidad sobre un evento en Amer Chem Society
RACI	2	Noticia sobre un premio de la *American Chemical Society* y noticia sobre la presentación de una conferencia
Chevron Phillips	1	Conferencia acerca de falta de profesionales en la industria
Chemistry World	1	Noticia sobre traspaso de científico reconocido
ChEnected AIChE	1	Publicidad sobre el *American Institute of Chemical Engineers*
ExxonMobil	1	Publicidad sobre innovaciones de la empresa para el medio ambiente y la sostenibilidad
Dow	1	Publicidad sobre la empresa y cómo valora la naturaleza para ahorrar dinero

Tabla 5.4.15 Descripción de la interpretación visual de los contenidos de los *tweets* de las cinco empresas o sociedades más relevantes mencionadas por los usuarios

Empresa o sociedad	**Número de *tweets***	**Descripción contenidos de los *tweets***
C&EN	42	Contenidos relacionados con la actividad humana y noticias de la sociedad
Amer Chem Society	33	Noticias sobre la sociedad y algunas relacionadas con el entorno educativo. También aparecen contenidos relacionados con la actividad humana
American Chemistry	25	Contenidos relacionados con la actividad humana
Chemistry World	23	Contenidos relacionados con la actividad humana y sobre la sociedad
Dow	20	Noticias sobre la empresa, sus productos y profesionales

Del análisis de los *tweets* emitidos por las empresas y sociedades y de aquellos de las cinco empresas y sociedades más mencionadas por los usuarios, podemos observar como la mayoría están relacionados con la actividad humana y con noticias sobre la empresa o sociedad.

En resumen, a partir del análisis de la frecuencia y el contenido de los *tweets* de la lista de empresas y sociedades potencialmente más relevantes en la química, a partir del análisis de la frecuencia de los *tweets* de los usuarios más mencionados o que más mencionan y la comparación de su autoría con las cuentas de la lista de empresas y sociedades potencialmente más relevantes en la química, y a partir del análisis gráfico y de las medidas de centralidad y topología de la red global y de las redes aisladas dentro de la red global, detectamos que las empresas y sociedades

potencialmente más relevantes en la química no se corresponden con los usuarios que más mencionan o que son más mencionados, que las empresas y sociedades están dentro de redes aisladas en la red global y no están conectadas con los usuarios que más mencionan o que son más mencionados y que mayoría de los *tweets* publicados de las empresas o sociedades o en los que son mencionados, están relacionados con la actividad humana y con noticias sobre la empresa o sociedad.

6 Discusión

En esta sección vamos a discutir las posibles contribuciones de la tesis en función del marco teórico analizado así como las limitaciones de esta y las potenciales líneas de investigación futuras que se pueden deducir.

6.1 Posibles contribuciones de la tesis

Los resultados obtenidos en el apartado 5.1 sugieren que Twitter contiene información que forma parte de la imagen pública de la química. La imagen pública de la química en Twitter está basada en contenidos y opiniones espontáneas a diferencia de la estudiada hasta la fecha, confeccionada a partir de encuestas y análisis de documentos, y es aparentemente más amplia debido a las diversas temáticas obtenidas (ver Tabla 5.1.9), Actividad Humana (AH), Conocimiento Científico (CC), Entorno Educativo (EE), Entretenimiento (E) y Relación Humana (RH) pero acotadas dentro del conjunto de datos de estudio.

La clasificación de los *tweets* de las temáticas realizadas por los expertos químicos proporcionan valores del estadístico kappa de Fleiss (ver Tabla 5.1.11) que sugieren que los niveles de concordancia entre estos expertos químicos que no son fruto de la casualidad, excepto en la temática de Conocimiento Científico (CC). Adicionalmente la concordancia en las temáticas de Entorno Educativo (EE) y Actividad Humana (AH) con el *benchmark* de Landis and Koch (1977) indica un acuerdo entre los expertos bastante bueno. Estas dos temáticas son las que se ha encontrado que contienen la mayoría de términos de la imagen pública de la química descrita en el marco teórico y, en concreto, indicios de quimiofobia.

Existe una fuerte presencia de *tweets* en las temáticas EE (35%) y AH (18%) (ver Tabla 5.1.9) pero limitada en el resto, como por ejemplo en la temática CC (5%). Por tanto, no parece ser Twitter un canal de comunicación elegido y utilizado dentro del ámbito científico para transmitir estos contenidos, hecho alineado con la bibliografía consultada.

El proceso de adquisición, limpieza y preparación de textos impacta significativamente al conjunto de datos (ver Tabla 5.1.3) que es posteriormente analizado pasando de un conjunto inicial de 256 833 a 76 242 *tweets* (30% de los iniciales). Queriendo obtener los *tweets* únicos en inglés para ser capaces de analizar temáticas únicas, las

operaciones que más afectan a esta disminución son la extracción de *retweets* y la detección con el idioma “english” (32% y 23% de reducción del número de *tweets* totales respectivamente).

Aunque la metodología de *clustering spherical k-means* junto con las heurísticas de búsqueda del mejor número de clústeres *elbow method* y *silhouette* y las de selección del mejor número de clústeres *L-method* y curvatura no proporcionen un número de clústeres iguales y concluyentes (ver Figura 5.4), parecen aportar un rango de valores del número de clústeres suficientemente adecuado para que los expertos químicos mediante el análisis visual de los clústeres consigan clasificar (ver Tabla 5.1.9) la mayoría de los clústeres (82%) y los *tweets* (78%) e interpretarlos con unos niveles de concordancia no fruto de la casualidad y estadísticamente significativos.

El resultado del análisis de sentimientos en las temáticas AH y EE (ver Tabla 5.2.4 y Tabla 5.2.5) sugiere un mayor porcentaje de *tweets* positivos que de negativos en ambas temáticas (56% frente a 36% en AH y 63% frente a 33% en EE) así como en el análisis de emociones (ver Figura 5.22 y Figura 5.23). Estos resultados parecen estar en línea con los estudios más recientes donde la percepción positiva de la imagen pública de la industria química pasaba de un 36% en 1996 a un 49% en 2010 (Hadhri, 2010), el 51% de los encuestados tenían una imagen positiva o neutra y el 59% creían que los beneficios de la química eran superiores a sus efectos dañinos (The Royal Society of Chemistry y TNS BMRB, 2015). Asimismo, la existencia de ambos sentimientos y emociones tanto positivos como negativos también parece estar en línea con estos resultados.

En la temática EE y debido a que términos más frecuentes en los *wordclouds* de *tweets* negativos también aparecían en los positivos (ver Figura 5.12 y Figura 5.13), se analizó visualmente el contenido de una muestra representativa de los términos más frecuentes en los *tweets* positivos (ver extracto en el Anexo 11 o el archivo “Clasificacion muestra tweets EE.csv” de la documentación electrónica de la tesis). Algunos *tweets* que fueron clasificados como positivos en el análisis de sentimientos parece que deberían ser clasificados como o bien neutros o bien negativos (ver Tabla 5.2.7). Además, la muestra también contenía ironías que parecen ser ligeramente superiores porcentualmente en los *tweets* clasificados visualmente como negativos o neutros (ver Tabla 5.2.8).

En los datos analizados de Twitter no parece que exista un posicionamiento neutro (ver ver Tabla 5.2.4 y Tabla 5.2.5), o bien los sentimientos de forma mayoritaria son positivos o bien negativos, ya que el número de *tweets* neutros con respecto a los *tweets* positivos o negativos son minoritarios, 7,9% en la temática AH y 3,1% en la temática EE. No obstante, atendiendo al análisis de una muestra de *tweets* en la temática EE, esta polarización no parecería tan distante debido a la existencia de un porcentaje potencialmente significativo de *tweets* clasificados como positivos y que parecen ser neutros. En cambio y de forma diferencial, el estudio de la Royal Society of Chemistry (2015) expone por ejemplo que el 20% de los encuestados tienen una percepción neutra de la química.

En la temática AH, los *wordclouds* obtenidos del análisis de sentimientos (ver Figura 5.10 y Figura 5.11) enfatizan los términos "chemical", relacionado con aspectos industriales, como por ejemplo productos obtenidos mediante procesos fabriles, y "chemistry", término entendido como la ciencia química de forma genérica, teniendo ambos términos connotaciones negativas y positivas respectivamente. Estos resultados parecen estar en consonancia con los de la Royal Society of Chemistry (2015).

En esta temática también parece existir actitudes quimiofóbicas en *tweets* clasificados tanto como positivos o negativos con la aparición de términos destacados como por ejemplo "attack", "syria"[25], "chemical attack", "syrian opposition", "chemical warfare", "toxic", "toxic chemical", "chemical fire" o "chemical leak" en los *tweets* negativos o "chemical free" y "used chemical" en los positivos. Este hecho sumado a que parece haber una falta de términos con connotaciones positivas contrapuestas a la quimiofobia, puede ayudar a crear o reforzar percepciones quimiofóbicas en usuarios de Twitter.

En la temática EE y mediante los *wordclouds* obtenidos del análisis de sentimientos (ver Figura 5.12 y Figura 5.13), la imagen de la química parece estar basada en elementos específicos de la educación química como por ejemplo los métodos de evaluación por la predominancia de términos como "final", "exam", "chem final", "final tomorrow", "quiz tomorrow", "lab", "lecture", "chem lab"o "chem lecture" más que en las temáticas de la comunicación del conocimiento químico en la educación y su influencia (Nicolas, 2006; Penagos and Lozano, 2009; Chamizo, 2011; Lacolla et al., 2013) o los

[25] Durante la época del muestreo había una guerra civil en Siria y se hablaba del uso de armas químicas en algunos ataques a la población civil

contenidos curriculares (Jiménez and Criado García-Legaz, 2005; Nicolas, 2006; Muñoz and Nardi, 2011; Linthorst, 2012; Piñeros and Parga, 2014), siendo una nueva contribución en este ámbito. Dentro de esta temática, no obstante, la imagen ha sido percibida como negativa por estudiantes de química (Yager and Penick, 1983; Furió Más, 2006), de forma similar a lo que parece se transmite dentro de la temática EE como por ejemplo con términos destacados relacionados con sentimientos negativos como "hate", "hard", "crying", "need help", "never understand" y "chemistry hard" en *tweets* clasificados como negativos, reforzados con la presencia de términos en *tweets* positivos como "someone help" y "help chemistry" y los potenciales *tweets* negativos clasificados como positivos detectados mediante la revisión visual de la muestra de *tweets* positivos. Esto sugiere la conveniencia de seguir investigando en la didáctica de la química para paliar estos sentimientos negativos y favorecer el aprendizaje.

Aunque en el análisis de sentimientos el porcentaje de unigramas de los *tweets* incluidos en el lexicón utilizado es elevado (72% de los unigramas totales de la temática AH y 75% de los unigramas totales de la temática EE), en el análisis de emociones este valor es menor (39% de los unigramas totales de la temática AH y 47% de los unigramas totales de la temática EE). Adicionalmente el número y el porcentaje de términos con alguna emoción en el lexicón utilizado para el análisis de emociones es limitado. Estos resultados sugieren que las emociones no deban interpretarse de forma separada según el tipo de emoción y con prudencia atendiendo a sus resultados.

En todo el conjunto de *tweets* y *retweets* los usuarios de Twitter están conectados mayoritariamente a un solo usuario medido por las menciones a otros usuarios (ver Tabla 5.4.2 y Tabla 5.4.6), débilmente conectados a otros usuarios (ver Tabla 5.4.11 y Tabla 5.4.12) y con 23 077 grupos de pocos usuarios conectados entre ellos y aislados de los demás (ver Tabla 5.4.13) de forma que los mensajes se quedan dentro de estos grupos siendo los contenidos de estos *tweets* poco mencionados. Esta distribución de los usuarios en Twitter está alineadas con los estudios revisados dentro del marco téorico (ver apartado 2.2).

De los cinco usuarios más mencionados (ver Tabla 5.4.5) solo es posible disponer de la información de sus cuentas de Twitter de tres de ellos por estar el resto o bien no activas o suspendidas. Estas cuentas son diversas con respecto a su origen y parecen pertenecer a una organización o a un particular, estando en este grupo *YouTube* portal de Internet usado para compartir y visualizar vídeos, *weyhrauchlaw* empresa de

consultoría centrada en la aviación y en la aplicación de la legislación aérea o *Michal5SOS* guitarrista y componente de la banda de música *5 Seconds of Summer*. Asimismo los usuarios más mencionados no se corresponden con los químicos a seguir e *influencers* sugeridos por el *Chemical & Engineering News* (ver Tabla 2.2.10) y *Feedspot* (ver Tabla 2.2.11) respectivamente.

Las cuentas activas de los cinco usuarios que más mencionan que han podido ser consultadas (ver Tabla 5.4.9) parecen ser pertenecientes a usuarios privados y no a organizaciones o empresas y no se ha podido realizar un análisis de la tipología del usuario a partir de la información de la cuenta proporcionada por Twitter.

Las empresas y sociedades potencialmente relevantes en el ámbito de la química no son los usuarios ni más mencionados ni los que más mencionan (ver Tabla 5.4.10). Aunque las sociedades tienen todas ellas cuenta en Twitter, no es así en el caso de las empresas, con un 28% del total de las 50 empresas consideradas relevantes sin cuenta. Esto hace suponer que no todas las empresas consideran Twitter como un medio de comunicación adecuado a sus objetivos.

La actividad de las empresas y sociedades es reducida tanto por los *tweets* publicados como por las menciones de otros usuarios (ver apartado 5.4). Solo ocho de las 80 que tienen cuenta emitieron 15 *tweets* en el periodo de estudio y 24 fueron mencionadas por 185 usuarios en 209 *tweets*. Las menciones por usuario son reducidas (ver Tabla 5.4.1), con el 75% de los usuarios realizando un *tweet* por mención, con un máximo de cinco. Las cinco empresas y sociedades con mayor número de menciones, que representan el 68% del total de las 209 menciones a empresas y sociedades, son mencionadas por un número pequeño de usuarios, en particular por 130 usuarios. Las empresas y sociedades en la red seleccionada de usuarios con más de dos menciones y sus menciones (ver Figura 5.25) están dentro de redes aisladas con un pequeño número de usuarios y no tienen conexión con los usuarios que tienen un número de menciones superior a 10. Parece por tanto que las empresas y sociedades potencialmente relevantes en la química no son de interés por los usuarios en el conjunto de *tweets* estudiado, ni tampoco parecen ser activas atendiendo al número de *tweets* publicados.

En los contenidos de los mensajes publicados por todas las empresas y sociedades (ver Tabla 5.4.14), las empresas publican contenidos sobre sus actividades para publicitarlas, para explicar cómo afectan positivamente sus acciones sobre el medio

ambiente y la sostenibilidad y para enfatizar la falta de profesionales en la industria. Las sociedades publican contenidos relativos a sus eventos y conferencias. Las temáticas de los contenidos publicados por empresas y sociedades están alineadas con los estudios revisados en el marco teórico (ver apartado 2.2). No obstante estos contenidos, el reducido número de *tweets* no nos permite generalizar estos resultados. En el caso de las cinco empresas y sociedades que son más mencionadas, los contenidos de los *tweets* donde son mencionadas (ver Tabla 5.4.15) están relacionados con noticias de la actividad humana, sobre noticias de la sociedad o empresa mencionada y en un caso sobre noticias del entorno educativo.

Aunque en el caso de las empresas y los contenidos que publican parecen estar relacionados con la promoción de sus actividades dentro de un ámbito favorable para el medio ambiente y la sostenibilidad, los mensajes están relacionados con sus actividades sin aparentemente existir una estrategia de comunicación común entre diferentes empresas, sociedades y asociaciones, utilizando Twitter como un canal de noticias y publicidad más que como un canal de educación sobre las virtudes y beneficios de la química y por tanto, no pareciendo contribuir las empresas y sociedades al cambio de la imagen pública histórica negativa de la química en esta red social.

En resumen,

- Twitter parece contener información que forma parte de la imagen pública de la química.
- Esta imagen es aparentemente más amplia que la imagen estudiada hasta la fecha, incluyendo las temáticas Actividad Humana (AH), Conocimiento Científico (CC), Entorno Educativo (EE), Entretenimiento (E) y Relación Humana (RH).
- Las temáticas AH y EE son las que se ha encontrado que contienen la mayoría de términos de la imagen pública de la química descrita en el marco teórico y, en concreto, indicios de la quimiofobia.
- Twitter no parece ser un canal de comunicación elegido y utilizado dentro del ámbito científico para transmitir sus contenidos, hecho alineado con la bibliografía consultada.
- El resultado del análisis de sentimientos y emociones en las temáticas AH y EE sugiere un mayor porcentaje de tweets positivos que de negativos en ambas temáticas. Estos resultados parecen estar en línea con los estudios más recientes.

- La existencia de ambos sentimientos y emociones tanto positivos como negativos también parece estar en línea con estos resultados.
- Parece existir un posicionamiento polarizado hacia sentimientos positivos o negativos pero no neutros, a diferencia del estudio de la Royal Society of Chemistry (2015).
- En la temática AH, el término “chemical”, relacionado con aspectos industriales y “chemistry”, entendido como la ciencia química de forma genérica presentan connotaciones negativas y positivas respectivamente, en consonancia con los resultados de la Royal Society of Chemistry (2015).
- En la misma temática parece existir actitudes quimiofóbicas en tweets clasificados tanto como positivos o negativos que pueden ayudar a crear o reforzar percepciones quimiofóbicas.
- En la temática EE, la imagen de la química parece estar basada en elementos específicos de la educación química, hecho diferencial con la bibliografía consultada.
- En la misma temática la imagen de la química parece percibida de forma negativa, en consonancia con los estudios revisados. Esto sugiere la conveniencia de seguir investigando en la didáctica de la química para paliar estos sentimientos negativos y favorecer el aprendizaje.
- Los usuarios de Twitter están conectados mayoritariamente y de forma débil a un solo usuario por su número de menciones formando un alto número grupos de pocos usuarios conectados entre ellos y aislados de los demás. Los mensajes se quedan dentro de estos grupos siendo los contenidos de estos tweets poco mencionados. Esto está en línea con con los estudios revisados dentro del marco téorico.
- Los cinco usuarios más mencionados no se corresponden ni con los químicos a seguir e influencers sugeridos por el *Chemical & Engineering News* y *Feedspot* respectivamente, ni con las empresas y sociedades potencialmente relevantes en el ámbito de la química.
- La actividad de las empresas y sociedades potencialmente relevantes es reducida tanto por los tweets publicados como por las menciones de otros usuarios, de forma que no son de interés por los usuarios en el conjunto de tweets estudiado, ni tampoco parecen ser activas atendiendo al número de tweets publicados.
- Las temáticas de los contenidos publicados por empresas y sociedades científicas están alineadas con los estudios revisados en el marco teórico, utilizando Twitter como un canal de noticias y publicidad sin una finalidad

educativa aparente sobre las virtudes y beneficios de la química y por tanto, no pareciendo contribuir al cambio de la imagen pública histórica negativa de la química en esta red social.

6.2 Limitaciones de la investigación

Las limitaciones que consideramos más influyentes en esta investigación son el perdiodo de tiempo en el que se recogieron los *tweets,* así como el conjunto de *tweets* obtenido y su número, su proceso de limpieza y preparación, el análisis de los clústeres mediante su visualización y los métodos de análisis de sentimientos y emociones utilizados.

Los contenidos de los *tweets* pueden diferir según el periodo de tiempo en el que son adquiridos. En este caso estaban focalizados, entre otros, en la utilización de armas químicas en la guerra de Siria en la temática AH o exámenes de química en la temática EE. Como se describe en el marco teórico, Twitter se parece más a un medio informativo donde los usuarios difunden contenidos a sus seguidores, se expresan pensamientos, opiniones y emociones y las empresas difunden información tanto de ellas mismas como de sus productos. Por tanto, si el periodo de tiempo cambia, las noticias y los contenidos probablemente cambiarán y algunas de las temáticas pueden diferir en porcentaje con respecto al resto. Con respecto a la temática EE, atendiendo a que un curso escolar habitualmente está comprendido entre los meses de septiembre u octubre hasta junio o julio, es probable que los contenidos no difieran tanto ya que los *tweets* fueron captados desde enero hasta junio. Entonces, si aumentamos el periodo de tiempo de captación de *tweets* así como su número permitiría obtener unos resultados y conclusiones probablemente más generalizables tanto en contenidos, porcentaje de *tweets* en cada temática, su clasificación positiva o negativa e incluso el descubrimiento de nuevas temáticas.

Las palabras claves de búsqueda “chemical”, ”chem*” o “chemistry” escogidas por considerarse no sesgadas con respecto a la imagen pública de la química, pueden limitar la recogida de otros tweets relacionados con la imagen pública. El utilizar otras palabras clave permitiría complementar la imagen, sus sentimientos y emociones y los usuarios más influyentes detectados en este estudio.

El no tener en cuenta los *retweets* para analizar contenidos únicos así como solo considerar los *tweets* detectados con el idioma “english” tiene un gran efecto sobre el

número de *tweets* útiles para ser clasificados y analizados posteriormente, con una disminución de 256 833 a 89 663, una reducción conjunta de un 65% (ver Tabla 5.1.3) y en particular de 175 149 a 89 663, una reducción de un 33% en el caso de solo considerar los *tweets* detectados con el idioma “english”. Aunque el no considerar los *retweets* seguirá afectando al número de *tweets* útiles ya que permite la detección de contenidos únicos, el probar otros algoritmos de detección de los *tweets* con idioma “english” podría ayudar a reducir el número de *tweets* descartados.

El método de selección automática de características utilizado disminuye fuertemente la dimensionalidad de la TDM de 302 637 a 864 bigramas (ver Tabla 5.1.5). Aunque ayuda al rendimiento del proceso de *clustering* (ver apartado 4.3), parte del número de bigramas que no se consideran pueden contener información relevante para el proceso de *clustering*, afectando por tanto al resultado de los clústeres obtenidos.

Aunque la visualización de los clústeres mediante *wordclouds* de unigramas y bigramas es un método extendido (ver apartado 4.5), no asegura que validen que el significado que proporcione un *wordcloud* sea el mismo que el significado del texto completo. No se han encontrado referencias ni a favor ni en contra.

El análisis de sentimientos y emociones están basados en lexicones que no tienen en cuenta el contexto (ver apartado 4.6 y apartado 4.7) en el que cada término está utilizado pudiendo cambiar el valor de sentimiento o emoción de cada término en un *tweet* y por tanto afectando al valor de sentimiento o emoción resultante del *tweet*. Asimismo no todos los términos de los *tweets* estaban incluidos en los lexicones, y en especial en el caso del análisis de emociones donde solo estaban incluidos el 39% de los unigramas totales de la temática AH y el 47% de los unigramas totales de la temática EE. Estas limitaciones pueden tener un impacto importante en la clasificación de los *tweets* según sus emociones y sentimientos. La combinación de varios lexicones o la construcción de un lexicón específico en el ámbito de la química junto con procesos que tengan en cuenta el contexto, como por ejemplo la detección de ironías, permitirían probablemente aumentar la fiabilidad de los resultados obtenidos.

6.3 Futuras líneas de la investigación

Las futuras líneas de investigación deberían estar focalizadas en mejorar las limitaciones descritas anteriormente así como el complementar más frecuentemente la percepción de la imagen pública de la química en Twitter con métodos más

tradicionales como encuestas tanto al público general como especializado o revisiones de documentos y noticias en medios de comunicación para obtener una percepción combinada y continuada que actualmente no se dispone. De esta forma, tanto científicos como profesionales podrían obtener una visión más amplia y actualizada de la imagen pública de la química que les permitiría desarrollar planes de acción mucho más detallados y focalizados para su mejor alineación con sus objetivos e intereses y a la vez, una disminución incremental y progresiva de la percepción negativa de la química. Esta disminución a su vez, les permitiría transmitir mejor sus mensajes debido a un cambio de percepción del público al cual se dirigen.

De forma más concreta, algunas de las futuras líneas de investigación propuestas son:

- Replicar el estudio realizado con un mayor número de *tweets* durante un periodo más extendido en el tiempo. Permitiría comprender con mayor profundidad la diversidad temática de la imagen pública de la química, tener una mayor confianza sobre los sentimientos y emociones y los contenidos referidos a cada uno de ellos, poder realizar un análisis en más profundiad del comportamiento humano, poder generalizar los usuarios más relevantes y el papel de empresas y sociedades profesionales así como poder evaluar la evolución de la imagen pública de la química.
- Replicar el estudio realizado incorporando *tweets* con idiomas diferentes al inglés. La inclusión de nuevos idiomas implicaría entender y seleccionar las herramientas a aplicar para traducirlos a un idioma común así como poder entender si la imagen pública de la química, los sentimientos y emociones que transmite y los usuarios más relevantes y el papel de empresas y sociedades científicas son iguales en los diferentes idiomas.
- Realizar un estudio sobre las características y herramientas actuales y deseables de la educación química y su transmisión incorporando los nuevos canales de comunicación como son las redes sociales *on-line* a partir de las opiniones de los docentes, para reducir los sentimientos y emociones negativas percibidas por los estudiantes.
- Realizar un estudio detallado sobre la quimiofobia en los medios de comunicación actuales tanto en el entorno educativo como en la actividad humana para entender los contenidos, temáticas, organizaciones y profesionales en los que se soporta con el objetivo de proporcionar tanto a docentes como a empresas y sociedades científicas líneas de contenidos y canales de comunicación más adecuados para poder combatirla.

7 Conclusiones

En este capítulo se presentan las principales conclusiones a las que se han llegado con los objetivos propuestos y a la vista de los resultados obtenidos del análisis de 256 833 *tweets* públicos en inglés que contienen la palabra clave "chemical", "chemistry" o "chem*" capturados durante el primer semestre de 2015.

En cuanto al primer objetivo de investigación, que consiste en conocer a qué se refieren los usuarios cuando hablan de la química en Twitter:

1) Se han detectado cinco temáticas generales relacionadas con la raíz "chem*" en Twitter, que corresponden a la actividad humana entendida como la presencia de la química en ámbitos de la producción o la industria, el entorno académico presente en un curso escolar, el conocimiento científico relacionado con conceptos químicos y entidades abstractas como por ejemplo fórmulas químicas, la relaciones humanas con respecto a los sentimientos o emociones entre personas y el entretenimiento relacionado con expresiones culturales y de medios.
2) Las temáticas más relevantes obtenidas a partir de la clasificación realizada por 18 expertos en química son las de entorno académico y actividad humana con un porcentaje de *tweets* clasificados del 35% y del 18% respectivamente, siendo los menos frecuentes los clasificados en las temáticas entretenimiento y conocimiento científico con porcentajes del 13% y del 5% respectivamente. Los valores de la kappa de Fleiss global sobre la fiabilidad intra-evaluadores de los ámbitos más relevantes cercana a 0,4 y los valores resultantes del *benchmarking* de Landis and Koch nos sugieren que los resultados obtenidos no son por azar, siendo la clasificación, adecuada.

Respecto al segundo objetivo de investigación planteado, centrado en conocer los sentimientos y emociones que se transmiten a través de la imagen pública de la química observada:

3) Solamente se han estudiado los sentimientos y emociones en las temáticas entorno educativo y actividad humana. Aparecen tanto sentimientos como emociones de los dos signos, positivos y negativos predominando los positivos, lo que está en línea con los estudios más recientes sobre la percepción pública de la química. Estos sentimientos y emociones parecen estar polarizados en Twitter, no siendo así en otros estudios anteriores.

4) En los contenidos de la temática actividad humana destacan términos que connotan o pueden inducir a quimiofobia, como por ejemplo “chemical attack”, o “toxic chemical” en *tweets* con sentimientos negativos o “chemical free” en los positivos. Asimismo destaca la connotación negativa del término “chemical” en comparación con la positiva de “chemistry”. Estas conclusiones están en línea con los estudios anteriores.
5) En los contenidos de la temática entorno educativo destacan términos relacionados con la enseñanza de la química en las aulas con connotaciones negativas y de dificultad con respecto al aprendizaje de la química, como por ejemplo “never understand” y “chemistry hard” en *tweets* con sentimientos negativos o “someone help” y “help chemistry” en los positivos. Estas conclusiones están en línea con los estudios anteriores.

Respecto al tercer objetivo de investigación planteado, centrado en conocer los usuarios más relevantes para la imagen pública de la química y su comportamiento en Twitter:

6) Los usuarios más destacados no se corresponden con las sociedades y empresas aparentemente más activas e influyentes de la química, que publican pocos *tweets*, son poco mencionadas por el resto de usuarios y difunden contenidos relacionados con la temática actividad humana y noticias de sus organizaciones.
7) Las redes formadas por los usuarios y sus menciones son numerosas y mayoritariamente con pocos usuarios en cada una de ellas. En aquellas compuestas por los usuarios más mencionados o que más mencionan, estos están poco relacionados entre ellos y con un usuario conectado fuertemente al resto, siendo difícil la dispersión de los contenidos al resto de los usuarios de la red. Todo esto indica que no se detectan líderes de opinión.

Para finalizar, se destaca que durante este estudio se ha detectado la existencia de una imagen pública de la química en Twitter, con preponderancia de las temáticas entorno académico y actividad humana, en las que los sentimientos están polarizados en los signos positivo y negativo, predominando los positivos, hechos que están en línea con lo conocido antes. Asimismo, sugiere la necesidad de seguir investigando en la didáctica de la química para reducir las connotaciones negativas que como materia parecen tener para los estudiantes y también parece necesaria una mayor actividad de divulgación de los riesgos y los beneficios de la química para la sociedad, actividad en

la que se podrían implicar especialmente las empresas, los profesionales y las sociedades científicas.

8 Referencias

Aarts, O., Van Maanen, P. P., Ouboter, T. y Schraagen, J. M. (2012) «Online social behavior in twitter: A literature review», *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*, pp. 739-746. doi: 10.1109/ICDMW.2012.139.

Adrian, F. M. y de Paula, F. (2013) *Propuesta de criterios para el análisis de la responsabilidad social corporativa en la industria química española*. Universitat Ramon Llull.

Aggarwal, C. C. (ed.) (2011) *Social Network Data Analytics*. Springer. doi: 10.1007/978-1-4419-8462-3.

Aggarwal, C. C. y Yu, P. S. (2000) «Finding generalized projected clusters in high dimensional spaces», *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 70-81. doi: 10.1145/342009.335383.

Aggarwal, C. C. y Zhai, C. (2012) *Mining Text Data*. Boston, MA: Springer Science & Business Media.

Ahmad Kharman Shah, N., Latif Shabgahi, S. y Cox, A. M. (2016) «Uses and risks of microblogging in organisational and educational settings», *British Journal of Educational Technology*, 47(6), pp. 1168-1182. doi: 10.1111/bjet.12296.

Alelyani Salem, Jiliang Tang y Huan Liu. (2013) «Feature Selection for Clustering: A Review.», *Data Clustering: Algorithms and Applications*, 29, pp. 110-121.

Allen, V. (2004) *The changing image of chemistry*, *Chemistry World*. Disponible en: https://www.chemistryworld.com/feature/the-changing-image-of-chemistry/1012666.article (Accedido: 1 de enero de 2016).

Antilla, L. (2010) «Self-censorship and science: A geographical review of media coverage of climate tipping points», *Public Understanding of Science*, 19(2), pp. 240-256. doi: 10.1177/0963662508094099.

Arquero, J. L. y Romero-Frías, E. (2013) «Using social network sites in Higher Education: an experience in business studies», *Innovations in Education and Teaching International*, 50(3), pp. 238-249. doi: 10.1080/14703297.2012.760772.

Baccianella, S., Esuli, A. y Sebastiani, F. (2010) «SentiWordNet 3 . 0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining SentiWordNet», *Analysis*, 0, pp. 1-12. doi: 10.1.1.61.7217.

Bekkerman, R. y Allan, J. (2004) «Using Bigrams in Text Categorization», *Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst*, pp. 1-10.

Bensaude-Vincent, B. y Simon, J. (2012) «Chemistry: the impure science». World Scientific Publishing.

Berinato, S. y Clark, J. (2010) «Six ways to find value in Twitter's noise». Harvard Business School Publishing Corporation, pp. 34-35.

Bicen, H. y Cavus, N. (2011) «Social network sites usage habits of undergraduate students: Case study of Facebook», *Procedia - Social and Behavioral Sciences*. Elsevier B.V., 28, pp. 943-947. doi: 10.1016/j.sbspro.2011.11.174.

Blaschke, L. M. (2014) «Using social media to engage and develop the online learner in self-determined learning», *Research in Learning Technology*, 22(1063519). doi: 10.3402/rlt.v22.21635.

Bonacich, P. (1972) «Factoring and weighting approaches to status scores and clique identification», *The Journal of Mathematical Sociology*. Routledge, 2(1), pp. 113-120. doi: 10.1080/0022250X.1972.9989806.

Borgatti, S. P. (2005) «Centrality and network flow», *Social Networks*, 27(1), pp. 55-71. doi: 10.1016/j.socnet.2004.11.008.

Bowman, T. D. (2015) «Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter», *ProQuest Dissertations and Theses*, (July), p. 273.

Boyd, D., Golder, S. y Lotan, G. (2010) «Tweet, tweet, retweet: Conversational aspects of retweeting on twitter», *Proceedings of the Annual Hawaii International Conference on System Sciences*. doi: 10.1109/HICSS.2010.412.

Boyd, D. M. y Ellison, N. B. (2007) «Social network sites: Definition, history, and scholarship», *Journal of computer-mediated Communication*, 13(1), pp. 210-230. doi: 10.9790/487X-0124852.

Bravo-Marquez, F., Mendoza, M. y Poblete, B. (2013) «Combining strengths, emotions and polarities for boosting Twitter sentiment analysis», *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '13*, pp. 1-9. doi: 10.1145/2502069.2502071.

Breslow, R. (1993) «Let 's Put An End to ` Chemophobia '», *The Scientist*, p. 1.

Brooks, B. A. (2011) «The Strength of Weak Ties», *University of Auckland Business Review*, 13, pp. 19-21. doi: 10.1016/j.mnl.2018.12.011.

Bull, G., Thompson, A., Searson, M., Garofalo, J., Park, J., Young, C. y Lee, J. (2008) «Connecting Informal and Formal Learning Experiences in the Age of Participatory Media», *Contemporary Issues in Technology and Teacher Education*, 8(2), pp. 100-107. doi: Sept 8, 2008.

Burden, T. (2014) «K-12 teachers uncertain about how to connect with students and parents via social media, reveals University of Phoenix survey», *University of Phoenix*.

Carlson, N. (2011) *How Twitter Was Founded - Business Insider*. Disponible en: https://www.businessinsider.com/how-twitter-was-founded-2011-4 (Accedido: 29 de noviembre de 2020).

Carpenter, J. (2015) «Preservice Teachers ' Microblogging : Professional Development via Twitter», *Contemporary Issues in Technology and Teacher Education*, 15(2), pp. 1-21.

Carpenter, J. P. y Krutka, D. G. (2014) «How and why educators use Twitter: A survey of the field», *Journal of Research on Technology in Education*, 46(4), pp. 414-434. doi: 10.1080/15391523.2014.925701.

Casadevall, A. y Fang, F. C. (2009) «Is peer review censorship?», *Infection and Immunity*, 77(4), pp. 1273-1274. doi: 10.1128/IAI.00018-09.

Casanella, S. (2019) «La química a Twitter. Anàlisi de mencions». Treball de Fi de Grau (Grau en Enginyeria Química). IQS, Universitat Ramon Llull.

Case, C. J. y King, D. L. (2010) «Cutting Edge Communication: Microblogging At the Fortune 200, Twitter Implementation and Usage», *Issues in Information Systems*, 11(1), pp. 216-223.

Case, C. y King, D. (2011) «Twitter Usage in the Fortune 50: A Marketing Opportunity?», *Journal of Marketing Development and Competitiveness*, 5(3), pp. 94-103.

Cedefop (2016) «Skill shortage and surplus occupations in Europe», *Briefing Note-9115 EN*, (November), pp. 1-4. doi: 10.2801/05116.

Chamizo, J. A. (2011) «La imagen pública de la química», *Educacion Quimica*, 22(4), pp. 320-331.

Chamizo, J. A., Sosa, P. y Zepeda, S. (2005) «Análisis de las ideas previas de la

química», *Enseñanza De Las Ciencias*, Extra, pp. 1-4.

Chaturvedi, I., Cambria, E., Welsch, R. E. y Herrera, F. (2018) «Distinguishing between facts and opinions for sentiment analysis: Survey and challenges», *Information Fusion*. Elsevier, 44(December 2017), pp. 65-77. doi: 10.1016/j.inffus.2017.12.006.

Chen, C. H., Lee, W. P. y Huang, J. Y. (2018) «Tracking and recognizing emotions in short text messages from online chatting services», *Information Processing and Management*. Elsevier, 54(6), pp. 1325-1344. doi: 10.1016/j.ipm.2018.05.008.

Chen, L. y Chen, T. L. (2012) «Use of Twitter for formative evaluation: Reflections on trainer and trainees' experiences», *British Journal of Educational Technology*, 43(2), pp. 49-52. doi: 10.1111/j.1467-8535.2011.01251.x.

Choi, B. C. K. y Pak, A. W. P. (2005) «A catalog of biases in questionnaires», *Preventing Chronic Disease*, 2(1), pp. 1-13.

Cidell, J. (2010) «Content clouds as exploratory qualitative data analysis», *Area*, 42(4), pp. 514-523. doi: 10.1111/j.1475-4762.2010.00952.x.

Clemence, M., Gilby, N., Shah, J. y Swiecicka, J. (2013) «Wellcome Trust Monitor Wave 2 Tracking public views on sicence», *Biomedical Research and Science Education, London: Research, Ipsos Mori*, (May), p. 148.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. y Qu, H. (2010) «Context-preserving, dynamic word cloud visualization», *IEEE Computer Graphics and Applications*, 30(6), pp. 42-53. doi: 10.1109/MCG.2010.102.

Curtis, L., Edwards, C., Fraser, K. L., Gudelsky, S., Holmquist, J., Thornton, K. y Sweetser, K. D. (2010) «Adoption of social media for public relations by nonprofit organizations», *Public Relations Review*, 36(1), pp. 90-92. doi: 10.1016/j.pubrev.2009.10.003.

Davenport, S. W., Bergman, S. M., Bergman, J. Z. y Fearrington, M. E. (2014) «Twitter versus Facebook: Exploring the role of narcissism in the motives and usage of different social media platforms», *Computers in Human Behavior*, 32, pp. 212-220. doi: 10.1016/j.chb.2013.12.011.

Declaración de la Química (2002). Disponible en: http://www.fundacionquimica.org/declaracion_de_quimica.php (Accedido: 22 de enero de 2021).

Denecke, K. (2009) «Are SentiWordNet scores suited for multi-domain sentiment classification?», *4th International Conference on Digital Information Management, ICDIM 2009*. IEEE, pp. 32-37. doi: 10.1109/ICDIM.2009.5356764.

DePaolo, C. A. y Wilkinson, K. (2014) «Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data», *TechTrends*, 58(3), pp. 38-44. doi: 10.1007/s11528-014-0750-9.

Dermentzi, E., Papagiannidis, S., Toro, C. O. y Yannopoulou, N. (2016) «Academic engagement: Differences between intention to adopt social networking sites and other online technologies», *Computers in Human Behavior*, 62, pp. 321-332.

Devitt, A. y Ahmad, K. (2007) «Sentiment polarity identification in financial news: A cohesion-based approach», *ACL 2007 - Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, (June), pp. 984-991.

Dhillon, I. S. y Modha, D. S. (2001) «Concept decompositions for large sparse text data using clustering», *Machine Learning*, 42(1-2), pp. 143-175. doi: 10.1023/A:1007612920971.

Donelan, H. (2016) «Social media for professional development and networking opportunities in academia», *Journal of Further and Higher Education*, 40(5), pp. 706-729. doi: 10.1080/0309877X.2015.1014321.

Dorronsoro, S. (2017) «La imagen pública de la química en Twitter. Evaluación de influencers.» Treball de Fi de Grau (Grau en Enginyeria Química). IQS, Universitat Ramon Llull.

Dron, J. y Anderson, T. (2009) «How the Crowd Can Teach. Handbook of Research on Social Software and Developing Ontologies London IGI Global (Vol. Handbook o, pp. 1-17)». IGI Global. Retrieved from http://www. igi-global. com/viewtitlesample. aspx.

Du, H., Hao, J.-X., Kwok, R. y Wagner, C. (2013) «Can a Lean Medium Enhance Large-Group Communication? Examining the Impact of Interactive Mobile Learning», *Journal of the American Society for Information Science and Technology*, 64(July), pp. 1852-1863. doi: 10.1002/asi.

Duffus, J. (1993) «Glossary for chemists of terms used in toxicology (IUPAC Recommendations 1993)», *Pure and Applied Chemistry*, 65(9), pp. 2003-2122. doi: 10.1351/pac199365092003.

Economics, O. (2019) *The Global Chemical Industry: Catalyzing Growth and Addressing Our World's Sustainability Challenges*.

Ekman, P. (1992) «An Argument for Basic Emotions», *Cognition and Emotion*, pp. 169-200. doi: 10.1080/02699939208411068.

Elías Pérez, C. (2006) «Influencia de los medios de comunicación en la elección ciencias-letras en bachillerato y universidad . El caso español : análisis del período 1988-2001», *Estudios sobre el mensaje periodíatico*, 12, pp. 253-274.

Entine, J. (2011) *How Chemophobia Threatens Public Health*. American Council on Science and Health.

Eurobarometer, S. (2017) «Attitudes of European citizens towards the environment QD1 QD6 Special Eurobarometer 468 Attitudes of European citizens towards the environment QD4», pp. 2-5.

European Commission (2013) *Flash Eurobarometer 361 - Chemicals*.

European Commission (2014) *Special Eurobarometer 416 Attitudes of European citizens towards the environment*, *Special Eurobarometer*. doi: 10.2779/25662.

European Commission (2015) «Science education for responsible citizenship», *Report to the European Commission of the Expert Group on Science Education*. doi: 10.2777/12626.

Eze, C., R C Nurse, J. y Jassim, H. (2013) «The use of web 2.0 technologies in marketing classes: key drivers of student acceptance», *Computers in Human Behavior*, 2, pp. 197-206. doi: 10.1002/cb.1444/abstract.

Feinerer, I., Hornik, K. y Meyer, D. (2008) «Text Mining Infrastructure in R», *Journal Of Statistical Software*, 25(5), pp. 1-54. doi: citeulike-article-id:2842334.

Feldman, R. y Sanger, J. (2007) *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. doi: 10.1017/CBO9780511546914.

Fellows, I. (2018) «wordcloud: Word Clouds». R package version 2.6. https://CRAN.R-project.org/package=wordcloud.

Fernández, I., Gil, D., Carrascosa, J., Cachapuz, A. y Praia, J. (2002) «Visiones deformadas de la ciencia transmitidas por la enseñanza», *Enseñanza De Las Ciencias*, 20(3), pp. 477-488.

Fleiss, J. L. (1971) «Measuring nominal scale agreement among many raters», *Psychological Bulletin*, 76(5), pp. 378-382. doi: 10.1037/h0031619.

Fleiss, J. L. (1981) «Balanced Incomplete Block Designs for Inter-Rater Reliability Studies», *Applied Psychological Measurement*, 5(1), pp. 105-112. doi:

10.1177/014662168100500115.

Fleiss, J. L., Levin, B. y Paik, M. C. (2003) *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken: John Wiley & Sons, Inc.

Forkosh-Baruch, A. y Hershkovitz, A. (2012) «A case study of Israeli higher-education institutes sharing scholarly information with the community via social networks», *Internet and Higher Education*. Elsevier Inc., 15(1), pp. 58-68. doi: 10.1016/j.iheduc.2011.08.003.

Furió Más, C. (2006) «La motivación de los estudiantes y la enseñanza de la Química. Una cuestión controvertida», *Educación Química*, 17(IV Jornadas Internacionales), pp. 222-227.

Galagovsky, L. R. (2005) «La enseñanza de la química pre-universitaria: ¿qué enseñar, cómo, cuánto, para quiénes?», *Química Viva*, 4(1), pp. 8-22.

Galagovsky, L. R. (2007) «ENSEÑAR QUÍMICA VS. APRENDER QUÍMICA: UNA ECUACIÓN QUE NO ESTÁ BALANCEADA», *Revista Química Viva*, mayo, pp. 1-13.

Galiano, J., López Pasquali, C. y Sevillano García, M. L. (2015) «Estrategias de enseñanza en la formación de profesores de química», *The Journal of the Argentine Chemical Society*, 102.

Gatti, L. y Guerini, M. (2012) «Assessing Sentiment Strength in Words Prior Polarities», *arXiv preprint arXiv:1212.4315*.

Gentry, J. (2015) «twitteR: R Based Twitter Client». R package version 1.1.9. https://CRAN.R-project.org/package=twitteR.

Gholipour Shahraki, A. (2015) *Emotion Mining from Text*, *Thesis*. University of Alberta.

Giachanou, A. y Crestani, F. (2016) «Like it or not: A survey of Twitter sentiment analysis methods», *ACM Comput Surv*, 49(2), p. Article 28; 1-41. doi: 10.1145/2938640.

Gibson, H., Faith, J. y Vickers, P. (2013) «A Survey of Two-Dimensional Graph Layout Techniques for Information Visualisation», *Information visualization*. Sage Publications Sage UK: London, England, pp. 324-357. doi: 10.1108/17410391111097438.

Gikas, J. y Grant, M. M. (2013a) «Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media», *Internet and Higher Education*. Elsevier Inc., 19, pp. 18-26. doi: 10.1016/j.iheduc.2013.06.002.

Gikas, J. y Grant, M. M. (2013b) «Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media», *Internet and Higher Education*, 19, pp. 18-26. doi: 10.1016/j.iheduc.2013.06.002.

Gottron, T. (2009) «Document word clouds: Visualising web documents as tag clouds to aid users in relevance decisions», *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5714 LNCS, pp. 94-105. doi: 10.1007/978-3-642-04346-8_11.

Greenhow, C. (2011) «Online social networks and learning», *On the horizon*. Emerald Group Publishing Limited, 19(1), pp. 4-12.

Greenhow, C. y Lewin, C. (2016) «Social media and education: reconceptualizing the boundaries of formal and informal learning», *Learning, Media and Technology*, 41(1), pp. 6-30. doi: 10.1080/17439884.2015.1064954.

Guerini, M., Gatti, L. y Turchi, M. (2013) «Sentiment analysis: How to derive prior polarities from SentiWordNet», *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1259-1269.

Guerris, M., Cuadros, J., Gonzalez, L. y Serrano, V. (2020) «Describing the public

perception of chemistry on twitter», *Chemistry Education Research and Practice*. Royal Society of Chemistry.

Hadhri, M. (2010) *CEFIC Facts and Figures 2010. The European Chemical Industry in a worldwide perspective*. CEFIC. Disponible en: https://es.scribd.com/document/44470516/Facts-and-Figures-2010-Report.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. y Witten, I. H. (2009) «The WEKA data mining software», *SIGKDD Explorations Newsletter*, 11(1), p. 10. doi: 10.1145/1656274.1656278.

Hansen, D., Shneiderman, B. y Smith, M. (2010) *Analyzing social media networks with NodeXL : insights from a connected world*. Morgan Kaufmann.

Hartigan, J. a. y Wong, M. a. (1979) «Algorithm AS 136: A K-Means Clustering Algorithm», *Journal of the Royal Statistical Society C*, 28(1), pp. 100-108. doi: 10.2307/2346830.

Hartings, M. R. y Fahy, D. (2011) «Communicating chemistry for public engagement.», *Nature chemistry*. Nature Publishing Group, 3(9), pp. 674-677. doi: 10.1038/nchem.1094.

Haustein, S. (2019) «Scholarly Twitter metrics», en *Springer handbook of science and technology indicators*. Springer, pp. 729-760.

Hayden, K., Youwen Ouyang, Scinski, L., Olszewski, B. y Bielefeldt, T. (2011) «Increasing Student Interest and Attitudes in STEM: Professional Development and Activities to Engage and Inspire Learners», *Contemporary Issues in Technology and Science Teacher Education*, 11(1), pp. 47-69.

Heil, B. y Piskorski, M. (2009) «New Twitter research: Men follow men and nobody tweets», *Harvard Business Review*, 1.

Heimerl, F., Lohmann, S., Lange, S. y Ertl, T. (2014) «Word cloud explorer: Text analytics based on word clouds», *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE, pp. 1833-1842. doi: 10.1109/HICSS.2014.231.

Hill, J. y Kumar, D. D. (2013) «Challenges for Chemical Education: Implementing the ‘ Chemistry for All ’ Vision», *Journal of the American Institute of Chemists*, 86(2), pp. 27-32.

Holland, J. H. (1992) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press.

Horn, F., Arras, L., Montavon, G., Müller, K.-R. y Samek, W. (2017) «Discovering topics in text datasets by visualizing relevant words», *arXiv preprint arXiv:1707.06100*, pp. 1-5.

Hornik, K., Feinerer, I., Kober, M. y Buchta, C. (2012) «Spherical k-Means Clustering», *Journal of Statistical Software*, 50(10), pp. 1-22.

Hornik, K., Mair, P., Rauch, J., Geiger, W., Buchta, C. y Feinerer, I. (2013) «The textcat Package for n-Gram Based Text Categorization in R», *Journal of Statistical Software*, 52(6), pp. 1-17. doi: 10.18637/jss.v052.i06.

Hu, M. y Liu, B. (2004) «Mining and summarizing customer reviews», *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177. doi: 10.1145/1014052.1014073.

Huang, J., Thornton, K. M. y Efthimiadis, E. N. (2010) «Conversational tagging in Twitter», *HT'10 - Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, pp. 173-177. doi: 10.1145/1810617.1810647.

Huberman, B. A., Romero, D. M. y Wu, F. (2009) «Social networks that matter Twitter under the microscope», *First Monday*, 14(1), pp. 1-9. doi: 10.5210/fm.v14i1.2317.

Hunter, J. D. y Caraway, H. J. (2014) «Urban Youth Use Twitter to Transform Learning and Engagement», *English Journal*, 103, pp. 76-82.

Hutto, C. J. y Gilbert, E. (2014) «Vader: A parsimonious rule-based model for sentiment analysis of social media text», *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), pp. 216-225.

ICIS Chemical Business (2016) *Special Report. ICIS Top 100 Chemical Companies*. Disponible en: https://goo.gl/ksuKpD (Accedido: 21 de septiembre de 2017).

Ioanid, A. y Scarlat, C. (2017) «Factors Influencing Social Networks Use for Business: Twitter and YouTube Analysis», *Procedia Engineering*, 181, pp. 977-983. doi: 10.1016/j.proeng.2017.02.496.

Jain, A. K. (2010) «Data clustering: 50 years beyond K-means», *Pattern Recognition Letters*. Elsevier B.V., 31(8), pp. 651-666. doi: 10.1016/j.patrec.2009.09.011.

Jansen, B. J., Sobel, K. y Zhang, M. (2009) «The Commercial Impact of Social Mediating Technologies : Micro-blogging as Online Word-of-Mouth Branding», *In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, (December), pp. 3859-3864.

Jansen, B. J., Zhang, M., Sobel, K. y Chowdury, A. (2009) «Twitter Power:Tweets as ElectronicWord of Mouth», *Journal of the American Society for Information Science and Technology*, 60(11), pp. 2169-2188. doi: 10.1002/asi.

Jiménez, J. B. y Criado García-Legaz, A. (2005) «Análisis de las actividades sobre la historia de la química en los libros de física y química del segundo ciclo de la eso», *Enseñanza De Las Ciencias*, (2000), pp. 1-6.

Judd, T. (2010) «Facebook versus email: Colloquium», *British Journal of Educational Technology*, 41(5), pp. 2009-2011. doi: 10.1111/j.1467-8535.2009.01041.x.

Junco, R., Elavsky, C. M. y Heiberger, G. (2013) «Putting twitter to the test: Assessing outcomes for student collaboration, engagement and success», *British Journal of Educational Technology*, 44(2), pp. 273-287. doi: 10.1111/j.1467-8535.2012.01284.x.

Junco, R., Heiberger, G. y Loken, E. (2011) «The effect of Twitter on college student engagement and grades», *Journal of Computer Assisted Learning*, 27(2), pp. 119-132. doi: 10.1111/j.1365-2729.2010.00387.x.

Kamada, T. y Kawai, S. (1989) «An algorithm for drawing general undirected graphs», *Information Processing Letters*, 31(April), pp. 7-15.

Kanavos, A., Perikos, I., Vikatos, P., Hatzilygeroudis, I., Makris, C. y Tsakalidis, A. (2014) «Conversation Emotional Modeling in Social Networks», *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2014-Decem, pp. 478-484. doi: 10.1109/ICTAI.2014.78.

Kaptein, R., Hiemstra, D. y Kamps, J. (2010) «How different are language models and word clouds?», *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5993 LNCS, pp. 556-568. doi: 10.1007/978-3-642-12275-0-48.

Kassens-Noor, E. (2012) «Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets», *Active Learning in Higher Education*, 13(1), pp. 9-21. doi: 10.1177/1469787411429190.

Kaufman, L. y Rousseeuw, P. J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P. y Silvestre, B. S. (2011) «Social media? Get serious! Understanding the functional building blocks of social media», *Business Horizons*, 54(3), pp. 241-251. doi: 10.1016/j.bushor.2011.01.005.

King, A. A. y Lenox, M. J. (2000) «Industry Self-Regulation without Sanctions.pdf», *Academy of Management Journal*, pp. 698-716.

Kiritchenko, S., Zhu, X. y Mohammad, S. M. (2014) «Sentiment analysis of short informal texts», *Journal of Artificial Intelligence Research*, 50, pp. 723-762. doi: 10.1613/jair.4272.

Knight, C. y Kaye, L. K. (2016) «"To Tweet or not to Tweet?": A comparison of academics' and students' usage of Twitter in academic contexts», *Innovations in education and teaching international*, 53(2), pp. 145-155.

Kobourov, S. G. (2012) «Spring Embedders and Force Directed Graph Drawing Algorithms», *arXiv preprint arXiv:1201.3011*, pp. 1-23.

Kodinariya, T. M. y Makwana, P. R. (2013) «Review on determining number of Cluster in K-Means Clustering», *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), pp. 90-95.

Krishna, K. y Murty, M. N. (1999) «Genetic K-means algorithm», *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 29(3), pp. 433-439. doi: 10.1109/3477.764879.

Krumpal, I. (2013) «Determinants of social desirability bias in sensitive surveys: A literature review», *Quality and Quantity*, 47(4), pp. 2025-2047. doi: 10.1007/s11135-011-9640-9.

Krutka, D. G. (2014) «Social media as a catalyst for convergence culture: Immersing pre-service social studies teachers in the social media terrain», *Digital social studies*. Information Age Publishing Charlotte, NC, pp. 271-302.

Krutka, D. G. y Carpenter, J. P. (2016) «Participatory Learning Through Social Media: How and Why Social Studies Educators Use Twitter», *Contemporary Issues in Technology and Teacher Education*, 16(1), pp. 38-59.

Kuo, B. Y. L., Hentrich, T., Good, B. M. y Wilkinson, M. D. (2007) «Tag clouds for summarizing web search results», *16th International World Wide Web Conference, WWW2007*, pp. 1203-1204. doi: 10.1145/1242572.1242766.

Kwak, H., Lee, C., Park, H. y Moon, S. (2010) «What is Twitter, a Social Network or a News Media?», en *Proceedings of the 19th international conference on World wide web*. AcM, pp. 591-600. doi: https://doi.org/10.1145/1772690.1772751.

De Laat, M., Lally, V., Lipponen, L. y Simons, R. J. (2007) «Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis», *International Journal of Computer-Supported Collaborative Learning*, 2(1), pp. 87-103. doi: 10.1007/s11412-007-9006-4.

Lacolla, L., Meneses Villagrá, J. A. y Valeiras, N. (2013) «Las representaciones sociales y las reacciones químicas: Desde las explosiones hasta Fukushima», *Educacion Quimica*, 24(3), pp. 309-315.

Landis, J. R. y Koch, G. G. (1977) «The Measurement of Observer Agreement for Categorical Data», *Biometrics*, 33(1), pp. 159-174. doi: 10.2307/2529310.

Larson, K. y Watson, R. (2011) «The value of social media: toward measuring social media strategies», *Thirty Second International Conference on Information Systems, Shanghai 2011*.

Laszlo, P. (2006) «On the self-image of chemists, 1950-2000», *International Journal for Philosophy of Chemistry*, 12(1), pp. 99-130. doi: 10.1142/9789812775856_0013.

Lazlo, P. y Greenberg, A. (1991) «Falacias acerca de la química», *Educación Química*, 2(1), pp. 29-35.

Lin, P.-C., Hou, H.-T., Wang, S.-M. y Chang, K.-E. (2013) «Analyzing knowledge

dimensions and cognitive process of a project-based online discussion instructional activity using Facebook in an adult and continuing education course», *Computers & Education*. Elsevier, 60(1), pp. 110-121.

Lin, P. C., Hou, H. T., Wang, S. M. y Chang, K. E. (2013) «Analyzing knowledge dimensions and cognitive process of a project-based online discussion instructional activity using Facebook in an adult and continuing education course», *Computers and Education*. Elsevier Ltd, 60(1), pp. 110-121. doi: 10.1016/j.compedu.2012.07.017.

Lin Pedersen, T. (2019) «ggraph: An Implementation of Grammar of Graphics for Graphs and Networks». R package version 2.0.0 https://CRAN.R-project.org/package=ggraph.

Liu, B., Hu, M. y Cheng, J. (2005) «Opinion Observer: Analyzing and Comparing Opinions on the Web», *Proceedings of the 14th International Conference on World Wide Web*, pp. 342-351. doi: 10.1145/1060745.1060797.

Liu, L., Kang, J. y Wang, Z. (2005) «A comparative study on unsupervised feature selection methods for text clustering», *2005 International Conference on Natural Language Processing and Knowledge Engineering*, 00, pp. 597-601. doi: 10.1109/NLPKE.2005.1598807.

Lloyd, E. K., Bondy, J. A. y Murty, U. S. R. (1976) *Graph Theory with Applications*. Macmillan London. doi: 10.2307/3617646.

Lovejoy, K., Waters, R. D. y Saxton, G. D. (2012) «Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less», *Public Relations Review*, 38(2), pp. 313-318. doi: 10.1016/j.pubrev.2012.01.005.

Lövheim, H. (2012) «A new three-dimensional model for emotions and monoamine neurotransmitters», *Medical Hypotheses*, 78(2), pp. 341-348. doi: 10.1016/j.mehy.2011.11.016.

Lowe, B. y Laffey, D. (2011) «Is twitter for the birds? Using twitter to enhance student learning in a marketing course», *Journal of Marketing Education*, 33(2), pp. 183-192. doi: 10.1177/0273475311410851.

Luhn, H. P. (1957) «A Statistical Approach to Mechanized Encoding and Searching of Literary Information», *IBM Journal of Research and Development*, 1(4), pp. 309-317. doi: 10.1147/rd.14.0309.

Luo, T. (2015) «Instructional guidance in microblogging-supported learning: insights from a multiple case study», *Journal of Computing in Higher Education*. Springer US, 27(3), pp. 173-194. doi: 10.1007/s12528-015-9097-2.

Luo, T. y Franklin, T. (2015) «ODU Digital Commons Tweeting and Blogging : Moving Towards Education 2 . 0», *International Journal on E-Learning*, 14(2), pp. 235-258.

Madge, C., Meek, J., Wellens, J. y Hooley, T. (2009) «Facebook, social integration and informal learning at university:'It is more for socialising and talking to friends about work than for actually doing work'», *Learning, media and technology*. Taylor & Francis, 34(2), pp. 141-155.

Madhulatha, T. S. (2012) «An overview on clustering methods», *IOSR Journal of Engineering*, 02(04), pp. 719-725. doi: 10.9790/3021-0204719725.

Madhusudhan, M. (2012) «Use of social networking sites by research scholars of the University of Delhi: A study», *The International Information & Library Review*. Taylor & Francis, 44(2), pp. 100-113.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. y Hornik, K. (2019) «cluster: Cluster Analysis Basics and Extensions». R package version 2.1.0 https://cran.r-project.org/web/packages/cluster/index.html.

Mahaffy, P., Ashmore, A., Bucat, B., Do, C. y Rosborough, M. (2008) «Chemists and

"the public": IUPAC's role in achieving mutual understanding (IUPAC technical report)», *Pure and Applied Chemistry*, 80(1), pp. 161-174. doi: 10.1351/pac200880010161.

Malaver, M., Pujol, R. y D'Alessandro Martínez, A. (2004) «Los Estilos De Prosa Y El Enfoque Ciencia-Tecnología-Sociedad En Textos Universitarios De Química General», *Educación Química*, 22(3), pp. 441-453.

Mammino, L. (2001) «Algunas reflexiones sobre la imagen de la Química», *Anales de la Real Sociedad Española de la Química*, 2, pp. 48-52.

Manca, S. y Ranieri, M. (2016) «Facebook and the others. Potentials and obstacles of Social Media for teaching in higher education», *Computers and Education*. Elsevier Ltd, 95, pp. 216-230. doi: 10.1016/j.compedu.2016.01.012.

Mandal, B. N. (2019) «ibd: Incomplete Block Designs». R package version 1.5.0 https://cran.r-project.org/package=ibd.

Manitz, J., Hempelmann, M., Kauermann, G., Kuechenhoff, H., Shao, S., Oberhauser, C., Westerheide, N. y Wiesenfarth, M. (2017) «samplingbook: Survey Sampling Procedures». R package version 1.2.2 https://CRAN.R-project.org/package=samplingbook.

Mäntylä, M. V., Graziotin, D. y Kuutila, M. (2018) «The evolution of sentiment analysis—A review of research topics, venues, and top cited papers», *Computer Science Review*. Elsevier Inc., 27, pp. 16-32. doi: 10.1016/j.cosrev.2017.10.002.

Marques, A. M., Krejci, R., Siqueira, S. W. M., Pimentel, M. y Braz, M. H. L. B. (2013) «Structuring the discourse on social networks for learning: Case studies on blogs and microblogs», *Computers in Human Behavior*. Elsevier Ltd, 29(2), pp. 395-400. doi: 10.1016/j.chb.2012.03.001.

Martin-Valdivia, M. T. y Martínez Camara, E. (2013) «Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches», *Expert Systems with Applications*. Elsevier, 40(10), pp. 3934-3942.

Mason, R. y Rennie, F. (2008) «Social networking as an educational tool», *E-learning and social networking handbook: Resources for higher education*, pp. 1-24.

Matthew L. Jockers (2015) «Syuzhet: Extract Sentiment and Plot Arcs from Text». R package v.1.0.4 https://github.com/mjockers/syuzhet.

McDermott, L. C. (1984) «Research on conceptual understanding in mechanics», *Physics Today*, 37, pp. 24-32.

Medhat, W., Hassan, A. y Korashy, H. (2014) «Sentiment analysis algorithms and applications: A survey», *Ain Shams Engineering Journal*. Faculty of Engineering, Ain Shams University, 5(4), pp. 1093-1113. doi: 10.1016/j.asej.2014.04.011.

Mercier, E., Julie, R. y Lavery, J. (2015) «Twitter in the collaborative classroom: micro-blogging for in-class collaborative discussions», *International journal of social media and interactive learning environments*, 3(2), pp. 83-99.

Michaelidou, N., Siamagka, N. T. y Christodoulides, G. (2011) «Usage, Barriers and Measurement of Social Media Marketing: An Exploratory Investigation of Small and Medium B2B Brands», *Industrial marketing management*, 40(7), pp. 1153-1159. doi: 10.1017/CBO9781107415324.004.

Michaelis, A. R. (1996) «Stop-chemophobia», *Interdisciplinary Science Reviews*. Taylor & Francis, 21(2), pp. 130-139.

Microsoft y Weston, S. (2019a) «doParallel: Foreach Parallel Adaptor for the "parallel" Package». R package version 1.0.15 https://CRAN.R-project.org/package=doParallel.

Microsoft y Weston, S. (2019b) «foreach: Provides Foreach Looping Construct». R package version 1.4.7 https://CRAN.R-project.org/package=foreach.

Mishori, R., Levy, B. y Donvan, B. (2014) «Twitter use at a family medicine conference: analyzing #STFM13», *Family medicine*, 46(8), pp. 608-614.

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. y Vespignani, A. (2013) «The Twitter of Babel: Mapping World Languages through Microblogging Platforms», *PLoS ONE*, 8(4). doi: 10.1371/journal.pone.0061981.

Mohammad, S. M. (2012) «#Emotional tweets», en *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics, pp. 246-255.

Mohammad, S. M. y Turney, P. D. (2010) «Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon», *CAAGET '10 Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, (June), pp. 26-34.

Mohammad, S. M. y Turney, P. D. (2013) «NRC Emotion Lexicon», pp. 1-234.

Mohammadi, E., Thelwall, M., Kwasny, M. y Holmes, K. L. (2018) «Academic information on Twitter: A user survey», *PLoS ONE*, 13(5), pp. 1-18. doi: 10.1371/journal.pone.0197265.

Mohey, D. y Hussein, E. M. (2016) «A survey on sentiment analysis challenges», *Journal of King Saud University - Engineering Sciences*. King Saud University. doi: 10.1016/j.jksues.2016.04.002.

Montejo-Ráez, A., Martinez-Cámara, E., Martin-Valdivia, M. T. y Ureña-López, L. A. (2014) «Ranked wordnet graph for sentiment polarity classification in twitter», *Computer Speech & Language*. Elsevier, 28(1), pp. 93-107.

Mora Penagos, W. M. (1997) «Naturaleza del conocimiento científico e implicaciones didácticas», *Revista Educación y Pedagogía*, 9(18), pp. 131-144.

Morales Vallejo, P. (2012) «Tamaño necesario de la muestra: ¿Cuántos sujetos necesitamos?», *Estadística aplicada*, 24(1), pp. 22-39.

Moreau, N. J. (2005) «Public Images of Chemistry», *Chemistry International*, (August), pp. 9-12.

Moreno, L. (2013) «Cartas al editor», *Anales de la Química*, 109(4).

Muñoz, L. y Nardi, R. (2011) «Las representaciones científicas en la formación inicial de profesores de química», *Encontro Nacional de Pesquisa em Educação em Ciências*, 8.

Neier, S. y Zayer, L. T. (2015) «Students' Perceptions and Experiences of Social Media in Higher Education», *Journal of Marketing Education*, 37(3), pp. 133-143. doi: 10.1177/0273475315583748.

Nepusz, T. y Csardi, G. (2006) «The igraph software package for complex network research», *InterJournal, complex systems*, 1695(5), pp. 1-9.

Network Science Corp. (2012a) *List of Federations*. Disponible en: https://goo.gl/j1Y8Cx (Accedido: 7 de abril de 2017).

Network Science Corp. (2012b) *List of Professional Societies for Analytical Chemistry*. Disponible en: https://goo.gl/jJWa88 (Accedido: 7 de abril de 2017).

Network Science Corp. (2012c) *List of Professional Societies for Chemistry*. Disponible en: https://goo.gl/hKE2We (Accedido: 17 de abril de 2017).

Newman, M. y Newman, M. E. J. (2010) «Mathematics of networks», *Networks*, pp. 109-167. doi: 10.1093/acprof:oso/9780199206650.003.0006.

Nicolas, E. (2006) «Aula y Laboratorio de Química La Química vista por 840 estudiantes de bachillerato», *Anales de la Química*, 102(4), pp. 64-67.

Nielsen, F. Å. (2011) «A new ANEW: Evaluation of a word list for sentiment analysis in microblogs», *CEUR Workshop Proceedings*, 718, pp. 93-98. doi: 10.1016/j.knosys.2015.06.015.

Noack, A. (2009) «Modularity clustering is force-directed layout», *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 79(2). doi: 10.1103/PhysRevE.79.026102.

Noorden, R. Van (2014) *A_battle_for_profiles*, *Nature*. Disponible en: http://www.nature.com/news/online-collaboration-scientists-and-the-social-network-1.15711 (Accedido: 10 de septiembre de 2014).

Oatley, K. y Johnson-Laird, P. N. (1987) «Towards a cognitive theory of emotions», *Cognition and emotion*. Taylor & Francis, 1(1), pp. 29-50.

Ortony, A., Clore, G. L. y Collins, A. (1988) «The cognitive structure of emotions», *Cambridge University Press*.

Oulasvirta, A., Lehtonen, E., Kurvinen, E. y Raento, M. (2010) «Making the ordinary visible in microblogs», *Personal and Ubiquitous Computing*, 14(3), pp. 237-249. doi: 10.1007/s00779-009-0259-y.

Palermo, A. (2018) «The future of the Chemical Sciences. Preparing for an Uncertain Future», *Chemistry World*, p. 6. doi: 10.1021/ed020p304.

Paoli, S. De y Rooy, A. La (2015) «Teaching with Twitter : reflections on practices , opportunities and problems», *EUNIS Journal of Higher Education IT*, 3.

Partington, J. R. (1951) *A short history of chemistry*. 2ª edición. Macmillan.

Parveen, F. (2012) «Impact Of Social Media Usage On Organizations», en *Pacific Asia Conference on Information Systems (PACIS)*.

Penagos, W. M. M. y Lozano, D. L. P. (2009) «La imagen pública de la química y su relación con la generación de actitudes hacia la química y su aprendizaje», *Tecné, Episteme y Didaxis: TED*, (27), pp. 67-93.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A. y Booth, R. J. (2007) «Linguistic Inquiry and Word Count (LIWC). Austin, TX: LIWC». Inc.

Piotrowski, C. (2015) «Emerging research on social media use in education: a study of dissertations», *Research in Higher Education Journal*, 27(January), pp. 1-12.

Plutchik, R. (1980) «A general psychoevolutionary theory of emotion», en *Theories of emotion*. Elsevier, pp. 3-33.

Popescu, A. M. y Jain, A. (2011) «Understanding the functions of business accounts on Twitter», *Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011*, pp. 107-108. doi: 10.1145/1963192.1963247.

Prestridge, S. (2014) «A focus on students' use of Twitter - their interactions with each other, content and interface», *Active Learning in Higher Education*, 15(2), pp. 101-115. doi: 10.1177/1469787414527394.

Quan-Haase, A., Martin, K. y McCay-Peet, L. (2015) «Networks of digital humanities scholars: The informational and social uses and gratifications of Twitter», *Big Data and Society*, 2(1), pp. 1-12. doi: 10.1177/2053951715589417.

R Core Team (2018) «R: A Language and Environment for Statistical Computing». Vienna, Austria.

RAE (2020) *químico, ca*. Disponible en: https://dle.rae.es/químico#Us0hxUS (Accedido: 20 de diciembre de 2020).

Reid, D. y Ostashewski, N. (2010) «Evolution of online teacher professional development in a social networking site: What's been working and what's not», en *EdMedia+ Innovate Learning*, pp. 1117-1122.

Reinhold, O. y Alt, R. (2012) «Social Customer Relationship Management: State of the Art and Learnings from Current Projects», en *Bled eConference, 26*.

Ribelles, R., Solbes, J. y Vilches, A. (1995) «Las interacciones C.T.S. en la enseñanza de las ciencias. Análisis comparativo de la situación para la Física y Química y la Biología y Geología», *Comunicación, Lenguaje y Educación*, pp. 135-143.

Richter, F. (2013) *Only 34% of All Tweets Are in English*. Disponible en: https://www.statista.com/chart/1726/languages-used-on-twitter/ (Accedido: 21 de abril de 2021).

Rinaldo, Shannon B., Tapp, S. y Laverie, D. A. (2011) «Learning by tweeting: Using twitter as a pedagogical tool», *Journal of Marketing Education*, 33(2), pp. 193-203. doi: 10.1177/0273475311410852.

Rinaldo, Shannon B, Tapp, S. y Laverie, D. A. (2011) «Learning by tweeting: Using Twitter as a pedagogical tool», *Journal of marketing education*. SAGE Publications Sage CA: Los Angeles, CA, 33(2), pp. 193-203.

Rinker, T. W. (2013) «qdapDictionaries: Dictionaries to Accompany the qdap Package». 1.0.7. University at Buffalo. Buffalo, New York.

Riquelme, F. y González-Cantergiani, P. (2016) «Measuring user influence on Twitter: A survey», *Information Processing and Management*, 52(5), pp. 949-975. doi: 10.1016/j.ipm.2016.04.003.

Roblyer, M. D., McDaniel, M., Webb, M., Herman, J. y Witty, J. V. (2010) «Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites», *The Internet and higher education*. Elsevier, 13(3), pp. 134-140.

Rodríguez Gómez, J. M. (2009) «Cambios metodológicos relacionados con el aprendizaje de las ciencias», *Revista Educación*, 33(1), pp. 61-73.

Rollini, R. (2020) *Chemophobia: a systematic review*. SISSA.

Rosen, A. y IKuhiro, I. (2017) *Giving you more characters to express yourself*. Disponible en: https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html.

Rousseeuw, P. J. (1987) «Silhouettes: A graphical aid to the interpretation and validation of cluster analysis», *Journal of Computational and Applied Mathematics*, 20(C), pp. 53-65. doi: 10.1016/0377-0427(87)90125-7.

Russell, J. A. (1980) «A circumplex model of affect.», *Journal of personality and social psychology*, 39(6), p. 1161.

Sabater, J. y Sierra, C. (2002) «Reputation and social network analysis in multi-agent systems», *Proceedings of the International Conference on Autonomous Agents*, (2), pp. 475-482. doi: 10.1145/544852.544854.

Sailer, M. O. (2013) «crossdes: Construction of Crossover Designs». R package v1.1.1. https://cran.r-project.org/package=crossdes.

Sailunaz, K. y Alhajj, R. (2019) «Emotion and sentiment analysis from Twitter text», *Journal of Computational Science*. Elsevier B.V., 36, p. 101003. doi: 10.1016/j.jocs.2019.05.009.

Sailunaz, K., Dhaliwal, M., Rokne, J. y Alhajj, R. (2018) «Emotion detection from text and speech: a survey», *Social Network Analysis and Mining*. Springer Vienna, 8(1), pp. 1-26. doi: 10.1007/s13278-018-0505-2.

Salmon, G., Ross, B., Pechenkina, E. y Chase, A. M. (2015) «The space for social media in structured online learning», *Research in Learning Technology*, 23(1063519), pp. 1-14. doi: 10.3402/rlt.v23.28507.

Salton, G. y Buckley, C. (1988) «Term-weighting approaches in automatic text retrieval», *Information Processing and Management*, 24(5), pp. 513-523.

Salvador, S. y Chan, P. (2004) «Determining the Number of Clusters / Segments in Hierarchical Clustering / Segmentation Algorithms», *16th IEEE international conference on tools with artificial intelligence*, pp. 576-584. doi: 10.1109/ICTAI.2004.50.

Sammer, T. y Back, A. (2011) «Towards microblogging success factors: an empirical survey on Twitter usage of Austrian universities», *MCIS 2011 Proceedings*, 40.

Schibeci, R. A. (1986) «Images of science and scientists and science education», *Science Education*. Wiley Online Library, 70(2), pp. 139-149. doi: 10.1002/sce.3730700208.

Schummer, J. (2004) «The public images of chemistry in the twentieth century», *International Conference, Paris, 17-18 September 2004*.

Schummer, J. B.-V. B. V. T. B. (2006) «Editorial : The Public Image of Chemistry , I», *International Journal for Philosophy of Chemistry*, 12(1), pp. 3-4.

Schummer, J., Bensaude-Vincent, B. y Van Tiggelen, B. (2007) *The Public Image of Chemistry*, *World Scientific Publishing*. World Scientific Publishing. doi: 10.1142/9789812775856.

Scott, J. y Carrington, P. J. (2011) «The SAGE handbook of social network analysis». SAGE publications. doi: https://doi-org.sare.upf.edu/10.1177%2F0038038588022001007.

Seaman, J. y Tinti-kane, H. (2013) «Social Media for Teaching and Learning», *Pearson*, (October), pp. 1-32.

Selwyn, N. (2007) «Web 2.0 applications as alternative environments for informal learning-a critical review», *Paper for CERI-KERIS international expert meeting on ICT and educational performance*, 16. doi: 10.5020/18061230.2018.7402.

Shaver, P., Schwartz, J., Kirson, D., O'Connor, C. y O'connor, G. (1987) «Emotion Knowledge: Further Exploration of a Prototype Approach», *Journal of personality and social psychology*, 52(6), p. 1061.

Sing, J. K., Sarkar, S. y Mitra, T. K. (2012) «Development of a novel algorithm for sentiment analysis based on adverb-adjective-noun combinations», *Proceedings - 2012 3rd National Conference on Emerging Trends and Applications in Computer Science, NCETACS-2012*. IEEE, 1, pp. 38-40. doi: 10.1109/NCETACS.2012.6203294.

Skeels, M. M. y Grudin, J. (2009) «When social networks cross boundaries: A case study of workplace use of facebook and linkedin», *GROUP'09 - Proceedings of the 2009 ACM SIGCHI International Conference on Supporting Group Work*, pp. 95-103. doi: 10.1145/1531674.1531689.

Smith, A., Lee, T. Y., Poursabzi-Sangdeh, F., Boyd-Graber, J., Elmqvist, N. y Findlater, L. (2017) «Evaluating Visual Representations for Topic Understanding and Their Effects on Manually Generated Topic Labels», *Transactions of the Association for Computational Linguistics*, 5, pp. 1-16. doi: 10.1162/tacl_a_00042.

Smith, M. A., Rainie, L., Shneiderman, B. y Himelboim, I. (2014) «Mapping Twitter Topic Networks : From Polarized Crowds to Community Clusters», *Pew Research Center*, 20, pp. 1-56.

Solbes, J. y Traver, M. (1996) «La utilización de la historia de las ciencias en la enseñanza de la Física y la Química», *Enseñanza de las Ciencias*, 14(1), pp. 103-112.

Solbes, J. y Vilches, A. (1992) «El modelo constructivista y las relaciones ciencia/técnica/sociedad», *Enseñanza de las Ciencias*, 10(2), pp. 181-186.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F. y Pantic, M. (2017) «A survey of multimodal sentiment analysis», *Image and Vision Computing*. Elsevier B.V., 65, pp. 3-14. doi: 10.1016/j.imavis.2017.08.003.

Statista (2019a) *Global active usage penetration of leading social networks as of February 2018*. Disponible en: https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/ (Accedido: 9 de abril de 2019).

Statista (2019b) *Most famous social network sites worldwide as of January 2019, ranked by number of active users (in millions)*. Disponible en: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (Accedido: 9 de abril de 2019).

Statista (2019c) *Share of social media users in the United States who have changed their views about a political or social issue because of something they saw on social media in the past year as of June 2018*. Disponible en: https://www.statista.com/statistics/244921/social-medias-influence-on-political-opinions-of-us-internet-users/ (Accedido: 9 de abril de 2019).

Statista (2019d) *Social media usage worldwide*, *Statista*. Disponible en: https://www-statista-com.gate3.library.lse.ac.uk/study/12393/social-networks-statista-dossier/ (Accedido: 19 de abril de 2019).

Statista (2020a) *Chemical industry worldwide 2020*. Disponible en: https://www.statista.com/markets/410/topic/445/chemical-industry/#overview (Accedido: 29 de noviembre de 2020).

Statista (2020b) *Social media usage worldwide*, *Statista*. Disponible en: https://www.statista.com/topics/1164/social-networks/ (Accedido: 29 de noviembre de 2020).

Statista (2020c) *Twitter: monthly active users worldwide* | *Statista*, *Statista*. Disponible en: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ (Accedido: 29 de noviembre de 2020).

Statista (2020d) *Twitter Statista Dossier*, *Statista*. Disponible en: https://www.statista.com/study/9920/twitter-statista-dossier/ (Accedido: 29 de noviembre de 2020).

Stekolschik, G., Draghi, C., Adaszko, D. y Gallardo, S. (2010) «Does the public communication of science influence scientific vocation? results of a national survey», *Public Understanding of Science*, 19(5), pp. 625-637. doi: 10.1177/0963662509335458.

Stevens, V. (2008) «Trial by Twitter: The rise and slide of the year's most viral microblogging platform», *TESL-EJ*, 12(1), pp. 1-14.

Stieglitz, S. y Dang-Xuan, L. (2013) «Social media and political communication: a social media analytics framework», *Social Network Analysis and Mining*, 3(4), pp. 1277-1291. doi: 10.1007/s13278-012-0079-3.

Stieglitz, S. y Krüger, N. (2011) «Analysis of sentiments in corporate Twitter communication - A case study on an issue of Toyota», *ACIS 2011 Proceedings - 22nd Australasian Conference on Information Systems*.

Stocklmayer, S. y K. Gilbert, J. (2002) «Informal Chemical Education», en W. Coben, W. y Tobin, K. (eds.) *Chemical education: Towards research-based practice*. Kluwer Academic Publishers, pp. 143-164. doi: 10.1007/0-306-47977-X_4.

Strapparava, C. y Valitutti, A. (2004) «WordNet-Affect: An affective extension of WordNet», *Proceedings of the 4th International Conference on Language Resources*

and Evaluation, LREC 2004, pp. 1083-1086.

Suárez, J. G., Cervantes, C. T. y García, E. R. (2015) «Twitter como recurso para evaluar el proceso de enseñanza universitaria», *RUSC Universities and Knowledge Society Journal*, 12(3), pp. 32-45. doi: 10.7238/rusc.v12i3.2092.

Sun, S., Luo, C. y Chen, J. (2017) «A review of natural language processing techniques for opinion mining systems», *Information Fusion*. Elsevier B.V., 36, pp. 10-25. doi: 10.1016/j.inffus.2016.10.004.

Swani, K., Milne, G. R. y Brown, B. P. (2014) «Should B2B Tweets differ from B2C Tweets? An analysis of Fortune 500 companies' Twitter communication», *Industrial Marketing Management*, 43(5), pp. 873-881.

Tago, K. y Jin, Q. (2018) «Influence analysis of emotional behaviors and user relationships based on Twitter data», *Tsinghua Science and Technology*, 23(1), pp. 104-113. doi: 10.26599/TST.2018.9010012.

Tamassia, R. (2013) *Handbook of graph drawing and visualization*. CRC Press.

Tang, Y. y Hew, K. F. (2017) «Using Twitter for education: Beneficial or simply a waste of time?», *Computers and Education*. Elsevier Ltd, 106, pp. 97-118. doi: 10.1016/j.compedu.2016.12.004.

Team, R. C. y others (2013) «R: A language and environment for statistical computing». Vienna, Austria.

The Editors of Encyclopaedia Britannica (2020) *Twitter | History, Description, & Uses | Britannica*. Disponible en: https://www.britannica.com/topic/Twitter (Accedido: 29 de noviembre de 2020).

The Royal Society of Chemistry y TNS BMRB (2015) «Public attitudes to chemistry». Research report. https://www.rsc.org/campaigning-outreach/campaigning/public-attitudes-chemistry/, pp. 1-78.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. y Kappas, A. (2010) «Sentiment strength detection in short informal text», *Journal of the American Society for Information Science and Technology*, 61(12), pp. 2544–2558.

Thet, T. T., Na, J. C. y Khoo, C. S. G. (2010) «Aspect-based sentiment analysis of movie reviews on discussion boards», *Journal of Information Science*, 36(6), pp. 823-848. doi: 10.1177/0165551510388123.

Thorndike, R. L. (1953) «Who belongs in the family», *Psychometrika*, 18(4), pp. 267-276.

Tourangeau, R. y Yan, T. (2007) «Sensitive Questions in Surveys», *Psychological Bulletin*, 133(5), pp. 859-883. doi: 10.1037/0033-2909.133.5.859.

Trinder, K., Guiller, J., Margaryan, A., Littlejohn, A. y Nicol, D. (2008) «Learning From Digital Natives: Bridging Formal and Informal Learning», *Higher Education*, 1(May), pp. 1-57.

Trozzolo, A. M. (1975) «The image of chemistry». Conference. https://www3.nd.edu/~atrozzol/Image-2.pdf, pp. 1-7.

Tur, G. y Marín, V. I. (2014) «Enhancing learning with the social media: student teachers' perceptions on Twitter in a debate activity», *Journal of New Approaches in Educational Research*, 4(1), pp. 46-43. doi: 10.7821/naer.2015.1.102.

Twitter Inc. (2016) *Política de Privacidad de Twitter*, *Twitter Inc.* Disponible en: https://twitter.com/privacy?lang=es (Accedido: 25 de abril de 2021).

Twitter Inc. (2020a) *Información sobre las API de Twitter*. Disponible en: https://help.twitter.com/es/rules-and-policies/twitter-api (Accedido: 6 de diciembre de 2020).

Twitter Inc. (2020b) *Tweet Object*. Disponible en: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/tweet-object (Accedido: 6 de diciembre de 2020).

University of California (2019) *Guidance for Reviewing Protocols that Include Online Sources or Mobile Devices*. Disponible en: https://research.uci.edu/compliance/human-research-protections/irb-members/reviewing-protocols-online-mobile.html (Accedido: 26 de abril de 2020).

University of Wisconsin (2019) *IRB Guidance: Technology & New Media Research*. Disponible en: https://kb.wisconsin.edu/page.php?id=42376&no_frill=1 (Accedido: 26 de abril de 2020).

Veletsianos, G. (2012) «Higher education scholars' participation and practices on Twitter», *Journal of Computer Assisted Learning*, 28(4), pp. 336-349. doi: 10.1111/j.1365-2729.2011.00449.x.

Virkus, S. (2008) «Use of Web 2.0 technologies in LIS education: Experiences at Tallinn University, Estonia», *Program*, 42(3), pp. 262-274. doi: 10.1108/00330330810892677.

Vivas-Reyes, R. (2009) «Filosofía de la química: un área ampliamente olvidada», *Revista Academica Colombiana de Ciencias*, 33(126), pp. 125-128.

Waters, R. D., Burnett, E., Lamm, A. y Lucas, J. (2009) «Engaging stakeholders through social networking: How nonprofit organizations are using Facebook», *Public Relations Review*, 35(2), pp. 102-106. doi: 10.1016/j.pubrev.2009.01.006.

Welch, B. K. y Bonnan-White, J. (2012) «Twittering to increase student engagement in the university classroom», *Knowledge Management & E-Learning: An International Journal*, 4(3), pp. 325-345. doi: 10.34105/j.kmel.2012.04.026.

Weller, K., Bruns, A., Burgess, J., Mahrt, M. y Puschmann, C. (2014) *Twitter and society [Digital Formations, Volume 89]*. Peter Lang Publishing.

West, B., Moore, H. y Barry, B. (2015) «Beyond the Tweet: Using Twitter to Enhance Engagement, Learning, and Success Among First-Year Students», *Journal of Marketing Education*, 37(3), pp. 160-170. doi: 10.1177/0273475315586061.

Whitley, D. (1994) «A genetic algorithm tutorial», *Statistics and Computing*, 4(2), pp. 65-85. doi: 10.1007/BF00175354.

Wickham, H. (2019) «stringr: Simple, Consistent Wrappers for Common String Operations». R package version 1.4.0 https://cran.r-project.org/package=stringr.

Wilkinson, C. y Weitkamp, E. (2013) «A case study in serendipity: Environmental researchers use of traditional and social media for dissemination», *PLoS ONE*, 8(12), pp. 1-9. doi: 10.1371/journal.pone.0084339.

Yadollahi, A., Shahraki, A. G. y Zaiane, O. R. (2017) «Current State of Text Sentiment Analysis from Opinion to Emotion Mining», *ACM Computing Surveys*, 50(2), pp. 1-33. doi: 10.1145/3057270.

Yager, R. E. y Penick, J. E. (1983) «Analysis of Current Problems in the US.pdf», *European Journal of Science Education*, 5(4), pp. 463-469.

Yang, Y. y Pedersen, J. O. (1997) «A comparative study on feature selection in text categorization», *Icml*, 97(412-420), p. 35.

Ye, S. y Wu, F. (2013) «Measuring message propagation and social influence on Twitter.com», *International Journal of Communication Networks and Distributed Systems*, 11(1), pp. 59-76. doi: 10.1504/IJCNDS.2013.054835.

Zaglia, M. E. (2013) «Brand communities embedded in social networks», *Journal of Business Research*. Elsevier Inc., 66(2), pp. 216-223. doi:

10.1016/j.jbusres.2012.07.015.

Zhang, Y., Mańdziuk, J., Quek, C. H. y Goh, B. W. (2017) «Curvature-based method for determining the number of clusters», *Information Sciences*, 415-416, pp. 414-428. doi: 10.1016/j.ins.2017.05.024.

Zhao, Q. (2012) *Cluster Validity in Clustering Methods*. Itä-Suomen yliopisto.

Zhong, S. (2005) «Efficient Online Spherical K-Means Clustering», *IEEE International Joint Conference on Neural Networks*, 5, pp. 3180-3185. doi: 10.1109/IJCNN.2005.1556436.

Zimbra, D., Abbasi, A., Zeng, D. y Chen, H. (2018) «The State-of-the-Art in Twitter Sentiment Analysis», *ACM Transactions on Management Information Systems*, 9(2), pp. 1-29. doi: 10.1145/3185045.

9 Anexos

Anexo 1. Artículo publicado en el *Chemistry Education Research and Practice*

Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Chem. Educ. Res. Pract.*, 2020, 21, 989

Describing the public perception of chemistry on twitter†

Manuel Guerris, * Jordi Cuadros, * Lucinio González-Sabaté and Vanessa Serrano

The public image of chemistry is a relevant issue for chemical stakeholders. It has been studied throughout history by means of document analysis and more recently through surveys. Twitter, a worldwide online social network, is based on spontaneous opinions. We tried to identify the public perception of chemistry on Twitter, what it explains, and which sentiments are perceived. We gathered 256 833 tweets between 1st January 2015 and 30th June 2015 containing the words "chemistry", "chemical" or "chem". We cleaned and filtered them down to 50 725 tweets with textual information in English and clustered them using spherical k-means. The resulting clusters were categorised according to six topics by 18 chemistry experts. The prevailing topics were the learning environment topic, related to activities and tasks in chemistry courses, and the human activity topic, referring to facts and news about the chemical industry. The scientific knowledge topic, concerning communication of chemistry knowledge, only accounted for a small percentage of the tweets. We classified the tweets of most relevant topics based on their sentiment values and obtained more positive than negative perceptions. Nevertheless, the analysis of the unigrams and bigrams word clouds revealed a significant presence of chemophobia-related terms in the human activity topic, both in positive and negative classified tweets. It also revealed specific elements of chemistry courses negatively perceived in the learning environment topic.

Received 11th December 2019,
Accepted 7th May 2020

DOI: 10.1039/c9rp00282k

rsc.li/cerp

Introduction

We are literally surrounded by chemistry. Its role in modern society cannot be underestimated. Chemistry and its applications affect and improve almost all aspects of people's lives. Human health, the environment, products and production processes are but a few areas in which chemistry plays a major part. The general public perceives its effects as positive or negative. Some people's perception of chemistry is so negative they may develop an irrational fear of chemical products (Duffus *et al.*, 2007) or "chemophobia", a term defined by the International Union of Pure and Applied Chemistry (IUPAC).

These perceptions are related to how chemistry communicates itself. Science communication helps citizenship to acquire the knowledge about science to participate actively and responsibly in, with and for society (Hazelkorn, 2015), plays a significant role in awakening vocations (Stekolschik *et al.*, 2010) and promoting scientific careers (Hayden *et al.*, 2011). A lack of vocation can cause an insufficient supply of graduates from upper-secondary and higher education to meet increasing demand across the EU, and a shortage of STEM (science, technology, engineering and mathematics) professionals (Cedefop, 2016). Additionally, the role of chemistry is not well understood by policymakers, funders and the chemistry community itself (Palermo, 2018). Therefore, the knowledge of chemistry's public image and its understanding in terms of its contents and sentiments perceived by the public matters to all chemistry stakeholders. It is critical to understand what is communicated and how it is perceived to define better policies to reduce this shortage.

There is scientific literature that investigates this public image at academic and social levels. At the academic level, studies reveal that the view perceived by students of chemistry (Yager and Penick, 1983; Furió Más, 2006) and of science (Schibeci, 1986) is generally negative because of a lack of clarity on its communication and a distorted perception of students (Nicolas, 2006; Penagos and Lozano, 2009; Chamizo, 2011; Lacolla *et al.*, 2013). Academic contents far from students motivation (Piñeros and Parga, 2014), the lack of historical and social perspectives in the curricula (Jiménez and Criado García-Legaz, 2005; Nicolas, 2006; Muñoz and Nardi, 2011; Linthorst, 2012) and the absence of science, technology and society relationships in the teaching of science (Solbes and Vilches, 1992; Ribelles *et al.*, 1995; Malaver *et al.*, 2004; Furió Más, 2006) are aspects that contribute to giving a

IQS Univ. Ramon Llull, Via Augusta 390, 08017 Barcelona, Spain.
E-mail: manuel.guerris@iqs.url.edu, jordi.cuadros@iqs.url.edu; Tel: +34 932672000

† Electronic supplementary information (ESI) available: Additional analysis results and figures. See DOI: 10.1039/c9rp00282k

poor image of chemistry far away from the real world impacting students in a negative way.

At the social level, the public image of chemistry seems to have inherited a negative perception due to its negative historical associations (Schummer and Spector, 2007; Schummer *et al.*, 2007) and a lack of efficient communication on behalf of chemists (Hartings and Fahy, 2011). For instance, sensationalist propaganda in global media used to associate the chemical industry and chemistry with pollution and environmental degradation (Trozzolo, 1975; Penagos and Lozano, 2009). Despite this negative perception, efforts to improve the image of chemistry have been implemented such as "Chemical for All" (Hill and Kumar, 2013), a global strategy to convince the public that chemistry provides health, comfort, and well-being.

It seems the public image of chemistry has always been negative, although some of the recent studies reviewed suggest a positive change. In 2004, IUPAC (Mahaffy *et al.*, 2008) found a negative public image of chemistry related to the misunderstanding of chemistry, chemists, chemicals and the chemical industry. In 2010, the European Chemical Industry Council (Hadhri, 2010) measured the public perception of the chemical industry in relation to several industries in the European Union. It suggested that chemistry had a favourable image approximately at the same level as it was in 2008 and has improved since the late 90s. The Royal Society of Chemistry analysed chemists' internal perceptions and the society's perception of chemists and chemical products in the UK (The Royal Society of Chemistry and TNS BMRB, 2015) and refuted the negative image. The results of the study showed a neutral or even positive image with 51% of the respondents being neutral and 19% happy, and 59% of the respondents answering that the benefits of chemistry were higher than its harmful effects. They mostly perceived chemistry as a solution to major global challenges such as oil dependence, food shortages, pollution and access to drinking water as well feeling a positive impact on well-being. Additionally, 21% of the general public associated chemistry to school or teachers with negative memories related to it but with mixed feelings about the chemistry that they learnt at school. 48% either agreed or were neutral that school had put them off chemistry, 45% disagreed that chemistry learnt at school had been useful in everyday life, and 52% agreed that they did not feel confident enough to talk about chemistry, with negative described perceptions of chemistry in comparison with science.

Several authors have proposed educational activities to improve this perception (Pratt and Yezierski, 2018; Ratamun and Osman, 2018; Molina and Carriazo, 2019; Tortorella *et al.*, 2019).

All the methods used to study the public image of chemistry were based on surveys and document analysis. They are not designed and are not able to capture spontaneous opinions. Social networks, on the other hand, collect ideas that are expressed spontaneously. Twitter with its 204 million monthly active users in the second quarter of 2015 (Clement, 2019), and 23% of total adult internet users (Duggan, 2015) demonstrates to be a relevant and significant online social network. Moreover, it is used by citizens to read news (Pew Research Center, 2019) and it is one of the leading social media platforms used by business-to-business (B2B) and business-to-consumer (B2C) marketers worldwide (Statista, 2019). On Twitter, users can communicate by exchanging short messages or tweets of up to 140 characters in real-time during the time span of this research, and can follow other users without any relationship between them. Public tweets can be freely gathered using Twitter Search API and analysed to get their sentiments (Sailunaz and Alhajj, 2019).

Tweets overcome survey challenges (Choi and Pak, 2005; Tourangeau and Yan, 2007; Krumpal, 2013) and document analysis challenges (Casadevall and Fang, 2009; Antilla, 2010) because tweets are part of conversations between Twitter users (Boyd *et al.*, 2010; Huang *et al.*, 2010; Smith *et al.*, 2014). These conversations give Twitter users the ability to express their thoughts, opinions (Kanavos *et al.*, 2014) and emotions (Tago and Jin, 2018) which are included in human social behaviour (Aarts *et al.*, 2012; Ye and Wu, 2013).

Therefore, Twitter seems an appropriate social network for this analysis because of its high number of users, the spontaneous contents written by its users, its use in different sectors and the ability to gather public users' messages and to analyse their sentiments. Consequently, it will complement existing literature about the public image of chemistry.

The objectives of this research are the following:

- Which topics related to chemistry can be found on Twitter?
- To what extent do Twitter messages portray positive and negative sentiments towards chemistry?
- What do users tweet about chemistry?

Experimental section

The methodology we used combines text mining with sentiment analysis techniques (Fig. 1) and is explained with specific details in the coming sections.

Just as an outline of the methods used, text mining enables deriving information extracted from written resources through computation (Gupta and Lehal, 2009) and includes techniques

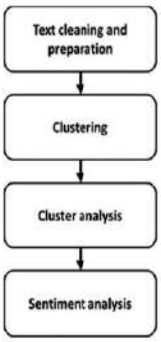


Fig. 1 Methodology used in the research.

and processes reported in the literature (Hearst, 1999; Hotho *et al.*, 2005; Berry, 2007; Feldman and Sanger, 2007; Delen and Crossland, 2008; Gupta and Lehal, 2009; Irfan *et al.*, 2015) to clean and obtain relevant information and to cluster those data according to topics. Clustering consists in grouping a set of objects based on their similarity and is useful with new or unlabelled objects (Jain *et al.*, 1999). There are different groups of clustering techniques (Fahad *et al.*, 2014), the partitioning-based methods such as the k-means algorithm (Jain, 2010) being the most popular and the most used.

We gathered twitter data during a period of time, cleaned and filtered them to eliminate non-relevant information and clustered them using a partitioning-based technique suitable to be applied to documents. Chemistry experts classified the clusters obtained into topics based on their contents. We did not opt for automatic classification methods which need text documents already classified into those topics (Hotho *et al.*, 2005) to classify new documents.

Sentiment analysis techniques (Yadollahi *et al.*, 2017) allow us to evaluate sentiments from terms, sentences, and documents. We classified the tweets related to the topics in the literature based on their sentiment value using a lexicon-based approach and analysed them by applying statistical and visual methods.

Text cleaning and preparation

To gather unbiased tweets about chemistry, we opted to limit our search to the terms "chemistry", "chemical" and "chem" admitting that, doing so, some other chemistry-related tweets could be missed. Public tweets containing these words were gathered between 1st January 2015 and 30th June 2015 using the twitteR package in R (Gentry, 2015).

Retweets, tweets written by a user and forwarded by another one, were discarded to avoid them possibly hiding in the clusters other tweets with a lower number and related to other contents. Expressions that did not add any meaning such as HTML tags, Unicode codes, Twitter account names, emails, URL addresses, non-letter symbols, and one and two-letter words were removed. Hashtags were kept because they could contain meaningful information.

English language tweets were selected with the textcat package in R (Hornik *et al.*, 2013). Stop words, commonly used words that do not add meaning to a document and provided by the tm package in R (Feinerer *et al.*, 2008), were removed. Based on previous tests, the terms "just", "now", "got", "will", "get", "much", "can" and "no", which did not contribute to the clustering process, as well as empty and duplicated tweets were also eliminated.

A bigram is a set of two consecutive words in a tweet. A bigram TDM (term-document matrix) is a matrix where each column corresponds to a tweet, a row to a bigram and each cell ij contains the number of times bigram i appears in tweet j. We built bigram TDM with cleaned tweets. We used bigrams instead of unigrams (single words) because there is no clear advantage in using unigrams in text categorization (Bekkerman and Allan, 2004) and bigrams are more accurate during cluster analysis. We did not want to lose information using n-grams, a sequence of n consecutive words from a given document being n an integer over three as their appearance in TDM could decrease.

We reduced the dimensionality of bigram TDM removing those bigrams with a low frequency to keep the most relevant ones and avoid those which might add noise to the information hampering the clustering process.

Clustering

Partitioning-based clustering techniques divide a set of objects into several partitions or clusters in such a manner that the objects in the same group are more similar to each other than to those in other clusters. We opted for a spherical k-means partitioning-based technique, a version of the k-means technique, because of its efficiency and effectiveness in text clustering (Dhillon and Modha, 2001; Zhong, 2005). We used an implementation of the skmeans package in R (Hornik *et al.*, 2012) to cluster the tweets automatically based on their similarity. Following Salton and Buckley (1988), we established a common baseline for the cosine similarity measure used in skmeans using term frequency-inverse document frequency (tf-idf).

We selected the number of clusters in skmeans based on two commonly used clustering validity indices, the elbow method (Madhulatha, 2012; Kodinariya and Makwana, 2013) and the silhouette method (Rousseeuw, 1987). Both methods are heuristic and used to determine the number of clusters visually. Then we quantitatively calculated it using the L-method algorithm (Salvador and Chan, 2004) and the curvature of a graph (Zhang *et al.*, 2017).

The different results obtained allowed us to choose a specific number of clusters trying to find a balance between the chemistry experts' capacity to manually classify the clusters and the closeness to the best solutions. With this number of clusters, we ran the skmeans clustering technique almost 10 000 times because of its stochastic behaviour, and we obtained the best solution. This optimal was the one in which the minimum value was calculated by the sum of the distance of each tweet to its respective prototype assigned to a cluster.

Cluster analysis

A word cloud is a graphical representation of the most important terms in a document where the size of a term is proportional to its frequency. We represented two word clouds per cluster, one with the 100 most frequent unigrams of tweets associated with the cluster and the other with the 100 most frequent bigrams. Our previous tests bigrams showed it is more useful for a cluster to be categorized by a chemistry expert. Still, we also decided to use unigrams word clouds to obtain more comprehensive information and to help the experts with their analysis.

The word clouds generated were visually interpreted by a group of chemistry experts. The topics obtained from our previous tests were the following:

- Human activity (HA): most terms are related to the presence of chemistry within the human activity such as production or the chemical industry.

- Scientific knowledge (SK): most terms are related to chemical concepts and abstract entities.
- Learning environment (LE): most terms are related to chemistry as a subject or course taught in class as well as student activities or exercises.
- Entertainment (E): most terms are related to cultural and media performances such as songs, musical groups, movies or TV series.
- Human relationships (HR): most terms are related to feelings between two or more people or emotions in general.
- Undefined (U): most terms either belong to several previous topics in which none of the terms predominate over others, or they belong to topics not defined in the list.

We defined a balanced incomplete block design (BIBD) (Fleiss, 1981) to assign clusters to be classified by chemistry experts. The experts were randomly divided into three groups where every expert was randomly assigned to one of the cluster groups defined in the BIBD. The order of the clusters analysed by every expert was randomized too. Each cluster was represented by its unigram and bigram word clouds.

We calculated the percentage of clusters and tweets assigned to each topic summing all the votes that a cluster received in each topic. The cluster was assigned to the topic with the highest number of votes. If several topics had the same number of votes, then the cluster was assigned to the U topic. Tweets belonging to the cluster inherited their cluster assignation.

We statistically analysed the results obtained using Fleiss' kappa (Fleiss, 1971). Fleiss' kappa and its significance level were calculated for every topic and the whole experiment (Fleiss *et al.*, 2003). The closer the value of kappa to one, the better the agreement. If the value was zero or below zero, the agreement was weaker than expected by chance. We used a common benchmark scale (Landis and Koch, 1977) to evaluate HA and LE Fleiss' kappa value results in addition to their statistical representativeness.

Sentiment analysis

We classified tweets from HA and LE topics, which were the most frequent ones, using a lexicon-based sentiment analysis method. A word sentiment is found using a lexicon and is measured by its polarity. A tweet polarity and thus its sentiment value is calculated adding its words polarities (Sun *et al.*, 2017). A polarity higher than zero means a positive sentiment, lower than zero a negative sentiment and equal to zero is considered neutral. We used bar charts to visualize the distribution of Twitter sentiment polarity.

The lexicon we used was based on SentiWordNet 3.0 lexicon (Baccianella *et al.*, 2010), which is commonly referenced in the literature (Medhat *et al.*, 2014; Mohey and Hussein, 2018; Sun *et al.*, 2017; Yadollahi *et al.*, 2017; Mäntylä *et al.*, 2018). This lexicon contains repeated words with different contexts and similar or different words which share polarity value because of similar contexts. We separated different words that shared a polarity value, assigned one polarity value per word and calculated an average polarity value for identical words. The result was a new list of single words with a single polarity value for each word.

A comparison word cloud is a graphical representation of terms from different documents represented in the same word cloud and differentiated by colour. The common terms are assigned to the document where the term has its maximum deviation calculated by its frequency in that document minus the average frequency in all the documents (Fellows, 2018). We built comparison word clouds of unigrams and bigrams with positive and negative tweets for the HA and LE topics and used them to interpret their main contents visually.

We selected representative samples of the tweets corresponding to some of the most frequent positive learning environment terms to understand their content better. Samples sizes were calculated using sample size for a proportion formula with a 0.95 confidence interval, 0.05 margin of error and 0.5 (worst case) expected sample proportion. We visually analysed the contents of the tweets samples and classified them into ironies or non-ironies and positive, negative and neutral sentiment.

Results

Text cleaning and preparation

We gathered a total of 256 833 tweets that ended up being 76 242 after text cleaning and preparation processes. Retweets and English language filtering were the operations that most affected them, reducing them to 89 663 (35% of the initial number of tweets).

We built the two bigram TDMs with their main characteristics described in Table 1. The selection of bigrams with a frequency over 29 caused the number to be reduced from 302 637 to 864 and the non-empty tweets from 76 242 to 50 725.

This reduction was due to the low frequency of most bigrams in tweets. At least 75% appeared once in all of them (Q3 = 1). We show an example of the processing of three tweets and part of the bigram TDM (Table 2).

Clustering

We calculated the tf-idf value instead of bigram frequencies in TDM. For each number of clusters between 2 and 285, we repeated the skmeans method 50 times and we selected the best solution from each one of the 50 repetitions. We calculated and graphically represented the clustering validity indexes of these best solutions for the elbow method and the silhouette method (Fig. 2A and B).

We observed that there was no clear elbow in the elbow method graph and no sharp change in the slope of the

Table 1 Main characteristics of bigrams TDMs

TDM	Tweets number	Bigrams number	Frequency				
			Min	Q1	Median	Q3	Max
Bigrams	76 242	302 637	1	1	1	1	3990
Bigrams with frequency over 29	50 725	864	30	36	48	81	3990

Table 2 Example of text cleaning for a tweet, its bigrams and bigram TDM

Original tweets	Tweet 1: "im dreading going back to college... especially bc that monday i have an 8am chem lab"			
	Tweet 2: "baking cookies for chem lab and accidentally used self rising flour OMG"			
	Tweet 3: "2 kumbe 1 chem journal... Copying time"			
Cleaned tweets	Tweet 1: "dreading going back college especially monday chem lab"			
	Tweet 2: "baking cookies chem lab accidentally used self rising flour omg"			
	Tweet 3: "kumbe chem journal copying time"			
Bigrams obtained from cleaned Tweet 1	"dreading going", "going back", "back college", "college especially", "especially monday", "monday chem", "chem lab"			
Bigram term document matrix		Tweet 1	Tweet 2	Tweet 3
	"college especially"	1	0	0
	"especially monday"	1	0	0
	"monday chem"	1	0	0
	"chem lab"	1	1	0

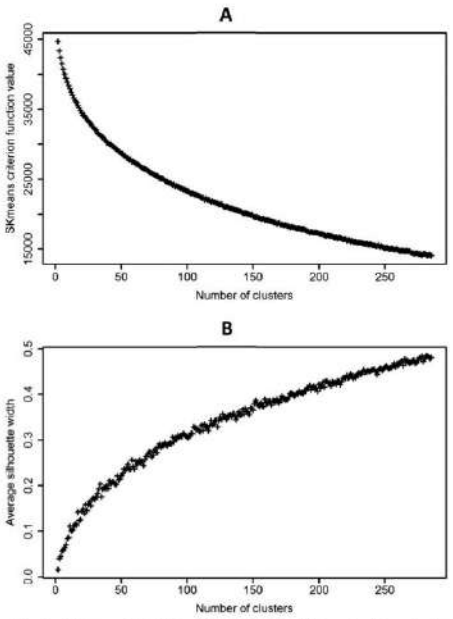


Fig. 2 Graphs to calculate the different methods for calculating the best number of clusters (A) elbow method graph. Skmeans criterion function value *vs.* number of clusters (B) silhouette method graph. Average silhouette width *vs.* number of clusters.

silhouette method graph, which did not let us visually determine the best number of clusters. We used the L-method and the curvature method to select the number of clusters numerically. The cluster corresponding to the minimum value in the L-method and the one corresponding to the maximum in the curvature method were considered to be the best ones. These results are included in ESI 1 (ESI†).

The elbow graph suggested using 78 and 193 clusters and the silhouette graph, 78 and 98. As there was no clear and exact solution for the number of clusters, we decided to round it off to 100 clusters. It is large enough to minimize mixed topic clusters and small enough to be classified by a chemistry expert.

We tried to run the skmeans implementation technique for 100 clusters 10 000 times, but we could only get 9723 due to technical issues. We selected the best solution with the best skmeans criterion value. In that solution, the minimum tweets per cluster were 95, the maximum 3476, the median 383 and Q1 and Q3 were 251 and 647 respectively.

Cluster analysis

Eighteen chemistry experts classified and assigned each cluster to a single topic. These experts were chemistry professors, all of them with a chemistry or chemical engineering university degree and holding a PhD in chemistry. They were selected based on their educational background, having more than ten years of professional experience and on the diversity of their expertise. To validate the classification, we designed a BIBD composed by six chemistry experts, 50 clusters per expert, six experts per cluster and 100 clusters.

With a total of 18 different experts, we replicated this design in each of the three groups with six experts per group, so each cluster was categorized by nine different experts. An example of one cluster with its graphical representation is shown in Fig. 3. The representations of all the clusters are included in ESI 2 (ESI†). These representations were used to classify the clusters obtaining the results shown in Table 3. The table with the specific cluster number assigned to each expert and the detailed classification results per expert are included in ESI 3 (ESI†).

It was possible to classify most clusters and tweets and only 14% of the clusters and 18% of the tweets were considered as undefined. LE and HA were the topics that obtained the largest numbers of clusters and tweets. LE represented 45% of the clusters and 39% of the tweets and HA 20% of the clusters and 18% of the tweets classified, whereas SK and E obtained the lowest ones. The SK topic, which concerns spreading scientific knowledge, only attained 6% of the clusters and 5% of the tweets classified.

We calculated Fleiss' kappa for each topic and the whole experiment. These values were statistically representative because of their very low p-values (less than 1×10^{-6}). We compared the Fleiss' kappa values for HA and LE (0.388 and 0.517 respectively) with Landis and Koch (1977) Kappa's benchmark scale. The obtained values show fair and moderate inter-rater reliabilities respectively.

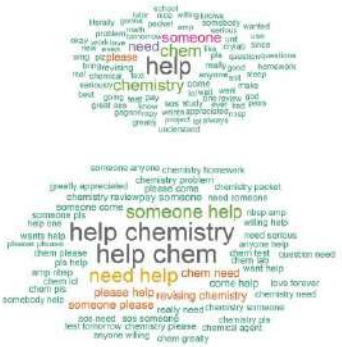


Fig. 3 Unigram and bigram word clouds of cluster number 8. This cluster was classified by experts 1, 2, 6, 8, 10, 11, 16, 17 and 18. Expert 4 classified it as human activity (HA) topic, experts 1, 8, 18 as human relationship (HR) and the rest as learning environment (LE). Based on the number of votes this cluster was classified in the LE topic.

Table 3 Experts classification results

	Percentage classified	
Topic	Clusters (%)	Tweets (%)
Human activity (HA)	20	18
Scientific knowledge (SK)	6	5
Learning environment (LE)	45	39
Entertainment (E)	5	13
Human relationship (HR)	10	7
Undefined (U)	14	18

Sentiment analysis

As it is explained in sentiment analysis of the experimental section, we modified the 117 659 polarities of SentiWordNet 3.0 to obtain a list of 146 842 and associated each one to a single word. We classified the HA and LE tweets into positive, neutral and negative tweets based on their polarity (Fig. 4). A higher percentage of positive than negative tweets was obtained. 72% and 75% of the total number of words from HA and LE tweets, respectively, were detected by the lexicon. The highest polarity tweets in the HA topic included "expert killed", "chemical free", "surveys marketresearchreports" and "forecasts marketing" bigrams. Their respective polarity classifications, as an example, are shown in Table 4.

It is worth noting how, for instance, the terms "free" and "acid" were valued as negative. Their polarity values depend on the different contexts provided by the lexicon, the polarity value of each context and the average polarity value calculated with all polarity values of a term. As an example, the different contexts of "acid" in the lexicon were "water-soluble compounds being able to damage water", "having the characteristics of an acid", "an acid reaction" and "being sour to the taste". Each context had a polarity value which was used to calculate the average polarity value of "acid".

Finally, we compared the positive and negative tweets of the HA and LE categories through comparative unigram and bigram word clouds (Fig. 5).

The visual analysis of HA comparison word clouds suggests that tweets classified as negative containing terms such as "attack", "syria", "chemical attack","syrian opposition" and "chemical warfare" and "toxic", "toxic chemical", "chemical fire" and "chemical leak" predominate over other negative tweet contents capable of fuelling chemophobia attitudes. The existence of terms such as "chemical free" and "used chemical" in tweets classified as positive might reinforce this effect. The term "chemical", seemingly understood as industry products, is considered mostly in negative tweets whereas "chemistry", apparently related as a physical science, appears in positive tweets.

The visual analysis of LE comparison word clouds indicates the difficulty of chemistry as a subject. Terms in tweets classified as negative related to academic activities such as "final", "exam", "chem final", "final tomorrow", "quiz tomorrow", "lab", "lecture", "chem lab", and "chem lecture" and terms related to feelings such as "hate", "hard", "crying", "need help", "never understand" and "chemistry hard" are predominant. This difficulty of learning chemistry is reinforced by the presence of terms in positive tweets such as "someone help" and "help chemistry", which can be due to several factors described in the literature.

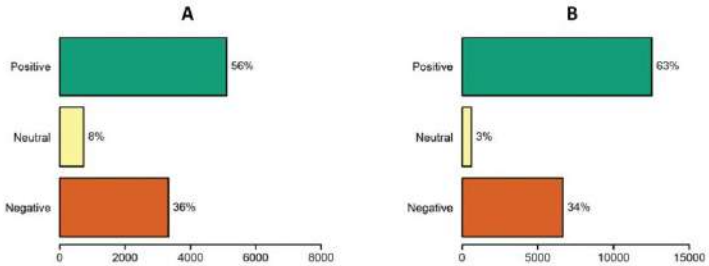


Fig. 4 Tweets classified by their polarity (A) human activity (HA) tweets (total number = 9159) (B) learning environment (LE) tweets (total number = 19 804).

Table 4 Examples of tweets' polarity classification. Positive words are coloured in green, negative words in red, neutral ones in orange and unfound words in black

Term	Clean tweet with words' polarity	Tweet Polarity
expert killed	chemical weapons expert killed coalition airstrike battle really begins know	Positive
chemical free	awesome sustainable design pollution busting billboard grows chemical free organic vegetables	Positive
surveys marketresearchreports	chemical phosphoric acid mono alkyl esters surveys marketresearchreports forecast mrx	Negative
forecasts marketing	chemical acid mordant brown surveys marketresearchreports forecasts marketing mrx market research	Negative

Despite these negative feelings about chemistry, terms such as "test", "chem test", "chemistry test", "test tomorrow", "teacher", "chem teacher" and "chemistry teacher" also appear in tweets classified as positive. We created two statistically representative samples of the tweets corresponding to the terms "test" and "teacher" to analyse them and understand their content. These terms included the rest of the most frequent positive terms. Sample sizes were 344 and 292 tweets from 3203 and 1208 corresponding to "test" and "teacher" terms respectively. We randomly selected these tweets and analysed and classified their contents as it was described in the experimental section part. Results obtained are included in ESI 4 (ESI†).

We found that many tweets in both samples seem to transmit either a neutral or a negative sentiment. There were also many ironies within negative classified tweets such as "First organic chemistry test tonight... Will someone start digging my grave now?" or "My Chem teacher looks like she could be a character on Phineas and Ferb". Despite been classified as positive by the sentiment lexicon used, these terms seem to reinforce the negative ones and thus increasing negative sentiment.

Discussion

Our findings based on the analysed data suggest that Twitter provides another, much broader side of the public image of chemistry which has not been studied so far. It is built upon spontaneous opinions in contrast to surveys and document analysis, which are limited to a few topics.

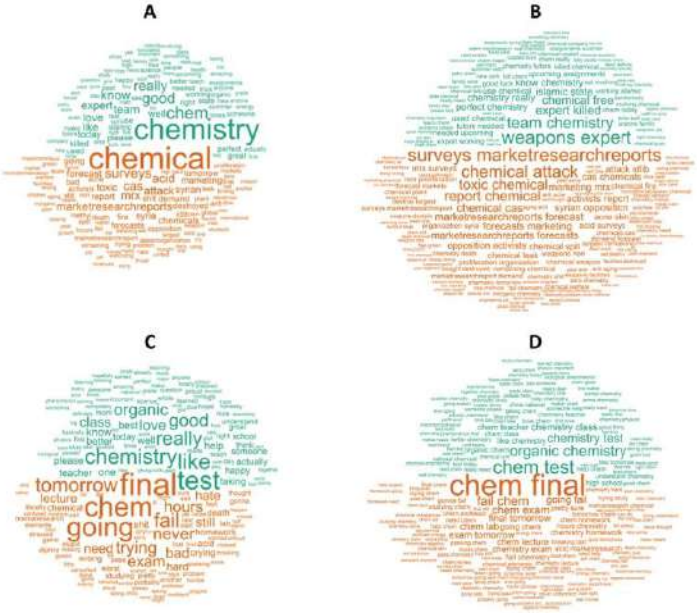


Fig. 5 Comparative word clouds of positive (green) and negative (red) tweets (A) human activity (HA) unigrams (B) human activity (HA) bigrams (C) learning environment (LE) unigrams (D) learning environment (LE) bigrams.

Chemistry-related public tweets containing the words "chemistry", "chemical" or "chem" reveal a strong presence of learning environment (LE) and human activities (HA) topics, but a limited one as far as transmitting chemistry knowledge is concerned.

The sentiment analysis results of HA tweets with a higher percentage of positive tweets than negative ones seem to follow the trend of the most recent studies (Hadhri, 2010; The Royal Society of Chemistry and TNS BMRB, 2015). The existence of both positive and negative tweets also seems to be in line with the contraposition between the positive and negative effects of chemistry described in The Royal Society of Chemistry and TNS BMRB (2015). There, 59% of the respondents answered that the benefits of chemistry were higher than its harmful effects and 51% perceived a neutral feeling. This contraposition seems to be highlighted by the terms "chemical", understood as industry products, and "chemistry", considered as a physical science, with a negative and positive connotations respectively. These results are in common with The Royal Society of Chemistry and TNS BMRB (2015). Chemophobia attitudes seem to be suggested by terms that appear in both negative and positive tweets related to HA. Many of these terms seem to be related to chemical war, chemical toxicity and chemical disasters. At the same time, these terms might create or reinforce chemophobia perceptions on Twitter users.

The sentiment analysis results of LE tweets also results in a higher percentage of positive tweets than negative ones. LE positive tweets, however, should be analysed deeper to review their sentiment values because of the context of the terms used in the sentiment lexicon. The contraposition between the positive and negative effects of chemistry is also present in this topic.

In LE, chemistry image seems to be based on specific elements of chemistry education such as evaluation methods and teachers rather than chemistry communication topics in academia and their influence (Nicolas, 2006; Penagos and Lozano, 2009; Chamizo, 2011; Lacolla *et al.*, 2013) and curricula contents (Jiménez and Criado García-Legaz, 2005; Nicolas, 2006; Muñoz and Nardi, 2011; Linthorst, 2012; Piñeros and Parga, 2014) being a new contribution on this topic. Similar to our research, this image has so far always been considered negative in the view perceived by students of chemistry (Yager and Penick, 1983; Furió Más, 2006).

Additionally, LE messages seem to contain words related to classroom elements such as lectures and exams, perceived negatively as well as expressing the difficulty of learning. These negative feelings about chemistry learning might favour the association between chemistry and negative memories, in agreement with the conclusions of The Royal Society of Chemistry and TNS BMRB (2015) study.

Conclusions

The collection of 256 833 tweets containing the words "chemistry", "chemical" or "chem" has allowed us to explore and analyse the public perception of chemistry on Twitter.

Text cleaning and preparation techniques reduced to 50 725 useful tweets. These were classified into six different topics. The two more frequent topics were activities and tasks related to chemistry courses (the learning environment topic, 39%) and facts and news related to the chemical industry and industry products (the human activity topic, 18%). Only a small percentage of tweets related to the transmission and communication of chemistry knowledge (the scientific knowledge topic, 5%) was found. The remaining tweets are either unclassified or belonging to categories less relevant to chemistry and chemistry education.

Sentiment analysis techniques helped us to observe many terms in the human activity topic suggesting chemophobia, whereas chemistry is perceived as difficult by Twitter users in the learning environment topic. These terms seem to relate to war, toxicity and disasters in the human activity topic. In the learning environment, most frequent terms seem to relate to classroom activities and students' sentiments about the chemistry subject.

These two topics contained both positive and negative sentiments aligned with the latest accepted vision of the public image of chemistry with chemophobia still present in the human activity topic. This observation and the negative feelings found in the learning environment topic suggest that there is still room for improvement in current practices in chemistry education, both in the formal and informal settings. These improvements may lead to better scientific communication and knowledge, enhancing and improving citizenship participation in science.

Limitations and further work

The main limitations of this study are the number of tweets, the text cleaning and preparation methods used, and the sentiment classification method. Contents and topics could differ depending on the period of time during which the tweets were gathered. Enhancing the time span and thus the number of tweets would provide more generalizable results while monitoring the evolution of positive and negative attitudes. At the same time, we could analyse special chemistry events and how attitudes differ between and after the events. Advanced natural language techniques to stem words, to interpret abbreviations, emojis and emoticons, once they become more accessible, could affect the number of different bigrams and might result in a better classification. The lexicon-based approach used in this research did not take into account the context of a word and was dependent on the lexicon words. A tweet's polarity, therefore, depended on the polarity value assigned to each word and the words found in the lexicon. This approach is not able, for instance, to evaluate ironies properly. Additionally, a small number of tweets might undergo a change in their sense because of some stop words removal, such as negative adverbs. The construction of a lexicon specifically focused on chemistry, combining several lexicons and the use of advanced sentiment techniques to evaluate words in their context might also help to obtain better generalizable results.

Further studies should focus on improving the main limitations described above as well as on monitoring the evolution of the public perception of chemistry on Twitter during a longer period. Thus, scientists and practitioners could obtain a wider view of this perception on Twitter as well as to be able to detect new topics and associated contents. This new knowledge will be helpful to chemistry stakeholders for improving the public image of chemistry.

Statement of ethics

All research has been conducted according to the ethics research guidelines in place at Univ. Ramon Llull.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

We acknowledge the help provided by all the chemistry experts who participated in this research.

References

Aarts O., Van Maanen P. P., Ouboter T. and Schraagen J. M., (2012), Online social behavior in twitter: a literature review, Proceedings – 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012, pp. 739–746, DOI: 10.1109/ICDMW.2012.139.

Antilla L., (2010), Self-censorship and science: a geographical review of media coverage of climate tipping points, *Public Understand. Sci.*, **19**(2), 240–256, DOI: 10.1177/0963662508094099.

Baccianella S., Esuli A. and Sebastiani F., (2010), SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining SentiWordNet, *Analysis*, 1–12, DOI: 10.1.1.61.7217.

Bekkerman R. and Allan J., (2004), Using Bigrams in Text Categorization, Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst, pp. 1–10.

Berry M. W., (2007), Survey of text mining: clustering, *Classification, and Retrieval*, ed. M. W. Berry and M. Castellanos, Springer, 2nd edn, DOI: 10.1007/978-1-84800-046-9.

Boyd D., Golder S. and Lotan G., (2010), Tweet, tweet, retweet: conversational aspects of retweeting on twitter', Proceedings of the Annual Hawaii International Conference on System Sciences, DOI: 10.1109/HICSS.2010.412.

Casadevall A. and Fang F. C., (2009), Is peer review censorship?, *Infect. Immun.*, **77**(4), 1273–1274, DOI: 10.1128/IAI.00018-09.

Cedefop, (2016), *Skill shortage and surplus occupations in Europe*, pp. 1–4, DOI: 10.2801/05116.

Chamizo J. A., (2011), La imagen pública de la química, *Educ. Quím.*, **22**(4), 320–331.

Choi B. C. K. and Pak A. W. P., (2005), A catalog of biases in questionnaires, *Prev. Chronic Dis.*, **2**(1), 1–13.

Clement J., (2019), Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019, Statista. Available at: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/, accessed: 28 January 2020.

Delen D. and Crossland M. D., (2008), Seeding the survey and analysis of research literature with text mining, *Expert Syst. Appl.*, **34**(3), 1707–1720, DOI: 10.1016/j.eswa.2007.01.035.

Dhillon I. S. and Modha D. S., (2001), Concept decompositions for large sparse text data using clustering, *Mach. Learn.*, **42**(1–2), 143–175, DOI: 10.1023/A:1007612920971.

Duffus J. H., Nordberg M. and Templeton D. M., (2007), Glossary of terms used in toxicology, 2nd edition (IUPAC Recommendations 2007)', *Pure Appl. Chem.*, **79**(7), 1153–1344, DOI: 10.1351/pac200779071153.

Duggan M., (2015), The Demographics of Social Media Users, Pew Research Center. Available at: https://www.pewresearch.org/internet/2015/08/19/the-demographics-of-social-media-users/, accessed: 28 January 2020.

Fahad A., Alshatri N., Tari Z., Alamri A., Khalil I., Zomaya A. Y., Foufou S. and Bouras A., (2014), A survey of clustering algorithms for big data: Taxonomy and empirical analysis, *IEEE Trans. Emerg. Top. Comput.*, **2**(3), 267–279, DOI: 10.1109/TETC.2014.2330519.

Feinerer I., Hornik K. and Meyer D., (2008), Text Mining Infrastructure in R, *J. Stat. Softw.*, **25**(5), 1–54, DOI: citeulike-article-id:2842334.

Feldman R. and Sanger J., (2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, DOI: 10.1017/CBO9780511546914.

Fellows I., (2018), wordcloud: Word Clouds', R package version 2.6, https://CRAN.R-project.org/package=wordcloud.

Fleiss J. L., (1971), Measuring nominal scale agreement among many raters, *Psychol. Bull.*, **76**(5), 378–382, DOI: 10.1037/h0031619.

Fleiss J. L., (1981), Balanced Incomplete Block Designs for Inter-Rater Reliability Studies, *Appl. Psychol. Meas.*, **5**(1), 105–112, DOI: 10.1177/014662168100500115.

Fleiss J. L., Levin B. and Paik M. C., (2003), *Statistical Methods for Rates and Proportions*, 3rd edn, Hoboken: John Wiley & Sons, Inc.

Furió Más C., (2006), La motivación de los estudiantes y la enseñanza de la Química. Una cuestión controvertida, *Educ. Quím.*, **17**(IV Jornadas Internacionales), 222–227.

Gentry J., (2015), twitteR: R Based Twitter Client, R package version 1.1.9, https://CRAN.R-project.org/package=twitteR.

Gupta V. and Lehal G. S., (2009), A survey of text mining techniques and applications, *J. Emerg. Technol. Web Intell.*, **1**(1), 60–76, DOI: 10.4304/jetwi.1.1.60-76.

Hadhri M., (2010), CEFIC Facts and Figures 2010. The European Chemical Industry in a worldwide perspective. CEFIC. Available at: https://es.scribd.com/document/44470516/Facts-and-Figures-2010-Report.

Hartings M. R. and Fahy D., (2011), Communicating chemistry for public engagement, *Nat. Chem.*, **3**(9), 674–7, DOI: 10.1038/nchem.1094.

Hayden K., Ouyang Y., Scinski L., Olszewski B. and Bielefeldt T., (2011) Increasing Student Interest and Attitudes in STEM: Professional Development and Activities to Engage and Inspire Learners, *Contemp. Issues Technol. Sci. Teach. Educ.*, **11**(1), 47–69.

Hazelkorn E., (2015), *Science education for responsible citizenship: report to the European Commission of the Expert Group on Science Education*, Publications Office of the European Union, p. 88, DOI: 10.2777/12626.

Hearst M. A., (1999), Untangling text data mining, Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, pp. 3–10, DOI: 10.3115/1034678.1034679.

Hill J. and Kumar D. D., (2013), Challenges for Chemical Education: Implementing the 'Chemistry for All' Vision, *J. Am. Inst. Chem.*, **86**(2), 27–32.

Hornik K., Feinerer I., Kober M. and Buchta C., (2012), Spherical k-Means Clustering, *J. Stat. Softw.*, **50**(10), 1–22.

Hornik K., Mair P., Rauch J., Geiger W., Buchta C. and Feinerer I., (2013), The textcat Package for n-Gram Based Text Categorization in R, *J. Stat. Softw.*, **52**(6), 1–17, DOI: 10.18637/jss.v052.i06.

Hotho A., Nürnberger A. and Paaß G., (2005), A Brief Survey of Text Mining, *J. Comput. Linguis. Lang. Technol.*, **20**, 19–62, DOI: 10.1111/j.1365-2621.1978.tb09773.x.

Huang J., Thornton K. M. and Efthimiadis E. N., (2010), Conversational tagging in Twitter, HT'10 – Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, pp. 173–177, DOI: 10.1145/1810617.1810647.

Irfan R., King C. K., Grages D., Ewen S., Khan S. U., Madani S. A., Kolodziej J., Wang L., Chen D., Rayes A., Tziritas N., Xu C. Z., Zomaya A. Y., Alzahrani A. S. and Li H., (2015), A survey on text mining in social networks, *Knowl. Eng. Rev.*, **30**(2), 157–170, DOI: 10.1017/S0269888914000277.

Jain A. K., (2010), Data clustering: 50 years beyond K-means, *Pattern Recogn. Lett.*, **31**(8), 651–666, DOI: 10.1016/j.patrec.2009.09.011.

Jain A. K., Murty M. N. and Flynn P. J., (1999), Data clustering: a review, *ACM Comput. Surv.*, **31**(3), 264–323, DOI: 10.1145/331499.331504.

Jiménez J. B. and Criado García-Legaz A., (2005), Análisis de las actividades sobre la historia de la química en los libros de física y química del segundo ciclo de la eso, *Enseñanza De Las Ciencias*, **2000**, 1–6.

Kanavos A., Perikos I., Vikatos P., Hatzilygeroudis I., Makris C. and Tsakalidis A., (2014), Conversation Emotional Modeling in Social Networks, Proceedings – International Conference on Tools with Artificial Intelligence, ICTAI, 2014-Decem, pp. 478–484, DOI: 10.1109/ICTAI.2014.78.

Kodinariya T. M. and Makwana P. R., (2013), Review on determining number of Cluster in K-Means Clustering, *Int. J. Adv. Res. Comput. Sci. Manage. Stud.*, **1**(6), 90–95.

Krumpal I., (2013), Determinants of social desirability bias in sensitive surveys: a literature review, *Qual. Quan.*, **47**(4), 2025–2047, DOI: 10.1007/s11135-011-9640-9.

Lacolla L., Meneses Villagrá J. A. and Valeiras N., (2013) Las representaciones sociales y las reacciones químicas: Desde las explosiones hasta Fukushima, *Educ. Quim.*, **24**(3), 309–315.

Landis J. R. and Koch G. G., (1977), The Measurement of Observer Agreement for Categorical Data, *Biometrics*, **33**(1), 159–174, DOI: 10.2307/2529310.

Linthorst J. A., (2012), The image of chemistry and curriculum changes, *Educ. Quim.*, **23**(2), 240–242, DOI: 10.1016/S0187-893X(17)30115-5.

Madhulatha T. S., (2012), An overview on clustering methods, *IOSR J. Eng.*, **02**(04), 719–725, DOI: 10.9790/3021-0204719725.

Mahaffy P., Ashmore A., Bucat B., Do C. and Rosborough M., (2008), Chemists and "the public": IUPAC's role in achieving mutual understanding (IUPAC Technical Report), *Pure Appl. Chem.*, **80**(1), 161–174, DOI: 10.1351/pac200880010161.

Malaver M., Pujol R. and D'Alessandro Martínez A., (2004), Los Estilos De Prosa Y El Enfoque Ciencia-Tecnología-Sociedad En Textos Universitarios De Química General, *Educ. Quim.*, **22**(3), 441–453.

Mäntylä M. V., Graziotin D. and Kuutila M., (2018), The evolution of sentiment analysis—A review of research topics, venues, and top cited papers, *Comput. Sci. Rev.*, **27**, 16–32, DOI: 10.1016/j.cosrev.2017.10.002.

Medhat W., Hassan A. and Korashy H., (2014), Sentiment analysis algorithms and applications: A survey, *Ain Shams Eng. J.*, **5**(4), 1093–1113, DOI: 10.1016/j.asej.2014.04.011.

Mohey D. and Hussein E. M., (2018), A survey on sentiment analysis challenges, *J. King Saud Univ. Sci.*, **30**(4), 330–338, DOI: 10.1016/j.jksues.2016.04.002.

Molina M. F. and Carriazo J. G., (2019), Awakening Interest in Science and Improving Attitudes toward Chemistry by Hosting an ACS Chemistry FeSTiVAl in Bogotá, Colombia, *J. Chem. Educ.*, **96**(5), 944–950, DOI: 10.1021/acs.jchemed.8b00670.

Muñoz L. and Nardi R., (2011), Las representaciones científicas en la formación inicial de profesores de química, *Encontro Nacional de Pesquisa em Educação em Ciências*, 8.

Nicolas E., (2006), Aula y Laboratorio de Química La Química vista por 840 estudiantes de bachillerato, *Anal. Quim.*, **102**(4), 64–67.

Palermo A., (2018), The future of the Chemical Sciences. Preparing for an Uncertain Future, *Chem. World*, **6**, DOI: 10.1021/ed020p304.

Penagos W. M. M. and Lozano D. L. P., (2009), *La imagen pública de la química y su relación con la generación de actitudes hacia la química y su aprendizaje*, Tecné, Episteme y Didaxis: TED, vol. 27, pp. 67–93.

Pew Research Center, (2019), News Use Across Social Media Platforms 2018, available at: http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/, accessed: 19 February 2019.

Piñeros Y. and Parga D., (2014), *Caracterización de los contenidos curriculares contextualizados para la enseñanza de la química*, Revista Tecné, Episteme y Didaxis: TED.

Pratt J. M. and Yezierski E. J., (2018), A novel qualitative method to improve access, elicitation, and sample diversification for enhanced transferability applied to studying chemistry outreach, *Chem. Educ. Res. Pract.*, **19**(2), 410–430, DOI: 10.1039/c7rp00200a.

Ratamun M. M. and Osman K., (2018), The Effectiveness Comparison of Virtual Laboratory and Physical Laboratory in Nurturing Students' Attitude towards Chemistry, *Creat. Educ.*, **9**(9), 1411–1425, DOI: 10.4236/ce.2018.99105.

Ribelles R., Solbes J. and Vilches A., (1995), Las interacciones C.T.S. en la enseñanza de las ciencias, Análisis comparativo de la situación para la Física y Química y la Biología y Geología, *Comunicación, Lenguaje y Educación*, pp. 135–143.

Rousseeuw P. J., (1987), Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, **20**(C), 53–65, DOI: 10.1016/0377-0427(87)90125-7.

Sailunaz K. and Alhajj R., (2019), Emotion and sentiment analysis from Twitter text, *J. Comput. Sci.*, **36**, 101003, DOI: 10.1016/j.jocs.2019.05.009.

Salton G. and Buckley C., (1988), Term-weighting approaches in automatic text retrieval, *Inform. Process. Manage.*, **24**(5), 513–523.

Salvador S. and Chan P., (2004), Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms', 16th IEEE international conference on tools with artificial intelligence, pp. 576–584, DOI: 10.1109/ICTAI.2004.50.

Schibeci R. A., (1986), Images of science and scientists and science education, *Sci. Educ.*, **70**(2), 139–149, DOI: 10.1002/sce.3730700208.

Schummer J., Bensaude-Vincent B. and Van Tiggelen B., (2007), *The Public Image of Chemistry*, World Scientific Publishing, DOI: 10.1142/9789812775856.

Schummer J. and Spector T. I., (2007), The visual image of chemistry: Perspectives from the history of art and science, *Int. J. Philos. Chem.*, **13**(1), 1–40.

Smith M. A., Rainie L., Shneiderman B. and Himelboim I., (2014), *Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters*, Pew Research Center, vol. 20, pp. 1–56.

Solbes J. and Vilches A., (1992), El modelo constructivista y las relaciones ciencia/técnica/sociedad, *Enseñanza de las Ciencias*, **10**(2), 181–186.

Statista, (2019) Leading social media platforms used by B2B and B2C marketers worldwide as of January 2018, Available at: https://www.statista.com/statistics/259382/social-media-platforms-used-by-b2b-and-b2c-marketers-worldwide/, accessed: 19 February 2019.

Stekolschik G., Draghi C., Adaszko D. and Gallardo S., (2010), Does the public communication of science influence scientific vocation? results of a national survey, *Public Underst. Sci.*, **19**(5), 625–637, DOI: 10.1177/0963662509335458.

Sun S., Luo C. and Chen J., (2017), A review of natural language processing techniques for opinion mining systems, *Inform. Fusion*, **36**, 10–25, DOI: 10.1016/j.inffus.2016.10.004.

Tago K. and Jin Q., (2018), Influence analysis of emotional behaviors and user relationships based on Twitter data, *Tsinghua Sci. Technol.*, **23**(1), 104–113, DOI: 10.26599/TST.2018.9010012.

The Royal Society of Chemistry and TNS BMRB, (2015), Public attitudes to chemistry', Research report, https://www.rsc.org/campaigning-outreach/campaigning/public-attitudes-chemistry/, pp. 1–78.

Tortorella S., Zanelli A. and Domenici V., (2019), Chemistry Beyond the Book: Open Learning and Activities in Non-Formal Environments to Inspire Passion and Curiosity, *Substantia*, **3**, 39–47, DOI: 10.13128/Substantia-587.

Tourangeau R. and Yan T., (2007), Sensitive Questions in Surveys, *Psychol. Bull.*, **133**(5), 859–883, DOI: 10.1037/0033-2909.133.5.859.

Trozzolo A. M., (1975), The image of chemistry. Conference, https://www3.nd.edu/~atrozzol/Image-2.pdf, pp. 1–7.

Yadollahi A., Shahraki A. G. and Zaiane O. R., (2017), Current State of Text Sentiment Analysis from Opinion to Emotion Mining, *ACM Comput. Surv.*, **50**(2), 1–33, DOI: 10.1145/3057270.

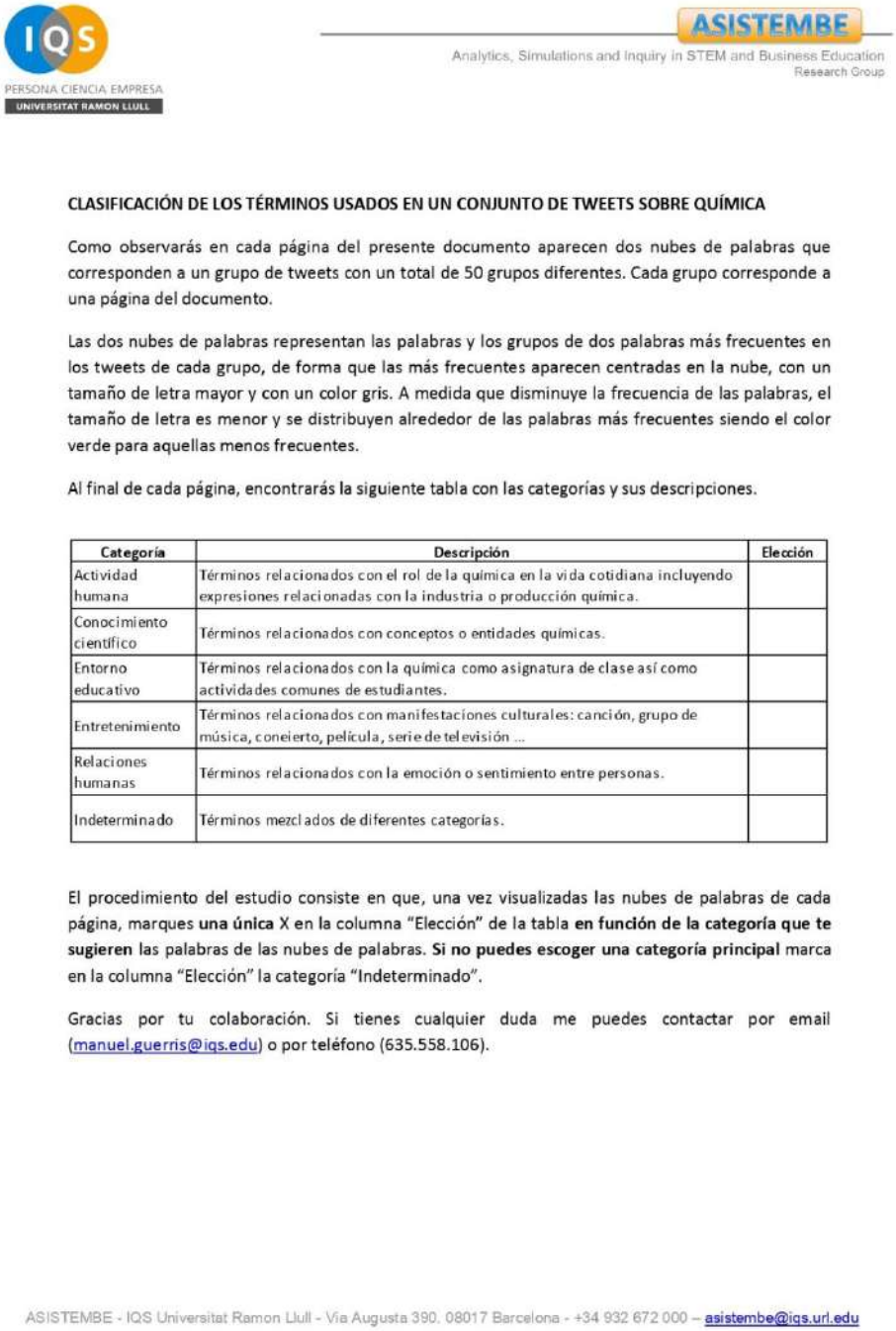
Yager R. E. and Penick J. E., (1983), Analysis of Current Problems in the US.pdf, *Eur. J. Sci. Educ.*, **5**(4), 463–469.

Ye S. and Wu F., (2013), Measuring message propagation and social influence on Twitter.com, *Int. J. Commun. Netw. Distri. Syst.*, **11**(1), 59–76, DOI: 10.1504/IJCNDS.2013.054835.

Zhang Y., Mańdziuk J., Quek C. H. and Goh B. W., (2017), Curvature-based method for determining the number of clusters, *Inform. Sci.*, **415–416**, 414–428, DOI: 10.1016/j.ins.2017.05.024.

Zhong S., (2005), Efficient Online Spherical K-Means Clustering, *IEEE Int. Joint Conf. Neural Netw.*, **5**, 3180–3185, DOI: 10.1109/IJCNN.2005.1556436.

Anexo 2. Ejemplo de extracto de documento entregado a los expertos de química



PERSONA CIENCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group

CLASIFICACIÓN DE LOS TÉRMINOS USADOS EN UN CONJUNTO DE TWEETS SOBRE QUÍMICA

Como observarás en cada página del presente documento aparecen dos nubes de palabras que corresponden a un grupo de tweets con un total de 50 grupos diferentes. Cada grupo corresponde a una página del documento.

Las dos nubes de palabras representan las palabras y los grupos de dos palabras más frecuentes en los tweets de cada grupo, de forma que las más frecuentes aparecen centradas en la nube, con un tamaño de letra mayor y con un color gris. A medida que disminuye la frecuencia de las palabras, el tamaño de letra es menor y se distribuyen alrededor de las palabras más frecuentes siendo el color verde para aquellas menos frecuentes.

Al final de cada página, encontrarás la siguiente tabla con las categorías y sus descripciones.

Categoría	Descripción	Elección
Actividad humana	Términos relacionados con el rol de la química en la vida cotidiana incluyendo expresiones relacionadas con la industria o producción química.	
Conocimiento científico	Términos relacionados con conceptos o entidades químicas.	
Entorno educativo	Términos relacionados con la química como asignatura de clase así como actividades comunes de estudiantes.	
Entretenimiento	Términos relacionados con manifestaciones culturales: canción, grupo de música, concierto, película, serie de televisión ...	
Relaciones humanas	Términos relacionados con la emoción o sentimiento entre personas.	
Indeterminado	Términos mezclados de diferentes categorías.	

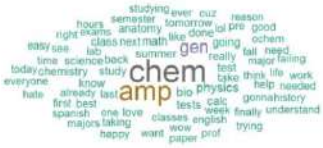
El procedimiento del estudio consiste en que, una vez visualizadas las nubes de palabras de cada página, marques **una única** X en la columna "Elección" de la tabla **en función de la categoría que te sugieren** las palabras de las nubes de palabras. **Si no puedes escoger una categoría principal** marca en la columna "Elección" la categoría "Indeterminado".

Gracias por tu colaboración. Si tienes cualquier duda me puedes contactar por email (manuel.guerris@iqs.edu) o por teléfono (635.558.106).

ASISTEMBE - IQS Universitat Ramon Llull - Via Augusta 390. 08017 Barcelona - +34 932 672 000 – asistembe@iqs.url.edu

Fig. 9.1 Primera página del documento entregado a cada uno de los expertos en química

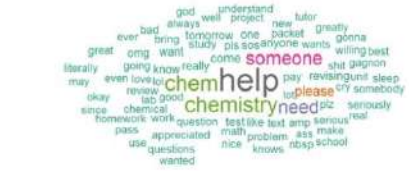
Cluster 1/50 (10)



Categoría	Descripción	Elección
Actividad humana	Términos relacionados con el rol de la química en la vida cotidiana incluyendo expresiones relacionadas con la industria o producción química.	
Conocimiento científico	Términos relacionados con conceptos o entidades químicas.	
Entorno educativo	Términos relacionados con la química como asignatura de clase así como actividades comunes de estudiantes.	
Entretenimiento	Términos relacionados con manifestaciones culturales: canción, grupo de música, concierto, película, serie de televisión ...	
Relaciones humanas	Términos relacionados con la emoción o sentimiento entre personas.	
Indeterminado	Términos mezclados de diferentes categorías.	

Fig. 9.2 Ejemplo de segunda página del documento entregado a uno de los expertos en química

Cluster 50/50 (8)



Categoría	Descripción	Elección
Actividad humana	Términos relacionados con el rol de la química en la vida cotidiana incluyendo expresiones relacionadas con la industria o producción química.	
Conocimiento científico	Términos relacionados con conceptos o entidades químicas.	
Entorno educativo	Términos relacionados con la química como asignatura de clase así como actividades comunes de estudiantes.	
Entretenimiento	Términos relacionados con manifestaciones culturales: canción, grupo de música, concierto, película, serie de televisión ...	
Relaciones humanas	Términos relacionados con la emoción o sentimiento entre personas.	
Indeterminado	Términos mezclados de diferentes categorías.	

Fig. 9.3 Ejemplo de última página del documento entregado a uno los expertos en química

Anexo 3. Lista de empresas y sociedades seleccionadas potencialmente más relevantes en el ámbito de la química

Listado de empresas:

Rango 2015	Compañía	Cuentas de Twitter	Número de cuentas
1	BASF	@BASF @BASFCorporation	2
2	Sinopec	@SinopecNews	1
3	Dow Chemical	@DowChemical	1
4	SABIC	@SABIC	1
5	Mitsubishi Chemical Holdings	-	0
6	LyondellBasell Industries	@LyondellBasell	1
7	ExxonMobil Chemical	@exxonmobil @ExxonMobil_EU @exxonmobil_sg	3
8	INEOS	@INEOS	1
9	DuPont	@DuPont_News @DuPont_ability	2
10	Linde Group	@The_Linde_Group	1
11	Toray	-	0
12	Sumitomo Chemical	-	0
13	Air Liquide	@airliquidegroup	1
14	Total	@Total	1
15	LG Chem	-	0
16	AkzoNobel	@AkzoNobel	1
17	Johnson Matthey	@Johnson_Matthey	1
18	PPG Industries	@PPG	1
19	Agrium	@agriuminc	1
20	Evonik	@Evonik	1
21	Merck KGaA	@EMDMilliporeBio@ MilliporeSigma @EMDSerono @ merckgroup	4
22	Ecolab	@Ecolab	1
23	Syngenta	@Syngenta	1
24	Covestro	@CovestroGroup	1
25	Yara International	@Yara	1
26	Reliance Industries	@flameoftruth	1
27	Solvay	@Solvay	1
28	Mitsui Chemicals	-	0
29	Braskem	@BraskemSA	1
30	Shin-Etsu Chemical	-	0
31	Sherwin-Williams	@SherwinWilliams	1
32	PTT Global Chemical	-	0
33	Praxair	@PraxairInc	1
34	Persian Gulf Petrochemical Industry	-	0
35	Huntsman	@Huntsman_Corp	1
36	Lotte Chemical	-	0
37	Air Products	@airproducts	1
38	Henkel Adhesive Technologies	@Henkel	1
39	Sekisui Chemical	-	0
40	DSM	@DSM	1
41	Eastman Chemical	@EstmanChemCo	1
42	Chevron Phillips Chemical	@chevronphillips	1
43	Sasol	@SasolSA	1

Rango 2015	Compañía	Cuentas de Twitter	Número de cuentas
44	Mosaic	@MosaicCompany	1
45	LANXESS	@LANXESS	1
46	Asahi Kasei	-	0
47	Borealis	-	0
48	Arkema	@Arkema_group	1
49	Teijin	-	0
50	Formosa Chemicals & Fibre (Taiwan)	-	0

Listado de sociedades:

Sociedad profesional	Cuenta de Twitter	Número de cuentas
American Chemical Society (ACS)	@AmerChemSociety	1
European Federation of Chemical Engineering (EFCE)	@EFCE_Comms	1
ACS Green Chemistry	@ACSGCI	1
American Chemistry	@AmChemistry	1
American Crystallographic Association, Inc. (ACA)	@ACAxtal	1
American Institute of Chemical Engineers (AIChE)	@ChEnected	1
American Oil Chemists' Society (AOCS)	@aocs	1
American Society for Biochemistry and Molecular Biology (ASBMB)	@ASBMB	1
American Society for Mass Spectrometry (ASMS)	@asmsnews	1
American Society of Brewing Chemists (ASBC)	@BrewingChemists	1
AOAC International	@AOACNews	1
Biochemical Society	@BiochemSoc	1
C&EN	@cenmag	1
Canadian Society of Clinical Chemists (CSCC)	@CSCC_CACB	1
Chemical Heritage Foundation	@ChemHeritage	1
Chemical Institute of Canada	@CIC_ChemInst	1
Chemistry World	@chemistryworld	1
Electrochemical Society	@ECSorg	1
European Association for Chemical and Molecular Sciences (EuCheMS)	@EuCheMS	1
European Chemical Industry Council (Cefic)	@Cefic	1
Federation of Analytical Chemistry and Spectroscopy Societies	@FACSSnetworking	1
Federation of the European Biochemical Societies (FEBS)	@FEBSnews	1
Indian Chemical Society	@indianchemsoc	1
Institute of Chemistry of Ireland (ICI)	@irishchemistry	1
Institution of Chemical Engineers (IChemE)	@IChemE	1
International Federation Of Clinical Chemistry and Laboratory Medicine (IFCC)	@EuroMedChem	1
International Mass Spectrometry Foundation (IMSF)	@TheIMSF	1
International Union of Crystallography (IUCr)	@IUCr	1
International Union of Pure and Applied Chemistry (IUPAC)	@IUPAC	1

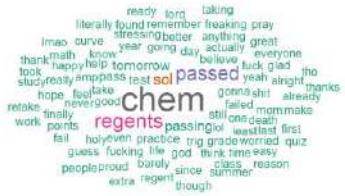
Sociedad profesional	Cuenta de Twitter	Número de cuentas
National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE)	@NOBCChE	1
Plastics and chemical industries asco	@AusChemistry	1
Royal Society of Chemistry (RSC)	@RoySocChem	1
Society of Chemical Industry (SCI)	@SCIupdate	1
Society of Chemical Manufacturers and Affiliates (SOCMA)	@socma	1
Society of Cosmetic Chemists (SCC)	@SocietyCosChem	1
The International Council of Chemical Associations (ICCA)	@ICCA_Chem	1
The Royal Australian Chemical Institute	@RACI_HQ	1

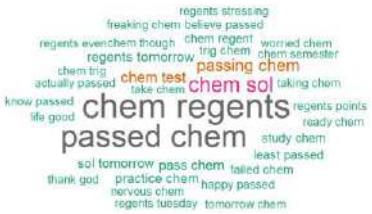
Anexo 4. Ejemplo del proceso de limpieza de *tweets* y detección de idioma

Número	text	texto limpio	idioma detectado en texto procesado
1	Drowning in AP Chem to ring in the new year. Ya nothing has changed...	drowning chem ring the new year nothing has changed	scots
2	RT @MCNocando: In 2014 I experienced the feeling that Goku felt on this episode of DBZ. I found Ubb I'm gonna trainâ€¦ http://t.co/usWzde6RBJ	experienced the feeling that goku felt this episode dbz found ubb gonna train	english
3	@chem_cake Ð□Ðµ.		*NA*
4	Ñ□ Ñ□Ð°Ð¾Ñ€Ð¾		*NA*
5	Ð¿Ð¾Ð¹Ð´Ñƒ Ñ□ÐµÐ±Ðµ Ñ‡Ð°Ð¹Ð°Ñƒ Ð·Ð°Ð²Ñ€Ð°Ðž Ñ□:		*NA*
6	RT @rtyourcrushx: Niall Horan http://t.co/EnaN4ybgRA	niall horan	afrikaans
7	RT @rtyourcrushx: Angelina Jolie http://t.co/pVl6kJVAzB	angelina jolie	dutch
8	im dreading going back to college ... especially bc that monday i have an 8am chem lab í ½íˌ„í ½í´«	dreading going back college especially that monday have chem lab	english
9	é§'æ¾¤ã□Œã□□ã□Ÿ		*NA*
10	baking cookies for chem lab and accidentally used self rising flour OMG	baking cookies for chem lab and accidentally used self rising flour omg	english

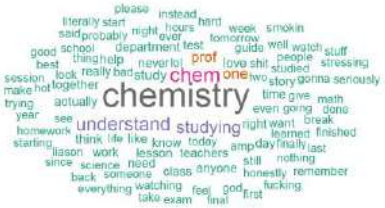
Anexo 5. Ejemplo de *wordclouds* de unigramas y bigramas obtenidos a partir del *clustering*

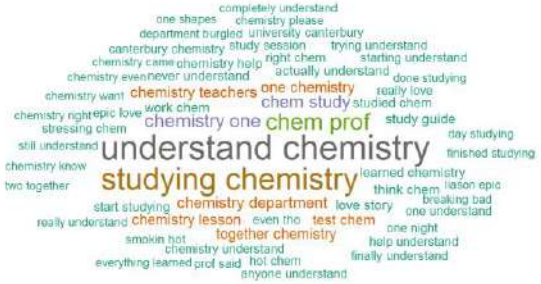
Clúster 1 número de *tweets* = 363 1/100



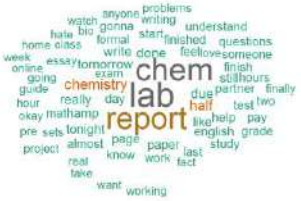


Clúster 2 número de *tweets* = 902 2/100

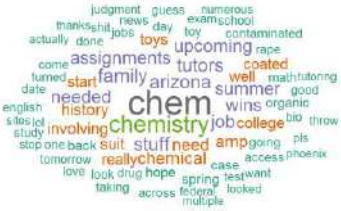




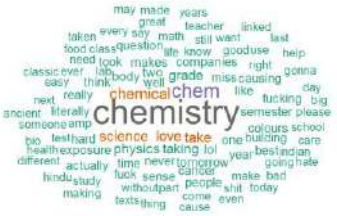
Clúster 3 número de *tweets* = 126 3/100

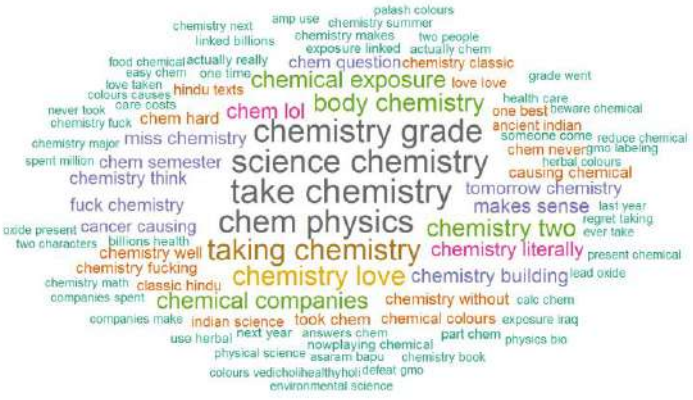


Clúster 4 número de *tweets* = 280 4/100



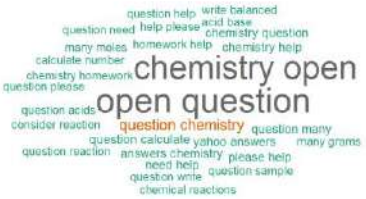
Clúster 5 número de *tweets* = 1235 5/100



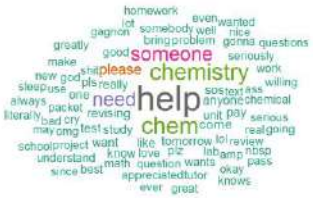


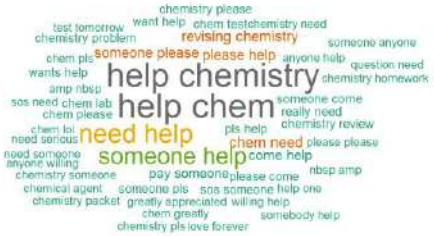
Clúster 6 número de *tweets* = 537 6/100

Clúster 7 número de *tweets* = 182 7/100

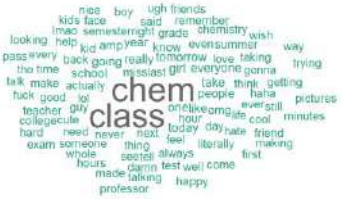


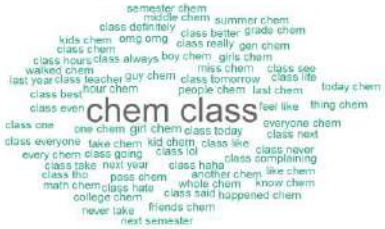
Clúster 8 número de *tweets* = 372 8/100



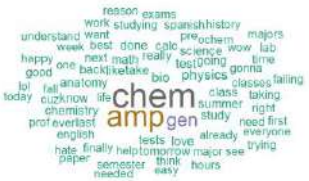


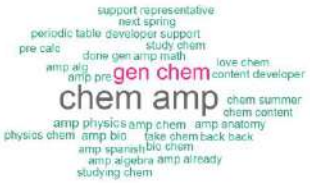
Clúster 9 número de *tweets* = 511 9/100





Clúster 10 número de *tweets* = 209 10/100





Anexo 6. Detalles de los expertos químicos seleccionados

Iniciales Experto	Especialidad 1	Especialidad 2	Carrera Universitaria	Master	Año PhD	Profesor desde
PB	Calidad	Analytical Chemistry	Chemical Engineer	Agro-food Chemistry Master	1992	2001
ES	Industrial Security		Chemical and Industrial Engineer		1999	2003
JJM	Numerical methods	Simulation	Chemical and Industrial Engineer		1988	1992
IB	Organic Chemistry	Pharmaceutical Chemistry	Chemical Engineer		1985	1994
RN	Industrial chemistry	Security	Chemical and Industrial Engineer		1982	1992
JA	Inorganic Chemistry	Industrial products analysis	Chemical Engineer		1997	2009
RP	Chemical and Biological Laboratories	Chemistry	Chemical Engineer	Master of Advanced Studies	2010	2018
OP	Thermodynamics	Energetic Technology	Chemical and Industrial Engineer		2009	2011
JT	Linear Algebra	Molecular Design	Chemical Engineer		1991	2005
JS	Projects	Functional design	Chemical and Industrial Engineer		1983	1992
RG	Environmental technologies	Technical draw	Chemical Engineer	Master in Environmental Engineering	2008	2014
XB	Integrated Laboratory	Process Chemistry	Chemical Engineer		2010	2018
DM	Inorganic Chemistry	Physics	Chemistry degree	Materials Science	2010	2015
LLC	Analytical Chemistry	Advanced Chromatography	Chemical Engineer	Masters in Chemistry	1978	1989
CC	Surface Chemistry	Nanoscience and Nanotechnology	Chemical and Industrial Engineer		1999	2010
JB	Analytical Chemistry Laboratory	Quality	Chemical Engineer	Environmental Master	2008	2004
RE	Mathematics	Molecular Design	Chemical Engineer	Master's degree in Applied Statistics	2011	2013
FLL	Processes control systems	Thermodynamics	Chemical Engineer	Master of Advanced Studies	2006	2015

Anexo 7. Asignación de clústeres a expertos

	Número de experto																	
Orden del clúster en el documento de clasificación	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	10	80	18	37	26	65	17	28	29	80	27	31	4	59	58	23	13	8
2	58	86	56	46	20	49	61	75	98	71	30	17	65	84	87	89	11	60
3	42	88	92	27	98	67	60	46	52	55	13	4	85	92	7	70	38	65
4	38	12	74	57	53	90	97	9	32	83	14	57	74	72	84	53	8	50
5	35	77	44	17	31	55	86	26	93	70	3	16	9	9	96	55	39	38
6	43	85	40	90	29	37	55	30	96	79	26	73	73	73	93	71	44	33
7	84	79	16	40	82	64	12	5	20	53	12	36	72	89	77	48	22	47
8	33	82	36	73	97	13	21	2	91	57	1	69	83	2	82	63	2	58
9	47	23	34	22	34	84	54	33	50	48	25	22	96	58	71	73	27	11
10	100	91	3	35	28	46	34	40	62	87	22	32	98	94	95	57	35	52
11	59	54	21	2	22	62	89	44	78	68	34	26	91	75	1	68	7	59
12	13	83	90	58	62	72	88	16	53	50	2	58	2	88	72	47	46	70
13	67	89	57	1	23	87	22	43	27	47	23	25	60	81	68	82	26	67
14	30	67	28	31	14	36	65	39	82	74	10	11	1	93	60	96	32	30
15	36	14	7	16	65	70	82	8	17	66	5	92	56	98	81	19	9	35
16	62	73	45	99	25	43	96	42	18	93	4	56	75	68	85	95	31	66
17	63	66	27	41	93	78	91	36	65	67	45	21	58	90	73	50	42	44
18	87	95	100	92	89	54	20	35	25	69	43	3	67	57	9	72	21	90
19	70	93	29	56	51	30	99	48	23	96	39	35	89	62	94	94	40	62
20	53	50	26	15	27	41	19	17	97	72	24	74	68	95	97	81	43	13
21	51	71	5	36	54	48	64	29	66	52	11	44	100	4	80	12	19	53
22	55	47	11	34	13	79	63	27	51	8	38	47	76	78	83	80	24	79
23	92	76	1	33	30	38	16	45	88	84	76	28	87	100	75	74	6	80
24	54	70	35	69	49	11	13	32	16	49	41	99	94	7	61	88	18	100
25	41	97	71	25	91	53	62	20	95	14	33	46	3	99	79	24	76	41
26	66	98	59	18	50	66	32	41	21	75	75	27	88	97	57	64	10	92
27	81	8	42	39	64	10	24	3	22	64	20	40	57	83	69	8	41	39
28	40	69	69	6	63	58	85	38	85	73	42	18	82	71	3	49	25	36
29	39	53	22	38	18	68	30	12	26	51	6	42	92	5	78	91	77	63
30	52	96	58	9	24	33	93	13	31	85	8	7	62	70	91	86	34	49
31	48	72	32	42	94	42	78	14	34	77	21	15	69	74	6	98	17	81
32	72	51	46	100	85	100	52	25	28	82	9	29	86	3	4	77	1	55
33	11	74	17	47	19	50	49	77	99	10	17	33	79	79	90	97	75	10
34	61	94	99	5	21	44	94	10	15	91	48	2	97	76	2	56	28	84
35	83	10	41	4	61	63	14	31	19	76	44	59	80	61	74	84	30	45
36	45	75	2	45	32	51	23	76	60	97	36	1	61	67	67	79	20	40
37	79	49	43	44	16	92	27	24	94	63	7	71	6	60	89	66	4	87
38	44	48	15	26	86	61	15	1	12	54	35	41	5	1	86	75	45	78
39	90	52	33	11	15	83	26	23	54	23	19	45	95	77	59	93	12	43
40	64	64	47	28	60	35	28	21	49	56	37	39	93	6	100	51	37	54
41	80	84	6	29	17	40	29	37	55	19	18	34	59	56	98	10	23	72
42	37	81	25	21	55	81	31	7	14	12	29	90	70	82	92	83	48	48
43	46	57	9	32	66	60	25	19	89	89	31	38	84	87	70	67	16	37
44	65	55	31	59	96	80	95	15	64	81	32	37	78	80	5	76	5	46
45	60	19	4	43	52	47	53	18	61	98	16	5	7	65	62	52	3	61
46	49	87	20	74	88	59	98	22	13	88	77	6	81	86	56	87	14	51
47	50	56	38	3	12	39	18	6	24	94	46	20	77	96	99	85	15	42
48	78	63	37	7	78	52	66	34	63	86	40	100	71	85	76	14	33	83
49	68	24	39	20	95	45	51	4	86	24	28	9	90	91	88	54	29	68
50	8	68	73	71	99	8	50	11	30	95	15	43	99	69	65	69	36	64

Anexo 8. Clasificación de clústeres por experto

	Número de experto																	
Número de clúster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1			EE	EE				CC			EE	EE	EE	EE	EE		AH	
2			EE	EE				EE			EE	EE	EE	EE	E		EE	
3			EE	AH				AH			EE	EE	EE	EE	EE		EE	
4			I	I				AH			EE	AH	EE	AH	EE		AH	
5			I	EE				EE			EE	I	I	RH	CC		AH	
6			AH	CC				AH			AH	AH	AH	AH	AH		CC	
7			CC	AH				EE			EE	EE	EE	EE	CC		EE	
8	RH	EE				AH		RH		EE	EE					EE	EE	RH
9			EE	EE				EE			EE	EE	EE	EE	EE		EE	
10	EE	I				CC		EE		EE	EE					EE	EE	CC
11	EE		RH	RH		I		RH			EE	I					E	RH
12		AH			AH		AH	CC	AH	AH	AH					AH	AH	
13	CC				AH	EE	AH	RH	AH		AH						AH	AH
14		EE			I		RH	I	I	CC	EE					CC	EE	
15			AH	EE	EE		RH	EE	CC		EE	EE					EE	
16			EE	EE	EE		EE	EE	EE		EE	EE					EE	
17			RH	RH	RH		AH	RH	RH		I	RH					E	
18			EE	RH	EE		CC	EE	I		EE	EE					EE	
19		AH			AH		CC	EE	I	CC	EE					AH	EE	
20			EE	EE	EE		EE	EE	EE		EE	EE					EE	
21			EE	I	EE		I	EE	EE		EE	EE					EE	
22			CC	EE	EE		EE	EE	AH		EE	EE					EE	
23		EE			EE		CC	EE	EE	EE	EE					EE	EE	
24		CC			AH		AH	CC	CC	AH	I					CC	CC	
25			AH	EE	EE		EE	EE	EE		EE	EE					EE	
26			I	I	I		CC	EE	I		EE	I					I	
27			CC	CC	CC		CC	CC	CC		CC	CC					CC	
28			AH	AH	AH		AH	AH	I		AH	AH					AH	
29			EE	EE	EE		EE	EE	EE		EE	EE					EE	
30	I				RH	AH	RH	AH	AH		AH						AH	AH
31			AH	AH	AH		CC	CC	AH		I	AH					AH	
32			CC	EE	EE		AH	EE	AH		I	EE					EE	
33	EE		CC	RH		E		RH			AH	I					RH	CC
34			I	AH	I		AH	AH	AH		AH	AH					AH	
35	AH		AH	AH		AH		AH			AH	AH					AH	AH
36	CC		AH	AH		AH		E			AH	AH					I	AH
37	EE		EE	EE		CC		EE			EE	EE					EE	CC
38	E		RH	I		E		AH			EE	EE					I	AH
39	CC		AH	AH		EE		CC			CC	I					AH	AH

	Número de experto																	
Número de clúster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
40	CC		AH	AH		I		I			AH	I					AH	AH
41	E		E	RH		RH		RH			RH	E					E	E
42	AH		AH	I		AH		AH			AH	I					AH	AH
43	EE		CC	AH		CC		EE			AH	EE					EE	EE
44	EE		EE	EE		EE		EE			EE	EE					EE	EE
45	EE		EE	EE		EE		EE			EE	EE					EE	EE
46	EE		EE	EE		EE		EE			EE	EE					EE	EE
47	EE	AH	I	RH		AH				I		I				AH		AH
48	EE	RH				AH		EE		RH	EE					EE	I	RH
49	CC	I			RH	AH	CC		I	AH						I		CC
50	RH	RH			E	RH	RH		RH	RH						RH		RH
51	EE	EE			EE	EE	EE		EE	EE						EE		EE
52	CC	EE			I	EE	EE		I	EE						CC		CC
53	EE	EE			EE	EE	EE		EE	EE						EE		EE
54	EE	CC			I	EE	I		I	AH						I		AH
55	EE	EE			EE	CC	EE		AH	EE						EE		EE
56		AH	AH	AH						AH		AH	AH	AH	CC	AH		
57		CC	CC	AH						CC		I	I	CC	CC	RH		
58	I		EE	RH		AH						I	EE	AH	CC			RH
59	EE		EE	EE		EE						EE	EE	EE	EE			EE
60	EE				EE	EE	EE		EE				EE	EE	EE			EE
61	EE				EE	EE	EE		EE				EE	EE	EE			CC
62	RH				RH	RH	RH		RH				I	E	E			E
63	EE	EE			I	I	I		EE	EE						EE		EE
64	I	EE			I	AH	EE		I	EE						AH		EE
65	EE				EE	I	CC		I				EE	I	EE			EE
66	I	AH			RH	AH	I		I	AH						RH		AH
67	EE	EE				EE				EE			EE	EE	EE	EE		EE
68	EE	CC				CC				EE			CC	EE	CC	EE		EE
69		AH	AH	CC						CC		CC	CC	CC	CC	CC		
70	E	E				EE				EE			I	AH	E	E		E
71		EE	I	I						EE		I	I	AH	EE	AH		
72	CC	I				RH				RH			RH	RH	I	E		RH
73		EE	EE	EE						EE		EE	EE	EE	CC	EE		
74		E	E	E						RH		E	E	E	E	E		
75		CC						EE		EE	I		I	CC	EE	I	AH	
76		AH						AH		AH	AH		AH	AH	CC	AH	AH	
77		EE						AH		EE	I		EE	EE	EE	EE	EE	
78	EE				I	AH	EE		AH				AH	I	AH			AH
79	EE	EE				RH				EE			EE	EE	EE	EE		EE
80	EE	CC				CC				CC			I	AH	CC	EE		CC

	Número de experto																	
Número de clúster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
81	EE	EE				EE				EE			EE	EE	EE	EE		EE
82		EE			EE		EE		EE	EE			EE	EE	EE	EE		
83	EE	EE				I				EE			I	EE	I	AH		EE
84	CC	AH				CC				CC			CC	CC	CC	CC		CC
85		EE			EE		AH		EE	EE			EE	EE	EE	EE		
86		I			AH		AH		AH	EE			I	AH	EE	AH		
87	E	RH				E				RH			E	E	E	E		E
88		RH			RH		RH		EE	RH			I	RH	RH	EE		
89		E			E		E		E	E			E	RH	E	E		
90	EE		I	EE		CC						I	I	AH	AH			AH
91		EE			EE		I		I	EE			EE	EE	I	EE		
92	EE		EE	EE		E						EE	I	AH	EE			EE
93		I			I		I		RH	I			I	AH	EE	RH		
94		AH			AH		EE		AH	AH			I	AH	I	AH		
95		EE			RH		RH		RH	AH			RH	EE	EE	EE		
96		RH			RH		RH		RH	RH			RH	RH	RH	RH		
97		RH			RH		AH		I	AH			RH	I	CC	RH		
98		CC			I		EE		I	CC			I	EE	EE	EE		
99			EE	RH	RH		EE		CC			I	I	EE	I			
100	RH		RH	RH		AH						EE	RH	EE	I			EE

AH = Actividad Humana
CC = Conocimiento Científico
EE= Entorno Educativo
E = Entretenimiento
RH = Relación Humana
I = Indefinido

Anexo 9. Extracto del lexicón utilizado en en análisis de sentimientos

Número	Palabra	Polaridad
11156	barite	0,000
11157	baritone	-0,031
11158	baritone_horn	0,000
11159	baritone_voice	0,000
11160	barium	0,000
11161	barium_dioxide	0,000
11162	barium_enema	0,000
11163	barium_hydroxide	0,000
11164	barium_monoxide	-0,125
11165	barium_oxide	-0,125
11166	barium_peroxide	0,000
11167	barium_protoxide	-0,125
11168	barium_sulfate	0,000
11169	barium_sulphate	0,000
11170	bark	-0,042
11171	bark-louse	0,000
11172	bark_beetle	0,000
11173	bark_louse	0,000
11174	barkeep	0,000
11175	barkeeper	0,000
11176	barker	0,000
11177	barking_deer	0,000

Anexo 10. Extracto del resultado del cálculo de polaridad de los *tweets* de la temática Actividad Humana (AH)

Temática AH		
Identificador_*tweet*	**Texto_limpio**	**Polaridad_*tweet***
550768971511656449	chem pre cal homework	0
550770874647732224	eat food still want chem notes love	0,277777778
550771122568847360	fresno state stop someone please oakland raiders see haha	0,210454545
550773234610630656	literally learned one thing chem used high	0,493055556
550773732688400384	full chem	0,019230769
550775521903382528	chandelier come memories come flooding back chem medlink medlinkdoc	0,439123377
550777145652695040	literal reaction saw grade chem smilethroughthepain	0,033791209
550777795387736064	literally want cry looking chem packet literally know none questions	-0,25
550779971447570434	rather million math chem problems writing friggin research paper fschool	-0,166666667
550782058982035457	holy shit chem kids question binding energy like honestly fuck	0,816355519
550785102503243776	chem engers going hard even though apart	-0,36525974
550785289275576322	hiring chemical process control engineer rahway processcontrolengr	-0,022267206
550785476589027328	please vote intelligent process control system chemical industry realtimechem chemtics	0,617316127
550788399238770692	nowplaying nothing chemistry version razortop solar brings tasty funk atmosphere one	0,364583333
550788868149350400	job hallandale chemistry tutors needed upcoming assignments summer well	0,366117647
550788903544713216	braid free weave free twist free chemical free free naturalhair	-0,056324111
550789062542376960	syria firstaid people protect sarin chemical weapons safety staid	0,270833333
550789115344863232	beautiful love story miss aperfectending amp chemistry unmatched love	1,25
550789198555668480	job sewickley college chemistry tutors needed upcoming assignments summer well	0,366117647

Anexo 11. Extracto de clasificación de la muestra de *tweets* positivos de la temática Entorno Educativo (EE) correspondiente a los *tweets* que contienen los términos “test” y “teacher”

Número	*Tweets* que contienen el término "test"	IRONÍA	POSITIVO	NEGATIVO	NEUTRO	SIN CLASIFICAR
1	Can someone please FaceTime me for like 10 mins and teach me how to do the Chem test				X	
2	He's gonna see my first test score and laugh í ½í¸í ½í¸ I NEED TO GET AN A ON MY CHEM TEST I AM SO DETERMINED TO IMPRESS HIM		X			
3	Positive that I'm failing this chemistry test tomorrow<ed><U+00A0><U+00BD><ed><U+00B8><U+008A><ed><U+00A0><U+00BD><ed><U+00B8><U+008A><ed><U+00A0><U+00BD><ed><U+00B8><U+008A>	X		X		
4	Chemistry test tomorrow and I have no clue what we've even been "learning"	X		X		
5	I NEED to good on this chem test				X	
6	Failing this chem test tomorrow			X		
7	I'm honestly really pissed off I got a 90 on my Chemistry test and not a 100 bye			X		
8	lol guessed on more than half the questions on my chem test				X	
9	97 on chem test is this even real life	X				X
10	@oMxrs i did great on my speech but failed my chem test lmao			X		
11	That chemistry test was some kinda ridiculous #prayformygrades	X		X		
12	I put avocado's constant on my chem test and my teacher didn't even notice	X			X	
13	My chem test was actually my professor playing a trick on me.	X			X	
14	I hope I'll have time to finish my Chem. test tomorrow, orthodontics are an inconvenience.			X		
15	Just so everyone knows, after my last Chem test I decided I'm going to TSTC to get certified as a welder. Ray Bans better be OSHA approved.				X	
16	Tfw you drive out to Barnes and Noble and deal with idiots on 19 only to find out that they're sold out of the ap chem test book #weeping				X	
17	Hehe chemistry test tmoro hehe can't wait to fail hehe forgot everything I revised Xxxxxxxxxzxx	X		X		
18	All I want is a tall guy with a sense of humor and to pass this chem test is that too much to ask foe				X	
19	Chem and eng test today				X	
20	WOO CHEMISTRY TEST TOMORROW YEP SO FUN I LOVE SCHOOL SO STRESS FREE AND EVERYTHING WOW		X			
21	I still haven't studied for this Chemistry Test <ed><U+00A0><U+00BD><ed><U+00B8><U+0085><ed><U+00A0><U+00BD><ed><U+00B8><U+0085><ed><U+00A0><U+00BD><ed><U+00B8><U+0085><ed><U+00A0><U+00BD><ed><U+00B8><U+0085><ed><U+00A0><U+00BD><ed><U+00B8><U+0085>				X	
22	When you have 6 hours of dance, calculus homework, and a chem test to study for :,) http://t.co/K2qHVRzyxU	X				X
23	“@hollyelainne: Me when I read the first question on my chemistry test.. http://t.co/v6KndKdjPq” thx @abigailsudduth	X				X

Número	*Tweets* que contienen el término "teacher"	IRONÍA	POSITIVO	NEGATIVO	NEUTRO	SIN CLASIFICAR
1	My chemistry teacher actually didn't say my first name wrong during attendance...I mean I told her before but still lol	X			X	
2	when u find out your chem teacher is a part time masseuse http://t.co/Z0xaddU3fe	X			X	
3	Speaking of Sudafed, it's essentially just really mild meth. Or so said my high school Chem teacher, Mr. Kadle.				X	
4	My chemistry teacher's handwriting, it makes me cringe! http://t.co/2XyPIA3eWA			X		
5	Next semester's goal: become bio/chem teacher's pet so I have good references/connections later on í ½í¸¼				X	
6	My chemistry teacher is making his own formulas and testing us on it... Is that allowed?				X	
7	@alexbracken I started making squealing noises in Chemistry today while reading Never Fade cause Chubs is back!.. Teacher doesn't approveí ½í¸			X		
8	I love how my chem teacher even knows about how much of a crazy person I am		X			
9	whY THE FUCK WOULD YOU GIVE US HOMEWORK THE DAY BEFORE A TEST THAT ISN'T EVEN RELEVANT TO ANYTHING WE HAVE LEARNED @ CHEM TEACHER			X		
10	My step-dad could win an award for pissing me off more than my chemistry teacher	X		X		
11	@cassieonthecob ok but the new semester started over a month ago and my Chem teacher still hasn't published grades I gtg				X	
12	Chemistry teacher really expects us to do this packet.				X	
13	IM GOING TO FUCKING KILL MYSELF BC I JUST REALIZED IM HAVING THE STUPID CHINESE CHEM TEACHER WHOSE CLASS IS SO HARD			X		
14	I also remember the time my chemistry teacher ripped up someone's work because they found the correct answer in the textbook				X	
15	My chem teacher started class by showing us the band that he's in Facebook page.				X	
16	So the chem teacher can't find his HCL for the demo tomorrow and I can't finish writing/printing the program until he does #principalprobs				X	
17	my chem teacher put my midterm grade on my report card and when my dad sees it im gonna get in so much trouble i hate my life			X		
18	*chem teacher talking about kids putting weed in rice krispies & eating them at school* "*gasp* Who would do such a profound thing????"	X		X		

Anexo 12. Extracto del lexicón NRC v.0.92

Término	Emoción	Valencia emoción 1= Sí, 0 = No
aback	anger	0
aback	anticipation	0
aback	disgust	0
aback	fear	0
aback	joy	0
aback	negative	0
aback	positive	0
aback	sadness	0
aback	surprise	0
aback	trust	0
abacus	anger	0
abacus	anticipation	0
abacus	disgust	0
abacus	fear	0
abacus	joy	0
abacus	negative	0
abacus	positive	0
abacus	sadness	0
abacus	surprise	0

Término	Emoción	Valencia emoción 1= Sí, 0 = No
abacus	trust	1
abandon	anger	0
abandon	anticipation	0
abandon	disgust	0
abandon	fear	1
abandon	joy	0
abandon	negative	1
abandon	positive	0
abandon	sadness	1
abandon	surprise	0
abandon	trust	0

Anexo 13. Extracto de resultados de evaluación de emociones para cada uno de los *tweets*

Identificador *Tweet*	Ira	Anticipacion	Disgusto	Miedo	Alegria	Tristeza	Sorpresa	Confianza
550802458520596	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550801268130603	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.33
550792495588982	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550785102503243	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550782058982035	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00
550779971447570	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550777795387736	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
550777145652695	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550775521903382	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550773732688400	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550773234610630	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550771122568847	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
550770874647732	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.33

Anexo 14. Resultado de *tweets* publicados y en los que son mencionados de las entidades pertenecientes a la lista de empresas y sociedades seleccionadas del ámbito de la química

***Tweets* publicados:**

nombre	cuenta	tweet
C&EN	cenmag	RT @ACSERC: In this week's @cenmag: @AmerChemSociety will host an #investor #pitch event for #chemistry #startups in San Diego: http://t.co…
C&EN	cenmag	Spring has sprung! And with it the beautiful (and poisonous) chemistry of daffodils http://t.co/fXKv8zBVzx http://t.co/z9v6N93VAO
Chemistry World	ChemistryWorld	Chemistry laureate Yves Chauvin who laid out olefin metathesis mechanism has died http://t.co/GlosOPymK9 http://t.co/maTihD6kzN
RACI	RACI_HQ	RT @MonashUni: Professor Milton Hearn has been awarded a prestigious award from the American Chemical Society: http://t.co/CROyMoV7vj
RACI	RACI_HQ	RT @SimonWLewis: Chemistry in the Bush. Dr Gavin Flematti presents the 2015 RACI Bayliss Lecture at Curtin Uni @RACI_HQ @reneewebs http://t…
Amer Chem Society	AmerChemSociety	Special recognition for PSU Student Affiliates of the American Chemical Society | News http://t.co/GJuHaSOfAH @ACSundergrad
Amer Chem Society	AmerChemSociety	RT @ACSGCI: Chemical Angel Network investor gives industry tips at @DESCAlliance â€˜Lunch and Learnâ€™ event: http://t.co/wULarKiBnA
Amer Chem Society	AmerChemSociety	RT @acswebinars: Set your alarm, peeps. Dr. Hartel will be taking your chocolate chemistry Qs in tomorrow's Reddit AMA at 12p ET! http://t.â€¦
Amer Chem Society	AmerChemSociety	In 12 min, you have 12 h to enter #chemchamps http://t.co/GkZ7hvm5J1 #scicomm glory
Amer Chem Society	AmerChemSociety	How are you shaping the future of ACS and the broader chemistry community? #YouShareWeLearn http://t.co/hWJFGVcHLa
Amer Chem Society	AmerChemSociety	RT @ACSReactions: Got a chemistry tattoo? Send them our way! We're currently working on a video about tattoos and we want YOU to be in it!
ChEnected AIChE	ChEnected	RT @CharlieSchwedle: Mid Michigan Section of the American Institute of Chemical Engineers @ChEnected #bullockcreek #STEM http://t.co/zt7ebT…
Chevron Phillips	chevronphillips	CPChem CEO Pete Cella to speak tomorrow @ The Woodlands Economic Outlook Conf about closing the labor shortage gap in the chemical industry.
ExxonMobil	exxonmobil	#Innovation in the chemical industry means benefits to the envt and sustainability. Learn how we're involved http://t.co/7T1kuLTVRa
Dow	DowChemical	RT @TheENDSReport: Dow Chemical seeks to save money by valuing nature http://t.co/TqKZN5gLdD

Tweets mencionados:

nombre	cuenta	usuario	tweet
C&EN	@cenmag	Ktilger_Chem	RT @cenmag: High-speed photography trick helps solve age-old question of why sodium and potassium explode in water. Video here: http://t.coâ€¦
C&EN	@cenmag	susanjainsworth	Job in 1 wd: service-oriented--Soosairaj Therese, chem prof at Bronx CC #cenoneword #acsdenver @cenmag @ACSNatlMtg http://t.co/AFPvrtlZAJ
C&EN	@cenmag	Chem_Consult	RT @cenmag: Not the bees' knees: Evidence mounts against neonicotinoid pesticides http://t.co/LD1Kej6gxy http://t.co/YxCcyHDKud
C&EN	@cenmag	Ktilger_Chem	RT @cenmag: What's Tony Stark's suit really made of? @acsreactions takes on the science of The @avengers https://t.co/XgncYOn4LY
C&EN	@cenmag	hony_2014	Chemical Safety: Explosion hazard in synthesis of azidotrimethylsilane http://t.co/L8lGIHvCPE via @cenmag
C&EN	@cenmag	arndt_eric	RT @cenmag: RT @DrRubidium: "I only buy chemical-free." #FiveWordsToRuinADate
C&EN	@cenmag	meganlatshaw	RT @cenmag: RT @DrRubidium: "I only buy chemical-free." #FiveWordsToRuinADate
C&EN	@cenmag	IBchemmilam	RT @cenmag: RT @DrRubidium: "I only buy chemical-free." #FiveWordsToRuinADate
C&EN	@cenmag	ACSDivCHED	RT @BibianaCampos: Dow donates $1million to support AACT http://t.co/6Med6PGvev via @cenmag @ACSpressroom #scienceed
C&EN	@cenmag	MistahKindah	RT @cenmag: Cover story: How custom chemical manufacturers help bring drugs (like cancer drug Imbruvica, shown) to market http://t.co/ZVKeaâ€¦
C&EN	@cenmag	laurenkwolf	Chemistry Quiz: For another preview of @cenmag 's story "When Chemicals Became Weapons of War," test your wits: http://t.co/VFAIFpClHA
C&EN	@cenmag	RSC_HSG	RT @compoundchem: Compelling reading: @saraheverts & @cenmag mark 100 years since advent of chemical warfare: http://t.co/PZZw4HZRKZ http:/â€¦
C&EN	@cenmag	acsdchas	@cenmag: Boosting Safety At Chemical Facilities: EPA weighs new approaches to process safety http://t.co/YRE82mDFxo http://t.co/RzORXpOmPZ
C&EN	@cenmag	markmackllc	Two Italian chemical firms have launched projects to produce levulinic acid from biomass. http://t.co/ubdTWAN2rC via @cenmag
C&EN	@cenmag	Catalysis_IMCN	RT @vigabalme: Great short biography of Fritz Haber and his relationship with war: http://t.co/eg0KmkkuyP via @cenmag
C&EN	@cenmag	Sciguy999	Union Carbide, now a Dow Chemical sub, to buy back W Virginia plant that once produced methyl isocyanate http://t.co/DzJpaXdQhC via @cenmag
C&EN	@cenmag	alexcunn	RT @cenmag: At the #Watchglass: Celebrating the Keeling Curve http://t.co/Kw7ooK2Znw @amerchemsociety @noaa @scripps_ocean http://t.co/1bG…
C&EN	@cenmag	vancew	RT @cenmag: At the #Watchglass: Celebrating the Keeling Curve http://t.co/Kw7ooK2Znw @amerchemsociety @noaa @scripps_ocean http://t.co/1bG…
C&EN	@cenmag	WendinhGarCalde	RT @cenmag: At the #Watchglass: Celebrating the Keeling Curve http://t.co/Kw7ooK2Znw @amerchemsociety @noaa @scripps_ocean http://t.co/1bG…
C&EN	@cenmag	dpchalasani	A generally optimistic cover story in this week's @cenmag about India's #chemical industry. @PMOIndia http://t.co/OR50LC1eAP
C&EN	@cenmag	AWinnr	RT @AmerChemSociety: Video: Bugs wage chemical warfare with butts and guts featuring @cenmag #speakingofchem http://t.co/tmwj7IvMJn http://…

nombre	cuenta	usuario	tweet
C&EN	@cenmag	tralfamadorable	These tats are SWEET. #chemistry Randa Roland http://t.co/2q3ObaoD0p via @cenmag
C&EN	@cenmag	m39618462	RT @cenmag: Even after you get out of the sun, aftereffects of UV light can cause cancerous damage to skin http://t.co/uJWAgzkzre http://t.…
C&EN	@cenmag	cenmag	RT @ACSERC: In this week's @cenmag: @AmerChemSociety will host an #investor #pitch event for #chemistry #startups in San Diego: http://t.co…
C&EN	@cenmag	ConnorChapek	RT @cenmag: Now, usually we don't do this but, uh, we thought we'd share a li'l preview of the Guinness. http://t.co/MHE7l2xRmh http://t.co…
C&EN	@cenmag	Chemkishan	RT @cenmag: Now, usually we don't do this but, uh, we thought we'd share a li'l preview of the Guinness. http://t.co/MHE7l2xRmh http://t.co…
C&EN	@cenmag	ErikaOltermann	RT @BerkeleyLab: Why do bubbles in a pint of @GuinnessIreland fall? http://t.co/R2rMCknfOh @cenmag #StPatricksDay #beer http://t.co/awbPrNd…
C&EN	@cenmag	susanjainsworth	Job in 1 wd: stimulating--principle investigator, NIH #cenoneword #acsdenver @ACSNatlMtg @cenmag #chemjobs #chemistry http://t.co/z9nCfzTJ68
C&EN	@cenmag	MoleculeWorld	RT @cenmag: New #3Dprinting method cuts print times to minutes--with #chemistry! http://t.co/f8utWHwIgT #Video via @carbon3D http://t.co/gT…
C&EN	@cenmag	mrnavas	RT @cenmag: New #3Dprinting method cuts print times to minutes--with #chemistry! http://t.co/f8utWHwIgT #Video via @carbon3D http://t.co/gT…
C&EN	@cenmag	Carbon3D	RT @cenmag: New #3Dprinting method cuts print times to minutes--with #chemistry! http://t.co/f8utWHwIgT #Video via @carbon3D http://t.co/gT…
C&EN	@cenmag	jv3sund	RT @LamResearch: Another #FunFactFriday! Super cool #infographic on the #chemistry of #Guinness via @cenmag! http://t.co/RvhIUYyMEB http://…
C&EN	@cenmag	Ritabril	RT @cenmag: New #3Dprinting method cuts print times to minutes--with #chemistry! http://t.co/f8utWHwIgT #Video via @carbon3D http://t.co/gT…
C&EN	@cenmag	wmgeil	RT @compoundchem: This month’s @cenmag graphic looks at daffodil chemistry: colour, aroma, poison & medicine: http://t.co/rvDQ28WeSY http:/…
C&EN	@cenmag	Chm107Downtown	RT @cenmag: Spring has sprung! And with it the beautiful (and poisonous) chemistry of daffodils http://t.co/fXKv8zBVzx http://t.co/z9v6N93V…
C&EN	@cenmag	ItsScienceMagic	RT @cenmag: Spring has sprung! And with it the beautiful (and poisonous) chemistry of daffodils http://t.co/fXKv8zBVzx http://t.co/z9v6N93V…
C&EN	@cenmag	josharellano	RT @cenmag: Spring has sprung! And with it the beautiful (and poisonous) chemistry of daffodils http://t.co/fXKv8zBVzx http://t.co/z9v6N93V…
C&EN	@cenmag	MellaNofri	RT @cenmag: Chemistry in the UK: Attitude is largely <ed><U+00A0><U+00BD><ed><U+00B8><U+0090>, not rotten nor vicious. #chemperceptions http://t.co/lg8eMjzXqt http://t.co/2R3ryu…
C&EN	@cenmag	nanocastsafety	RT @cenmag: From the Safety Zone: A chemistry fire at @uutah, with lessons for the future http://t.co/h8h1qxamDw #chemsafety @jkemsley
C&EN	@cenmag	MSScienceBlog	RT @compoundchem: Summer strawberry chemistry is the theme in this month’s #PeriodicGraphics in @cenmag: http://t.co/1I1Xj4WDoO http://t.co…
C&EN	@cenmag	AtomicUniverse	@cenmag Please support and retweet. #science #education in #comics for #teaching https://t.co/bqdHwFjfWt http://t.co/ND9ALkFVfZ

nombre	cuenta	usuario	tweet
C&EN	@cenmag	liam_d_odonnell	RT @compoundchem: Summer strawberry chemistry is the theme in this month's #PeriodicGraphics in @cenmag: http://t.co/1I1Xj4WDoO http://t.co…
Chemistry World	@ChemistryWorld	snapdannyboy	RT @ChemistryWorld: Sad inditement of US policy on science? Cash for toughest chem synthesis problem runs out http://t.co/qXbTwK2s7k http:/â€¦
Chemistry World	@ChemistryWorld	oundle_chem	RT @ChemistryWorld: These water droplets mixed with food colour chase each other. We now know why #dotrythisathome http://t.co/5pIjQZiEad h…
Chemistry World	@ChemistryWorld	ericscerri	is this cool or what? world population clock http://t.co/foAjZoW1aC http://t.co/sxpYebLPmH .@angew_chem.@ChemistryWorld.@NYT.@WSJ.@econ
Chemistry World	@ChemistryWorld	JEPG24	RT @ChemistryWorld: Waves of self-catalysing chemical reactions can be controlled using short strands of DNA http://t.co/sllJu1L9Jj
Chemistry World	@ChemistryWorld	The_Frasian	RT @ChemistryWorld: Just meteorite dust, a simple chemical and the solar wind can form DNA bases http://t.co/iuo5BnP4iM http://t.co/1bQhg1V…
Chemistry World	@ChemistryWorld	glitter_hole	RT @ChemistryWorld: Just meteorite dust, a simple chemical and the solar wind can form DNA bases http://t.co/iuo5BnP4iM http://t.co/1bQhg1V…
Chemistry World	@ChemistryWorld	aleradar	RT @ChemistryWorld: Just meteorite dust, a simple chemical and the solar wind can form DNA bases http://t.co/iuo5BnP4iM http://t.co/1bQhg1V…
Chemistry World	@ChemistryWorld	CoC_guy	RT @ChemistryWorld: Just meteorite dust, a simple chemical and the solar wind can form DNA bases http://t.co/iuo5BnP4iM http://t.co/1bQhg1V…
Chemistry World	@ChemistryWorld	whatsbobgonnado	RT @ChemistryWorld: Scientists claim evidence for a new type of bond, 1st proposed 30 years ago http://t.co/tXlfrMNkXF http://t.co/ITuvAEzT…
Chemistry World	@ChemistryWorld	morrisresearch	RT @ChemistryWorld: US chemical industry body accused over links to discredited fire safety group http://t.co/qM0QJTMv0V http://t.co/wBdvOW…
Chemistry World	@ChemistryWorld	chemparrot	RT @ChemistryWorld: A lot of people have been reading this one: Toughest synthesis challenge nears completion http://t.co/qXbTwK2s7k http:/â€¦
Chemistry World	@ChemistryWorld	Furness_Ent	RT @GSKUlverston: If you have the right chemistry, this job at GSK Ulverston might be for you. @ChemistryWorld : http://t.co/c6K2donJAp
Chemistry World	@ChemistryWorld	Fadders10	RT @ChemistryWorld: Chemistry laureate Yves Chauvin who laid out olefin metathesis mechanism has died http://t.co/GlosOPymK9 http://t.co/maâ€¦
Chemistry World	@ChemistryWorld	dbfulton	RT @ChemistryWorld: Chemistry laureate Yves Chauvin who laid out olefin metathesis mechanism has died http://t.co/GlosOPymK9 http://t.co/maâ€¦
Chemistry World	@ChemistryWorld	borderbend	RT @artologica: Oooh, special Chemistry and Art themed issue of @ChemistryWorld http://t.co/CociEQIV4o via @stuartcantrill
Chemistry World	@ChemistryWorld	AidanBaker	RT @Clare_Sansom: Thankful for splendid @ChemistryWorld science communication awards meeting with 10 excellent finalists and chemistry them…
Chemistry World	@ChemistryWorld	Clare_Sansom	Thankful for splendid @ChemistryWorld science communication awards meeting with 10 excellent finalists and chemistry themed lunch
Chemistry World	@ChemistryWorld	Tzublal	RT @ChemistryWorld: Massive open online courses are letting millions learn. What's in it for unis? http://t.co/QDsy9YPsj8 #MOOC http://t.co…
Chemistry World	@ChemistryWorld	Westaeus	RT @ChemistryWorld: Review: 'Chemistry: a very short introduction'. A slim yet essential volume http://t.co/XdmiZqKW8w http://t.co/GNjCmjoL…
Chemistry World	@ChemistryWorld	GBoissonnat	RT @abithramadevan: Cart full of solvents on the way to hood... Chemists satisfaction #chemistry #research @RealTimeChem @ChemistryWorld ht…

nombre	cuenta	usuario	tweet
Chemistry World	@ChemistryWorld	sherfranklin	@ChemistryWorld thanks for sharing Chemistry World, have a great Thursday :) (insight by http://t.co/I48EZLrNkN)
Chemistry World	@ChemistryWorld	EllenMellon_88	RT @ChemistryWorld: How does the public feel about chemistry? @markpeplow looks at the evidence #chemperceptions http://t.co/R0wjEa7P4j htt…
Chemistry World	@ChemistryWorld	thebiolady	RT @ChemistryWorld: Survey time. Do you enjoy reading Chemistry World? Tell us why, so we can make it even better https://t.co/yYviaqONv3 h…
ACS Green Chemistry	@ACSGCI	AmerChemSociety	RT @ACSGCI: Chemical Angel Network investor gives industry tips at @DESCAlliance â€˜Lunch and Learnâ€™ event: http://t.co/wULarKiBnA
ACS Green Chemistry	@ACSGCI	Kidzroom	Coming? "@ACSGCI: Design of Safer Chemicals and Products: the Nexus of Toxicology and Chemistry: http://t.co/kE05WGw7ho via @NWGreenChem"
ACS Green Chemistry	@ACSGCI	ktann83	http://t.co/C45ZIlJitM this gives a real positive direction for the future of pharmaceutical chemistry. #cnmchem2810 thanks @ACSGCI
American Chemistry	@AmChemistry	MunkhturTMT	RT @EnergyNation: .@AmChemistry: U.S. chemical industry #shale investment hits $135B http://t.co/Xol6ZC7X0p http://t.co/cHhA8kXiaB
American Chemistry	@AmChemistry	sime0n	@AmChemistry @bizroundtable @chemsafetyboard Gross mismanagement of the investigations by the chairman and director: http://t.co/RLCaX9noOB
American Chemistry	@AmChemistry	OffGridOnGrid	RT @AmChemistry New #Trade Report: U.S. #chemical #exports linked to #ShaleGas could double by 2030: http://t.co/csYbWW5AMf #NatGas http://â€¦
American Chemistry	@AmChemistry	ideawomen	RT @AmChemistry: New #Trade Report: U.S. #chemical #exports linked to #ShaleGas could double by 2030: http://t.co/esTe2BV1St #NatGas http:/…
American Chemistry	@AmChemistry	nratrh	RT @AmChemistry: When fluorine and carbon atoms join together, they create a powerful #chemical bond vital to many industries! #PFAS http:…
American Chemistry	@AmChemistry	MagloVillalba	RT @AmChemistry: When fluorine and carbon atoms join together, they create a powerful #chemical bond vital to many industries! #PFAS http:…
American Chemistry	@AmChemistry	fcoroac	RT @AmChemistry: When fluorine and carbon atoms join together, they create a powerful #chemical bond vital to many industries! #PFAS http:…
American Chemistry	@AmChemistry	dewi_darmawati	RT @AmChemistry: When fluorine and carbon atoms join together, they create a powerful #chemical bond vital to many industries! #PFAS http:…
American Chemistry	@AmChemistry	bizcochopat	RT @AmChemistry: When fluorine and carbon atoms join together, they create a powerful #chemical bond vital to many industries! #PFAS http:…
American Chemistry	@AmChemistry	polymerchemical	RT @AmChemistry: If you support helping #chemical manufacturers drive massive #job growth, you support #TPA: http://t.co/NeNrxbbRTi http://…
American Chemistry	@AmChemistry	skoolycool	@AmChemistry I don't support absolving chemical companies of their environmental responsibilities.
American Chemistry	@AmChemistry	Rdday91191	RT @AmChemistry: If you support unleashing massive growth potential for US #chemical exports, you support #TPA: http://t.co/6uFDyqPM6k http…
American Chemistry	@AmChemistry	TaylorVenture7	RT @AmChemistry: If you support helping #chemical manufacturers drive massive #job growth, you support #TPA: http://t.co/NeNrxbbRTi http://…
American Chemistry	@AmChemistry	baldeilys	RT @AmChemistry: Learn about #FluoroTechnology innovations and use of short-chain #PFAS: http://t.co/ojGmLV7VRj http://t.co/zJdEHlhygH http…
American Chemistry	@AmChemistry	featherlou	@AmChemistry Sure, that's just what Big Chemistry would say.

nombre	cuenta	usuario	tweet
American Chemistry	@AmChemistry	silanoldep	Chemistry!!! is out! http://t.co/jR4binWyPa Stories via @AmChemistry @stuartcantrill @ACSPublications
American Chemistry	@AmChemistry	smokva200	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	drskharlamov	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	IsraelFonseca11	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	Magallanes35	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	Scarlet37236746	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	zouxingjian	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	AlexaSkia	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	mrizzo64	RT @AmChemistry: Learn about #FluoroTechnology innovations and important use of short-chain #PFAS: http://t.co/bmEKyUKhnb #Chemistry http:/…
American Chemistry	@AmChemistry	davidhth	RT @CaptG2: See, I told you so... "@AmChemistry lied about lobbying role on flame retardants, consultant says" http://t.co/IKuOpyREfS via @…
RACI	@RACI_HQ	RACI_HQ	RT @SimonWLewis: Chemistry in the Bush. Dr Gavin Flematti presents the 2015 RACI Bayliss Lecture at Curtin Uni @RACI_HQ @reneewebs http://t…
RACI	@RACI_HQ	arri_aus	RT @CurtinMedia: The @RACI_HQ Titration Comp for high school students is off and running in the Curtin chemistry labs http://t.co/SpDiXhfq0…
SCI	@SCIupdate	chitinette	RT @SCIupdate: Thinking about starting a career in #chemical and #lifesciences? Attend #dayofscienceandcareers #Scotland on 1 June: http://…
Amer Chem Society	@AmerChemSociety	AndyNguyenTC	@UTArlington Chem Prof Sandy Dasgupta honored by @AmerChemSociety for Excellence In Education. http://t.co/iA8J4T1dAs
Amer Chem Society	@AmerChemSociety	ACSPublications	RT @AmerChemSociety: #ACSdenver Fred Kavli Innovation in Chem Lecture from Dr Laura L Kiessling 5:30pm today Bellco Theater, Conv Ctr http:…
Amer Chem Society	@AmerChemSociety	UNLGradStudies	Article in the Journal of @AmerChemSociety feature recent CHEM PhDs and Postdoc! @UNLChemistry @NSF @UNLPostdoc #UNL
Amer Chem Society	@AmerChemSociety	ruderalcoop	We highly commend @AmerChemSociety for including a @Cannabis_Chem Committee! http://t.co/z86SyIZ9qg
Amer Chem Society	@AmerChemSociety	234chem	Jun Nishikawa got FIRST place 4 Cinci Section of @AmerChemSociety Ralph E. Oesper 1st Level Chem Exam! Recognized tonight. @SycamoreSchools
Amer Chem Society	@AmerChemSociety	234chem	Josh Pelberg was 1 of 12 local qualifiers 4 the US Nat'l Chem Olympiad test. Recognized tonight at @AmerChemSociety mtg. @SycamoreSchools
Amer Chem Society	@AmerChemSociety	tinypetite99	RT @AmerChemSociety: BSC student receives leadership award by American Chemical Society http://t.co/Z0e6vCad50 @acsundergrad
Amer Chem Society	@AmerChemSociety	JTCCoach	RT @AACTconnect: Thermodynamics & Chemical Engineering lesson: http://t.co/DTC8s09BiE Find more teaching resources from @AmerChemSociety Laâ€¦

nombre	cuenta	usuario	tweet
Amer Chem Society	@AmerChemSociety	E4SmartVillages	Silk could be new ‘green’ material for next-generation batteries - American Chemical Society http://t.co/t5cJF9xfZC via @AmerChemSociety
Amer Chem Society	@AmerChemSociety	OwlerAlerts	Thomas M. Connelly joined American Chemical Society (@AmerChemSociety) as CEO https://t.co/ve8AvvUUUw
Amer Chem Society	@AmerChemSociety	Locanzeco	RT @Brain_Facts_org: One minor chemical change can turn the sweet #smell of a #rose sour http://t.co/BXsFFWZ03F video via @AmerChemSociety …
Amer Chem Society	@AmerChemSociety	jeffjeff2007	RT @NOAA: CO2 record at NOAA observatory named a Nat'l Historic Chemical Landmark by @AmerChemSociety http://t.co/VOPHFqADOD http://t.co/D…
Amer Chem Society	@AmerChemSociety	takavl	RT @NOAA: CO2 record at NOAA observatory named a Nat'l Historic Chemical Landmark by @AmerChemSociety http://t.co/VOPHFqADOD http://t.co/D…
Amer Chem Society	@AmerChemSociety	inherentquality	RT @NOAA: CO2 record at NOAA observatory named a Nat'l Historic Chemical Landmark by @AmerChemSociety http://t.co/VOPHFqADOD http://t.co/D…
Amer Chem Society	@AmerChemSociety	AWinnr	RT @AmerChemSociety: Video: Bugs wage chemical warfare with butts and guts featuring @cenmag #speakingofchem http://t.co/tmwj7IvMJn http://…
Amer Chem Society	@AmerChemSociety	JustinGood	â€œ@AmerChemSociety: The remarkable chemistry of pizza (video) http://t.co/7qaZwtfQu5â€ Hey @emilydorsey -- this might be good for biochem.
Amer Chem Society	@AmerChemSociety	CatalentRnD	RT @AmerChemSociety: RT @ACSpressroom: #Beer compound could help fend off #Alzheimerâ€™s and Parkinsonâ€™s diseases http://t.co/KKYo6HRQrP #cheâ€¦
Amer Chem Society	@AmerChemSociety	OvaLombok	RT @VISTAScience: The chemistry of romance: http://t.co/lmge3rBnz7 @AmerChemSociety #chemistry #vday
Amer Chem Society	@AmerChemSociety	cenmag	RT @ACSERC: In this week's @cenmag: @AmerChemSociety will host an #investor #pitch event for #chemistry #startups in San Diego: http://t.co…
Amer Chem Society	@AmerChemSociety	HarryMacTough	RT @WordChem: Chemistry PhD - worth the effort? I look at the factors involved. inChemistry @AmerChemSociety http://t.co/nRDXzIchpF http://…
Amer Chem Society	@AmerChemSociety	NeilLHeckman	RT @WordChem: Chemistry PhD - worth the effort? I look at the factors involved. inChemistry @AmerChemSociety http://t.co/nRDXzIchpF http://…
Amer Chem Society	@AmerChemSociety	SigmaAldrich	RT @SAGlobalCit_JW: It's coming up #green at @AmerChemSociety w/ the @SigmaAldrich #Green #Chemistry session in #Denver http://t.co/a0Zk8hP…
Amer Chem Society	@AmerChemSociety	melvekrog	@peterhartlaub @burritojustice This awesome front page even refers to the @AmerChemSociety ! #GoGiants #Chemistry
Amer Chem Society	@AmerChemSociety	RissaChem	RT @AmerChemSociety: Millikin chemistry students combine fun, science - Quad-Cities Online: Q-C News http://t.co/GyTt8SCGi7 @ACSundergrad
Amer Chem Society	@AmerChemSociety	OChemJulie	A6: going to a chemistry investor event @ACSERC working with the @AmerChemSociety We have chemistry!! @KatrinaStevens1 #EdTechBridge
Amer Chem Society	@AmerChemSociety	heydebigale	RT @AmerChemSociety: Under 40 hours left to enter #chemchamps. Don't let one of the 8 spots for free #scicomm training pass you by http://t…
Amer Chem Society	@AmerChemSociety	silanoldep	Chemistry!!! is out! http://t.co/15WC92WqC4 Stories via @Chemicalweek @AmerChemSociety
Amer Chem Society	@AmerChemSociety	ruderalcoop	RT @Cannabis_Chem: The Cannabis Chemistry Committee is now OFFICIALLY a part of @AmerChemSociety and @ACSSCHB! Big thanks to everyone that …
Amer Chem Society	@AmerChemSociety	WMlsACS	RT @bhgross144: Today, Liebig's Kaliapparat is familiar to most US chemists thanks to @AmerChemSociety's logo. #histSTM #chemistry http://t…

nombre	cuenta	usuario	tweet
Amer Chem Society	@AmerChemSociety	royal_rebello	RT @stevelevine: Thanks for this review @WordChem: My book review for "The Powerhouse" @AmerChemSociety website http://t.co/JvhjWwmtR8 http…
Amer Chem Society	@AmerChemSociety	silanoldep	Chemistry!!! is out! http://t.co/sVaLLIoglI Stories via @AmerChemSociety @elsevierscience @3DMicroFactory
Amer Chem Society	@AmerChemSociety	NE14NaCl_aq	RT @AmerChemSociety: RT @ACSReactions: Got a chemistry tattoo? Send them our way! We're currently working on a video about tattoos and we w…
Amer Chem Society	@AmerChemSociety	GrandboisMatt	RT @DowChemical: #NEWS: @AmerChemSociety has named several of our scientists among the 2015 “Heroes of Chemistry!" Read more: http://t.co/K…
ChEnected AIChE	@ChEnected	Niallmacdowell	RT @EnergyTechnol: An excellent article for chemical engineers working in energy research http://t.co/JW1fDSdGEK @DowChemical @ChEnected
ChEnected AIChE	@ChEnected	ChEnected	RT @CharlieSchwedle: Mid Michigan Section of the American Institute of Chemical Engineers @ChEnected #bullockcreek #STEM http://t.co/zt7ebT…
ChEnected AIChE	@ChEnected	flprin	RT @CharlieSchwedle: Mid Michigan Section of the American Institute of Chemical Engineers @ChEnected #bullockcreek #STEM http://t.co/zt7ebT…
ChemHeritage	@ChemHeritage	NextCenturySean	RT @DrRubidium: The cool stuff you find around the ol' chem department! cc @ChemHeritage http://t.co/yIRhr8gE1W
ChemHeritage	@ChemHeritage	future_ish	RT @DrRubidium: The cool stuff you find around the ol' chem department! cc @ChemHeritage http://t.co/yIRhr8gE1W
ChemHeritage	@ChemHeritage	bhgross144	On the 50th anniversary of #MooresLaw learn how its namesake got his scientific start! (H/T @ChemHeritage) http://t.co/27Dza15KMv #histSTM
ChemHeritage	@ChemHeritage	MuseumsCouncil	RT @ChemHeritage: Celebrating the 50th anniversary of #MooresLaw with the story of Moore's scientific start: a chemistry set! #HistSTM http…
ChemHeritage	@ChemHeritage	FryeDwight	RT @ChemHeritage: Celebrating the 50th anniversary of #MooresLaw with the story of Moore's scientific start: a chemistry set! #HistSTM http…
ChemHeritage	@ChemHeritage	LauraEVentura	RT @ChemHeritage: Celebrating love + chemistry, and #MarriageEquality! #LoveWins http://t.co/aaQdV3fxO0
ChemHeritage	@ChemHeritage	RKPriestley	RT @ChemHeritage: Celebrating love + chemistry, and #MarriageEquality! #LoveWins http://t.co/aaQdV3fxO0
ECS	@ECSorg	NatureEnergyJnl	Just completed registration for @ECSorg meeting and very excited to get my 1st set of #chemistry trading cards! http://t.co/OXd75s5fUi
IChemE	@IChemE	HUCES_UOH	RT @IChemE: IChemE course - Mentoring for Chemical Engineers, 18 March 2015 - Book now: http://t.co/5aaNOs64gq
IChemE	@IChemE	HUCES_UOH	RT @IChemE: IChemE Course - Creativity for Chemical Engineers, 17 March 2015 - Book now: http://t.co/2zZoi3FuMb
IChemE	@IChemE	wilianto1	RT @IChemE: For all chemical engineers in industry & aged under 30 - could you be an IChemE award winner? http://t.co/UfwQjlJ3S6 http://t.c…
IUPAC	@IUPAC	khchem	Got a kid who likes #science and cartoons? @IUPAC Phys Chem Cartoon Competition. $100 to 4 winners. http://t.co/o8U3fF4huL #STEMedu
IUPAC	@IUPAC	khchem	@IUPAC 2015 Distinguished Women in #Chemistry or Chemical Engineering -- Call for Nominations. 2/15/15 deadline. http://t.co/AdHAX50CnG
IUPAC	@IUPAC	khchem	@IUPAC 2015 Distinguished Women in Chemistry or Chemical Engineering -- Call for Nominations. 2/15/15 deadline. http://t.co/AdHAX50CnG
IUPAC	@IUPAC	CSIROPublishing	RT @DrMaggieHardy: Nominations for the @IUPAC Distinguished Women in #Chemistry or #ChemicalEngineering #Award close 15 February: http://t.â€¦

nombre	cuenta	usuario	tweet
IUPAC	@IUPAC	howitt_julia	RT @DrMaggieHardy: Nominations for the @IUPAC Distinguished Women in #Chemistry or #ChemicalEngineering #Award close 15 February: http://t.â€¦
IUPAC	@IUPAC	larkened	RT @DrMaggieHardy: Nominations for the @IUPAC Distinguished Women in #Chemistry or #ChemicalEngineering #Award close 15 February: http://t.â€¦
IUPAC	@IUPAC	khchem	RT @drkeegansawyer: Apply Now. US Nat'l Cmte Young Observers Program. Attend 2015 @IUPAC #Chemistry Congress in South Korea @NASciences htâ€¦
IUPAC	@IUPAC	khchem	@IUPAC 2015 #Chemistry Congress travel awards for Busan, South Korea. Open to early-career chemists around the world. http://t.co/dD1raQH6X1
IUPAC	@IUPAC	DrMaggieHardy	Nominations for the @IUPAC Distinguished Women in #Chemistry or #ChemicalEngineering #Award close 15 February: http://t.co/JDVmxnlBal
IUPAC	@IUPAC	facrespoalvarez	RT @NatureNews: Contrary to #chemistry lore the @IUPAC hasn't ruled on what elements belong in that little gap http://t.co/WTk3MK1dx8 http:…
Royal Soc. Chemistry	@RoySocChem	NUITResearch	Congrats Professor Marks for awards from @RoySocChem & the Italian Chemical Society: http://t.co/Asqv1TLDBd. @WeinbergCollege @NU_McCormick
Royal Soc. Chemistry	@RoySocChem	ccddublin	RT @ucddublin: UCD student wins @RoySocChem Royal Society of Chemistry online video competition http://t.co/uhdFY0jj0W http://t.co/9abcYnjt…
Royal Soc. Chemistry	@RoySocChem	LucyFaithEvans	RT @jayeshnavin: Great summary of @RoySocChem study on atts to chemistry by @chiara_ceci: http://t.co/XD0wlEwk8U. Look fwd to digesting ful…
Royal Soc. Chemistry	@RoySocChem	KarenAvocado	RT @chiara_ceci: .@RoySocChem celebrates diversity in science with 175 faces of chemistry http://t.co/lxlhR8H3es #SciComm15
Royal Soc. Chemistry	@RoySocChem	gardav05	RT @BHSSciencegeeks: Thinking of chemistry for A-level or at university? Sign up to ChemNet from The @RoySocChem here http://t.co/rmvaPHbGLP
Cefic	@Cefic	KevinFaircloth1	RT @Cefic: #Water as the strangest chemical explained by @guardian video: http://t.co/eWSIFJYgQU #CeficWatermatters http://t.co/R7JW5kURap
Chevron Phillips	@chevronphillips	AtlOBrien	RT @chevronphillips: CPChem CEO Pete Cella to speak tomorrow @ The Woodlands Economic Outlook Conf about closing the labor shortage gap in â€¦
Eastman Chemical Co.	@EastmanChemCo	CircularEco	RT @SustainBrands: .@EastmanChemCo 1st U.S. member of new group that is supporting sustainable #supplychains in the global chem industry ht…
Syngenta	@Syngenta	JSTR_1	@Blackfang108 @TakeThatGMOs @AgaroSegel Screw @MonsantoCo @Syngenta @DowChemical @DuPont_News & any other Chem Co trying to "feed world!"
Syngenta	@Syngenta	NCLK23	@Syngenta organomineral and organic NPK fertilizer must be used instead of chemical NPK
Syngenta	@Syngenta	TakeThatGMOs	RT @JSTR_1: @MonsantoCo @GMOAnswers @TakeThatGMOs @Syngenta @DuPont_News Chemical companies need to stop making GMOs & lying about their sâ€¦
Syngenta	@Syngenta	JSTR_1	@MonsantoCo @GMOAnswers @TakeThatGMOs @Syngenta @DuPont_News Chemical companies need to stop making GMOs & lying about their safety â‹…world
Syngenta	@Syngenta	cfoxthomas	RT @OrionGrassroots: Hawaiian activists confront @Syngenta on its Swiss home turf over pesticide overuse in Hawaii http://t.co/UHDwYgLA8o v…

nombre	cuenta	usuario	tweet
Syngenta	@Syngenta	sepia3C	RT @OrionGrassroots: Hawaiian activists confront @Syngenta on its Swiss home turf over pesticide overuse in Hawaii http://t.co/UHDwYgLA8o v…
Syngenta	@Syngenta	RTCreatresstv	RT @OrionGrassroots: Hawaiian activists confront @Syngenta on its Swiss home turf over pesticide overuse in Hawaii http://t.co/UHDwYgLA8o v…
Syngenta	@Syngenta	carm2158	RT @OrionGrassroots: Hawaiian activists confront @Syngenta on its Swiss home turf over pesticide overuse in Hawaii http://t.co/UHDwYgLA8o v…
Yara International	@yara	abeeremad__	@yara_bader2 hahahahahah<ed><U+00A0><U+00BD><ed><U+00B8><U+0082><ed><U+00A0><U+00BD><ed><U+00B8><U+0082><ed><U+00A0><U+00BD><ed><U+00B8><U+0082><ed><U+00A0><U+00BD><ed><U+00B8><U+0082>goodluckk 7ub<U+2764><U+FE0F><U+2764><U+FE0F>2al chem 2al dont even remind me<ed><U+00A0><U+00BD><ed><U+00B8><U+0082><ed><U+00A0><U+00BD><ed><U+00B8><U+0082>
Yara International	@yara	Ghfran_Sy	RT @yaraGHkhoury: the chemistry between the judges<3 #MBCTheXFactor http://t.co/ItSm1sG5oV
Yara International	@yara	RanimHossam	@yaraehabsoliman la enty hatshofeny b3d el chemistry <ed><U+00A0><U+00BD><ed><U+00B8><U+0082><ed><U+00A0><U+00BD><ed><U+00B8><U+0082><ed><U+00A0><U+00BD><ed><U+00B2><U+0083><ed><U+00A0><U+00BD><ed><U+00B2><U+0083>
Evonik	@Evonik	larice_chamber	RT @EvonikCareersNA: Chem. Engineers & Scientists, key parts to our current & future success! Exploring opportunities. Growing together. ht…
ExxonMobil	@exxonmobil	DR___White	RT @exxonmobil: Introduction to Chemical Engineering | Lecture 1: http://t.co/zPKj0Likw6 via @YouTube @Stanford
ExxonMobil	@exxonmobil	CGLRGreatlakes	RT @exxonmobil: #Innovation in the chemical industry means benefits to the envt and sustainability. Learn how we're involved http://t.co/7T…
ExxonMobil	@exxonmobil	Fortune500_Inc	RT @exxonmobil: #Innovation in the chemical industry means benefits to the envt and sustainability. Learn how we're involved http://t.co/7T…
ExxonMobil	@exxonmobil	GTMGQ	RT @exxonmobil: #Innovation in the chemical industry means benefits to the envt and sustainability. Learn how we're involved http://t.co/7T…
BASF North America	@BASFCorporation	LinkWorxSeo	RT @BASFCorporation: See how better #farming techniques are creating prosperous communities: http://t.co/ntAaNj1ScN http://t.co/J4Mce75uC1
BASF North America	@BASFCorporation	ChainGangGirl88	RT @BASFCorporation: See how BASF is creating chemistry to tackle the challenges of sustainable food: http://t.co/yxIb7no9bw #SCS_CH http:/…
Dow	@DowChemical	silanoldep	Chemistry!!! is out! http://t.co/b1Caj3UtEf Stories via @DowChemical @Chem_Info @quimica_es
Dow	@DowChemical	JSTR_1	@Blackfang108 @TakeThatGMOs @AgaroSegel Screw @MonsantoCo @Syngenta @DowChemical @DuPont_News & any other Chem Co trying to "feed world!"
Dow	@DowChemical	CChrisafides	RT @AZPacPlastics: Our favorite Tweeter of the Day goes to DOW Chem @DowChemicalPU Big Thanks in every language! AZ Pacific Plastics http:…
Dow	@DowChemical	SmartOrangeOwl	New Company Profile Available: @DowChemical. Access the profile here: http://t.co/oG6uCivgu5
Dow	@DowChemical	Niallmacdowell	RT @EnergyTechnol: An excellent article for chemical engineers working in energy research http://t.co/JW1fDSdGEK @DowChemical @ChEnected
Dow	@DowChemical	NWAnalytics	RT @PlantServices: .@DowChemical sees big savings from #bigdata thanks to a global analytics tool that helps make sense of it all. http://t…
Dow	@DowChemical	uqalumni	RT @UQNorthAmerica: #UQ alumnus Dr Andrew Liveris, CEO of @DowChemical, will speak @michiganstateu commencement next month http://t.co/UHGi…

nombre	cuenta	usuario	tweet
Dow	@DowChemical	dharmaLin	Tell @DowChemical to shelve its newest chemical cocktail #EnlistDuo & help save #monarchs from collapse. Sign now: http://t.co/tIxh8Tvf2I
Dow	@DowChemical	amerlekelly	RT @RecklessDreamN: Tell @DowChemical to shelve its newest chemical cocktail #EnlistDuo & help save #monarchs from collapse. Sign now: http…
Dow	@DowChemical	ConsiderThis1	RT @RecklessDreamN: Tell @DowChemical to shelve its newest chemical cocktail #EnlistDuo & help save #monarchs from collapse. Sign now: http…
Dow	@DowChemical	Greeenguy111	RT @RecklessDreamN: Tell @DowChemical to shelve its newest chemical cocktail #EnlistDuo & help save #monarchs from collapse. Sign now: http…
Dow	@DowChemical	ShisoSocial	Do <3 love butterflies? Tell @DowChemical to shelve its newest chemical cocktail #EnlistDuo! TAKE ACTION: http://t.co/FIRwn13Z2C #wildlife
Dow	@DowChemical	ruthmen	RT @all4feet: Tell @DowChemical to shelve its newest chemical cocktail #EnlistDuo & help save #monarchs from collapse. Sign now: http://t.c…
Dow	@DowChemical	keepmving	RT @cscmp: Using #IT as a competitive weapon: @DowChemical, @PepsiCo and the #IoT http://t.co/VTFirqFCnJ (via @logisticsviewpt)
Dow	@DowChemical	WeylandR	RT @DowChemical: #DYK: We were the first chemical company to achieve the @HRC 100% for #LGBT inclusion in the workplace. Now 10 years & cou…
Dow	@DowChemical	mindaykay	RT @DowChemical: #DYK: We were the first chemical company to achieve the @HRC 100% for #LGBT inclusion in the workplace. Now 10 years & cou…
Dow	@DowChemical	Antonpjm	RT @UQNorthAmerica: Congrats to @uqalumni Dr Andrew Liveris, CEO of @DowChemical, on winning the ICIS Kavaler Award http://t.co/BZuQi487aG …
Dow	@DowChemical	uqalumni	RT @UQNorthAmerica: Congrats to @uqalumni Dr Andrew Liveris, CEO of @DowChemical, on winning the ICIS Kavaler Award http://t.co/BZuQi487aG …
Dow	@DowChemical	silanoldep	Chemistry!!! is out! http://t.co/zyXNC1Ucx4 Stories via @Chemicalweek @DowChemical @nchazarra
Dow	@DowChemical	GrandboisMatt	RT @DowChemical: #NEWS: @AmerChemSociety has named several of our scientists among the 2015 “Heroes of Chemistry!" Read more: http://t.co/K…
DuPont News	@DuPont_News	JSTR_1	@Blackfang108 @TakeThatGMOs @AgaroSegel Screw @MonsantoCo @Syngenta @DowChemical @DuPont_News & any other Chem Co trying to "feed world!"
DuPont News	@DuPont_News	TakeThatGMOs	RT @JSTR_1: @MonsantoCo @GMOAnswers @TakeThatGMOs @Syngenta @DuPont_News Chemical companies need to stop making GMOs & lying about their sâ€¦
DuPont News	@DuPont_News	JSTR_1	@MonsantoCo @GMOAnswers @TakeThatGMOs @Syngenta @DuPont_News Chemical companies need to stop making GMOs & lying about their safety â‹…world
DuPont News	@DuPont_News	Marlore2	RT @Eng_Materials: Safety, efficiency and reliability for chemical industry https://t.co/xjDh9CCkL7 @DuPont_News @ACHEMAworldwide http://t.…
BASF	@BASF	GrainTradeAus	#NWPGP Thanks #Aust #Grain Storage & Protection Conf Silver Sponsors @BASF_Agro_Au FABTECH, Kotzur @NatRoadTranAsn & Sumitomo Chemical
BASF	@BASF	MrTox2000	RT @BASFPerforMatls: #Chemistry in the Great Outdoors|YouTube @accpolyurethane http://t.co/ifqeY0sBSU
BASF	@BASF	LinkWorxSeo	RT @BASFCorporation: See how better #farming techniques are creating prosperous communities: http://t.co/ntAaNj1ScN http://t.co/J4Mce75uC1

nombre	cuenta	usuario	tweet
BASF	@BASF	pope_john3	RT @BASF: Discover Hövding, airbag for cyclists that protects against head injuries #CreatingChemistry. http://t.co/VqYw3JUOD2 http://t.co/…
BASF	@BASF	sethunairNY	An #innovator in the field of #Chemistry @BASF | My sister Jaya Mohanan on her inspiring life and work! >> http://t.co/4wBT0pAJtQ #STEMWomen
BASF	@BASF	ChainGangGirl88	RT @BASFCorporation: See how BASF is creating chemistry to tackle the challenges of sustainable food: http://t.co/yxIb7no9bw #SCS_CH http:/…

Anexo 15. Número de menciones de usuarios a empresas y sociedades.

Número de *tweets* en las que empresas o sociedades están mencionadas (columnas) en función de los usuarios que mencionan (filas):

(1) = C&EN; (2) = Amer Chem Society; (3) = American Chemistry; (4) = Chemistry World; (5) = Dow; (6) = IUPAC; (7) = Syngenta; (8) = ChemHeritage; (9) = BASF; (10) = Royal Soc. Chemistry; (11) = ExxonMobil; (12) = DuPont News; (13) = ACS Green Chemistry; (14) = ChEnected AIChE; (15) = IChemE; (16) = Yara International; (17) = RACI; (18) = BASF North America; (19) = SCI; (20) = Evonik; (21) = Cefic; (22) = Chevron Phillips; (23) = Eastman Chemical Co.; (24) = ECS;

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total general
JSTR_1					1		2					2													5
silanoldep		2	1		2																				5
khchem						5																			5
Niallmacdowell					1									1											2
TakeThatGMOs							1					1													2
cenmag	1	1																							2
GrandboisMatt		1			1																				2
LinkWorxSeo									1									1							2
HUCES_UOH															2										2
ruderalcoop		2																							2
AWinnr	1	1																							2
susanjainsworth	2																								2
uqalumni					2																				2
ChainGangGirl88									1									1							2
Ktilger_Chem	2																								2
234chem		2																							2
RissaChem		1																							1
MellaNofri	1																								1
vancew	1																								1
abeeremad__																1									1
NeilLHeckman		1																							1
baldeilys			1																						1
sime0n			1																						1
bhgross144								1																	1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total general
Locanzeco		1																							1
bizcochopat			1																						1
MrTox2000									1																1
borderbend				1																					1
oundle_chem				1																					1
Carbon3D	1																								1
Scarlet37236746			1																						1
carm2158							1																		1
TaylorVenture7			1																						1
CatalentRnD		1																							1
AtomicUniverse	1																								1
Catalysis_IMCN	1																								1
MagloVillalba			1																						1
ccddublin										1															1
MoleculeWorld	1																								1
CChrisafides					1																				1
nanocastsafety	1																								1
acsdchas	1																								1
NUITResearch										1															1
cfoxthomas							1																		1
pope_john3									1																1
CGLRGreatlakes											1														1
RSC_HSG	1																								1
ACSDivCHED	1																								1
sherfranklin				1																					1
Chem_Consult	1																								1
snapdannyboy				1																					1
Chemkishan	1																								1
tralfamadorable	1																								1
chemparrot				1																					1
whatsbobgonnado				1																					1
ChEnected														1											1
liam_d_odonnell	1																								1
chitinette																			1						1
m39618462	1																								1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total general
Chm107Downtown	1																								1
Marlore2												1													1
CircularEco																							1		1
mindaykay					1																				1
Clare_Sansom				1																					1
mrizzo64			1																						1
CoC_guy				1																					1
MunkhturTMT			1																						1
ConnorChapek	1																								1
NCLK23							1																		1
ConsiderThis1					1																				1
amerlekelly					1																				1
CSIROPublishing						1																			1
OChemJulie		1																							1
davidhth			1																						1
OwlerAlerts		1																							1
dbfulton				1																					1
RanimHossam																1									1
dewi_darmawati			1																						1
RKPriestley								1																	1
dharmaLin					1																				1
AndyNguyenTC		1																							1
dpchalasani	1																								1
sepia3C							1																		1
DR___White											1														1
SigmaAldrich		1																							1
DrMaggieHardy						1																			1
SmartOrangeOwl					1																				1
drskharlamov			1																						1
takavl		1																							1
E4SmartVillages		1																							1
thebiolady				1																					1
EllenMellon_88				1																					1
UNLGradStudies		1																							1
ericscerri				1																					1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total general
Westaeus				1																					1
ErikaOltermann	1																								1
wmgeil	1																								1
facrespoalvarez						1																			1
laurenkwolf	1																								1
Fadders10				1																					1
AmerChemSociety													1												1
fcoroac			1																						1
LucyFaithEvans										1															1
featherlou			1																						1
Magallanes35			1																						1
flprin														1											1
markmackllc	1																								1
Fortune500_Inc											1														1
meganlatshaw	1																								1
FryeDwight								1																	1
melvekrog		1																							1
Furness_Ent				1																					1
MistahKindah	1																								1
future_ish								1																	1
morrisresearch				1																					1
gardav05										1															1
mrnavas	1																								1
GBoissonnat				1																					1
MSScienceBlog	1																								1
Ghfran_Sy																1									1
MuseumsCouncil								1																	1
glitter_hole				1																					1
NatureEnergyJnl																								1	1
GrainTradeAus									1																1
NE14NaCl_aq		1																							1
ACSPublications		1																							1
NextCenturySean								1																	1
Greeenguy111					1																				1
nratrh			1																						1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total general
GTMGQ											1														1
NWAnalytics					1																				1
HarryMacTough		1																							1
OffGridOnGrid			1																						1
heydebigale		1																							1
OvaLombok		1																							1
hony_2014	1																								1
polymerchemical			1																						1
howitt_julia						1																			1
RACI_HQ																	1								1
AidanBaker				1																					1
Rdday91191			1																						1
IBchemmilam	1																								1
Ritabril	1																								1
ideawomen			1																						1
royal_rebello		1																							1
inherentquality		1																							1
RTCreatresstv							1																		1
IsraelFonseca11			1																						1
ruthmen					1																				1
ItsScienceMagic	1																								1
Sciguy999	1																								1
jeffjeff2007		1																							1
sethunairNY									1																1
JEPG24				1																					1
ShisoSocial					1																				1
josharellano	1																								1
Antonpjm					1																				1
aleradar				1																					1
skoolycool			1																						1
JTCCoach		1																							1
smokva200			1																						1
JustinGood		1																							1
arndt_eric	1																								1
jv3sund	1																								1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total general
arri_aus																	1								1
KarenAvocado										1															1
The_Frasian				1																					1
zouxingjian			1																						1
tinypetite99		1																							1
KevinFaircloth1																					1				1
Tzublal				1																					1
AlexaSkia			1																						1
AtlOBrien																						1			1
Kidzroom													1												1
WendinhGarCalde	1																								1
ktann83													1												1
WeylandR					1																				1
alexcunn	1																								1
wilianto1															1										1
larice_chamber																				1					1
WMlsACS		1																							1
larkened						1																			1
LauraEVentura								1																	1
keepmving					1																				1
Total general	42	33	25	23	20	10	8	7	6	5	4	4	3	3	3	3	2	2	1	1	1	1	1	1	209

Lista de empresas o sociedades y número de menciones:

Mencionado	Número de menciones
C&EN	42
Amer Chem Society	33
American Chemistry	25
Chemistry World	23
Dow	20
IUPAC	10
Syngenta	8

Mencionado	Número de menciones
ChemHeritage	7
BASF	6
Royal Soc. Chemistry	5
DuPont News	4
ExxonMobil	4
ACS Green Chemistry	3
ChEnected AIChE	3
IChemE	3
Yara International	3
BASF North America	2
RACI	2
Cefic	1
Chevron Phillips	1
Eastman Chemical Co.	1
ECS	1
Evonik	1
SCI	1

Lista de usuarios y número de *tweets* en los que mencionan a empresas o sociedades:

Usuario	Número de menciones
JSTR_1	5
khchem	5
silanoldep	5
234chem	2
AWinnr	2
cenmag	2
ChainGangGirl88	2
GrandboisMatt	2
HUCES_UOH	2
Ktilger_Chem	2
LinkWorxSeo	2
Niallmacdowell	2

Usuario	Número de menciones
ruderalcoop	2
susanjainsworth	2
TakeThatGMOs	2
uqalumni	2
abeeremad__	1
acsdchas	1
ACSDivCHED	1
ACSPublications	1
AidanBaker	1
aleradar	1
AlexaSkia	1
alexcunn	1
AmerChemSociety	1
amerlekelly	1
AndyNguyenTC	1
Antonpjm	1
arndt_eric	1
arri_aus	1
AtlOBrien	1
AtomicUniverse	1
baldeilys	1
bhgross144	1
bizcochopat	1
borderbend	1
Carbon3D	1
carm2158	1
CatalentRnD	1
Catalysis_IMCN	1
ccddublin	1
CChrisafides	1
cfoxthomas	1
CGLRGreatlakes	1
Chem_Consult	1
Chemkishan	1
chemparrot	1

Usuario	Número de menciones
ChEnected	1
chitinette	1
Chm107Downtown	1
CircularEco	1
Clare_Sansom	1
CoC_guy	1
ConnorChapek	1
ConsiderThis1	1
CSIROPublishing	1
davidhth	1
dbfulton	1
dewi_darmawati	1
dharmaLin	1
dpchalasani	1
DR___White	1
DrMaggieHardy	1
drskharlamov	1
E4SmartVillages	1
EllenMellon_88	1
ericscerri	1
ErikaOltermann	1
facrespoalvarez	1
Fadders10	1
fcoroac	1
featherlou	1
flprin	1
Fortune500_Inc	1
FryeDwight	1
Furness_Ent	1
future_ish	1
gardav05	1
GBoissonnat	1
Ghfran_Sy	1
glitter_hole	1
GrainTradeAus	1

Usuario	Número de menciones
Greeenguy111	1
GTMGQ	1
HarryMacTough	1
heydebigale	1
hony_2014	1
howitt_julia	1
IBchemmilam	1
ideawomen	1
inherentquality	1
IsraelFonseca11	1
ItsScienceMagic	1
jeffjeff2007	1
JEPG24	1
josharellano	1
JTCCoach	1
JustinGood	1
jv3sund	1
KarenAvocado	1
keepmving	1
KevinFaircloth1	1
Kidzroom	1
ktann83	1
larice_chamber	1
larkened	1
LauraEVentura	1
laurenkwolf	1
liam_d_odonnell	1
Locanzeco	1
LucyFaithEvans	1
m39618462	1
Magallanes35	1
MagloVillalba	1
markmackllc	1
Marlore2	1
meganlatshaw	1

Usuario	Número de menciones
MellaNofri	1
melvekrog	1
mindaykay	1
MistahKindah	1
MoleculeWorld	1
morrisresearch	1
mrizzo64	1
mrnavas	1
MrTox2000	1
MSScienceBlog	1
MunkhturTMT	1
MuseumsCouncil	1
nanocastsafety	1
NatureEnergyJnl	1
NCLK23	1
NE14NaCl_aq	1
NeilLHeckman	1
NextCenturySean	1
nratrh	1
NUITResearch	1
NWAnalytics	1
OChemJulie	1
OffGridOnGrid	1
oundle_chem	1
OvaLombok	1
OwlerAlerts	1
polymerchemical	1
pope_john3	1
RACI_HQ	1
RanimHossam	1
Rdday91191	1
RissaChem	1
Ritabril	1
RKPriestley	1
royal_rebello	1

Usuario	Número de menciones
RSC_HSG	1
RTCreatresstv	1
ruthmen	1
Scarlet37236746	1
Sciguy999	1
sepia3C	1
sethunairNY	1
sherfranklin	1
ShisoSocial	1
SigmaAldrich	1
sime0n	1
skoolycool	1
SmartOrangeOwl	1
smokva200	1
snapdannyboy	1
takavl	1
TaylorVenture7	1
The_Frasian	1
thebiolady	1
tinypetite99	1
tralfamadorable	1
Tzublal	1
UNLGradStudies	1
vancew	1
WendinhGarCalde	1
Westaeus	1
WeylandR	1
whatsbobgonnado	1
wilianto1	1
wmgeil	1
WMlsACS	1
zouxingjian	1

Matriz de número de *tweets* de usuarios que mencionan a empresas o sociedades y empresas o sociedades mencionadas:

Nombre de usuario	Nombre de empresa o sociedad																							Total general
	Total	BASF	C&EN	Amer Chem Society	Syngenta	IUPAC	American Chemistry	Chemistry World	SABIC I ثابساس	Merck	ACS Green Chemistry	Royal Soc. Chemistry	ChemHeritage	IUCr	Ecolab	AkzoNobel	ECS	Cefic	Henkel	Dow	Yara International	DSM Company	DuPont News	
maka_shizyuzi	4																							4
acsdchas			4																					4
khchem						3																		3
sethunairNY		3																						3
234chem				2																				2
susanjainsworth			2																					2
AppCrier		2																						2
JSTR_1					1																		1	2
skoolycool							1																	1
melvekrog				1																				1
laurenkwolf			1																					1
angew_chem	1																							1
OwlerAlerts				1																				1
AceNewsServices					1																			1
tralfamadorable			1																					1
bhgross144													1											1
AlbqBonita					1																			1
BrowardSTEM				1																				1
NCLK23					1																			1
BuyBookstore	1																							1

Nombre de usuario	Nombre de empresa o sociedad																							Total general
	Total	BASF	C&EN	Amer Chem Society	Syngenta	IUPAC	American Chemistry	Chemistry World	SABIC I ثيباس	Merck	ACS Green Chemistry	Royal Soc. Chemistry	ChemHeritage	IUCr	Ecolab	AkzoNobel	ECS	Cefic	Henkel	Dow	Yara International	DSM Company	DuPont News	
Sciguy999			1																					1
ChemDraw				1																				1
AmySkalicky	1																							1
ChemistSays			1																					1
werkvacature									1															1
chemoutlook																		1						1
levoneW1		1																						1
Clare_Sansom								1																1
maurbelle								1																1
ClassTrips				1																				1
MyHarmReduction																						1		1
Cleggan1	1																							1
NUITResearch												1												1
CoatingIndustry		1																						1
rnomics														1										1
deirdrelockwood			1																					1
ShisoSocial																				1				1
denise_geary	1																							1
stuartfox21	1																							1
dhara_chv	1																							1
Suzannehaley	1																							1
dpchalasani			1																					1
vacaturejobs									1															1
DrGvanK										1														1

Nombre de usuario	Nombre de empresa o sociedad																								Total general
	Total	BASF	C&EN	Amer Chem Society	Syngenta	IUPAC	American Chemistry	Chemistry World	SABIC I ‌ثيباس	Merck	ACS Green Chemistry	Royal Soc. Chemistry	ChemHeritage	IUCr	Ecolab	AkzoNobel	ECS	Cefic	Henkel	Dow	Yara International	DSM Company	DuPont News	
AndyNguyenTC				1																				1
DrMaggieHardy						1																		1
left4bread_						1																		1
DuttonAlejandro	1																							1
loveacupotea										1														1
E4SmartVillages				1																				1
markmackllc			1																					1
edinashu	1																							1
mauritaniafrica				1																				1
ericscerri								1																1
mkeenan23	1																							1
exp_RD_FE																			1					1
NatureEnergyJnl																	1							1
FangirlForAri	1																							1
nigerian_herald																1								1
FawcettSharon					1																			1
OChemJulie				1																				1
featherlou							1																	1
rajaaelbiyad1		1																						1
ForEffectiveGov					1																			1
ruderalcoop				1																				1
freedomforthwin					1																			1
Ames_Jobs_USA		1																						1
GerardoRich		1																						1

Nombre de usuario	Nombre de empresa o sociedad																							Total general
	Total	BASF	C&EN	Amer Chem Society	Syngenta	IUPAC	American Chemistry	Chemistry World	SABIC I ثيباس	Merck	ACS Green Chemistry	Royal Soc. Chemistry	ChemHeritage	IUCr	Ecolab	AkzoNobel	ECS	Cefic	Henkel	Dow	Yara International	DSM Company	DuPont News	
sime0n							1																	1
jamespositive		1																						1
socialentrep_dp		1																						1
JawaadZaheer																					1			1
Studio4News					1																			1
joshsilberg	1																							1
SustainOurEarth		1																						1
JR_son															1									1
TaigaCompany		1																						1
zelaiikha						1																		1
UNLGradStudies				1																				1
123PR113	1																							1
werk_vacature									1															1
Kidzroom											1													1
WorldUSNews					1																			1
KirstenEdgar						1																		1
ktann83											1													1
Total general	**18**	**14**	**13**	**13**	**9**	**7**	**3**	**3**	**3**	**2**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**99**

Anexo 16. Extracto de resultados de tablas de menciones

Todos los *tweets y retweets*

Mencionado	**Identificador de *Tweet***	**Nombre en pantalla de usuario en Twitter**
@MCNocando	550803520564121600	_CHEM_TRAIL_
@rtyourcrushx	550803344131100672	chem_98
@RelatableQuote	550802273555988481	chem_98
@Nero	550801899780177920	phyto_chem
@chantarose	550801523996127233	rltor_prn
@A_LENSS	550801506270994432	ZeCrepz
@ZeCrepz	550801506270994432	ZeCrepz
@CuteDecorations	550801506270994432	ZeCrepz
@ReallyHighIdeas	550801506270994432	ZeCrepz
@theage	550801268130603008	phyto_chem
@b0yp0wer	550801268130603008	phyto_chem
@junkphilosophe	550800964077113344	dylb010
@dylb010	550800964077113344	dylb010
@ddlovato	550800692940922880	chem_98
@Tactilepoet	550800423381393408	lehimesa

Anexo 17. Descripción y organización de la documentación electrónica de la tesis

El código R, los ficheros de datos y los documentos de clasificación de los *wordclouds* por los expertos pueden consultarse en en la dirección web https://github.com/mguerris/Tesis-Doctoral.git dentro de las carpetas “R”, “Datos” y “EvaluacionWordclouds” respectivamente.

Código R desarrollado

Nombre	Descripción	Referencias a apartados de este documento	Referencias a tablas y figuras de este documento
01_Estadisticas_tweets_brutos.R	Búsqueda del número de tweets brutos con *hashtags*, menciones, urls, *retweets*, emoticonos, *emojis*	4.2, 5.1	Tabla 5.1.2
02_Estadisticas_Tweets_Brutos_y_Netos.R	Búsqueda del número de *tweets* brutos y netos que contienen las palabras clave de búsqueda	4.2, 5.1	Tabla 5.1.1, Tabla 5.1.4
03_Tratamiento_Tweets_Tdm.R	Procesos de limpieza de textos de los *tweets* y obtención de la matriz TDM de bigramas	4.2, 5.1	
04_Calculo_dtm.R	Selección de bigramas con una frecuencia superior a 29, cálculo del factor Tf-Idf normalizado y obtención de la matriz traspuesta de la matriz TDM	4.2,4.3, 5.1	
05_PCalculo_Elbow.R	Implementación en paralelo del método *spherical k-means* de K=2 a K=285 clústeres 50 veces para cada K	4.3, 5.1	
06_Selec_Stats_Elbow.R	Selección de la mejor repetición en cada K correspondiente al menor valor de la función criterio Q	4.3, 5.1	
07_Grafica_Elbow_para_tesis.R	Confección de las gráficas de la función criterio Q en función de K, y del coeficiente de silueta en función de K, cálculo de los valores del algoritmo *L-method* y coeficiente de silueta, confección de las gráficas de los valores del algoritmo *L-method* y los valores del módulo de curvatura en función de K para las funciones criterio Q y para el coeficiente de silueta	4.3, 5.1	Figura 5.2, Figura 5.3, Figura 5.4
08_PCalculs_skmeans_100cluster_Paralel_Serie.R	Cálculo en paralelo y en serie del método *spherical k-means* con K*=100 clústeres 9723 veces	4.3, 5.1	
09_Cerca_plot_ analisis_millor_resultat_100clusters.R	Búsqueda de la mejor repetición, con un menor valor, de la función criterio Q, confección de la gráfica y boxplot de los valores de la función criterio Q para K*=100 clústeres en función del número de repetición, confección de la estadística descriptiva y boxplot del número de tweets por clúster de la mejor solución obtenida por el método spherical k-means	4.3, 5.1	Figura 5.5, Figura 5.6
10_Calculo_BIBD.R	Generación del diseño en bloques incompletos y balanceados, chequeo de cumplimiento de las condiciones de cada BIBD, confección de la distribución de clústeres BIBD asignados por experto y asignacion del orden de visualización del número de clúster a cada uno de los expertos de forma aleatoria	4.5, 5.1	
11_Extraccion Graficos.R	Confección de los documentos para cada uno de los 18 expertos según la	5.1	

Nombre	Descripción	Referencias a apartados de este documento	Referencias a tablas y figuras de este documento
	asignación de clústeres por experto y el orden de visualización del número de clúster en el documento		
12_Calculo_Estadisticas_Categorizacion_Kappa.R	Confección de los resultados de la clasificación porcentual de clústers y *tweets* de todos los experos en función de la temática asignada y resultados del estadístico kappa de Fleiss para cada temática y para el experimento global	4.5, 5.1	Tabla 5.1.9, Tabla 5.1.11
13_Calculo_lexicon_from_SentiWordNet30.R	Confección del lexicón utilizado en el análisis de sentimientos Sentiwor_mod.Rdata a partir del lexicón SentiWordNet 3.0, cáclulo del número y porcentaje de palabras en función de su polaridad, estadística descriptiva de las palabras con polaridad positiva o negativa, histograma de la distribución de las palabras con polaridad positiva y negativa	4.5, 5.2	Figura 5.9, Tabla 5.2.1, Tabla 5.2.2
14_Analisis_Sentimientos_Tweets_Seleccionados.R	Detección de los unigramas de los *tweets* de las temáticas AH y EE en el lexicón utilizado, cálculo de la polaridad de los *tweets* de las temáticas AH y EE, cálculo del número de *tweets* de AH y EE según su polaridad, distribución de la polaridad de los *tweets* de las temáticas AH y EE, wordclouds comparativos de unigramas y bigramas de los *tweets* positivos y negativos de cada una de las temáticas AH y EE, wordclods de unigramas y bigramas de *tweets* con polaridad netura de las temáticas AH y EE	4.6, 5.2	Figura 5.10, Figura 5.11, Figura 5.12, Figura 5.13, Figura 5.14, Figura 5.15, Figura 5.16, Figura 5.17, Tabla 5.2.4, Tabla 5.2.5, Tabla 5.2.6
15_Muestra_terminos_EE_tweets.R	Obtención de las muestras representativas de los *tweets* positivos que contenían los términos "test" y "teacher"	4.6, 5.2	
16_Estadisticas_NRC.R	Análisis del lexicón NRC v0.92 según el número de palabras con y sin emoción, el número de palabras según el número de emociones y según su polaridad y el número de emociones	4.6, 5.3	Figura 5.18, Figura 5.19, Figura 5.20, Figura 5.21, Tabla 5.3.1, Tabla 5.3.2, Tabla 5.3.3, Tabla 5.3.4
17_Analisis_Emociones_Tweets_Seleccionados_Bruto.R	Detección de las palabras coincidentes del lexicón NRC v0.92 con las palabras de los *tweets* de las temáticas AH y EE, cálculo de las emociones de cata *tweet*, distribución del número de *tweets* normalizados por emoción y tipo de sentimiento positivo o negativo por cada temática	4.7, 5.3	Figura 5.22, Figura 5.23
18_Busqueda_infuencers_y_contenidos.R	Obtención de *tweets* emitidos por empresas y sociedades y *tweets* en los que son mencionados, confección de la distribución del número de *tweets* por usuarios en los que se mencionan empreas y sociedades y del número de menciones por usuarios	4.7, 5.4	Tabla 5.4.1
19_Calculo_menciones_popularidad_para_SNA.R	Confección de la tabla compuesta por los usuarios mencionados, el identificador de *tweet* en el que se mencionan y los usuarios que mencionan, estadística descriptiva de la frecuencia de menciones de los usuarios mencionados, obtención de usuarios más mencionados y su número de menciones, estadística descriptiva de la frecuencia con la que los usuarios mencionan a otros, obtención de usuarios que con mayor frecuencia mencionan a otros y sus respectivas frecuencias	4.8, 5.4	Tabla 5.4.2, Tabla 5.4.3, Tabla 5.4.4, Tabla 5.4.6, Tabla 5.4.7, Tabla 5.4.8

Nombre	Descripción	Referencias a apartados de este documento	Referencias a tablas y figuras de este documento
20_SNA_Tematica_ALL.R	Confección de la red completa de usuarios que realizan alguna mención o son mencionados, cálculo de métricas de centralidad y topología, obtención de redes aisladas dentro de la red completa, distribución estadística de las medidas de centralidad y topología de las redes aisladas, representaciones gráficas de la red completa con el método Kamada and Kawai	4.8, 5.4	Figura 5.24, Figura 5.25, Tabla 5.4.11, Tabla 5.4.12, Tabla 5.4.13

Ficheros de datos y su descripción

Fichero de datos	Descripción
CyClCm.RData	Dataframe de los *tweets* adquiridos
CyClCm_ff.RData	Dataframe de los *tweets* adquiridos y resultantes después de los procesos de limpieza y preparación
Corpus_CyClCm_f.RData	Corpus creado con los *tweets* resultantes después de los procesos de limpieza y preparación
Tdm_f.RData	Matriz TDM de bigramas
Tdm_ff.RData	Matriz TDM de bigramas con una frecuencia absoluta superior a 29
Dtm_f.RData	Traspuesta de la matriz TDM de bigramas con una frecuencia absoluta superior a 29 y con el factor Tf-Idf normalizado
ElbowNUMERODECLUSTERS_rep50.Rdata	Ficheros resultado de la repetición del método *spherical k-means* de K=2 hasta K=285 clústeres 50 veces para cada K
Best_skm_rep50.RData	Resultado de la mejor repetición en cada K del método *spherical k-means* de K=2 hasta K=285 clústeres 50 veces correspondiente al menor valor de la función criterio Q
Resum_silouette.RData	Valores de silueta promedio para la mejor repetición en cada K del método spherical k-means de K=2 hasta K=285 clústeres 50 veces correspondiente al menor valor de la función criterio Q
Clusters_skm_rep50.Rdata	Número de clústeres desde K=2 hasta K=285
Stats_skm_rep50.Rdata	Valores mínimo, Q1, mediana, promedio, Q3 y máximo de las 50 repeticiones del *método spherical k-means* de K=2 hasta K=285 clústeres
Skmeans_ncluster_100_repNUMERODEREPETICION_ncores15.Rdata Skmeans_Serie1_ncluster_100_repNUMERODEREPETICION.Rdata Skmeans_Serie2_ncluster_100_repNUMERODEREPETICION.Rdata Skmeans_Serie3_ncluster_100_repNUMERODEREPETICION.Rdata	Ficheros resultado de las 9723 repeticiones del método *spherical k-means* para K*=100 clústeres
Bestskm_100clusters.RData	Resultado de la mejor repetición del método *spherical k-means* para K=100 clústeres correspondiente al menor valor de la función criterio Q
Rater_Total.pdf	Wordclouds de unigramas y bigramas de los *tweets* pertenecientes a cada uno de los clústeres obtenidos de la mejor

Fichero de datos	**Descripción**
	repetición del método *spherical k-means* para K=100 clústeres correspondiente al menor valor de la función criterio Q
BIBd_clustersrater.Rdata	Distribución de clústeres asignados a cada uno de los 18 expertos según el BIBD escogido
Entrada datos evaluadores.csv	Tabla para entrar manualmente la clasificación de los clústeres asignados a cada experto
Resultados clasificacion.xlsx	Resultados y estadísticos de la clasificación de los clústeres realizada por los expertos
SentiWordNet_3.0.0_20130122.txt	Lexicón SentiWordNet 3.0
Sentiword_mod.RData	Lexicón final confeccionado a partir del lexicón SentiWordNet 3.0
Polaridad_AH.csv	Resultado de la polaridad de los *tweets* de la temática Actividad Humana (AH)
Polaridad_EE.csv	Resultado de la polaridad de los *tweets* de la temática Entorno Educativo (EE)
id_tweets_AH.Rdata	Número identificador (campo Id) de los *tweets* de la temática Actividad Humana (AH)
id_tweets_EE.Rdata	Número identificador (campo Id) de los *tweets* de la temática Entorno Educativo (EE)
Clasificacion muestra tweets EE.csv	Resultado de la clasificación de la muestra de *tweets* con polaridad positiva de Entorno Educativo (EE) que contienen los términos "test" y "teacher" según su sentimiento y su ironía
NRC-Emotion-Lexicon-Wordlevel-v0.92.txt	Lexicón de valencia de emociones NRC v0.92
Emociones_AH.RData	Resultado de los sentimientos de los *tweets* de la temática Actividad Humana (AH)
Emociones_EE.RData	Resultado de los sentimientos de los *tweets* de la temática Entorno Educativo (EE)
usuariosInspección.Rdata	Cuentas de Twitter y nombre de las empresas y sociedades presumiblemente relevantes en la química
01_Tweets_Screenname_Empresas.xlsx	Tweets realizados por la lista de empresas y sociedades presumiblemente relevantes en la química
02_Tweets_lista_mencionados.xlsx	Tweets realizados usuarios que mencionan a cuentas de la lista de empresas y sociedades presumiblemente relevantes en la química
Tabla de menciones TOTAL.csv	Tabla de los usuarios mencionados, el identificador del *tweet* en el que se mencionan y los usuarios que mencionan
Usuarios mencionados TOTAL.csv	Menciones únicas de usuarios mencionados por el resto de usuarios
Usuarios que mencionan TOTAL.csv	Usuarios únicos que mencionan a otros usuarios

Documentos de clasificación de los *wordclouds* y su descripción

Nombre del documento	Descripción
01_Rater_01.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 1 según el BIBD escogido
02_Rater_02.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 2 según el BIBD escogido
03_Rater_03.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 3 según el BIBD escogido
04_Rater_04.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 4 según el BIBD escogido
05_Rater_05.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 5 según el BIBD escogido
06_Rater_06.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 6 según el BIBD escogido
07_Rater_07.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 7 según el BIBD escogido
08_Rater_08.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 8 según el BIBD escogido
09_Rater_09.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 9 según el BIBD escogido
10_Rater_10.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 10 según el BIBD escogido
11_Rater_11.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 11 según el BIBD escogido
12_Rater_12.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 12 según el BIBD escogido
13_Rater_13.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 13 según el BIBD escogido
14_Rater_14.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 14 según el BIBD escogido
15_Rater_15.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 15 según el BIBD escogido
16_Rater_16.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 16 según el BIBD escogido
17_Rater_17.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 17 según el BIBD escogido
18_Rater_18.pdf	Wordclouds de unigramas y bigramas de la distribución de clústeres asignado al experto 18 según el BIBD escogido

Ficheros de datos y su relación con el código R desarrollado

Nombre del fichero de datos	Nombre del fichero de código R																			
	01_Estadisticas_tweets_brutos.R	02_Estadisticas_Tweets_Brutos_y_Netos.R	03_Tratamiento_Tweets_Tdm.R	04_Calculo_dtm.R	05_PCalculo_Elbow.R	06_Selec_Stats_Elbow.R	07_Grafica_Elbow_para_tesis.R	08_PCalculs_skmeans_100cluster_Paralel_Serie.R	09_Cerca_plot_analisis_millor_resultat_100clusters.R	10_Calculo_BIBD.R	11_Extraccion Graficos.R	12_Calculo_Estadisticas_Categorizacion_Kappa.R	13_Calculo_lexicon_from_SentiWordNet30.R	14_Analisis_Sentimientos_Tweets_Seleccionados.R	15_Muestra_terminos_EE_tweets.R	16_Estadisticas_NRC.R	17_Analisis_Emociones_Tweets_Seleccionados_Bruto.R	18_Busqueda_infuencers_y_contenidos.R	19_Calculo_menciones_popularidad_para_SNA.R	20_SNA_Tematica_ALL.R
CyClCm.RData (*)	X	X	X															X	X	
CyClCm_ff.RData (*)		X	X								X	X		X	X		X			
Corpus_CyClCm_f.RData			X								X	X		X						
Tdm_f.RData			X	X																
Tdm_ff.RData				X																
Dtm_f.RData				X	X		X	X												
ElbowNUMERODECLUSTERS_rep50.Rdata (**)					X	X														
Best_skm_rep50.RData						X	X													
Resum_silouette.RData							X													
Clusters_skm_rep50.Rdata						X	X													
Stats_skm_rep50.Rdata						X	X													
Skmeans_ncluster_100_repNUMERODEREPETICION_ncores15.Rdata (***) Skmeans_Serie1_ncluster_100_repNUMERODEREPETICION.Rdata (***) Skmeans_Serie2_ncluster_100_repNUMERODEREPETICION.Rdata (***) Skmeans_Serie3_ncluster_100_repNUMERODEREPETICION.Rdata (***)								X	X											

Nombre del fichero de datos	Nombre del fichero de código R																			
	01_Estadisticas_tweets_brutos.R	02_Estadisticas_Tweets_Brutos_y_Netos.R	03_Tratamiento_Tweets_Tdm.R	04_Calculo_dtm.R	05_PCalculo_Elbow.R	06_Selec_Stats_Elbow.R	07_Grafica_Elbow_para_tesis.R	08_PCalculs_skmeans_100cluster_Paralel_Serie.R	09_Cerca_plot_analisis_millor_resultat_100clusters.R	10_Calculo_BIBD.R	11_Extraccion Graficos.R	12_Calculo_Estadisticas_Categorizacion_Kappa.R	13_Calculo_lexicon_from_SentiWordNet30.R	14_Analisis_Sentimientos_Tweets_Seleccionados.R	15_Muestra_terminos_EE_tweets.R	16_Estadisticas_NRC.R	17_Analisis_Emociones_Tweets_Seleccionados_Bruto.R	18_Busqueda_infuencers_y_contenidos.R	19_Calculo_menciones_popularidad_para_SNA.R	20_SNA_Tematica_ALL.R
Bestskm_100clusters.RData									X		X	X		X						
Rater_Total.pdf											X									
BIBd_clustersrater.Rdata										X	X	X								
NUMERODEEXPERTO_Rater_NUMERODEEXPERTO.pdf (****)											X									
Entrada datos evaluadores.csv												X								
Resultados clasificacion.xlsx												X								
SentiWordNet_3.0.0_20130122.txt													X							
Sentiword_mod.RData													X	X						
Polaridad_AH.csv														X						
Polaridad_EE.csv														X						
id_tweets_AH.Rdata														X						
id_tweets_EE.Rdata														X	X					
Clasificacion muestra tweets EE.csv															X					
NRC-Emotion-Lexicon-Wordlevel-v0.92.txt																X	X			
Emociones_AH.RData																	X			
Emociones_EE.RData																	X			

	Nombre del fichero de código R																			
Nombre del fichero de datos	01_Estadisticas_tweets_brutos.R	02_Estadisticas_Tweets_Brutos_y_Netos.R	03_Tratamiento_Tweets_Tdm.R	04_Calculo_dtm.R	05_PCalculo_Elbow.R	06_Selec_Stats_Elbow.R	07_Grafica_Elbow_para_tesis.R	08_PCalculs_skmeans_100cluster_Paralel_Serie.R	09_Cerca_plot_analisis_millor_resultat_100clusters.R	10_Calculo_BIBD.R	11_Extraccion Graficos.R	12_Calculo_Estadisticas_Categorizacion_Kappa.R	13_Calculo_lexicon_from_SentiWordNet30.R	14_Analisis_Sentimientos_Tweets_Seleccionados.R	15_Muestra_terminos_EE_tweets.R	16_Estadisticas_NRC.R	17_Analisis_Emociones_Tweets_Seleccionados_Bruto.R	18_Busqueda_infuencers_y_contenidos.R	19_Calculo_menciones_popularidad_para_SNA.R	20_SNA_Tematica_ALL.R
usuariosInspección.Rdata																		X		
01_Tweets_Screenname_Empresas.xlsx																		X		
02_Tweets_lista_mencionados.xlsx																		X		
Tabla de menciones TOTAL.csv																			X	X
Usuarios mencionados TOTAL.csv																			X	
Usuarios que mencionan TOTAL.csv																			X	

(*) Por temas de confidencialidad, ya que los usuarios tienen derecho a borrar sus *tweets* publicados aunque los *tweets* son públicos, los ficheros no están en la documentación electrónica en la dirección web https://github.com/mguerris/Tesis-Doctoral.git y están a disposición previa petición por e-mail al autor de este estudio.

(**) Los ficheros resultado de la implementación en paralelo del método spherical k-means de K=2 a K=285 clústeres 50 veces para cada K no están en la documentación electrónica en la dirección web https://github.com/mguerris/Tesis-Doctoral.git por su gran tamaño y están a disposición previa petición por e-mail al autor de este estudio.

(***) Los ficheros resultado del cálculo en paralelo y en serie del método spherical k-means con K*=100 clústeres 9723 veces no están en la documentación electrónica en en la dirección web https://github.com/mguerris/Tesis-Doctoral.git por su gran tamaño y están a disposición previa petición por e-mail al autor de este estudio.

(****) Los ficheros corresponden a los documentos pdf entregados a los expertos para su evaluación (ver apartado 4.5). Estos pueden consultarse en la carpeta “Documentos de evaluación de Wordclouds” dentro de la documentación electrónica de la tesis.